EDITED BY
ABIGAIL C. COHN
CÉCILE FOUGERON
MARIE K. HUFFMAN

The Oxford Handbook *of* LABORATORY PHONOLOGY

THE OXFORD HANDBOOK OF

LABORATORY PHONOLOGY

OXFORD HANDBOOKS IN LINGUISTICS

The Oxford Handbook of Applied Linguistics
Second Edition
Edited by Robert B. Kaplan

The Oxford Handbook of Case
Edited by Andrej Malchukov and Andrew Spencer

The Oxford Handbook of Cognitive Linguistics
Edited by Dirk Geeraerts and Hubert Cuyckens

The Oxford Handbook of Comparative Syntax
Edited by Gugliemo Cinque and Richard S. Kayne

The Oxford Handbook of Compounding
Edited by Rochelle Lieber and Pavol Štekauer

The Oxford Handbook of Computational Linguistics
Edited by Ruslan Mitkov

The Oxford Handbook of Compositionality
Edited by Markus Werning, Edouard Machery, and Wolfram Hinzen

The Oxford Handbook of Field Linguistics
Edited by Nicholas Thieberger

The Oxford Handbook of Grammaticalization
Edited by Heiko Narrog and Bernd Heine

The Oxford Handbook of Japanese Linguistics
Edited by Shigeru Miyagawa and Mamoru Saito

The Oxford Handbook of Laboratory Phonology
Edited by Abigail C. Cohn, Cécile Fougeron, Marie K. Huffman

The Oxford Handbook of Language Evolution
Edited by Maggie Tallerman and Kathleen Gibson

The Oxford Handbook of Language and Law
Edited by Lawrence Solan and Peter Tiersma

The Oxford Handbook of Linguistic Analysis
Edited by Bernd Heine and Heiko Narrog

The Oxford Handbook of Linguistic Interfaces
Edited by Gillian Ramchand and Charles Reiss

The Oxford Handbook of Linguistic Minimalism
Edited by Cedric Boeckx

The Oxford Handbook of Linguistic Typology
Edited by Jae Jung Song

The Oxford Handbook of Translation Studies
Edited by Kirsten Malmkjaer and Kevin Windle

THE OXFORD HANDBOOK OF

LABORATORY PHONOLOGY

Edited by

ABIGAIL C. COHN,
CÉCILE FOUGERON,

and

MARIE K. HUFFMAN

with assistance from

MARGARET E. L. RENWICK

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First published 2012

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
Data available
Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
CPI Group (UK) Ltd, Croyden, CR0 4YY
ISBN 978–0–1–9–957503–9
1 3 5 7 9 10 8 6 4 2

To Pat Keating

who has had a profound influence on each of us
and inspired us to work together on this volume

This page intentionally left blank

Contents

PART III: MULTIDIMENSIONAL REPRESENTATIONS OF KNOWLEDGE OF SOUND STRUCTURE

PART IV: INTEGRATING DIFFERENT PERSPECTIVES: INSIGHTS FROM PRODUCTION PERCEPTION, AND ACQUISITION

PART V: METHODOLOGIES AND RESOURCES

Acknowledgments

First we thank John Davey for inviting us to undertake this project, for his foresight that there was a need for this handbook, and his trust in giving us a free hand in how to carry it out.

We thank all the contributors to this volume who agreed to play ball with us as we embarked on this intensely collaborative interpretation of what a handbook should be. We also deeply thank all of the reviewers who lent their critical eyes and precious time to make the end result more accessible, precise, and helpful.

We owe an enormous debt to Peggy Renwick who in addition to being a linguistics graduate student, has served as the editorial manager for this project. Without Peggy's organization, efficiency, insight, enthusiasm, and humor, the project would not have come together. Peggy, we can't thank you enough! Thanks to the Cornell Department of Linguistics for financial support through faculty research funds to make Peggy's assistance possible.

Thanks to Julia Steer for seeing the project through to completion with us, and to all of the OUP staff who assisted along the way. Thanks also to Richard Shedenhelm for his special help and critical intervention with the references and index and to Henning Reetz for his help cross-checking references.

We thank Mary E. Beckman and John Kingston and Cambridge University Press for permission to reprint "Introduction," in J. Kingston and M. Beckman (1990, eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (Cambridge: Cambridge University Press, 1-16); and Janet B. Pierrehumbert, Mary E. Beckman, and D. Robert Ladd and Oxford University Press for permission to reprint "Conceptual foundations of phonology as a laboratory science," in N. Burton-Roberts, P. Carr, and G. Docherty (2000, eds.). *Phonological Knowledge: Conceptual and Empirical Issues* (Oxford: Oxford University Press, 273–304).

Thanks to the community of people who developed the technologies (Skype, Gmail, Google Docs, Dropbox . . .) that made our long distance collaboration feasible. It is just over four years since we started to discuss this project and it has been an enormously fun collaboration. We will miss our weekly meetings.

We each thank our departments and colleagues for the various ways they have lent support to this endeavor and finally we each thank our families (to Wilson and Hannah and Sarah, to Vassilis and Eleni, to Bililo and Bazika) for their patience and support.

We hope that the end result reflects at least in some measure how enjoyable and productive it has been for us to work with each other and the community of scholars brought together in this book.

The Contributors

Adam Albright is Associate Professor at MIT. His research interests include phonology, morphology, and learnability, with an emphasis on using computational modeling and experimental techniques to investigate issues in phonological theory.

Arto Anttila is Associate Professor of Linguistics at Stanford University and Adjunct Professor (Dosentti) of General Linguistics at the University of Helsinki, Finland. His research focuses on phonology, morphology, and language variation.

Amalia Arvaniti is Associate Professor of Linguistics at the University of California, San Diego and co-organizer of LabPhon IV. Her research focuses on the realization and representation of intonation and speech rhythm, and the interaction of intonation and pragmatics cross-linguistically.

Harald Baayen is Professor of Quantitative Linguistics at the University of Alberta. His research addresses the role of probability in grammars, particularly its role in morphology and morphological processing in visual and auditory comprehension and in speech production.

Mary E. Beckman is Professor of Linguistics at the Ohio State University, and co-founder of the Laboratory Phonology conference series. She has done research on word accent and intonation, articulatory modeling of prosody, and phonological development.

Paul Boersma is Professor of Phonetic Sciences at the University of Amsterdam. His research focuses on the development of a computational multilevel model of phonological and phonetic production and comprehension, including their acquisition and evolution across the generations.

Yiya Chen is Lecturer at Leiden University Center for Linguistics (LUCL) and affiliated member at Leiden Institute for Brain and Cognition (LIBC). Her research mainly focuses on prosody and prosodic variation, with particular attention to tonal languages.

Ioana Chitoran is Associate Professor of Linguistics in the Linguistics and Cognitive Science Program at Dartmouth College. Her research focuses on the

phonetics-phonology interface, and the relation between temporal variability and phonological structure, particularly in Caucasian and Romance languages.

Cynthia G. Clopper is Assistant Professor of Linguistics at the Ohio State University. Her research interests include the relationship between linguistic variation and speech processing and prosodic variation within and across languages.

Andries W. Coetzee is Assistant Professor of Linguistics at the University of Michigan. He specializes in phonology, focusing on the intersection between theoretical and laboratory phonology, and with particular attention to the role of variation in speech production and processing.

Abigail C. Cohn is Professor of Linguistics at Cornell University. Her research focuses on the relationship between phonology and phonetics and is informed by laboratory phonology approaches.

Jennifer Cole is Professor in the Department of Linguistics and Beckman Institute at the University of Illinois at Urbana-Champaign, co-organizer of LabPhon 9, and editor of *Laboratory Phonology*. Her research interests include phonological theory, prosody, phonetic variability, learning, and speech processing and recognition.

Marie-Hélène Côté is Associate Professor and Chair of Linguistics at the University of Ottawa. Her research addresses the role of perceptual factors in phonological processes, the status of the syllable, and treatment of variation, with particular attention to French.

Lisa Davidson is Associate Professor of Linguistics at New York University. Her main areas of interest are cross-language speech production and perception, foreign language acquisition, articulatory phonology, and loanword phonetics and phonology.

Katherine Demuth is Professor of Linguistics at Macquarie University in Sydney, Australia. Her research explores children's development of phonological and morphological representations, including the acoustic and articulatory correlates of early speech planning processes.

Mariapaola D'Imperio is Professor of Phonetics and Phonology at the University Aix-Marseille I and co-organizer of LabPhon 10. Her main research interests are intonational phonology, prosody and meaning in Romance languages, and speech production and perception.

Gerard Docherty is Professor of Phonetics at Newcastle University and co-organizer of LabPhon II. His recent work has focused on sociophonetic variability and its implications for our understanding of the phonetics-phonology interface.

Jan Edwards is Professor of Communicative Disorders at the University of Wisconsin-Madison, and Research Scientist at the Waisman Center. Her research

focuses on the interactions between phonological and lexical development in early childhood.

Mirjam Ernestus is Associate Professor at the Center for Language Studies, Radboud University Nijmegen and associate editor of *Laboratory Phonology*. Her research focuses on production and comprehension of pronunciation variation, especially in spontaneous speech.

Paola Escudero is Senior Researcher and Lecturer at MARCS Auditory Laboratories, University of Western Sydney. Her research focuses primarily on phonetic, phonological, and lexical development in L1, L2, bilingual, and multilingual speakers.

Ashley Farris-Trimble is a Post-Doctoral Research Fellow at the University of Iowa. Her research examines issues in spoken language comprehension including whether listeners are sensitive to phonological alternation during online word recognition, and lexical competition dynamics in cochlear implant users.

Cécile Fougeron is Research Scientist in experimental phonetics at CNRS/University of Paris 3 and co-organizer of LabPhon 10. Her research interests include the segmental manifestation of prosodic organization and the phonetic characteristics of speech disorders.

Stefan A. Frisch is Assistant Professor in the Department of Communication Sciences and Disorders at the University of South Florida. His research focuses on the representation and processing of words in the mental lexicon, and speech articulation, particularly the articulatory patterns of speech errors.

Sónia Frota is Assistant Professor in the Department of Linguistics at the Universidade de Lisboa and Director of the Phonetics Laboratory and Lisbon Baby Lab. Her research focuses on prosodic and intonational phonology, the syntax-phonology interface, and acquisition and processing of prosody.

Adamantios Gafos is Professor of Linguistics at New York University and Senior Scientist at Haskins Laboratories. His interests lie at the intersection of phonology and cognitive science.

Louis Goldstein is Professor of Linguistics at the University of Southern California, Senior Scientist at Haskins Laboratories, and co-organizer of LabPhon 8. His main research interest is in dynamical modeling of articulatory organization in speech production, phonetics, and phonology.

Helen M. Hanson is Assistant Professor in Electrical and Computer Engineering, Union College, and Research Affiliate at MIT's Research Laboratory of Electronics. Her research addresses quantal and enhancement theory, the acoustics of developing speech, speech respiration and prosody, and speech synthesis.

Jonathan Harrington is Professor of Phonetics at the Institute of Phonetics and Speech Processing, University of Munich. His research is in laboratory phonology, with a particular emphasis on the phonetic bases of sound change.

Mark Hasegawa-Johnson is Associate Professor in the Department at Electrical and Computer Engineering and Beckman Institute at the University of Illinois at Urbana-Champaign and associate editor of *Laboratory Phonology*. His field of interest is speech production and recognition by humans and computers.

Sarah Hawkins is Director of Research in Speech and Music Science at the University of Cambridge. She explores processing of phonetic detail that indicates linguistic structure, but not necessarily phonemic category, and is extending her interests into musical interaction.

Jennifer Hay is Associate Professor in the Department of Linguistics at the University of Canterbury, Director of the New Zealand Institute of Language, Brain, and Behaviour, and co-organizer of LabPhon 11. Her research interests are in laboratory phonology, sociophonetics, and morphology.

Lori L. Holt is Professor of Psychology at Carnegie Mellon University. Her research uses human and non-human perceptual and learning paradigms, neuroimaging, and computational models to investigate the perceptual and cognitive mechanisms contributing to speech perception.

Philip Hoole is Senior Lecturer at the Institute of Phonetics and Speech Processing, Munich University. His main research interests include linguistic phonetics, speech motor control, and laryngeal articulation.

Marie K. Huffman is Associate Professor of Linguistics at Stony Brook University. Her research focuses on the acoustic analysis of speech, especially its temporal structure.

William Idsardi is Professor of Linguistics at the University of Maryland, College Park. His research interests span across phonological theory, poetic meter, and phonological acquisition, including computational modeling and neurophysiological studies.

Khalil Iskarous is a Research Scientist at Haskins Laboratories. His research focuses on speech production and its relation to speech perception, dynamical theories of linguistic structure, effects of contrast and prosody on coarticulation, articulatory-acoustic relations, and speech development and maturation.

Paul Iverson is Reader in Speech Science at University College London. His primary research focus is plasticity in speech perception, particularly in first- and second-language acquisition, accent adaptation, and phonetically degraded conditions (e.g. noise and hearing impairment).

John Kingston is Professor of Linguistics at the University of Massachusetts, Amherst and co-founder of the Laboratory Phonology conference series. His principal interests are the interface between phonetics and phonology, and the interaction between auditory processing, linguistic knowledge, and articulatory control.

Robert Kirchner is Associate Professor of Linguistics at the University of Alberta. His research interests include the phonetics-phonology interface and computational modeling of phonological learning.

Barbara Kühnert is Assistant Professor of English Linguistics and member of the CNRS at the University of Paris 3, and co-organizer of LabPhon 10. Her research interests lie in the area of speech production, in particular its temporal organization.

D. Robert Ladd is Professor of Linguistics at the University of Edinburgh, co-organizer of LabPhon II, and first President of the Association for Laboratory Phonology. His research focuses on intonation, and more recently on phonological representation and interfaces with phonetics.

Aditi Lahiri is Professor of Linguistics and Fellow of Somerville College, University of Oxford. Her research interests include phonology, phonetics, historical linguistics, psycholinguistics, and neurolinguistics.

Dan Loehr is Principal Engineer, computational linguist, and manager of an artificial intelligence department at the MITRE Corporation. His research interests include the theory and implementation of multimodal conversational systems, especially the contributions of gesture and intonation to discourse.

Jessica Maye is Lecturer in the Department of Linguistics at Northwestern University. Her research interests include speech perception by adults and infants, early language acquisition, and language processing in bilinguals.

Bob McMurray is Associate Professor of Psychology at the University of Iowa. His work uses eye tracking, developmental techniques, phonetic analysis, and computational models to examine the processes of online spoken word recognition in adults, children, and impaired populations.

Norma Mendoza-Denton is Associate Professor of Linguistic Anthropology at the University of Arizona. Her research addresses sociophonetic variation in discourse, the construction of social meaning through phonetic variants, language and gesture, and interfaces with biological and visual anthropology.

Jeff Mielke is Assistant Professor of Linguistics at the University of Ottawa. His research interests include laboratory and computational approaches to phonology.

Benjamin Munson is Associate Professor of Speech-Language-Hearing Sciences at the University of Minnesota, Minneapolis. His research addresses the processes through which children learn phonological categories, and also the learning and processing of socially stratified variation in spoken language.

Noël Nguyen is Professor of Speech and Language Sciences and head of the Laboratoire Parole et Langage, Aix-Marseille Université and CNRS, Aix-en-Provence, France. His research focuses on the production, perception, and understanding of spoken language.

Francis Nolan is Professor of Phonetics in the Department of Theoretical and Applied Linguistics, University of Cambridge. His research addresses aspects of phonetic theory including connected speech processes, speaker characteristics, variation in English, prosody, and forensic phonetics.

Janet B. Pierrehumbert is Professor of Linguistics at Northwestern University and co-organizer of LabPhon V. Her research interests include prosody and intonation, statistical models of phonological representation and processing, and interaction of cognitive and social factors in shaping language systems.

David Poeppel is Professor of Psychology and Neural Science at New York University. His research seeks to uncover the brain bases of speech perception, lexical representation, and lexical processing from both the linguistic and neurobiological points of view.

Brechtje Post is Lecturer in Phonetics and Phonology in the Department for Theoretical and Applied Linguistics, University of Cambridge. Her research interests include intonational phonetics and phonology, speech processing, prosodic phonology, and the acquisition of prosody.

Marianne Pouplier is Principal Investigator at the Institute of Phonetics and Speech Processing, Ludwig-Maximilians University Munich. Her main research interests include the phonetics-phonology interface and speech production.

Pilar Prieto is ICREA Research Professor affiliated to the Universitat Pompeu Fabra. Her research interests focus on the interaction between phonology and phonetics in intonation, audiovisual prosody, and the acquisition of prosody.

Henning Reetz is Chair for Phonetics and Phonology at the University of Frankfurt. His research interest is human speech perception and its modeling.

Niels O. Schiller is Professor of Psycho- and Neurolinguistics, Leiden University, and board member of the Leiden Institute for Brain and Cognition (LIBC). His research interests include word form encoding, reading aloud, and phonetic aspects of speech production.

James M. Scobbie is Professor of Speech Science at Queen Margaret University, Edinburgh and associate editor of *Laboratory Phonology*. His research addresses phenomena where phonology and phonetics overlap, using a mixture of methodologies, with a focus on articulatory phonetic data and cross-dialectal comparisons.

Christine H. Shadle is Senior Research Scientist at Haskins Laboratories. Her research focuses on the acoustics and aeroacoustics of speech production, especially of fricative consonants; she has used and developed different methods of vocal-tract imaging and signal analysis.

Rajka Smiljanic is Assistant Professor of Linguistics at the University of Texas at Austin. Her research focuses on experimental phonetics, cross-language and second-language speech production and perception, and laboratory phonology.

Jae Yung Song is Assistant Professor at the University of Wisconsin-Milwaukee. Her research interests focus on the articulatory and acoustic characteristics of child and child-directed speech, and their implications for the development of the phonological representation of words.

Shari R. Speer is Professor of Linguistics and Director of the Psycholinguistics Laboratories at the Ohio State University. Her primary research interest is the influence of prosodic structure, particularly intonation, on the production, comprehension, and acquisition of spoken languages.

Jane Stuart-Smith is Reader in English Language at the University of Glasgow. Her current research focuses on synchronic and diachronic approaches to variation in speech, linking research in phonetics and sociolinguistics based on investigations of socially stratified speech corpora.

Alice Turk is Professor of Linguistic Phonetics at the University of Edinburgh. Her research interests are in speech production, speech perception, and prosodic organization, with a particular focus on speech timing.

Linda Van Guilder is a Computational Linguist at Abraxas Corporation. Her interests lie with Natural Language Processing technologies that make language data accessible to end users focused particularly on natural language systems engineering.

Natasha Warner is Associate Professor of Linguistics at the University of Arizona and co-organizer of LabPhon 7. Her research addresses speech reduction, speech acoustics and perception, and the interface of phonetics, phonology, and psycholinguistics, and language revitalization.

Paul Warren is Associate Professor of Linguistics at Victoria University of Wellington, and co-organizer of LabPhon 11. His research interests include spoken word

recognition, the use of intonation in sentence processing, and the phonetics and phonology of English varieties.

Elizabeth C. Zsiga is Associate Professor of Linguistics at Georgetown University. Her research interests focus on the phonetic and phonological patterns that occur at and across word boundaries in connected speech.

Abbreviations

·	syllable boundary
ADS	Adult-directed speech
ANOVA	Analysis of Variance
Anti-EM	Anticipatory Eye Movement
APriL	Acquisition of Prosody in a first Language
ASCII	American Standard Code for Information Interchange
ASR	Automatic Speech Recognition
ATR	Advanced Tongue Root
AX	Perceptual identification task
AXB	Perceptual discrimination task
C	Consonant
CCC	Change, Chance, Choice—Blevins's model
CF	Center frequency
CHASE	Comfortable Head Anchor for Sonographic Examinations
CHILDES	Child Language Data Exchange System
CHT	Conditioned Head-Turn
CP paradigm	Categorical Perception Paradigm
CT	Cricothyroid
CV	Consonant-Vowel
CVC	Consonant-Vowel-Consonant
dB	Decibels
DFT	Discrete Fourier Transform
DFT	Dispersion-Focalization Theory
DRM	Distinctive Region Model
DTW	Dynamic Time Warping
EEG	Electroencephalography
EGG	Electroglottography
EMA	Electromagnetic Articulography
EMG	Electromyography
EMMA	Electromagnetic Midsagittal Articulography
EP	European Portuguese
EPG	Electropalatography

ERC	Elementary Ranking Condition
ERP	Event-related Potential
f0	Fundamental frequency
F1	First formant frequency
F2	Second formant frequency
F3	Third formant frequency
FDIF	Frequency Domain Inverse Filtering
fMRI	Functional Magnetic Resonance Imaging
FPA	Firthian Prosodic Analysis
FPD	Fine Phonetic Detail
FUL	Featurally Underspecified Lexicon
GLA	Gradual Learning Algorithm
GVS	Great Vowel Shift
H	High tone
H & H Theory	Hyper- and Hypo- Articulation Theory
HAS	High Amplitude Sucking
HATS	Head and Transducer Support system
HG	Harmonic grammar
HL	High-Level synthesis
HMM	Hidden Markov Model
HOCUS	Haskins Optically Corrected Ultrasound System
HPP	Headturn Preference Procedure
Hz	Hertz, or cycles per second.
ICE-GB	International Corpus of English (British English data subset)
IDS	Infant-directed Speech
If0	Intrinsic f0
INDSCAL	Individual Differences Scaling
INTSINT	International Transcription System for Intonation
IP	Intonational Phrase
IPA	International Phonetic Alphabet
IPLP	Intermodal Preferential Looking Procedure
IRED	Infrared Emitting Diodes
IViE	Intonational Variation in English
kHz	kiloHertz, or thousand samples per second. 1 kHz = 1000 Hz.
L	Low tone
L1	First language or native language
L2	Second language or non-native language
LAFF	Lexical Access From Features

LIFG	Left Inferior Frontal Gyrus
LPC	Linear Predictive Coding
MDS	Multidimensional Scaling
MEG	Magnetoencephalography
MF	Modulation Frequency
MMN	Magnetic Mismatch Negativity
Momel	Modeling Melody
MRI	Magnetic Resonance Imaging
MRPA	Machine Readable Phonetic Alphabet
NLM	Native Language Magnet model
NTSC	National Television System Committee
O/E	Observed over Expected
OCP	Obligatory Contour Principle
OT	Optimality Theory
PAL	Phase Alternating Line
PAM	Perceptual Assimilation Model
PEBLS	Phonological Exemplar-Based Learning System
PET	Positron Emission Tomography
PhP	Phonological Phrase
P_{oral}	Intraoral air pressure
PSOLA	Pitch Synchronous OverLap Add
PW	Prosodic Word
PWG	Prosodic Word Group
QT	Quantal Theory
RLD	Romance Languages Database
RMS	Root Mean Squared
RT	Reaction Time measures
SAMPA	Speech Assessment Methods Phonetic Alphabet
SLM	Speech Learning Model
SPICE	Simulation Program with Integrated Circuit Emphasis
SS ANOVA	Smoothing Spline Analysis of Variance
StOT	Stochastic Optimality Theory
SVLR	Scottish Vowel Length Rule
T-order	Typological order
TAD	Theory of Adaptive Dispersion
TBU	Tone-bearing unit
TIMITBET	Texas Instruments / Massachusetts Institute of Technology alphaBET

TMS	Transcranial Magnetic Stimulation
ToBI	Tones and Break Indices
U	Volume velocity, in units of volume per unit time
UR	Underlying Representation
UTF-8	Unicode Transformation Format, 8-bits
UTI	Ultrasound Tongue Imaging
V	Vowel
VC	Vowel-Consonant
VFP	Visual Fixation Procedure
VOT	Voice Onset Time
VWP	Visual World Paradigm
WSP	Weight-to-Stress Principle

PART I

INTRODUCTION

This page intentionally left blank

CHAPTER 1

INTRODUCTION

ABIGAIL C. COHN,
CÉCILE FOUGERON, AND
MARIE K. HUFFMAN

1.1 Background

Over the past few decades researchers interested in linguistic aspects of human speech have made a concerted effort to strengthen the empirical foundation of their work by incorporating the methodologies and perspectives of the traditionally experimentally oriented fields such as phonetics, sociolinguistics, language acquisition, speech science, and psycholinguistics. This integrated and dynamic approach has led to fruitful collaborations across research specialties, and the current state of the field is diverse and intellectually stimulating. This volume offers a detailed picture of this increasingly influential research perspective called *laboratory phonology*.

In a narrow sense, laboratory phonology is associated with an approximately biennial conference (*LabPhon*), each of which has resulted in a published volume of papers. Starting with LabPhon 11, conference papers appear in the recently founded Association of Laboratory Phonology's journal, *Laboratory Phonology*. In a broader sense, laboratory phonology is a scientific perspective of an expansive community of scholars who are dedicated to bringing multidisciplinary approaches to bear on the critical questions concerning how spoken language is structured, learned, and used. Laboratory phonology is not a specific theory. Rather, researchers with this perspective draw on theories and tools from various branches of the cognitive

and natural sciences to elucidate the nature of human speech (see Pierrehumbert et al. 2000/this volume for fuller discussion of this point; see also Croot 2010). Thus *laboratory* is understood here in a very broad sense, representing systematic experimental approaches.

The laboratory phonology approach has advanced our understanding of questions about human speech that have commonly been framed in terms of traditional definitions of phonetics and phonology. Central insights into the nature of these questions arise by placing them in the wider context of cognitive and biological systems, particularly through integrated investigation of production, perception, and acquisition. Laboratory phonology is an intellectual space. Embracing a variety of theoretical approaches leads to innovative research questions, and this eclectic outlook means that techniques are ever evolving. The focus of this volume is on where laboratory phonology is headed. In this introduction, we first briefly review the development of laboratory phonology in the context of the motivation for this volume (Section 1.2). We then turn to the goals and structure of the volume (Section 1.3).

1.2 Motivation for the volume and brief history of laboratory phonology

While handbooks exist for phonology, phonetics, sociolinguistics, and psycholinguistics, there is to date no handbook for laboratory phonology. Although not a field or subfield in the strict sense, with a twenty-five-year history, the growing body of literature within the laboratory phonology approach and the rich set of results in this shared endeavor deserves a state-of-the-art assessment. In addressing this need, this handbook presents research results and methodological approaches, while reflecting on them in light of broader themes and directions concerning the study of human speech.

The term *laboratory phonology* was coined by Janet Pierrehumbert in the planning stages of the first conference which took place at The Ohio State University, in June 1987. LabPhon I was co-organized by Mary Beckman and John Kingston, with one of the central goals being to bridge the distinct subfields and subcultures of phonology and phonetics. It also established the foundational premise that progress would be achieved more successfully through integrated methodologies, as stated in the introduction to the LabPhon I volume (Beckman and Kingston 1990/this volume p. 13) and repeated here:

> Therefore, we ask: how can we use the physical models and experimental paradigms of phonetics to construct more viable surface phonological representations? Con-

versely, what can we learn about underlying phonetic representations and processes from the formal cognitive models and computational paradigms of phonology? Determining the relationship between the phonological component and the phonetic component demands a hybrid methodology.

Since that time, the LabPhon conferences have brought together an increasingly large community of scholars with diverse backgrounds, but shared interests, addressing the fundamental question of the nature of human speech and phonological systems. Over the years, the importance of the conference and the impact of related work have grown, as attested to by the strong attendance at the biennial conferences, now including a truly international audience (see Pierrehumbert and Clopper 2010 for a network analysis of the increasing intellectual influence of this work). In this volume, we use the term *laboratory phonology* to refer to this body of research and we use the term *LabPhon* to refer to the conferences.

Over the past quarter century, there has been an evolution of issues and themes central to laboratory phonology (see Cohn 2010 for more detailed discussion). As mentioned above, the first LabPhon conference set out to bridge the gap between phonology and phonetics, to redefine the questions being asked, and to promote more integrated methodologies. For the first several meetings, the questions and methodologies were defined in terms of phonology and phonetics. Laboratory phonology brought into focus some of the central questions of the time, such as the nature of the interface, and the language-specific nature of phonetics (contra the view espoused by Chomsky and Halle 1968 *The Sound Pattern of English*), leading to the now commonly accepted concept of phonetic knowledge.

A central result is an enriched awareness of variation, not just in terms of the details of physical realization (implicit in phonetics), but in terms of variation in all dimensions of language use, leading to the question of the role of variation in the knowledge of sound systems. Traditional linguistics has ignored some types of variation and divided up the rest according to sometimes rigid categorizations—variation has often been assumed to be either systematic or unpredictable, regular or random. These divisions were inspired in part by common views of the division between competence and performance. The laboratory phonology perspective eschews this division, acknowledging that the full range of variation is central to an understanding of linguistic representations and processing of speech. Attention to phonetics, both production and perceptual processes, as well as to sociolinguistic and diachronic detail, have revealed the closely integrated nature of language competence and performance. Laboratory phonology has played a critical role in showing that only with greater attention to fine detail in our empirical studies will we be able to develop adequate models.

There has also been increased attention to questions that are central to psycholinguistics, including concerted attention to language acquisition and the lexicon, and highlighting the role of stochastic generalization in the organization and knowledge of sound systems (a critical alternative to viewing phonology and phonetics

as separate modules related by mapping). Recent work continues to strengthen these themes while drawing in new empirical domains such as signed language, second-language acquisition, and disordered systems. These threads of research have led to an enriched understanding of the complexity and the multiplicity of representations.

The focus on integrated methodologies in laboratory phonology has meant encouraging phonologists to extend their methods beyond the analysis of what Kenstowicz and Kisseberth (1979) term corpus-internal evidence; that is, the impressionistic transcription of a corpus of utterances. First, the increased attention to experimental data has highlighted the ways that relying on impressionistic data is both inadequate and misleading. The range of experimental methodologies employed highlights the complexity of linguistic behavior which is under speaker control. This shift also meant enhancing the quantitative, experimental methodologies of phonetics with more formal analysis and modeling. Experimental approaches expanded from linguistic phonetics to include psycholinguistic experimental approaches resulting in an integration of questions related to processing and cognition. These methodological shifts were fundamental to breaking down the way that practice in phonology and phonetics respectively led to "the assumed division of labor . . . [creating] a harmful illusion that we can compartmentalize phonological facts from phonetics facts" (Beckman and Kingston 1990/this volume p. 15).

Laboratory phonology's "coming of age," in terms of its successful development as a recognized approach to the investigation of human speech and sound systems, is attested to by the founding of the Association of Laboratory Phonology, celebrated at the recent LabPhon 12 meeting. The central goals of greater dialog across subfields, greater integration of methodology, and greater collaboration, have remained the hallmarks of laboratory phonology. The success of laboratory phonology is that those issues which at first were defined in an effort to bridge phonology and phonetics are now understood more broadly, as truly interdisciplinary questions, bridging linguistics with neighboring fields within the broader context of cognitive science. It is this rich research endeavor that is the focus of this volume.

1.3 Goals and structure of the volume

This handbook is designed to serve as a guide to the results, mechanics, and philosophy of the laboratory phonology approach. It is meant to illustrate this field of research and the many ways to harvest it. It aims to introduce in-depth discussion of critical questions facing the field, as well as some of the important outcomes, while

also surveying the many investigative approaches and tools that may be brought to bear on these questions. The goal is not only to characterize the current state of the art, but to lay the groundwork for future directions.

The thematic organization of early LabPhon conferences reflected the specific concern of integrating the study of phonology and phonetics, accomplished in part by focusing on particular themes. Yet, as discussed by Cohn (2010) and Pierrehumbert and Clopper (2010) and reviewed above, the field has evolved and the thematic coverage of research undertaken within the laboratory phonology approach has diversified. As a consequence, this handbook attempts neither to recapitulate this process of intellectual development, nor to comprehensively review each of these threads of research. Rather, a selection of major research topics and areas of active and promising research are highlighted. We have intentionally chosen *not* to compartmentalize the volume into thematic parts as is often done in such works. Fundamental issues are interwoven throughout the volume, and are treated through different lenses in the different parts of the book. For example, questions related to prosodic constituents are covered in multiple chapters and from different points of views, including their effect on segmental variation, their status in a theory of prosodic representations, and the many analytic tools used to investigate their phonetic reality. Language acquisition also receives attention throughout the book, in terms of how sound structures and processes are acquired, current methods used to study such questions, and how results from the language acquisition literature inform other questions about the nature of sound structure, its representation, and human speech more generally. This integrated structure offers multiple access points for learning about and learning to do laboratory phonology.

In planning the specific contents and structure of the volume, our goals were to gather a wide range of perspectives, across subfields and disciplines, and to highlight the complementarity of different approaches and backgrounds. We have emphasized multidisciplinarity and encouraged co-authorship, both within specific contributions, and often within collections of two or more contributions constituting individual chapters. The volume thus affirms the benefits gained through collaborative effort in the advancement of science.

The handbook is organized into five main parts. **Part I:** *Introduction*, sets the stage with this introduction and the reprinting of two foundational pieces: Beckman and Kingston's (1990) introduction to *Papers in Laboratory Phonology I* and Pierrehumbert's et al. (2000) "Conceptual foundations in phonology as a laboratory science." These together aim to provide a conceptual and historical framework upon which the rest of the work stands. The core of the volume is divided into **Parts II–IV** covering topics addressing fundamental issues (what and why) and **Part V** covering methodologies and resources (how) that together constitute laboratory phonology approaches to these questions.

In Parts II–IV, the key to the organization is conceptual coverage. A broad range of topics are covered, such as phonological contrast and representation, prosodic organization, structure and role of the lexicon, phonological and phonetic variation, sociophonetics, typological generalizations, speech perception and production, language acquisition, and historical change. Questions related to these topics are addressed from the perspective of how laboratory phonology approaches have provided insight into human speech and language structure. Authors were asked to frame the essential questions, review contribution of work in laboratory phonology, and identify current developments and promising approaches.

In Part II: *Nature and types of variation: Their interpretation within a laboratory phonology perspective*, the contributions examine different sources of variation in speech, speaker-related, message-related, and system-related. This set of contributions demonstrates the critical importance of engaging with this multifaceted and complex variation. They highlight the benefits of an integrated approach, rather than separating some aspects of variation as fundamental and other aspects as outside the domain of investigation.

In Part III: *Multidimensional representations of knowledge of sound structure*, the contributions examine the content, access, and evolution of representations of speech, with special attention to the variety and richness of linguistic representations, relationships between levels of representation, and the challenges these offer to current and future models of linguistic representations. This part starts with different perspectives on the nature of lexical representations and then turns to how different aspects of what have been traditionally understood as phonological structure are represented and organized, acquired, and change over time.

In Part IV: *Integrating different perspectives: Insights from production, perception, and acquisition*, the contributions offer insight into how laboratory phonology has informed traditional themes within the study of phonology, phonetics, and human speech more generally. How do the methodologies and ways of framing testable research questions through a laboratory phonology approach advance our understanding of these long-standing questions? Rich multidisciplinary work offers varied perspectives on the issues of the nature of production and perception and their integration, as well as the relationship between language acquisition and the human capacity for language.

In Part V: *Methodologies and resources*, the topics treated in the other parts are approached through a direct consideration of methodologies, paradigms, tools, and resources that together constitute how people *do* laboratory phonology. Authors were asked to present a critical overview of available techniques, including examples of research using these techniques and how they have increased our understanding of human speech. The contributions highlight the multidisciplinarity and diversity of methods that are the essence of the laboratory phonology perspective. This part presents selected methodologies and resources that have proven useful,

with attention to the types of theoretical issues to which these approaches have been (and can be) appropriately applied.

This volume is intended for a wide audience: graduate students, faculty, and other researchers in phonetics, phonology, psycholinguistics, and sociolinguistics. We hope it proves equally useful to those whose work does not directly address theoretical linguistic issues regarding speech, but which is critically concerned with speech and language, such as computational linguists, speech pathologists, neurolinguists, biologists, and anthropologists. We hope that this handbook will serve both the novice and the more advanced researcher alike, with its cross-cutting approach to themes and methodologies providing a fabric enabling readers to understand and engage with the material at a variety of levels.

The book is designed to be read either from start to finish, or by focusing on specific parts and chapters, and we hope the reader will use the abundant cross-referencing to take advantage of the connections between the many pieces. Since the book is not linear in its organization, the table of contents and index become all the more important as tools.

We invite the reader to jump in!

CHAPTER 2

INTRODUCTION, PAPERS IN LABORATORY PHONOLOGY I: BETWEEN THE GRAMMAR AND PHYSICS OF SPEECH (REPRINT)*

MARY E. BECKMAN AND
JOHN KINGSTON

* We thank Mary E. Beckman and John Kingston and Cambridge University Press for permission to reprint "Introduction," in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (Cambridge: Cambridge University Press, 1–16).

This chapter is a condensed and mildly edited version of the original first chapter to *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*. The condensing was done by removing the paragraphs that introduced each of the other twenty-seven chapters of the original volume. These paragraphs have been replaced by listings of the chapter titles and authors. All other edits are corrections of typos in the original.

While each of the papers in this volume has its specific individual topic, collectively they address a more general issue, that of the relationship between the phonological component and the phonetic component. This issue encompasses at least three large questions. First, how, in the twin processes of producing and perceiving speech, do the discrete symbolic or cognitive units of the phonological representation of an utterance map into the continuous psychoacoustic and motoric functions of its phonetic representation? Second, how should the task of explaining speech patterns be divided between the models of grammatical function that are encoded in phonological representations and the models of physical or sensory function that are encoded in phonetic representations? And third, what sorts of research methods are most likely to provide good models for the two components and for the mapping between them?

Previous answers to these questions have been largely unsatisfactory, we think, because they have been assumed *a priori*, on the basis of prejudices arising in the social history of modern linguistics. In this history, phonology and phonetics were not at first distinguished. For example, in the entries for the two terms in the *Oxford English Dictionary* each is listed as a synonym for the other; *phonology* is defined as "The science of vocal sounds (= PHONETICS)" and *phonetics as* "The department of linguistic science which treats of the sounds of speech; phonology." The subsequent division of this nineteenth-century "science of sounds" into the two distinct subdisciplines of phonology and phonetics gave administrative recognition to the importance of the grammatical function of speech as distinct from its physical structure and also to the necessity of studying the physical structure for its own sake. But this recognition was accomplished at the cost of creating two separate and sometimes mutually disaffected scientific subcultures.

We can trace the origin of this cultural fissure to two trends. One is the ever-increasing reliance of phonetic research on technology, rather than on just the analyst's kinesthetic and auditory sensibilities. This trend began at least in the first decade of this century, with the use of the X-ray to examine vowel production and the adoption of the kymograph for examining waveforms. With such technical aids, phoneticians could observe the physical aspects of speech unfiltered by its grammatical function. With this capability, phonetics expanded its subject matter far beyond the taxonomic description of "speech sounds" found in phonological contrast, to develop a broader, domain-specific attention to such extra-grammatical matters as the physiology of speech articulation and the physics of speech acoustics, the peripheral and central processes of speech perception, and the machine synthesis and recognition of speech.

The other trend that led to the separation of the two subdisciplines was the development of more complete formal models of the grammatical function of speech than are instantiated in the International Phonetic Alphabet. This trend had its initial main effect in the 1930s, with the emergence of distinctive feature theory, as elaborated explicitly in Prague Circle phonology (Trubetzkoy 1939) and implicitly

in the American structuralists' emphasis on symmetry in analyzing phonological systems (Sapir 1925). Distinctive feature theory effectively shifted the focus of twentieth-century phonology away from the physical and psychological nature of speech sounds to their role in systems of phonemic contrast and morphological relatedness.

Both of these trends undermined the alphabetic model that underlay the nineteenth-century synonymy between phonetics and phonology, but they did so in radically different ways. The analysis of "vocal sounds" into their component units of phonological contrast eventually led to new non-alphabetic representations in which phonological features were first accorded independent commutability in different rows of a matrix and then given independent segmentation on different autosegmental tiers. The use of new technology, on the other hand, questioned the physical basis originally assumed for alphabetic segmentation and commutability, by revealing the lack of discrete sequential invariant events in articulation or acoustics that might be identified with the discrete symbols of the IPA. These radically different grounds for doing away with a strictly alphabetic notation for either phonological or phonetic representations produced an apparent contradiction.

Modeling the cognitive function of speech as linguistic sign requires two things: first, some way of segmenting the speech signal into the primitive grammatical entities that contrast and organize signs and second, some way of capturing the discrete categorical nature of distinctive differences among these entities. A direct representation of these two aspects of the grammar of speech is so obviously necessary in phonological models that it is hardly surprising that the early, rudimentary phonetic evidence against physical segmentation and discreteness should elicit the reaction that it did, a reaction caricatured in Trubetzkoy's declaration that "Phonetics is to phonology as numismatics is to economics." A more benign form of this prejudice recurs in the common assumption among phonologists that nonautomatic, language-specific aspects of phonetic representations and processes should share the discrete segmental nature of phonological symbols and rules.

This apparent contradiction induced also a complementary prejudice on the part of phoneticians. Instrumentally aided investigation of speech has resulted in decades of cumulative progress in phonetic modeling, including the monumental achievement of the acoustic theory of speech production (Fant 1960). A great deal of this research has necessarily been concerned with the details of mapping from one extra-grammatical system to another—for example, from acoustic pattern to cochlear nerve response or from motor excitation to articulatory pattern. This research into the relationships among different phonetic subcomponents has derived little direct benefit from advances in phonological theory. As a result, it has often been assumed that arguments about phonological representations and processes are irrelevant to the phonetic component as a whole, a prejudice that could be expressed in its most malignant form as "phonology is to phonetics as astrology is to astronomy."

We have caricatured these prejudices at some length because we feel that they are a major impediment to answering our three questions concerning the relationship between phonology and phonetics. They distort our pictures of the two linguistic components and of the shape of the mapping between them. One set of theories describes the mapping as a trivial translation at the point where the linguistically relevant manipulations of discrete symbolic categories are passed to the rote mechanics of production and perception. Another set of theories places the dividing line at the point where the arbitrary taxonomy of linguistic units yields to experimentally verifiable models of speech motor control, aerodynamics, acoustics, and perception.

Such distortions are inevitable as long as the relegation of aspects of sound patterns between the two linguistic components is guided by unquestioned assumptions about what research methods are appropriate to which field. Therefore, we ask: how can we use the physical models and experimental paradigms of phonetics to construct more viable surface phonological representations? Conversely, what can we learn about underlying phonetic representations and processes from the formal cognitive models and computational paradigms of phonology? Determining the relationship between the phonological component and the phonetic component demands a hybrid methodology. It requires experimental paradigms that control for details of phonological structure, and it requires observational techniques that go beyond standard field methods. The techniques and attitudes of this hybrid laboratory phonology are essential to investigating the large group of phonic phenomena which cannot be identified *a priori* as the exclusive province of either component.

An example of such a phenomenon is fundamental frequency downtrend. It is a common observation that f0 tends to fall over the course of an utterance. Phonologists have generally assumed that this downtrend belongs to the phonological component. They have postulated simple tone changes that add intermediate tone levels (e.g. McCawley's 1968 rule lowering High tones in Japanese to Mid tone after the first unbroken string of Highs in a phrase), or they have proposed hierarchical representations that group unbroken strings of High tones together with following Lows in tree structures that are interpreted as triggering a downshift in tonal register at each branch (e.g. Clements 1981). Phoneticians, on the other hand, have typically considered downtrend to belong exclusively to the phonetic component. They have characterized it as a continuous backdrop decline that unfolds over time, independent of the phonological tone categories. They have motivated the backdrop decline either as a physiological artifact of decaying subglottal pressure during a "breath group" (e.g. Lieberman 1967), or as a phonetic strategy for defining syntactic constituents within the temporal constraints of articulatory planning (e.g. Cooper and Sorensen 1981).

Each of these models is circumscribed by our notions about what research methods are appropriate to which linguistic subcomponent. If the observed downtrend

in a language is to be in the province of phonological investigation, it must be audible as a categorical tone change or register difference, and its immediate cause must be something that can be discovered just by examining the paradigm of possible phonological environments. If the downtrend is to be in the province of phonetic investigation, on the other hand, it must be quantifiable as a response to some physically specifiable variable, either by correlating fundamental frequency point-by-point to subglottal air pressure or by relating fundamental frequency averages for syllables to their positions within phonologically unanalyzed utterances of varying length. Each sort of model accounts for only those features of downtrend which can be observed by the methods used. Suppose, however, that the downtrend observed in a given language is not a single homogeneous effect, or suppose that it crucially refers both to discrete phonological categories and to continuous phonetic functions. Then there will be essential features of the downtrend that cannot be accounted for in either model. Indeed these features could not even be observed, because the research strategy attributes downtrend *a priori* either to manipulations of phonological representations or to phonologically blind phonetic processes.

In recent examples of the hybrid methods of laboratory phonology, Pierrehumbert (1980) has argued with respect to English and Poser (1984) and others (e.g. Pierrehumbert and Beckman 1988) regarding Japanese that downtrend is just such a heterogeneous complex of different components, many of which are generated in the mapping between phonological and phonetic representations. In both English and Japanese, certain phrase-final tones trigger a gradual lowering and compression of the pitch range as a function of the distance in time from the phrase edge. This component of downtrend is like the phonologically blind declination assumed in earlier phonetic models in that it seems to be a gradual backdrop decline. Yet it is unlike them in that it refers crucially to phonological phrasing and phrase-final tone features. Also, in both languages, certain other, phrase-internal, tonal configurations trigger a compression of the overall pitch range, which drastically lowers all following fundamental frequency values within some intermediate level of phonological phrasing. This largest component of downtrend is like the intermediate tone levels or register shifts in earlier phonological models in that it is a step-like change triggered by a particular phonological event, the bitonal pitch accent. Yet it is unlike them in that it is implemented only in the phonetic representation, without changing the phonological specification of the affected tones. If these characterizations are accurate, then downtrend cannot be modeled just by reference to the phonological or the phonetic structure. Indeed neither of these two components of downtrend can even be observed without instrumental measurements of fundamental frequency values in experiments that control for phonological tone values and phrasal structures. The phenomenon of downtrend seems to require such hybrid methods.

We think, moreover, that the list of phenomena requiring such hybrid methods and models is much larger than hitherto supposed. We believe that the time has come to undo the assumed division of labor between phonologists and other speech scientists; we believe this division of labor creates a harmful illusion that we can compartmentalize phonological facts from phonetic facts. At the very least, we maintain that the endeavor of modeling the grammar and the physics of speech can only benefit from explicit argument on this point. In support of this thesis, we present to you the papers in this volume.

Most of these papers were first presented at a conference we held in early June of 1987 at the Ohio State University. To this conference we invited about thirty phonologists and phoneticians. The papers at the conference were of two sorts. We asked some of the participants to report on their own research or ideas about some phenomenon in this area between phonology and phonetics. We asked the other participants to present papers reacting to these reports, by showing how the research either did or did not consider relevant phonological structures or phonetic patterns, and by reminding us of other research that either supported or contradicted the results and models proposed. By structuring the conference in this way we hoped to accomplish two things. First, we wanted to show the value of doing research in this area between phonology and phonetics, and second, we wanted to provoke phonologists and phoneticians into talking to each other and into thinking about how the methods and aims of the two fields could be united in a hybrid laboratory discipline tuned specifically to doing this sort of research. After the conference, we commissioned both sets of participants to develop their presentations into the papers which we have grouped in this volume so that the commentary papers follow immediately upon the paper to which they are reacting.

The specific topics that these groups of papers address fall into several large categories. First are papers which focus on suprasegmental phenomena in language: 2. Where phonology and phonetics intersect: the case of Hausa intonation, Sharon Inkelas and William R. Leben; 3. Metrical representation of pitch register, D. Robert Ladd; 4. The status of register in intonation theory: comments on the papers by Ladd and by Inkelas and Leben, G. N. Clements; 5. The timing of prenuclear high accents in English, Kim E. A. Silverman and Janet B. Pierrehumbert; 6. Alignment and composition of tonal accents: comments on Silverman and Pierrehumbert's paper, Gösta Bruce; 7. Macro and micro f0 in the synthesis of intonation, Klaus J. Kohler; 8. The separation of prosodies: comments on Kohler's paper, Kim E. A. Silverman; 9. Lengthenings and shortenings and the nature of prosodic constituency, Mary E. Beckman and Jan Edwards; 10. On the nature of prosodic constituency: comments on Beckman and Edwards's paper, Elisabeth Selkirk; 11. Lengthenings and the nature of prosodic constituency: comments on Beckman and Edwards's paper, Carol A. Fowler; 12. From performance to phonology: comments on Beckman and Edwards's paper, Anne Cutler.

The next group of papers addresses the question of the relationship between phonological representations and phonetic structures more generally: 13. The Delta programming language: an integrated approach to nonlinear phonology, phonetics, and speech synthesis, Susan R. Hertz; 14. The phonetics and phonology of aspects of assimilation, John J. Ohala; 15. On the value of reductionism and formal explicitness in phonological models: comments on Ohala's paper, Janet B. Pierrehumbert; 16. A response to Pierrehumbert's commentary, John J. Ohala.

The last group of papers addresses various aspects of segmental organization and coordination among segmental tiers: 17. The role of the sonority cycle in core syllabification, G. N. Clements; 18. Demisyllables as sets of features: comments on Clements' paper, Osamu Fujimura; 19. Tiers in articulatory phonology, with some implications for casual speech, Catherine P. Browman and Louis Goldstein; 20. Toward a model of articulatory control: comments on Browman and Goldstein's paper, Osamu Fujimura; 21. Gestures and autosegments: comments on Browman and Goldstein's paper, Donca Steriade; 22. On dividing phonetics and phonology: comments on the papers by Clements and by Browman and Goldstein, Peter Ladefoged; 23. Articulatory binding, John Kingston; 24. The generality of articulatory binding: comments on Kingston's paper, John J. Ohala; 25. On articulatory binding: comments on Kingston's paper, Louis Goldstein; 26. The window model of coarticulation: articulatory evidence, Patricia A. Keating; 27. Some factors influencing the precision required for articulatory targets: comments on Keating's paper, Kenneth N. Stevens; 28. Some regularities in speech are not consequences of formal rules: comments on Keating's paper, Carol A. Fowler.

The papers in this volume [. . .] represent a wide range of views on the issue of the relationship between phonology and phonetics. We trust that they also reflect the excitement and congenial argumentation that characterized the conference. And we hope that they will spark further inquiry into and discussion about topics in laboratory phonology.

CHAPTER 3

CONCEPTUAL FOUNDATIONS OF PHONOLOGY AS A LABORATORY SCIENCE (REPRINT)*

JANET B. PIERREHUMBERT,
MARY E. BECKMAN, AND
D. ROBERT LADD

3.1 Introduction

The term "laboratory phonology" was invented more than a decade ago as the name of an interdisciplinary conference series, and all three of us have co-organized

* This chapter is a reprinting of a paper that originally appeared in *Phonological Knowledge: Conceptual and Empirical Issues* (Burton-Roberts et al. 2000). We thank Janet B. Pierrehumbert, Mary E. Beckman, and D. Robert Ladd and Oxford University Press for permission to reprint this chapter.

That paper was a substantially reworked version of a position paper on laboratory phonology that was first published in *Current Trends in Phonology I* (Durand and Laks 1996). For the current reprinting, the authors have gone through the text to remove typos and to provide updated bibliographic information for citations that were "forthcoming" or "in press" at the time of the original printing. The following acknowledgments from the original paper still hold: For comments on previous drafts of this paper, we are grateful to Ann Bradlow, John Coleman, Jacques Durand, Jan Edwards, Stefan Frisch,

laboratory phonology conferences. Since then, the term has come into use not only for the conference series itself, but for the research activities exemplified by work presented there. In this paper, we give our own perspective on how research in laboratory phonology has shaped our understanding of phonological theory and of the relationship of phonological theory to empirical data.

Research activities within laboratory phonology involve the cooperation of people who may disagree about phonological theory, but who share a concern for strengthening the scientific foundations of phonology through improved methodology, explicit modeling, and cumulation of results. These goals, we would argue, all reflect the belief that phonology is one of the natural sciences, and that all of language, including language-specific characteristics and sociolinguistic variation, is part of the natural world. In what follows, we explore the ramifications of this position for the relationship of data and methods to phonological theory; for the denotations of entities in that theory; and for our understanding of Universal Grammar (UG) and linguistic competence.

3.2 Who and what

The Conference in Laboratory Phonology series was launched at the Ohio State University in 1987 by Beckman and Kingston to provide a forum for people doing laboratory research in phonology. The proceedings of this meeting also inaugurated a book series from Cambridge University Press. Subsequent conferences were hosted by University of Edinburgh, the University of California at Los Angeles, Oxford University, Northwestern University, and the University of York (UK), with the seventh conference to be held in 2000 at the University of Nijmegen. The conference has attracted people from very diverse intellectual backgrounds. American nonlinear phonology has been well represented by scholars such as Clements, Hayes, Leben, McCarthy, Selkirk, Steriade, and Vogel; Articulatory Phonology by Browman, Fowler, Goldstein, and Zsiga; Declarative Phonology by Broe, Coleman, Local, and Scobbie, and Optimality Theory by Steriade and Gussenhoven. Many of the participants—such as Cutler, Kohler, Ladefoged, Marslen-Wilson, Munhall, Nolan,

Jen Hay, Patricia Keating, Chris Kennedy, John Kingston, and Moira Yip. Although none of them are likely to agree with everything we have said here, we have benefited greatly from their suggestions about both substance and exposition. We are particularly grateful to David Hull, for fruitful discussion of the philosophy of science, and to the readers of *Current Trends in Phonology I* and the audience at Current Trends in Phonology II, for their responses to the earlier version of this paper. Work on the paper was supported by NSF Grant No. BNS-9022484 to Northwestern University; and by an Ohio State University Distinguished Scholar award and NIH Grant No. 1 RO1 DC02932-01A2 to Mary Beckman. Part of D. Robert Ladd's work on the paper was carried out while a visiting scholar at the Max-Planck Institute for Psycholinguistics, Nijmegen.

Shattuck-Hufnagel, Stevens, and Werker—are not associated with any particular school of phonological theory. About two thirds of the participants are phonologists or phoneticians affiliated with linguistics departments. Most of the rest are affiliated with departments of psychology, electrical engineering and computer science, or communication sciences and disorders.

Despite the diverse backgrounds of the participants, a number of common goals and values have been reflected in the papers delivered at the conference. Papers have either reported experimental research on the mental representation of sound structure and its physical correlates, or else built on such research in a substantial way. The goal of such research is to address issues in phonology that are not effectively addressed using traditional types of data (namely, field transcriptions, informant judgments, and symbolic records of morphological alternations). The research presented at the meeting has drawn heavily on results and methodological advances in related sciences, including psychology, life sciences, and acoustics.

3.3 Linguistics and the scientific study of language

Laboratory phonologists are scientists who use laboratory methods to discover and explain the sound structure of human language. Their philosophical stance is generally that of researchers in the mature sciences, such as biology and physics. Specifically, most laboratory phonologists have abandoned the doctrine of dualism. They view language as a phenomenon of nature, albeit a particularly complex one. Language as a cognitive system imputed to individuals is thus to be explained in terms of general facts about the physical world (such as the fact that the resonances of an acoustic tube are determined by its shape); in terms of specific capabilities of the human species that arose through evolution (including both gross anatomical properties, such as the position of the larynx, and neurophysiological properties); and in terms of the interactions of the organism with its environment during development. In this view, social interaction is subsumed under the same umbrella, as a phenomenon of nature. Human societies, like all other mammalian social groups, are natural collections of individuals. And social interactions form part of the natural environment for the species, which influence individual members through natural (physical) mechanisms, such as propagation of sound and light waves, physical contact, and pheromones.

On the basis of this viewpoint, we reject the traditional distinction between knowledge of natural phenomena and knowledge of social conventions (with social

conventions differing from natural phenomena in being arbitrary).[1] We hold that social conventions ARE natural phenomena, so that there is no inconsistency in viewing language both as a social phenomenon and as a cognitive capability of the human species that is instantiated in individuals. Though social conventions vary considerably and surprisingly, so do the phenomena produced by many other physical systems, such as the weather. This does not mean that the variation is unbounded or that no relevant scientific laws can ever be formulated. Tools for building theories of such systems include statistics and stability theory, and we believe that these tools will play a significant role in our future theories of language.

Laboratory phonologists tend to believe that the scientific study of language both should and can progress. One reflection of this expectation is the long citation times for key works, such as Chiba and Kajiyama (1941) for perturbation analysis of vowel formants, and Fant (1960) for the linear acoustic theory of speech production. The idea that science progresses is very controversial in the philosophical literature. We would like to touch on this controversy because the relativists' position in it has been so influential amongst the leaders of generative linguistics. Much work by relativists, such as Kuhn (1962) and Feyerabend (1975), leaves the impression that shifts in scientific thinking are arbitrary outcomes of individual taste and power struggles within the scientific community. Espousal of Kuhnian thought has done much to glamorize conceptual upheavals within linguistics. Pullum (1991) acidly documents a climate in which authors of research papers take no responsibility for either facts or theoretical claims presented in prior work. This situation often provokes indignation amongst phoneticians and psycholinguists, and can lead them to moralistic invocations of work by positivists, such as Carnap, who espouse the traditional ideal of progress in science. However, as Laudan (1996) points out, positivists tend to define progress so narrowly that even the most successful sciences fail to live up to their definitions. For example, the suggestion (by Putnam 1978 and others) that real science is strictly cumulative, with each new framework subsuming all of the successes of its predecessors, would leave humankind with no

[1] The best-known type of arbitrariness in language is de Saussure's *l'arbitraire du signe*, or the apparently arbitrary association of lexemes (word sound patterns) with word meanings. *L'arbitraire du signe* bears some discussion in connection with the point we are making here. Clearly, the association of word forms with word meanings is not determinate; different languages use extremely different lexemes for highly analogous concepts. Even onomatopoeic terms differ across languages. However, de Saussure was incorrect in assuming that any non-determinate relationship is arbitrary. In a stochastic system, non-determinacy still obeys laws, when the probability distributions of outcomes are examined. As online tools begin to make possible large-scale research into lexical structure, we expect that discoveries into the laws of lexeme-meaning associations will become available. For example, Willerman (1994) develops a model of why function words are disproportionately composed of unmarked phonemes in many languages (cf. Swadesh 1971). In a similar vein, we would not be surprised to learn that basic-level categories are typically denoted by shorter words.

extant example of a real science, not even physics or chemistry. Naive positivism is not a useful guide to productive scientific activity.

Our stance on this issue is a highly pragmatic one. Over its history, science has proved successful. A comparison between the state of scientific knowledge now and its state when it was closer to its beginnings (for example, at the time of Roger Bacon in the thirteenth century) reveals overall progress, in terms of the diversity of phenomena for which predictive theories exist, the detail and accuracy of the predictions, and the contributions of scientific knowledge to people's ability to thrive in their environment. Kuhn fails to explain the successes of science, by failing to explain how even two people—let alone humankind in general—can come to an agreement on matters such as the theory of electromagnetism or the germ theory of disease. Recent work in the positivist tradition, such as Quine (1954/1966, 1960, 1961), also fails to account for the evident progress in science, through overemphasis on the logical underdetermination of scientific theories and the elusiveness of the ultimate truth. Therefore, we do not subscribe to either the relativist or the positivist position on science. We are more impressed by more recent work in philosophy of science, such as Hull (1988, 1989) and Laudan (1983, 1996), which treats science as an adaptive human activity. Both of these works reflect intimate familiarity with the everyday conduct of science, and seek to elucidate how scientists actually do cooperate to advance the state of human knowledge despite the logical and social impediments discussed by the relativists and the positivists.

Some of the hallmarks of successful scientific communities that Hull and Laudan discuss are particularly relevant to the laboratory phonology community. One is cooperation within a group of critical size and diversity. Like biological populations, scientific communities atrophy and ultimately fail if they are too small or too homogeneous. Achieving such critical size and diversity was a primary goal of the founders of the Laboratory Phonology conference series. A second hallmark of a successful community is maintenance of a common vocabulary—which can be used by opposing parties in an argument—even at the expense of gradual drift in both the meanings of technical terms and the empirical domain under discussion. As documented in Hull (1989), this was one of the chief reasons for the success of Darwinism over creationism. A third is the existence of "auxiliary theories"—such as theories about how particular instruments work—which are also shared amongst people with different theories or research priorities. The laboratory phonology community has benefited from a plethora of auxiliary theories—covering matters from acoustic transmission to psychological distance, in areas from statistics and probability to physiology and neuroscience—which have permitted substantial agreement on the validity of experimental results and constructive debate about the relationship of these results to theory. Lastly, successful scientific communities recognize the value of mathematical formulation and use mathematics to make precise theoretical predictions. We develop this idea further in the next section.

3.4 Formalism and modeling

Formalizing theories mathematically is a crucial step in making them predictive. The field of mathematics is generally divided into two major areas, discrete mathematics and continuous mathematics. Discrete mathematics includes logic and formal language theory. Continuous mathematics includes calculus. When generative linguistics was launched by Noam Chomsky and his mentor Zellig Harris, it relied exclusively on discrete mathematics. Chomsky is in fact responsible for important results in formal language theory, which are widely applied in computer science. Much of his early work makes natural language seem like computer languages, and poses for natural language the type of questions that arise in designing programming languages, compilers, and other discrete algorithms. The identification of formal linguistics with linguistics formalized by discrete mathematics persists to the present day.

We believe that the identification of formalism with discrete formalism is erroneous and is deeply misleading in its influence on research strategy. The laboratory phonology community uses both discrete mathematics and continuous mathematics. It continually debates and evaluates what type of formalism is most apt and incisive for what types of linguistic phenomena. One reason for this stance is the strong ties of the community to research in speech synthesis. About one third of the authors of papers in the Laboratory Phonology books have worked on speech synthesis systems, and many continue to be active in speech synthesis research. The first speech synthesis was made possible by simultaneous breakthroughs in the acoustic theory of speech production and in the application of formal language theory to phonological description. The acoustic theory of speech production uses Laplace transforms (which belong to continuous mathematics) to model vocal tract transfer functions; Fant (1959) is noteworthy for its elegant discussion of how this particular tool supports deep understanding of the physical situation. The first comprehensive formalization of phonology—using discrete mathematics—is due to Chomsky and Halle (1968), with key concepts already developed in Hockett (1953, 1954) and Chomsky (1964). These two ingredients—a well-behaved characterization of the speech signal and a comprehensive and mathematically coherent system for encoding the phonology—are prerequisites for any viable synthesis system.

Although the synthesis systems just sketched involve a discrete phonology and a continuous acoustic phonetics, subsequent and related work, which we review below, has substantially eroded this division of labor. The relevance of continuous mathematical tools for the classical question of phonology ("What is a possible language sound system?") is shown by work on phonetic grounding of phonology, by work on the role of statistical knowledge in adult phonological competence, and by work on the development of phonology in the child. There are thus both

continuous and discrete aspects to the problems presented by language sound structure, even at the level of phonotactics and morphophonological alternations. We do not understand why most work in generative phonology declines to employ the tools of continuous mathematics.

It is widely recognized in the history and philosophy of science that formalization not only tests and consolidates theories; it also drives empirical exploration. Work on the articulatory and acoustic nature of phonological categories uses a methodology adopted from physics, in which the behavior of the basic equations of the theory is explored with respect to issues such as stability, linearity, invertability, and effects of boundary conditions. This exploration guides the selection of cases to be examined instrumentally. Cases in point include studies of the stability of vowel targets under natural and artificial perturbations (e.g. Lindblom 1963; Lindblom et al. 1979; Maeda 1991; Edwards 1992); explorations of non-linearities in the articulatory-to-acoustics mapping (e.g. Keating 1984; Stevens 1989; Kingston 1990); and explorations of the invertibility of this mapping (e.g. Atal et al. 1978; Badin et al. 1995; Loevenbruck et al. 1999). The collected fruits of this research strategy have supported every one of the many Laboratory Phonology papers that interpret acoustic data or that use speech synthesis to create controlled stimuli.

There has been a similar give-and-take between formal models of the categorical aspects of sound structure, and empirical investigation. Almost all synthesis systems up through the 1980s used the phonological formalization of the *SPE* approach, because it was the only fully formalized model available. Its very exactness made it possible to identify the scientific penalties for ignoring non-local aspects of phonological representation. In the decade after it appeared, evidence about non-local dependencies was provided both by theoretical phonologists working on stress, tone, vowel harmony, and non-concatenative morphology (such as Goldsmith 1976; Liberman and Prince 1977; McCarthy 1985) and by experimentalists working on syllable structure, fundamental frequency, and duration (such as 't Hart and Cohen 1973; Klatt 1976; Bruce 1977; Fujimura and Lovins 1977; Bell and Hooper 1978; K. Harris 1978). This body of evidence in the end led to formal models of "non-linear phonology." Although the formalization of non-linear phonology by linguists was initially sketchy, limitations of the *SPE* approach for morphophonemic parsing and for synthesizing reflexes of prosodic structure and intonation drove efforts for more complete formalization. A formalization of non-linear intonational phonology, with related fundamental frequency synthesis algorithms, was published in Pierrehumbert and Beckman (1988). Additional work on formalizing non-linear phonology for purposes of segmental synthesis was carried out independently by Hertz (1990, 1991) and by Coleman and Local (Coleman 1992, 1994; Coleman and Local 1992). Other work on formalizing non-linear phonology includes Hoeksema (1985), Bird and Klein (1990), Kornai (1991), Scobbie (1991/1999), Bird (1995), and Coleman (1998).

3.5 Methods, frameworks, and issues

The recent history of phonological theory has been marked by the invention of many frameworks, such as Lexical Phonology, Declarative Phonology, Government Phonology, and Optimality Theory. Frameworks are packages of assumptions about the fundamental nature of language, and the research strategy for empirical investigation is driven by top-down reasoning about the consequences of the framework. Frameworks correspond to paradigms in the Kuhnian view of science. One framework can replace another via a paradigm shift, if incorporating responses to successive empirical findings makes the prior framework so elaborate and arcane that a competitor becomes more widely attractive.

In contrast, laboratory phonology is not a framework. As we pointed out in Section 3.2, it is a coalition amongst groups of people, with some working in one or another of the various current frameworks, and others working in no phonological framework at all. As we mentioned in Section 3.3, the Kuhnian view of science is not prevalent among the members of the coalition as a whole, and our own view is that the Kuhnian attitude is at best an unhelpful guide to the conduct of laboratory work. Here we would like to develop some further consequences of this fact for the relationship among methodology, issues, and theories.

When a phonology student first embarks on experimental research, one of the most important lessons to assimilate is the need to operate both below and above the level of abstraction of a typical linguistic framework. On the one hand, the descriptive issues are extremely minute compared to those usually discussed by phonologists working in a particular framework. For example, a phonologist might begin with the observation that English, German, and Polish all exhibit a contrast between voiced and voiceless stops. In a laboratory experiment, the exact extent of the voicing, its statistical variation, and the dependence of these factors on structural position would all be at issue, as may be seen from the example of Keating (1984). An observation made in a few minutes in the field might suggest a hypothesis whose evaluation requires months of work in the laboratory.

On the other hand, almost any substantial fragment of a phonological framework turns out to be too specific and too rich in assumptions to be experimentally tested as such. For example, Feature Geometry packages together at least four assumptions that could in principle stand or fall separately (see Clements 1985; McCarthy 1988). The articulatory characterization, rather than the acoustic or aerodynamic characterization, is implied to be primary. The inventory of relevant articulatory features and feature combinations is held to be finite and universal. The features are held to be organized into a tree (rather than a directed graph or a lattice). Subclassification and markedness are related to a single underlying mechanism. A single suite of laboratory experiments on features could not test all of these specific claims simultaneously. To develop a research program in the general area

of Feature Geometry, the laboratory researcher must instead identify and unbundle the framework's leading ideas.

Similarly, particular proposals about metrical or autosegmental theory, such as Goldsmith (1976), Liberman and Prince (1977), Selkirk (1984), Halle and Vergnaud (1987), and Hayes (1995), all package together many assumptions about the representation of phonological patterns and about the way that phonological representations interact in determining individual outcomes. No one has run experiments designed to test any of these frameworks; it would not be possible to do so. However, a comparison of these five frameworks brings out the fact that they make related, but not identical, claims about the kinds of non-local interactions that are available in natural languages. The interplay of local and non-local factors in speech production and perception is very much amenable to experimental investigation, as shown (for example) by Beckman, Edwards, and Fletcher (1992), Pierrehumbert and Talkin (1992), Choi (1995), and Smith (1995) for production; and Miller and Dexter (1988), Johnson (1990), Huffman (1991), and Ladd et al. (1994) for perception.

Given the rapid pace of change in theoretical linguistics, and the great expense and labor of laboratory research, the shrewd experimentalist will not devote an experiment to even the most central claim of any single linguistic framework. Instead, he or she will look for a topic that represents a source of tension across many frameworks, or that has remained unsolved by traditional methods over many decades.

One class of topics that lend themselves to advances using experimentation are *theoretical issues*. In using this term, we do not mean issues that arise as corollaries of the main assumptions of individual frameworks. Rather, we mean the issues that can be formulated after a deep and sustained effort to compare different frameworks. Issues at this level of abstraction that have been tackled using laboratory methods include: the interaction of local and non-local aspects of the cognitive representation of sound structure (e.g. Bruce 1977, 1990; Kubozono 1992; Coleman 1994; also the references two paragraphs ago to the experimental investigation); the coherence and independence of putative levels of representation (e.g. Lindblom 1963; K. Harris 1978; Rialland 1994); the extent and objective consequences of underspecification (e.g. Pierrehumbert and Beckman 1988; Keating 1990b; Odden 1992; Choi 1995); the relation of qualitative and quantitative aspects of phonological competence (e.g. Keating 1984, 1990a; Pierrehumbert and Beckman 1988). In fact, easily half of the papers in the Laboratory Phonology books have some connection to the issues just listed.

Methodological advances can be just as important as theoretical ones in the progress of science. Established sciences use diverse methods. As pointed out in Laudan (1983), people who disagree theoretically may still share methods. These shared methods are one reason why research paradigms in the established sciences are not as incommensurate as Kuhn claims, and they contribute to the cohesion of research communities that are diverse enough for long-term vitality. In addition,

theories that unify results from many methods are more robust and more predictive, on the average, than those based on fewer methods, much as the five-prong chair base is more stable than the three-legged chair, which is in turn more stable than the one-legged chair. Overcoming the confining reliance of phonological research on the single method of internal reconstruction has been a high priority goal for many laboratory phonologists. Research in this field uses an extreme diversity of methods, including: acoustic analysis of speech productions under various elicitation conditions in the field or the laboratory; judgments and reaction times obtained during identification, discrimination, or prototypicality ratings of natural or synthetic stimuli; direct measurements of articulator movements using electropalatography (EPG), X-ray microbeams, and other recently developed articulatory records; measurements of brain activity; statistical analysis of lexicons; longitudinal analysis of speech produced by children with speech disorders; novel word games; induction of speech errors; priming patterns in lexical decision and other psycholinguistic tasks; patterns of attention in babies.

Related to the idea of a method is the idea of an *auxiliary theory*. Auxiliary theories are established theories, whether broad or modest in scope, to which debate at the forefront of research can uncontroversially refer. Theories of how particular instruments work provide examples. Probably the single most important auxiliary theory in our field is the acoustic theory of speech production. This theory relates critical aspects of speech articulation to eigenvalues of the vocal tract, which can in turn be related to peaks in the spectrum. It is thanks to this theory that two researchers can compare the formant values of the vowels in their experiments, agreeing on observations such as "The /i/ in Swedish is more peripheral in the vowel space than its closest counterpart in English." Such agreement can in turn provide the basis for experimental work directed towards more abstract issues. For example, it provides the basis for current research on the role of general learning mechanisms in phonological acquisition (Kuhl et al. 1992; Guenther and Gjaja 1996; Lacerda 1998; Lotto et al. 2000).

In connection with the goals of the present volume [Burton-Roberts et al. 2000], we would like to point out that auxiliary theories help to provide denotations for phonological terms, along the lines suggested by Kripke (1972) and Putnam (1973) for scientific vocabulary in general. Putnam takes up the issue of the reference of scientific terms in common use, such as "electricity" or "vaccination." As he points out, ordinary people do not in any deep sense understand the reference of these terms; however, the denotations are sufficiently established by access to experts who do have the requisite knowledge that they can also be everyday lay terms. In a similar sense, the denotation of the word "vowel" is provided by the acoustic theory of speech production, and related work on vowel perception and the like. The denotation of the term "articulatory gesture" is provided by the scientific community's present expertise in measuring articulatory events and relating them in a rigorously predictive way to their acoustic consequences. Insofar as we know

the denotation of the term "syllable," it is provided by work such as Bell and Hooper (1978), Derwing (1992), Treiman et al. (2000).

We would also like to adopt from the medical world the concept of a *syndrome*, defined (as in the *OED*) as "a characteristic combination of opinions, behaviors, features, social factors." In the history of the life sciences, discovery of a medical syndrome has repeatedly anticipated and shaped scientific theory by perspicuously uniting facts that point towards deeper conclusions. For example, the documentation of the Broca's and Wernicke's aphasia syndromes led the way towards present neurolinguistic theory.

One of the major contributions of laboratory phonology to the field of phonology has been the careful documentation of syndromes in language sound structure. The diverse and opportunistic methodology of this community has permitted its documentation of syndromes to be both novel and thorough. One type of contribution is that a more accurate documentation of a previously reported syndrome can render moot a theoretical dispute by showing that the supposed facts driving the dispute are not true. For example, armchair impressions about the applicability of the English Rhythm Rule fueled disputes in the various frameworks of metrical theory, such as Liberman and Prince (1977) and Hayes (1984). However, these impressions have been superseded by far more detailed instrumental studies, such as Shattuck-Hufnagel et al. (1994) and Grabe and Warren (1995). These studies both demonstrate that the Rhythm Rule applies in more contexts than reported in the previous phonological literature, and also suggest that the classic cases in English are as much a matter of accent placement as of stress or rhythm as such. This careful documentation of the syndrome at once vitiated Cooper and Eady's (1986) earlier skepticism and allowed laboratory phonologists to isolate those cases in which stress shift might be more purely a matter of rhythm (e.g. Harrington et al. 1998).

Documenting a new syndrome can raise new theoretical issues. For example, Pierrehumbert (1994a), Beckman and Edwards (2000b), Frisch (2000), Treiman et al. (2000), and Hay et al. (2003) all document a syndrome relating lexical statistics, well-formedness judgments (which are opinions), and behaviors on various speech tasks. As discussed in Dell (2000), this syndrome reveals the limitations of an entire class of phonological frameworks, including all standard generative models.

A syndrome that has considerable theoretical importance at the present time is that of the semi-categorical process. Repeatedly, experiments have shown that facultative or phrase-level processes that are transcribed as categorical in the traditional literature actually require continuous mathematics if examined in detail. Browman and Goldstein (1990a) discuss examples in which putatively categorical fast speech rules are shown through X-ray microbeam studies to be cases of gradient gestural overlap. Both Silverman and Pierrehumbert (1990) and Beckman, Edwards, and Fletcher (1992) show that lengthening and tonal realignment at

prosodic boundaries are better handled by a quantitative description than by the phonological beat addition rules proposed in Selkirk (1984). Zsiga (1995) used electropalatagraphic data to show that the palatalization of /s/ in sequences such as *miss you* is not categorical, thereby contrasting with the categorical alternation found in pairs such as *confess*, *confession*. Silva (1992) and Jun (1994b) use acoustic and electroglottographic data to evaluate a post-lexical rule of lenis stop voicing proposed in Cho (1990). They show that apparent voicing at phrase-internal word edges is an artifact of the interaction of independent phonetic factors, which govern the precise timing of the laryngeal features in general.

One way of interpreting such results is as an indication that phonology proper covers less, and phonetic implementation covers more, than traditional approaches supposed. Papers from the first few Laboratory Phonology conferences suggest an implicit consensus in favor of this interpretation. More recently, however, many laboratory phonologists (including us) have begun to interpret these results differently. The steady encroachment of gradience into the traditional domain of phonology raises a number of more fundamental issues: how gradient processes are represented in the mind, how they relate to less gradient processes, whether any processes are truly categorical, and how categoriality—insofar as it exists—actually originates. We take up these issues in the next section.

3.6 Categoriality

Most, though not all, standard phonological frameworks presuppose a modular decomposition of phonology and phonetics in which one module (phonology) is categorical and free of gradient cumulative effects. Thus it is to be formalized using discrete mathematics. The other module (phonetics) has continuous variation, it exhibits gradient cumulative effects, and it is to be formalized using continuous mathematics. The two modules are related by a discrete-to-continuous mapping called the 'phonetic implementation rules'. Pierrehumbert and Beckman (1988) provide a very thorough development of this modular framework for the case of tone and intonation. Pierrehumbert (1994b), in a subsequent reassessment of her earlier stance, assigns it the acronym MESM (Modified Extended Standard Modularization).[2]

The MESMic approach is adopted, in different ways, in at least two papers in the present volume (those by Myers [i.e. Myers 2000] and by Harris and Lindsey

[2] The acronym MESM is an allusion to the syntactic framework of Revised Extended Standard Theory that Chomsky launched (Chomsky 1977) and subsequently abandoned in proposing first Government and Binding theory, and then Minimalism.

[i.e. Harris and Lindsey 2000]), as well as in Bromberger and Halle's earlier (1992) paper. Myers endorses MESM and seeks to develop its typological consequences. Bromberger and Halle (1992) take the categorical entities of phonology to be mental entities, and the continuous spatiotemporal events of phonetics to be in the world. Phonological entities thus denote classes of entities in the world, in the same way that words (such as 'dog') denote classes of physical objects in the world in the extensional treatments of semantics developed by philosophers such as Tarski and Quine. Other work developing the denotional relationship of phonology to phonetics includes Pierrehumbert (1990b) and Coleman (1998). When embedded in this approach, phonetic implementation rules represent an explicit mathematical model of reference, within the limited domain of language sound structure, by encoding the expert scientific understanding of the denotations of the elements of the description. Phonetic implementation rules can seem complicated and elaborate, and many speech researchers have held the hope that the right conceptual framework would render the mapping between phonology and phonetics direct and transparent. But this hope, we would argue, is not well founded. Although the relationship between a sound percept and a phonological category may seem very direct to an individual listener, it still presents to the scientist a dazzling degree of complexity and abstractness. It requires powerful mathematical tools to formalize this relationship.

To appreciate the problems with the assumption that it is possible to define a direct mapping that is somehow simpler or less abstract than phonetic implementation, consider a layman's versus a scientist's understanding of the basic terms of color perception. The percept of "red" or "green" may appear intuitively to be "direct." One might imagine that such color terms correspond directly to particular light spectra. However, detailed experimental studies show that the correspondence is mediated by the exact frequency response of the cone cells in the retina, by the behavior of the optical nerve in integrating responses from cone cells of different types, and by sophisticated higher-level cortical processing that evolved to permit constancy of color percepts under varying conditions of illumination (Thompson et al. 1992). The color terms of specific languages in turn involve a learned categorization of this perceptual space; just as with vowel inventories, this category system is neither arbitrary nor universal (Berlin and Kay 1991; Lucy 1996). A complete scientific model of the meanings of color terms would need to describe the interaction of these factors. The intuitive "directness" of our perceptions does not relate to any particular simplicity in the scientific theory, but rather to the unconscious and automatic character of the neural processing involved.

The modularization of phonetics and phonology that was still assumed by most laboratory phonologists up through the early 1990s is no longer universally accepted, and we ourselves believe that the cutting edge of research has moved beyond it. A series of problems with MESM arises because the two types of representations it employs appear to be completely disparate. The approach thus fails to provide

leverage on central problems of the theory, notably those relating to the *phonetic grounding* of phonology. It has been accepted since Jakobson et al. (1952) that phonological categories are phonetically grounded. However, every effort to detail this grounding comes up against an apparent paradox, arising from the fact that phonological categories are at once natural and language-specific.

Phonological categories are natural in the sense that the actual phonetic denotation of each category shapes its patterning in the sound system. For example, as exhaustively documented by Steriade (1993) and Flemming (1995), neutralization of distinctive pre-nasalization or distinctive voicing typologically affects stops in unreleased positions, where bursts are not available as cues to the nasal contour or the voicing contrast. That is, the phonological rules that affect the stops (or, in a more modern formulation, the positional licensing constraints for the stops) reflect their actual phonetic character. Similarly, high vowels tend to participate in alternations with glides whereas low vowels do not. High vowels have a closer, or more consonant-like, articulation than low vowels and this phonetic property is what exposes them to being contextually interpreted as consonants.

The phonological categories are also natural in the sense that physical non-linearities—in both articulation and acoustics—have the result that phonetics is already quasi-categorical. These non-linearities appear to be exploited as the foundations of phonemic inventories. For specific proposals of this nature, see Stevens (1972, 1989), Browman and Goldstein (1990b), and Kingston and Diehl (1994).

But phonological categories are also language-specific. Despite the similarities of the vocal apparatus across members of the species—and the ability of people of any genetic background to acquire any language—phoneme inventories are different in different languages. It is easy to think of languages that simultaneously display unusual phonemes while lacking certain typologically more typical phonemes. For example, Arabic displays an unusual series of pharyngeal consonants but lacks a /p/. More theoretically trenchant, however, is the fact that analogous phonemes can have different phonological characterizations in different languages. For example, the phoneme /h/ patterns with obstruents in some languages (such as Japanese, where it alternates with geminate /p/ and with /b/), but with sonorants in others. Some languages (such as Taiwanese) treat /l/ as a stop, whereas others (such as English) treat it as a continuant.

Experimental studies also show that there are no two languages in which the implementation of analogous phonemes is exactly the same. When examined in sufficient detail, even the most common and stereotypical phonetic processes are found to differ in their extent, in their timing, and in their segmental and prosodic conditioning.[3] For example, Bradlow (1995) shows that the precise location of Spanish vowels in the acoustic space is different from that for English vowels,

[3] Arguably, there are even no two idiolects in which the implementation of analogous phonemes is exactly the same. Here, however, we emphasize the systematic characteristics which are shared amongst members of a speech community, because these necessarily represent some kind of implicit

even for typologically preferred point vowels. Laeufer (1992) shows that French and English differ in the extent of vowel lengthening before voiced stops (or vowel shortening before voiceless stops). Moreover, though the interaction of the effect with prosodic position is broadly similar for the two languages, there are also differences in detail relating to the allophonic treatment of syllable-final obstruents. Zsiga (2000) demonstrates a difference between Russian and English in the extent of subcategorical palatal coarticulation across word boundaries. Caramazza and Yeni-Komshian (1974) demonstrate that Québecois and European French differ not only in the well-known assibilation of /d/ and /t/ before high vowels, but also in the modal VOT values of all voiced versus voiceless stops, including the dentals before non-high vowels. Hyman (2001) discusses the strong tendency for a nasal to induce voicing of a following oral stop closure in nasal contour segments and in nasal-stop sequences (cf. Maddieson and Ladefoged 1993), but shows that, despite this tendency, some languages instead devoice stops after nasals.

Results such as these make it impossible to equate phonological inventories across languages; there is no known case of two corresponding phonemes in two languages having fully comparable denotations. Therefore phonological inventories only exhibit strong analogies. In fact, we would argue that there is no symbolic representation of sound structure whose elements can be equated across languages; the overwhelming body of experimental evidence argues against anything like Chomsky and Halle's (1968) phonological surface representation. In Chomsky and Halle (1968) and more recent work such as Chomsky (1993), Chomsky and Lasnik (1995), and Chomsky (1998), this representation (now known as 'PF' for 'Phonetic Form') is conceived of as symbolic, universal, and supporting a uniform interface to the sensorimotor system (Chomsky 1995: 21). Similar criticisms apply to the IPA if this is taken to be a technically valid level of representation in a scientific model (rather than the useful method of note-taking and indexing that it most assuredly is). The theoretical entities that can be absolutely equated across languages are the continuous dimensions of articulatory control and perceptual contrast. Languages differ in how they bundle and divide the space made available by these dimensions.

In view of such results, what is the character of the "implicit knowledge" that the linguist imputes to the minds of individual speakers in order to explain their productive use of language? Obviously, anything that is language-particular must be learned and thus represents implicit knowledge of some kind. Since languages can differ in arbitrarily fine phonetic detail, at least some of this knowledge is intrinsically quantitative. This should not come as a shock, since learned analog representations are known to exist in any case in the area of motor control (e.g. Bullock and Grossberg 1988; Saltzman and Munhall 1989; Bailly et al. 1991).

knowledge that emerges during language acquisition. Idiolectal differences could result from idiosyncratic anatomical or neural properties.

Although MESM asserts that the relationship of quantitative to qualitative knowledge is modular, this assertion is problematic because it forces us to draw the line somewhere between the two modules. Unfortunately, there is no place that the line can be cogently drawn. On the one hand there is increasing evidence that redundant phonetic detail figures in the lexical representations of words and morphemes (see Fougeron and Steriade 1997 on French schwa; Bybee 2000 on word-specific lenition rates; Frisch 1996 on phonotactics). Thus phonology has a distinctly phonetic flavor. But, on the other hand, the detailed phonetic knowledge represents the result of learning, and therefore has a distinctly phonological flavor. Also non-linearities in the domains of articulation, acoustics, and aerodynamics mean that even the physical speech signal already has a certain categorical nature.

In short, knowledge of sound structure appears to be spread along a continuum. Fine-grained knowledge of continuous variation tends to lie at the phonetic end. Knowledge of lexical contrasts and alternations tend to be more granular. However, the sources of categoriality cannot be understood if these tendencies are simply assumed as axiomatic in the definitions of the encapsulated models, as in MESM. A more pragmatic scientific approach is to make the factors that promote categoriality a proper object of study in their own right, without abandoning the insight that lexical contrasts and morphological alternations are more granular than phonetics alone requires. One way to do this is to view the discrete (or quasi-discrete) aspects of phonology as embedded in a continuous description, arising from cognitive processes that establish preferred regions in the continuous space and that maximize the sharpness and distinctness of these regions. That is, instead of viewing the discreteness of phonology as simply *sui generis*, we view it as a mathematical limit under the varied forces that drive discretization. The complexity of phonological categories can then be appreciated as fully as we appreciate the complexity of color perception.

Some specific factors contributing to discretization are already under active exploration. First, there is the idea that phonology prefers to exploit non-linearities in the physical system; the nature of the preference is however controversial. Stevens (1989) proposes that languages prefer vowels whose acoustics remain stable under small changes in articulation; Lindblom and his colleagues (Liljencrants and Lindblom 1972; Lindblom et al. 1984) hold, in contrast, that languages prefer vowel systems for which minimal articulatory effort produces maximal contrasts. Similarly, Pisoni (1977) argues that the preference for voiceless stops effectively exploits psychoacoustic non-linearities that render the stop bursts both objectively distinctive and psychologically salient; Summerfield (1981) and others, by contrast, point to boundary shifts with place of articulation, as well as to the attested integration of the Voice Onset Time (VOT) cue and the F1 cutback cue, as evidence for language-specific articulatory habits as the source of the discretization of the VOT continuum. (See Benkí 1998 for a recent review of these two

opposing views, and Damper 1998 for new evidence on the role of psychoacoustic non-linearities.)

Second, the use of speech sounds to contrast meanings requires that the sounds be robustly discriminable. This factor does not define any single region of the phonetic space as preferred, but it tends to push apart preferred regions in relation to each other. Results related to this factor include the finding by Johnson, Flemming, and Wright (1993) that the "best" vowels are more extreme than the most typical vowels, and a substantial body of work by Lindblom and colleagues on deriving vowel inventories from considerations of contrastiveness (see Lindblom 1992 for a summary review of successive refinements to the original "dispersion" model over the last two decades).

Third, connectionist modeling demonstrates the generic tendency of neural networks to warp the parameter space that is being encoded. Guenther and Gjaja (1996) show that when a neural network is trained on steady-state vowel tokens selected from Gaussian distributions centered on the average F1/F2 values for a language's distinct vowel categories, a language-specific warping of the F1/F2 space occurs in the perceptual map even with unsupervised learning—that is, even when the vowel categories are not provided as the output nodes in training and testing. Makashay and Johnson (1998) show that when this sort of network is trained on a more natural distribution of tokens (that is, steady-state vowels that reflect normal inter-gender variability), there is less clear convergence to vowel "prototypes"; however, distinct vowel categories re-emerge if f0 is included in the parameters of the space, to allow the model (in effect) to correlate inter-token variability with speaker identity. Damper and Harnad (2000) show related results for neural network modeling of VOT categories. They demonstrate that the sharp S-shaped boundary that is a hallmark of classical "categorical perception" is exhibited by a broad class of connectionist models, when the model is trained on tokens that cluster around the endpoints of the continuum. However, as Damper (1998) shows, the input to the model must be spectra that have been passed through an auditory front-end in order for the boundary to shift with place instead of falling at the center of the continuum (as predicted for perceptual learning in general by Macmillan et al. 1987).

Last, we may consider issues of cognitive complexity. Lexical contrasts and morphological alternations involve knowledge not of sounds alone, but of the relationship between sounds and meanings in the lexicon. As discussed in Werker and Stager (2000), children begin to master the association between word form and lemma at about 14 months by manipulating extremely coarse-grained phonetic contrasts. This is so despite their exquisite sensitivity to speech sounds as such, and despite a pattern of response to fine phonetic detail that is already language-specific at 11 months, as demonstrated by Werker and Tees (1984a), among others. Given the amount of neural circuitry that must be established to encode the relationships between word forms and word meanings, there may be limits on the ultimate extent

of phonological differentiation possible. (See Beckman and Pierrehumbert 2003 for further development of these ideas.)

3.7 Competence and performance

In the previous section, we developed a picture of implicit knowledge of sound structure that marks a significant departure from the most phonetically sophisticated generative model—namely, MESM. This picture has important consequences for the understanding of linguistic competence and the competence/performance distinction. The following quote from Chomsky (1995: 14) may serve to introduce our discussion of this issue:

> We distinguish between Jones's competence (knowledge and understanding) and his performance (what he does with the knowledge and understanding). The steady state constitutes Jones's mature linguistic competence.
>
> A salient property of the steady state is that it permits infinite use of finite means, to borrow Wilhelm von Humboldt's aphorism. A particular choice of finite means is a particular language, taking a language to be a way to speak and understand, in a traditional formulation. Jones's competence is constituted by the particular system of finite means he has acquired.

We find much to agree with in this quotation. Language does put finite means to infinite use. To explain the diverse and productive linguistic behavior that people exhibit, we impute abstract, implicit, and synoptic knowledge of language to individuals. The ability to acquire and apply such knowledge is a hallmark of the human species. However, the concept of linguistic competence carries with it in the generative literature a number of further axiomatic assumptions to which we take strong exception.

One assumption concerns the relationship of the various types of data gathered by linguists to theories of linguistic competence. Much of the generative literature assumes that well-formedness judgments provide the most direct and revealing data about competence, with other types of data presenting difficulties of interpretation that compromise their relevance. This assumption is articulated particularly clearly in an essay by Soames (1984), who undertakes to define linguistics proper in an *a priori* fashion on the basis of the data it deals with. However, studies in the sociolinguistics and psycholinguistics literature (e.g. Labov 1973; Bard et al. 1996) cast serious doubt on the reliability and predictiveness of well-formedness judgments. Well-formedness judgments are opinions. They are high-level metalinguistic performances that are highly malleable. They do not represent any kind of direct tap into competence, but are rather prone to many types of artifacts, such as

social expectations, experimenter bias, response bias, and undersampling. Hence, well-formedness judgments are just one type of evidence among many, and not a particularly good type of evidence as currently used (see the constructive criticisms of Bard et al. 1996).

All data about language come from performance, and all present difficulties of interpretation relating to the nature and context of the performance. Like scientists in other fields, we must assess the weight to assign to various types of data; statistics provides one tool for making such an assessment. But no matter how we weight the data, we must acknowledge that all data ultimately originate in performance. The notion that some data represent 'mere performance' does not in itself constitute sufficient grounds for discarding data.

A second assumption involves universals. Discussion in Chomsky (1995) articulates his conception of linguistic competence in terms of a UG: UG provides an overarching description of what all mature human languages have in common; simultaneously, it is claimed to describe the initial state of the child who embarks on language acquisition. This dual characterization of UG forces the view that language acquisition is a process of logical instantiation. UG provides logical schemata that describe all languages, and the child, armed with the schemata, instantiates the variables they contain so as to achieve a grammar of a particular language.

This understanding of UG is not logically necessary, nor is it supported by the available results on acquisition of phonology. At its root is the assumption that to achieve a formal model of language, the model must be formalized using the resources of logic. However, it is clear that phonetics must be formalized using continuous mathematics, and the experimental literature on phonological development makes it clear also that phonological knowledge depends in an inextricable fashion on phonetic skills, including the gradual acquisition of spatial and temporal resolution and coordination (see e.g. Elbers and Wijnen 1992; Locke and Pearson 1992; Edwards et al. 1999). As speakers acquire more practice with a category, the variance in their productions of the category gradually reduces, and this process continues well into late childhood (Lee et al. 1999). When children are first acquiring a phonological contrast, they often fail to reproduce an adult-like phonetic expression of the contrast. For example, Finnish children often produce disproportionately long geminate consonants. When children are acquiring the American English or Taiwanese Chinese contrast between aspirated and unaspirated initial stops, the VOT values for the aspirated stops contrast may be exaggerated, or they may be so small as to appear to fall into the unaspirated category (Macken and Barton 1980; Pan 1994). Similarly, an adult-like control of the spectrum of /s/ that differentiates it robustly from both /θ/ and /ʃ/ in English may not be achieved until five years of age, or even later in children with phonological disorder (Baum and McNutt 1990; Nittrouer 1995). As discussed in Scobbie et al. (2000), the trajectory from insufficient (or 'covert') phonetic contrasts to robust mature contrasts is a gradual

one. Hence it cannot be modeled as a process of logical instantiation, but only using statistics over a continuous space.

Additional patterns in acquisition that demand a statistical treatment are provided by investigations of babbling and early word productions, as well as by patterns of perceived substitution in children with and without phonological disorder. For example, vowel qualities in the earliest stages of variegated babbling show the impact of the frequencies of different vowels in the vowel space of the ambient adult language (de Boysson-Bardies et al. 1989). Consonants in later stages of variegated babbling that are concurrent with the acquisition of the first twenty-five words in production reflect cross-language differences in the relative frequencies of different places and manners of articulation (de Boysson-Bardies and Vihman 1991). Also, coronals are more frequent than either labials or dorsals in both English and Swedish, and children acquiring these languages already show language-specific differences in the fine acoustic details of coronal stops by the age of 30 months (Stoel-Gammon et al. 1994). This is so even though they may not yet have learned to robustly differentiate the spectra for dorsal place from coronal place of contact, making /t/ for /k/ one of the most commonly perceived substitutions in English-acquiring children (Edwards et al. 1997). Finally, although infants at the reduplicated babbling stages universally produce multisyllabic productions with simple CV alternations, children acquiring English (but not those acquiring French) show a marked increase in monosyllabic babbles, and in babbling productions ending with consonants, beginning at the first word stages (Vihman 1993). This difference reflects the predominant shapes of the most frequent words in the two languages.

In connection with these observations, we would reiterate our opposition to dualism. A mature language is instantiated in individual brains. The physical state of these brains represents an equilibrium state that is reached from an initial condition—the human genetic endowment—through interactions with the physical environment. For physical systems in general, it is a conceptual error to equate the initial conditions with generalizations over the equilibrium states that may evolve from these conditions. For example, the current state of our solar system (with nine planets moving nearly on the same plane on elliptical orbits around the sun) is an equilibrium state. In so far as this solar system is typical—with its sun, its small number of discrete planets, and its orbital plane—one might imagine a kind of "meta-grammar" of equilibrium states of the form:

(1) Solar system –> Sun, planet+

With a binding condition for orbital planes:

(2) For all i, Plane(planet[i]) = Plane(planet[i + 1])

However, the initial condition for our solar system was an unformed cloud of debris containing a mixture of heavy elements from a previous supernova explosion.

Neither (1) and (2) nor any discrete abstraction of them sensibly describes an unformed cloud of debris; nor is the current state sensibly viewed as the logical instantiation of the parameters of such a cloud. Describing how the planets arose from the debris requires gravitational field theory. That is, the discreteness of our own solar system does not arise from logical instantiation of the discrete elements of a meta-grammar. Instead, it arises as the discrete limit of continuous processes, much as we have shown for the case of phonological acquisition.

A third objection to Chomsky's conception of competence is its continued reliance on the assumption of an idealized uniform speaker-hearer community. According to Chomsky, this idealization is justified by the obvious absurdity of imagining that language acquisition would proceed better in a varied speech community than in a uniform community. However, there is much evidence that uniformity impedes the process of language acquisition, and that variability facilitates it, yielding exactly the result that Chomsky believes to be absurd. This evidence comes from several areas of research. Experiments on second-language learning show that learners who are exposed to varied examples of a phonemic category learn the category better than those who are exposed repeatedly to the same example (Logan et al. 1991). The variation in examples permits the learners to generalize to new cases and to transfer perceptual learning to production (Bradlow et al. 1997). Research on first-language acquisition of affixal categories similarly points to the role of variability in the morphological context—for example, the role of exposure to a sufficient number of different roots before the affix can be abstracted away as a productive independent morpheme. Thus, for the English past-tense affix, Marchman and Bates (1994) show that (contra the model and claims of Pinker and Prince 1988), the single best predictor of when over-regularized past-tense forms begin to appear is the number of different verbs that the child has acquired. That is, acquiring a large variety of regular past-tense verb forms permits the child to project the principles of regular past-tense formation, overpowering the high token frequency of some irregular verbs. Derwing and Baker (1980) similarly show that the syllabic plural allomorph is acquired later than the two consonantal allomorphs, in keeping with its lower type frequency.

Such results gain an intuitive interpretation when one reflects that *variability causes the need for abstraction*. The entire point of an abstraction such as the morpheme *-ed* or the phoneme /i/ is that it represents the same thing across differences in the root to which it is affixed or in the speaker's larynx size and vocal tract length, the speech style and effort of articulation, the segmental and prosodic context, and other kinds of systematic token-to-token variability. If these sources of variability did not exist, then lexical items could be encoded directly in terms of invariant phonetic templates. Abstractions are cognitively expensive. They are learned because variability makes them necessary. There is no reason why they should be learned in the absence of variability.

Laboratory phonologists share with other phonologists the aim of developing an explanatory theory of language. Overall, the issue is where the deep structural regularities of language come from. Work in the Chomskian tradition has emphasized the possibility that humans have a genetically innate predisposition to language, which is manifested through logical instantiation of the universal schemata of UG. However, there are also a number of other potential sources of deep, abstract, and universal characteristics of language. These include necessary or optimal properties of communication systems as such (as explored by Wiener 1948 in his work on cybernetics; also much subsequent work in information theory); objective consequences of the characteristics of the human vocal and auditory apparatus; and general cognitive factors (such as general facts about categorization, memory, and temporal processing). For the laboratory phonology community as a whole the interplay amongst these various possible factors is treated as an open question.

APPENDIX

The fundamental similarity between the PF representation of current Minimalist theory and the surface phonological representation of Chomsky and Halle (1968) can be deduced from quotations such as the following:

> Let us recall again the minimalist assumptions that I am conjecturing can be upheld: all conditions are interface conditions; and a linguistic expression is the optimal realization of such interface conditions. Let us consider these notions more closely.
> Consider a representation π at PF. PF [*sic*] is a representation in universal phonetics, with no indication of syntactic elements or relations among them. . . To be interpreted by the performance systems A[rticulatory]-P[erceptual], π must be constituted entirely of *legitimate PF objects*, that is elements that have a uniform language-independent interpretation at the interface [to the articulatory-perceptual system]. . .
>
> To make ideas concrete, we must spell out explicitly what are the legitimate objects at PF and LF. At PF, this is the standard problem of universal phonetics.
>
> (Chomsky 1993: 26–7; emphasis in the original)

This characterization of PF involves objects that are categorical and that support a universal phonetic interpretation. These assumptions are critical to some work in the Minimalist framework, such as Halle and Marantz's (1993) theory of Distributed Morphology (DM). DM claims that the PF level is the result of instantiating the lexical items in the morphological representation with phonological segments and features that can be manipulated by categorical rules and constraints. Thus, it presupposes the modular division between a language-specific categorical component and a universal quantitative phonetics that is clearly non-viable.

It is possible, however, to read much of the Minimalist literature in a different light—as an abdication of Chomsky and Halle's original claim that sound structure

as such has a "grammar," in the sense of an abstract computational system that is capable of generating novel forms. As Jackendoff (1997: 15) points out, in the Minimalist program "the fundamental generative component of the computational system is the syntactic component; the phonological and semantic components are 'interpretive' " (see also the discussion in Burton-Roberts, this volume [Burton-Roberts 2000]). One almost might interpret this research program as acknowledgment in advance by its proponents of some of the problems we raise regarding efforts to explain implicit knowledge of sound structure in terms of a categorical phonological module. At the same time, the Minimalist Program appears to make no pretense that its key concepts (such as grammaticality, UG, or linguistic competence) in any way pertain to language sound structure, and we are possibly being unfair in attacking these concepts as if they were claimed to pertain. However, this interpretation strikes us as regrettable, for many reasons.

First, it leads one to disregard the ways in which phonology and phonetics are grammar-like, enabling the speaker to create morphological neologisms, to make additions to the lexicon, and to produce regular allophonic patterns when saying novel phrases and sentences. To the extent that there are abstract parallels in sound structure across languages, these suggest the kind of deep universals that are the traditional target of linguistic theory. Even if these quasi-grammatical properties of phonology are embedded in an understanding of the physical world and of general cognitive capabilities, they are still scientifically important and tell us something about the human capacity for language.

Second, it leads one to disregard the ways in which morphological and syntactic relationships are echoed in quantitative effects in the phonetics (e.g. Sereno and Jongman 1995; Fougeron and Steriade 1997; Hay et al. 2003), which surely are the reflexes of the fact that phonetic knowledge is intertwined with the linguistic system rather than being decoupled from it.

Third, the interpretation undermines the effort to find parallels between phonology and syntax in the way that they relate to physical events in the world and to the language user's conceptualization of these events. It may turn out that, thanks to its restricted physical domain and advanced instrumentation, phonology is simply in the lead in an enterprise in which syntax will eventually catch up. If the relationship of syntax to this "world understanding" is eventually proven to resemble that of phonology (as we have described it here), then the Minimalist Program will have been carried through to its logical—truly "minimalist"—conclusion.

This page intentionally left blank

PART II

NATURE AND TYPES OF VARIATION: THEIR INTERPRETATION WITHIN A LABORATORY PHONOLOGY PERSPECTIVE

The goal of this part is to review sources of variation in speech, including speaker-related, message-related, and system-related variation, to understand what they tell us about phonological questions concerning representations and processes, phonological systems, phonetic correlates of abstract representations, the richness of stored information, etc. Together the totality of these data highlights the complexity and multifaceted nature of human speech and the need for integrated methodologies and models.

This page intentionally left blank

CHAPTER 4

SPEAKER-RELATED VARIATION– SOCIOPHONETIC FACTORS

GERARD DOCHERTY AND
NORMA MENDOZA-DENTON

In this chapter the authors provide a concise and rich review of the sociolinguistic literature on variation, including developments in the interpretation of such variation and the methods used to study it. It is argued that the concerns of sociolinguists and laboratory phonologists are increasingly converging on a deeper appreciation of the ways that "the social-indexical channel is embedded within speech processing and representation" (p. 56).

4.1 Introduction

Since building up momentum in the late 1980s, laboratory phonology has stood not for a unitary theoretical position, but rather for an approach which draws together a diverse group of scholars united in the belief that understanding of spoken communication can best be developed by integrating methods and concepts from research traditions (e.g. phonetics, phonology, psycholinguistics) which in the past have not always been effectively joined up (Cohn 2010; Pierrehumbert and Clopper 2010). A perusal of the LabPhon volumes over two decades provides ample evidence of the many facets of the laboratory phonology enterprise and highlights the defining theoretical questions which have driven the integration of methods and concepts characterizing its development. These include, for example, the nature of lexical and phonological representation, the extent to which representation of sounds and words in memory is governed by/reflects what we know about the processes of speech production, perception, and acquisition, and how much of what is observed empirically can be accounted for by general principles of cognition, motor control, physiology (and what aspects require some special modality-specific explanatory framework).

A common thread through work presented under the laboratory phonology banner is that advances in theory have been rigorously grounded on quantitative analysis of the performance of individuals, either as speakers or as listeners. However, it is only in the latter years of the development of laboratory phonology that members of this community have paid significant attention to the fact that the vast majority of utterances produced naturally by speakers and processed by listeners are situated in an interactional context in which the substance of speech is shaped by the social factors playing out within that interaction as well as by the propositional content transmitted between the interlocutors (see overview contributions by Local 2003; Docherty 2007a; Mendoza-Denton 2007; Huffman 2007; Foulkes 2010; Munson 2010; Foulkes et al. 2010).

The integration of a sociophonetic dimension into the laboratory phonology "project" raises a profound (but, as we suggest below, not irresolvable) tension. In general, the more we discover about socially-situated speech, the more we are confronted with the central role that the social-indexical channel plays in the natural performance of speakers/listeners, but equally, the more conspicuous becomes the absence of an account of how this channel of information is integrated into speech alongside the lexical-propositional channel which, to date, has been predominant in the development of theoretical stances around speech production, perception, and learning (though see McMurray and Farris-Trimble, this volume). This chapter presents a critical overview of these issues, highlighting some of the key ways in which social factors impact on the performance of speakers and listeners, and then

reviewing how this social-indexical dimension is starting to shape thinking within the laboratory phonology community.

4.2 Speaking the same language

One of the symptoms of the long-standing disconnect between investigators working on socially correlated aspects of phonological variation and those from within the subject areas drawn together under the laboratory phonology banner is that a good deal of what we know about speech as a social phenomenon has been developed with a different frame of reference, terminology, and analytic methodology from that which applies more generally within the laboratory phonology community. One example of this is the predominant role played by Varbrul analysis in sociolinguistic studies of phonological variation (Sankoff and Labov 1979; Sankoff 1988; and critically evaluated by Mendoza-Denton et al. 2003; Pierrehumbert 2006a; Johnson 2009; and Coetzee, this volume). But perhaps the clearest example of this is the notion of the *phonological variable* which is deployed as an analytic tool within a great deal of sociolinguistic research (see Chambers 1995 and Milroy and Gordon 2003 for a thorough evaluation of the application of this method). In the sociolinguistic analysis of phonological variation and change, phonological variables are segmental loci of socially structured variability, broadly equating to a phonemic level of abstraction (and in the case of vowels defined by Labov 2001: xvii as "abstract phonological elements that define historical word classes"). Unlike a conventional analysis of allophonic realization which targets what is hypothesized to be the "same" phonemic element across different contexts (e.g. word-initial vs. word-final), in the sociolinguistic analysis of a phonological variable the aim is to systematically track within- and across-speaker variability in a single context with a view to identifying the extent to which such variability is governed by diverse social factors. Examples of variables which have been the focus of relatively recent studies of English are (t)[1] (e.g. Docherty et al. 1997), (ing) (Labov 2001), (th) (Stuart-Smith and Timmins 2006), and a range of vowel variables including (ay), (aw), and (aeh) (Labov 2001).

[1] Note the use of parentheses to denote a phonological variable within sociolinguistic research; although this usage is not consistently applied. For vowel variables there is a difference between investigators such as Labov (1994, 2001) who do use this notation, and others who refer to vocalic variables by using the "lexical sets" proposed by Wells (1982); thus Labov (1994) refers to (ay), whereas Kerswill et al. (2008) refer to the PRICE lexical set.

In order to track the variants of a phonological variable, in many cases the analysis proceeds by scoring the occurrence of a set of auditorily identified variants; for example, in investigations of -t/-d deletion, investigators typically track the presence/absence of the plosive; for (ing)—whether the nasal is alveolar or velar; for (th)—whether the fricative is dental or bilabial. In many instances, and particularly for consonantal variables, investigators approach this analysis task with a preconceived set of variants (based on previous studies or from a pilot investigation), but it is incumbent on them to identify every variant which is encountered in a particular context including those which are unexpected or which occur with only very low frequency. (For example Docherty et al. 1997 unexpectedly found voiced [t$_{\text{v}}$] tokens in their study of (t) in Tyneside English, and for some speakers this turned out to be a salient characteristic of their realization of (t).) While the use of acoustic methods to analyze consonantal variables has increased over recent years (Docherty and Foulkes 1999; Stuart-Smith 2007b; Foulkes et al. 2010), they have for many years been established as the conventional method for analyzing vocalic variables (at least as far back as Labov et al. 1972). Typically, vowel variants are plotted in F1/F2 space usually following some form of normalization in order to minimize the risk of any cross-speaker differences being attributed sociolinguistic significance when they might simply arise from differences in vocal tract length. The relative merits of different types of normalization for tracking sociophonetic variability in vowel production are amply discussed by Labov (2001, 2006), Adank et al. (2004), and Watt and Fabricius (2002). On the whole, consonantal variables have tended to be looked at independently of one another, whereas for vowels there has been a greater attempt to consider a number of variables alongside one another in order to identify any mutual interaction (driven largely by the hypothesis that variation and change in one part of the vowel space can give rise to a chain reaction of shifts in vowel quality for particular lexical sets, a view which is strongly encapsulated in Labov's extensive analyses of vowel chain-shifts; 1994, 2001).

This overall methodology is well established and has undoubtedly enabled many fundamental insights into socially correlated phonological variability in speech performance. However, when viewed in light of what we know from experimental phonetic research about the properties of speech production and perception and how these shape phonological systems, there are dimensions of this work which are potentially problematic. One salient issue relates to whether, when investigators track discrete consonantal variants of a particular variable, the actual speech behavior being tracked is really as discrete as the analyst-imposed categories being deployed in the analysis. For example, in British English there are a number of accounts of the variable realization of (r) as an alveolar or labial variant, but acoustic analysis of these various realizations (e.g. Foulkes and Docherty 2000) suggests that the variability is best captured by considering any particular token to be positioned in a continuous (acoustic, and therefore articulatory) space between [ɹ] and [ʋ]; meaning that it would be a simplification to base any theoretical development on

this variable if it was being handled simply as having binary realizations.[2] This issue of the status of discrete segmental categories is of course a very familiar one to members of the laboratory phonology community, having been a key focus of work carried out under the rubric of Articulatory Phonology (Browman and Goldstein 1986; Pierrehumbert and Talkin 1992), and it represents an important source of tension between the perspectives on variation in phonetic realization provided by the fields of study being considered here.

A second issue is that the large volume of studies of socially correlated phonological variation is heavily skewed towards the analysis of one language (English), and recurrently focuses on a subset of variables, partly as a consequence of the adoption of the phonological variable methodology (leading effectively to a focus on segmental variation), and partly arising from the fact that certain variables (e.g. -t/-d deletion, (ing), and certain vowel variables) have been recurrently investigated in order to test particular hypotheses regarding variation and change, such as the notion of variable rules or vowel chain-shifting. It is true that particularly in recent years there have been studies which point to a wider range of phonetic parameters taking on a social-marking role (e.g. Esling 1978; Stuart-Smith 1999 on voice quality; Daly and Warren 2001 on the realization of f0 contours), but, in contrast to the understanding built up over many decades of the phonetic parameters deployed across languages in support of lexical-phonological contrast (e.g. Ladefoged and Maddieson 1996), we are still a long way short of a similar understanding of the cross-language range of phonetic parameters associated with the social-indexical channel.

A third issue relates to the analysis of vowel variables. While it is positive that acoustic methods are the standard in sociolinguistic studies of vowels (including for some investigators normalization into auditory space in order to gauge the perceptual relevance of differentiation in vowel realizations—e.g. Warren et al. 2007), there is some ambiguity about the extent to which many investigators equate the two-dimensional space within which tokens of vowels are typically plotted and compared with the multi-dimensional articulatory space within which vowels are articulated. Thus, a set of vowels which is distributed such that F2 is higher than another set will often be referred to as more "fronted," and likewise if the difference is an overall lower F1, the vowels will said to be "raised." These terms may well simply be serving as a means of capturing relative positioning within acoustic/auditory vowel space, but what they plainly cannot do is to reflect the complex relationship between articulatory and acoustic properties of vowels, and the danger is that they are interpreted as relating to the latter when in fact they can only reliably relate to the former. For example, realizational variants in vowel quality are only relatively

[2] Note too that discrete variables are the only type which investigators have been able to accommodate within the predominant analytic tool Varbrul, although this constraint has recently been loosened through Johnson's (2009) recent work on the development of R-Brul.

rarely attributed to different degrees of lip-rounding/protrusion, even though this articulatory parameter can have a major influence on formant frequencies, and does indeed participate in socially correlated variation (see Kerswill et al.'s 2008 analysis of the fronting and loss of rounding on GOAT vowels in British English). As pointed out by Foulkes et al. (2010), the focus on F1/F2 space has also drawn attention away from interesting sociophonetic differences in duration (e.g. Scobbie et al.'s 1999 work on the Scottish Vowel Lengthening Rule), formant dynamics, and contributions to perceived vowel quality made by F3 and the higher formants.

As indicated above, sociolinguistic analysis of phonological variables attempts to minimize positionally generated variation in order to capture significant inter-/intra-speaker variation in the same context. However, a further area in need of elaboration is the extent to which this form of analysis of realization variants is sufficiently sensitive to the range of phrasal and other prosodic features which work in the laboratory phonology community has shown to be closely associated with the magnitude and timing of articulatory gestures (e.g. Keating et al. 2003; Cho and McQueen 2005; Keating 2006). In general, the analysis of socially correlated variation has not controlled for factors such as prosodic constituency, the structure of conversational interaction, or speech rate. That factors such as these may well be important in tracking socially correlated variation is evident in studies such as Docherty et al. (1997) who found significant differences in word-final (t) realization depending on whether the token was in pre-pausal position or not; Local (2003), who described how phonetic detail can be used to denote key landmarks within a conversational interaction (such as turn transitions or conversational repair); and Docherty's (2007b) finding that speech rate influences cross-speaker variability in the realization of (t), even in a fairly formal reading style.

The areas just identified as potentially problematic with respect to conventional approaches to the analysis of socially correlated realization variation suggest that there is much to be gained from the greater methodological and theoretical refinement which would be engendered by a more productive dialogue across the laboratory phonology and sociolinguistics communities, and indeed, this process is well under way. (See, for example, the 2006 *Journal of Phonetics* special issue on Modelling Sociophonetic Variation, and the thematic orientation of the eleventh LabPhon Conference towards "Social information in the lexicon," dealing with questions such as "Is phonetic information in the lexicon accompanied by social information?", "How do social expectations about a speaker affect speech perception?", "Is speaker-specific detail stored in the lexicon?").

Thus, notwithstanding the need for this dialogue to develop much further in order to refine our understanding of sociophonetic variation, there is now a clear recognition of the significance of the social-indexical channel for speakers/listeners and for models of how speakers and listeners plan and execute their participation in spoken communication. The present discussion now moves on to explore some of

the key features of sociophonetic variability and its key points of contact to debates within laboratory phonology.

4.3 Social-indexical variation

The history of quantitative sociolinguistics is largely an attempt to understand how patterns characterizing a speech community emerge from and relate to individuals' linguistic production and perception, specifically in relation to phonological and other linguistic variables. A classic definition of the speech community reads as follows:

> The speech community has been defined as an aggregate of speakers who share a set of norms for the interpretation of language, as reflected in their treatment of linguistic variables: patterns of social stratification, style-shifting, and subjective evaluations. This orderly heterogeneity normally rests on a uniform structural base: the underlying phrase structure, the grammatical categories, the inventory of phonemes, and the distribution of that inventory in the lexicon. (Labov 1989a: 2)

This definition encompasses many of the defining characteristics and central assumptions of mainstream sociolinguistic research. Historically, the speech community was defined primarily by shared patterns of subjective evaluation (Labov 1972a), then redefined as "sharing a set of norms," which has largely been interpreted as *using* variables in similar ways rather than just *assessing* them in the same way—thereby eliding some of the complexity of production-perception relationships (Keating 1987; Johnson, Flemming, and Wright 1993; Liberman and Whalen 2000). Key in the sharing of norms are consistent patterns of social stratification, meaning that all segments of the population evaluate and use a particular form as more prestigious than another; methodologically, this entails that a community (already aggregated) must ordinarily be stratified by class, age, ethnicity, gender, etc., in order to be studied. In the above definition, style-shifting also emerges as central, albeit with a definition of style that posits a continuum between formal (word list) and informal (conversational) styles based on attention paid to speech: the more attention is paid, the more formal speech becomes (Chambers 1995). Making sense of "orderly heterogeneity" then becomes *the* puzzle in itself, since it is assumed that any stratified population replicates the history of language change and carries within it the seeds of further development. There is also an assumption that underneath the heterogeneity, the population is quite uniform and shares a "grammar," in the classic, generative sense of the term. Accordingly, if they didn't share a grammar (or a phonological inventory, or a set of evaluations) they would be a different speech community altogether. This in turn leads to claims that an individual's patterns of variation mirror those of her community grammar

in terms of the statistical ordering of factors that determine the variation (Guy 1980; Poplack and Tagliamonte 1991; Poplack 2001), and that if different factor orderings obtain among different subgroups for a specific linguistic variable in a population, then we must be dealing with different speech communities. Auger and Villeneuve (2008), for instance, make exactly this claim in their argument that Picard and French, two neighboring varieties, are distinct languages because their constraints on the morphophonological factors affecting *ne* deletion are differently ordered.

Most of the studies in this vein are quantitative (using the Varbrul method referred to above), but are not laboratory-based. The use of a stable, replicable methodology has nevertheless enabled a measure of control in the study of naturalistic speech. A canonical sociolinguistic interview divides its time between (1) demographic questions and background information; (2) a series of question-prompts that aims for relaxation and involvement as measured by unself-conscious storytelling on the part of the interviewee, with some fairly set topics (see Feagin 2002) that work more or less cross-culturally (though Wolfson 1976 and Mendoza-Denton 2008 offer some critiques); (3) a word list; and (4) a minimal-pair reading task. These last two tasks attempt to elicit the most self-conscious and formal genres in the speaker's repertoire, while the second storytelling task aims for the opposite: to capture the interviewee's speech at its most "natural" and "relaxed." It is these different levels of questioning that provide a control for interviewing protocols, and which define the styles (formal vs. informal) that are compared across subjects, interviewers, and even dialects/languages. Thus the methodology itself provides control across many different interviewing situations occurring in different cultural contexts (though many have subsequently noted crucial interviewer and contextual priming effects: Rickford and McNair-Knox 1994; Hay et al. 2009).

Traditionally, the emphasis in sociodemographically based sociolinguistics (Mendoza-Denton 2002) has been on understanding how language change arises from linguistic variation (the classical problem in Weinreich et al. 1968: actuation, transmission and diffusion of change in the speech community). Early studies were already tilted toward stratification in terms of sociological attributes and the styles (formal vs. informal) elicited as the independent variables. This iterative division in the samples yielded sociodemographically based correlations with linguistic variable use (e.g. as a function of social class, age, ethnicity, and gender),

Some of classical sociolinguistics' most notable findings (overwhelmingly driven by the study of phonology) include what Labov (1972a) called the Lower Middle Class crossover effect: the finding that in a population stratified by class, use of a phonological variable by the lower middle class will overshoot the norm of the upper class in the most formal styles. In Labov's case, this was demonstrated with (r) in New York City, and replicated early on by Trudgill (1974) in Norwich, England with the variable (ing). Curvilinear patterns showing that a group in

the center of the socioeconomic hierarchy is leading in a linguistic *change in progress* have been found in Philadelphia (Labov 2001), New York City (Labov 1966), Norwich (Trudgill 1974), Panama City (Cedergren 1973), and Cairo (Haeri 1996), *inter alia*. The explanation of the crossover effect is that the variable in question has gained an association with some desirable social distinction, so that status-sensitive groups overshoot what might otherwise pattern as parallel variation by class and style (known as *stable variation)*. As Labov (2002) remarks, "Sociolinguistic variation is parasitic upon linguistic variation. It is an opportunistic process that reinforces social distinctions by associating them with particular linguistic variant."

One of the most durable constructs in sociolinguistics is that of *apparent time* (Bailey et al. 1991), the assumption that if one slices the population into age brackets, the resulting distribution of variation will show changes in progress spreading through the population, with the speech of the youngest speakers reflecting the most innovative version of the community grammar and the speech of the oldest reflecting a more conservative version. The hypothesis of apparent time has two strong assumptions: one is that of the critical period, where it is assumed that speakers' phonology has been acquired and has stabilized by the teenage years (see Flege 2006); the other assumption is that speakers' linguistic systems are relatively stable and do not change as they age (but see Harrington, Palethorpe, and Watson 2000). One of the most complete studies so far to test the apparent time construct against real-time panel data, following the same speakers, is Sankoff and Blondeau (2007), who analyzed the community shift in the pronunciation of /r/ in Montreal French by comparing data collected in 1971 and 1984. They concluded that: "To the extent that older speakers change in the direction of change in progress during their adult lives, apparent time underestimates the rate of change" (Sankoff and Blondeau 2007: 582).

Findings in the area of ethnicity have tended to focus on the convergence/divergence question of black and white vernaculars in the USA (Labov and Harris 1986; Ash and Myhill 1986; Wolfram and Thomas 2002) and in transplant African-American communities such as those of Nova Scotia and Samaná (Poplack and Tagliamonte 1991; Poplack 2001), though the latter tend to focus on syntax rather than phonology. The bulk of sociophonetics work on ethnicity in the USA has historically concentrated on African Americans, immigrant non-whites, and their relationship to the changes in progress taking place in the majority community. In Europe, only recently have studies of bilingual immigrant communities taken a greater role as demographic changes show increasing participation of previously unexamined groups in the creation of new forms of the vernacular (Hewitt 1986; Rampton 1995; Kotsinas 1998; Heselwood and McChrystal 2000; Khan 2006; Khattab 2007; Cheshire et al. 2008; Alam 2009; Jannedy and Martins 2008). Other studies involving a single language with an "ethnicity" dimension (e.g. studies of language use in Northern Ireland by Milroy 1987a, McCafferty 1998; of varieties

of Bahraini Arabic by Holes 1986; and of varieties in Russia by Kochetov 2006c; and in China by Zhang 2005) are typically couched in terms of religion, culture, region, or other kinds of affiliation, and are similar to more canonical in-migration studies such as Kerswill (1994), Lane (2000), and Dyer (2002). Where ethnicity with concomitant multilingualism is the norm, studies tend to be classified as "language contact" (Flege 2006 on bilingual accommodation; Mesthrie 1992 on South African English; Devonish 2007 on Jamaican English; Holmes 1997 on New Zealand Maori English; papers in Meyerhoff and Nagy 2008).

The interpretation of gender differences in speech communities has been an enduring source of debate within sociolinguistic research. Eckert (1989) challenged the then-prevailing notion (linked primarily to Labov and Trudgill) that sound changes from above the level of consciousness were led by linguistically conservative, status-conscious women (see for instance Holmquist 1985), whereas sound changes from below the level of consciousness were started by working-class, covertly prestigeful men, and then taken over by women who became the leaders of change (for the ensuing debate, see Labov 1990; Coates 1993; Gordon 1997). A number of reasons had been suggested for what was perceived as social fact: because of status differentials, women were more linguistically insecure than men. (Note that this is also the kind of account that was used to explain why the lower middle class had the crossover effect.) Based on ethnographic work in an ethnically homogeneous high school close to Detroit, Eckert showed that social class and gender interacted within specific social structures in the field setting: at Belten High, the social landscape was dominated on the one hand by *jocks*, who were both establishment- and supra-locally oriented, had middle-class backgrounds, and a school-based social life; and on the other hand by *burnouts*, who were of working-class background, were locally oriented, rebelled against the school's *in loco parentis* role, and did not take part in school activities. Participation in the Northern Cities Chain Shift was led by burnouts, but within that category, it was the burnout girls who surpassed the burnout boys in iconic changes such as raising and backing of the nucleus of /ay/, while among the jocks the girls trailed the boys in this change. Clearly, more subtle explanations were needed than simply lumping all men and all women together in their participation in linguistic change.

This work paved the way for a major shift in the understanding of how change proceeds in communities and the role of individuals, and led to new ways of thinking about language and social meaning, especially the social indexicality of variables. By opening up the inner workings of communities in the late 1980s and early 1990s, both the social networks (Milroy 1987a; Milroy and Milroy 1985, 1993) and communities of practice frameworks (Lave and Wegner 1991; Eckert and McConnell-Ginet 1992) contributed to a sea change in how we understand the spread of variation and what it means to the speakers who are adopting it. Instead of looking at large communities from above, as disembodied analysts cutting up the social landscape into census tracts, researchers began trying to understand

communities from the participants' insider perspectives, and trying to uncover the social categories that may be meaningful within the community. An early harbinger of the problems in the traditional concepts of class, for example, was the work of Rickford (1986) who called for new, conflict- and power-based understandings of class because the social structure in Canewalk, Guyana, was not easily divided into composite class indices traditional in consensus-based sociolinguistics models. In Canewalk, language variation was governed by the categories of estate- and non-estate class, social divisions that were the result of Guyana's plantation history. Studies of class and gender as conglomerates of practices have led researchers deeper into the social histories of communities and life histories of individuals (Milroy and Milroy 1985; Johnstone and Bean 1997; Labov 2001; Mendoza-Denton 2008) to predict which individuals are the leaders in language change, and how that change might be structured in terms of language use, individuals' phonological systems, and the deployment of phonetic detail.

During the 1980s another large change took place in the way that sociophoneticians thought about style/register. As mentioned above, early studies linked individual style to community-wide stratification through the construct of attention paid to speech (Labov 1972a), but later studies such as Bell (1984) and Coupland (1980) looked at individuals' deployment of sociophonetic variables and posited that individuals, in crafting their temporary implementation of their linguistic styles, were responding to specific audiences, and often matching their phonological production to present, implied, or imagined audiences (this perspective is broadly known as audience design; Bell 1984).

The breaking apart of (a) strict demographic categories and (b) the perceivedly linear stratification of style has had significant consequences for sociophonetic research. In the area of gender, for instance, researchers are looking at gendered expectations in speech perception (Johnson et al. 1999); at the production of gendered speech toward children (Foulkes and Docherty 2006); and at non-binary gender situations (Pierrehumbert et al. 2004; Crocker and Munson 2006). In matters of style, laboratory phonology researchers in the communities of practice approach (exponents include Eckert 2000; Zhang 2005; Rose 2006; Stuart-Smith 2007b; Mendoza-Denton 2008; Alam 2009; Drager 2009; Lawson 2009) continue to develop ideas of styles as practices and constellations of behaviors (Eckert 2005; Podesva 2007, 2008), and of iconic personae that bring these styles together into salience and relevance in communities (see for instance Zhang's 2008 study of Beijing "smooth operator" speech, which involves strong rhotacization, the description of which is deeply rooted in Chinese literature—going all the way back to the Qing Dynasty).

Work in this vein suggests that for individual speakers the motivation for adopting particular socially marked patterns of phonetic realization seems to be chiefly about the construction and performance of identity or identities relating both to themselves as individuals and to their affiliation to (or dissociation from) the

diverse social groups with whom they interact. Where identities and ideologies regarding the use of language coincide within a community of speakers, this can provide the conditions conducive to the sorts of collective patterns of phonetic realization identified in conventional sociolinguistic studies. But where consistent differential patterning *is* observed across socially defined groups of speakers, it is perhaps not surprising that the distributions of variants across groups are rarely if ever categorical, given the readiness with which speakers adapt their performance to meet what they perceive as the social-indexical demands of particular communicative situations.

The overall picture emerging from sociolinguistic studies of phonological variation, then, is of individuals drawing on a wide range of phonetic parameters to index social affiliation and differentiation, and being able to do so flexibly on a moment-to-moment basis in line with the perceived demands of a particular communicative situation. While there is some evidence (e.g. Sangster 2002[3]) that stylistic adaptations can be under the conscious control of speakers, research into inter-speaker accommodation and convergence (Giles 1984; Bell 1984; Giles et al. 1991a; Coupland 2007) and anecdotal observation suggests that, more typically, shifting of this sort takes place without an explicit intention being formulated on the part of the speaker. And, of course, this all appears to be underpinned by a very significant process of learning and understanding of the community-specific social-indexical value of phonetic variation, and an ability to make instantaneous interpretations of the same.

4.4 The intersection of sociophonetics and laboratory phonology

As indicated at the start of this chapter, these aspects of speech communication have until relatively recently not been seen as a central concern of laboratory

[3] Sangster studied phonological variation in the performance of undergraduate students who had relocated to Oxford University from Liverpool in the North-West of England (an area with particularly marked accent features—Watson 2007), uncovering the ways in which such variation was tied in to how individuals (in some cases quite overtly) managed their identity as Liverpudlians in an environment in which there were very few people from Liverpool. For example, one student is quoted as follows: "When I first came here I was more broad than I was normally because when you get there and everyone's like [posh voice] "oh yes I come from wherever" and then when you hear people speak like that—I think it's an unconscious thing that you just make yourself sound more Scouse because they like it, and almost everyone speaks the same, and it's good to be different, it's not a different bad-different, it's a good, happy sort of everyone-likes-it different" (Sangster 2002).

phonology; in part reflecting the walls around subdisciplines which affect linguistics research no less than other areas, but also reflecting an orientation to speech production/perception/learning which has focused predominantly on the lexical-contrastive information carried by the speech signal. However, what renders this disconnect particularly problematic is the fact that the speech signal is the channel through which, at one and the same time, speakers phonetically realize the phonological system acting as the foundation for lexical contrast *and* project the social-indexical features appropriate for particular communicative contexts (Docherty et al. 2006). Thus, from the point of view of the individual speaker-listener, the transmission of the lexical and social-indexical channels of meaning appears to be an integrated process, which suggests that any attempt to account for or model one of these channels without accounting for how it integrates with the other will necessarily be incomplete. Thus, as pointed out above, for the sociolinguistics community, there are gains to be made by giving greater consideration to a number of hitherto largely neglected factors which may well have an influence on the distribution of variants found across a sample of speakers or which might paint a more realistic view of the nature of the variants themselves. Likewise, for the laboratory phonology community, one of the key impacts of sociolinguistic studies of phonological variation is the realization that in drawing on the details of speech performance as a means of refining theories of (for example) lexical representation, it is simply not possible to filter out the social-indexical meaning which will also be conveyed within a particular utterance and which constitutes a key factor responsible for the phonetic shape of an utterance.

This message has perhaps come home most strongly for the laboratory phonology community as a consequence of an increasing number of speech perception studies which have shown that social factors shape the processing and interpretation of speech signals in ways which are not foreseen within conventional models which (if not by design, certainly by default) have not made any allowance for the social-indexical channel in production/perception (see Nguyen, this volume, for further discussion and details). For example, Lachs et al. (2003) and Nygaard (2005) review evidence showing how speaker-specific characteristics influence listeners' responses in various types of listening task. Two other particularly insightful studies are those by Strand (1999) showing that gender stereotypes shape listeners' responses to an [s ~ ʃ] continuum (concluding that "higher level relatively complex social expectations might have an influence on such low-level basic processes as phonological categorization of the speech signal"; p. 93), and by Niedzielski (1999) showing that Detroit listeners' judgments of vowel quality in the same stimuli are dependent on whether they believe the speakers are from Detroit or Canada. More recently, a set of similar studies has been carried out by Hay and colleagues (e.g. Hay, Nolan, and Drager 2006) showing differential perception of the same stimulus material by listeners depending on the (implicit) beliefs that they have about social factors relating to the material that they are being asked to respond to. Crucially,

as well as showing that speech perception is not independent of social-indexical information, these studies also highlight that listeners have acquired knowledge of the typical associations between specific features of speech performance and the characteristics of individual speakers, and, perhaps most relevant for this chapter, of groups of speakers of various sorts (e.g. males vs. females, young vs. old, USA vs. Canada, etc.).

With evidence pointing to the importance of integrating social-indexical phonetic properties into accounts of production and perception, there is no doubt that we also need to consider their role in phonological acquisition, not only from the point of view of how a child begins to learn the value of and make use of the socially governed variants within her/his speech community, but also from the point of view of understanding the social-indexical properties of child-directed speech and how this differs from adult-directed speech within the same community (bearing in mind that from a very young child's point of view, the speech community may well be made up simply of the immediate family). For further discussion of this area, see Foulkes et al. (2005), Foulkes and Docherty (2006), Foulkes (2010). A key question is to what extent, when setting out on the path of acquiring knowledge of the sound pattern of the ambient language, a child can separate out from within the input that she/he is exposed to from birth those features of the speech signal which are lexically contrastive and those which are social-indexical. While this may well happen at a later stage of development (as described by Foulkes 2010), it seems likely (Foulkes et al. 2005; Docherty et al. 2006) that the process of phonological acquisition is at one and the same time a means for learning the building blocks of the native language lexicon and for learning how to sound like a member of the immediate speech community. It is an empirical question how this integration is achieved and for how long it is maintained.

In sum, as mentioned above, the more we learn about the social-indexical channel in speech performance and the extent to which it is integral to the performance of speakers and the processing of speech by listeners, the stronger becomes the need to account for how the social-indexical channel is embedded within speech processing and representation. Historically, this was simply not possible given that most models of the latter had a clear focus on seeking to sustain the hypothesis of underlying representational invariance in the face of abundant surface variability. An example of this approach can be seen in the work on relational invariance underpinning the quantal theory of speech production (e.g. Stevens and Blumstein 1978; Stevens 2002; Stevens and Keyser 2010; see Hawkins 2004 for an overview) in which variability is cast as "noise" which needs to be minimized in order for the underlying invariants to be discerned (see also Lahiri, this volume, for an overview of the Featurally Underspecified Lexicon model which applies some of the same principles). This approach is also reflected in the quantitative methodologies adopted by many researchers in which conclusions are drawn from reports of central tendencies characterizing a sample of speakers as a whole without reporting

either at all, or in any detail, the extent to which the overall findings reflect the performance of individual members of the sample.

At the simplest level, progress towards bridging this gulf can begin to be made in quantitative production/perception/learning studies by simply factoring in to experimental designs some of the key factors which sociolinguistic research has shown to be relevant in accounting for speaker performance. For example, there is a growing awareness of the fact that in describing the accent background of experimental subjects, it is not sufficient to simply refer to the name of the language spoken by those subjects (e.g. "10 speakers of American English," or "10 speakers of French," etc.); at the very least there is a need to know something about the geographical provenance of the speakers, about the nature of the particular variety or varieties which are represented in the sample of data, and about the extent to which any cross-speaker variation is likely to impact on the focus of the study. But it is also possible to begin to develop "joined-up" accounts of speaker performance by factoring into the analysis non-linguistic factors which are relevant in accounting for the criterial dependent variables. A good example is Scobbie's (2006) study of VOT in the Shetland variety of English. This study involved a word-list task with twelve subjects aged 16–30, six males/six females, all born in the Shetlands, all of whom had lived there all their lives, all from the same geographical area on the Islands, half attending the same school, and mostly known to each other (i.e. a highly controlled group of speakers which might not unreasonably be assumed to provide a homogeneous sample). Overall VOT distribution was extremely variable across speakers (/p/ ranged from 0 ms to 112 ms, /b/ from −190 ms to 41 ms). However, analysis of individuals' performance revealed individual realization strategies which imposed some structure on the group findings, but also brought to light that an important factor in accounting for the variability found in the VOT results was the place of origin of the speakers' parents; parents of Shetland origin were associated with shorter VOTs for /p/ and more pre-voiced /b/s; other Scottish parents with higher VOTs for /p/ and fewer cases of pre-voicing; English-parented subjects were more variable across the VOT continuum. Thus, despite having sampled speakers in a way that by almost any measure would appear to be a very good basis for generalizing across speakers of Shetland English, it was only by considering differences between individual members of that community that a more informed and theoretically more challenging account emerged.

A key development over the last decade has been the elaboration of a more far-reaching theoretical platform for addressing the issues identified above. The exemplar approach to phonological representation, in which knowledge of phonological patterning is based on a multifaceted, phonetically rich representation in memory derived from and continually shaped by an individual's experiences as a speaker-listener, has opened the doors to a model in which the integration of the lexical and social-indexical would be entirely predictable and natural (Goldinger 1997; Johnson 1997b; Pierrehumbert 2001a, 2006a; Hawkins 2003; Foulkes and Docherty 2006).

Notwithstanding that there are a number of dimensions of this particular approach which remain to be developed (Docherty and Foulkes forthcoming), a key characteristic is its assertion that, in building up knowledge of the systematic aspects of sound patterning from their experience with spoken communication, speaker-listeners automatically and simultaneously map out associations between signal properties and both linguistic and non-linguistic aspects of experienced stimuli (Johnson 1997b; Foulkes and Docherty 2006; Pierrehumbert 2006a; Hay, Warren, and Drager 2006; Mendoza-Denton 2007; Foulkes 2010). Since social-indexical information is systematically intertwined with other channels of meaning within the speech signal, by hypothesizing an integrated, probabilistic, and experience-driven representation the exemplar approach provides a conceptualization of how these various channels can be fundamentally integrated in speech processing and representation as suggested above. And in doing so, its advocates argue that it is not incompatible with the sorts of abstract phonological representations which have predominated to date in work on phonological representation (Pierrehumbert 2006a). Indeed, there is an emerging consensus behind the concept of a hybrid model of representation incorporating both abstract and exemplar representations with the balance between the two now constituting something of a new focus of experimental work (Goldinger 2007). McLennan (2007: 69) summarizes this debate saying that "the field has entered into a new phase in which, rather than debating over abstract versus episodic representations, efforts are now focused on determining the ideal framework that can account for their coexistence."

Tellingly, Pierrehumbert (2006a) refers to the conceptual framework offered by a phonetically rich probabilistic representation as a "toolkit," correctly reflecting the fact that while there is some way to go before the details of this framework are fully tested and evaluated, nevertheless it does allow for the framing of questions which hitherto would have struggled to find a theoretical "hook." For example, there has been a growth in interest in the dynamic nature of an individual's phonological knowledge; an exemplar model of representation predicts that phonological knowledge continues to evolve through the life span, shaped by individual experience, contrasting with the conventional view that the acquisition of phonology is focused in the early years of development and is from that point stable across speakers of the "same" variety. This is exemplified by Harrington and colleagues' study of the phonetic characteristics of the British Queen over fifty years' recordings of the annual Christmas Day Queen's Speech (Harrington et al. 2000, 2005; Harrington, this volume) which provides a particularly detailed real-time account of life-span changes in the speech performance of an individual positioned very much at the conservative pole of language use. As pointed out by Labov (2006), however, while evidence of change in adulthood such as this does indeed reinforce the view that phonological knowledge can continue to evolve through life, there is a need to devise an explanation for this which also accounts for the fact that this dynamism appears to be much less marked in adults than it is in younger speakers. This is

presumably in large part due to the connection between phonological patterning and identity formation referred to above, but this is one area which is in need of much further investigation.

A closely connected question is that of what takes place when speakers with different sociolects come into contact. An integrated theoretical framework should make it possible to devise an elegant interface between sociolinguistic models of dialect contact (e.g. Trudgill 1986; Britain and Trudgill 1999; Kerswill 2002) and what we know of how individual listeners' phonological representations are influenced by exposure to phonetic realizations which they previously had little experience of, as revealed, for example, in studies of "perceptual learning" following exposure to novel phonetic realizations (Norris et al. 2003; Kraljic et al. 2008; Cutler et al. 2010) and by work on the plasticity of phonological categories and how this can be associated to speakers' different levels of exposure to particular varieties of English (e.g. Evans and Iverson 2004). And increasingly there is potential to imbue models of dialect contact/change/formation with an understanding of the factors pertaining to conversational interaction which influence the behavior of individual speakers of different varieties when they interact (e.g. see Delvaux and Soquet's 2007 account of passive speech imitation in speakers of Flemish, studies by Pardo 2006 and Babel 2009 of phonetic convergence between interlocutors, and work by Wedel and Volkinburg 2009 and Pierrehumbert, this volume to model the consequences for a community of speakers of this sort of inter-interlocutor phonetic entrainment).

What these studies exemplify is that, while it remains an area of intense debate, the exemplar-model "toolkit" has brought to the fore the question of how speaker-listeners manage the multiple channels of information interwoven into the phonetic properties of the speech signal. This is now a central area of theoretical development and debate, and one which sits straightforwardly alongside the other questions which have drawn together the laboratory phonology community. Thus, the historical gap between models of production/perception/acquisition and what we know of how social-indexical meaning is conveyed and interpreted within speech has started to be bridged. And with this line of investigation showing every sign of developing further, the importance of addressing some of the methodological issues referred to earlier in this chapter cannot be overestimated.

4.5 Concluding comments

As indicated in the introduction to this chapter, in recent years there has been a significant and quite rapid change in the extent to which the social-indexical properties of speech have figured within debates on the key questions around which

the laboratory phonology community is unified. Arguably the most important factor in this development has been the postulation that phonological knowledge is phonetically rich and is shaped and defined by an individual's experience, thereby emphasizing an intimate connection between the properties of language and the speech performance and processing by users of language. This view is amply explored by Pierrehumbert (2006a: 516) who begins by asserting that "language is a collective behavior" and that it is formed "in populations, as people match their language systems to each other, and group themselves into social networks of people who share the same language." In its formative years, not least under the influence of Ohala's groundbreaking work (e.g. 1983, 1990c), the laboratory phonology community readily embraced the need to understand what aspects of phonological patterning could be accounted for by factors such as vocal tract physiology or general principles of motor control. But there is now a firmly established strand of activity within the laboratory phonology "project" which extends this to consider how factors arising from the social orientation of users of phonological systems account for the nature of those systems. This strand of work is at a relatively early stage of development, and, as pointed out above, is in need of further methodological and theoretical refinement. But, notwithstanding these points, a key attraction of this vein of research is that of allowing, in due course, the emergence of models of phonological knowledge with a more rounded understanding of the role of speakers and listeners in the acquisition and maintenance of that knowledge.

It is also important to emphasize that none of the above is necessarily out of line with the stance taken by many theoretical phonologists. Coetzee (this volume, p. 62) points out that, as the result of a growing interest in variation on the part of theoretical phonologists (largely driven by exploring the extent to which particular theoretical frameworks can deal with realizational gradience and variability), "phonology is now more ready than ever to integrate the apparently disparate approaches of theoretical and laboratory phonology." While, to date, sociophonetic variation has taken a somewhat secondary role in the exploration of this integration, the work reviewed here suggests that it is a nettle which needs to be grasped more firmly and which has the potential to deepen our understanding of how individuals' orientation to the social context of "real-life" speech communication impacts on the nature and characteristics of the speech signal from which that phonological knowledge is derived, and ultimately how it shapes the nature of phonological representation itself.

CHAPTER 5

INTEGRATING VARIATION IN PHONOLOGICAL ANALYSIS

VARIATION: WHERE LABORATORY AND THEORETICAL PHONOLOGY MEET

ANDRIES W. COETZEE

MODELING PHONOLOGICAL VARIATION

ARTO ANTTILA

In this chapter, the authors consider how the rich variability of speech can be integrated into more formal phonological models. Coetzee highlights the critical importance of experimental data, in all its richness and complexity, for developing adequate theoretical models of phonology. Anttila further explores the modeling of such variation within a constraint-based approach building on more traditional work within Optimality Theory.

5.1 Variation: where laboratory and theoretical phonology meet

Andries W. Coetzee

5.1.1 Introduction

As evidenced by this volume, laboratory approaches to phonology have increased in importance over the past several decades. In fact, at least since the first LabPhon conference (Kingston and Beckman 1990), laboratory techniques have been accepted as not only valid but necessary by many phonologists. These approaches, however, were not universally received this positively, and have been slow to make inroads into traditional theoretical phonology. Over the past decade, this has changed such that now, even for many researchers within traditional theoretical phonology, laboratory experimentation has become an established research method.[1]

Coetzee et al. (2009a: 2) identify the development of increasingly mathematically sophisticated models of grammar as one reason for this change. Laboratory data are characterized by variability, and therefore challenge grammatical models that are not architecturally suited to handle variation. It is now the case that the most widely used models of phonological grammar are much better suited to handle such data than were earlier models from the classical generative tradition. See, for instance, Anttila (1997, this chapter) and Boersma and Hayes (2001) for examples of how Optimality Theory (henceforth OT; Prince and Smolensky 1993/2004) deals with variable data. In more recent developments, Harmonic Grammar (henceforth HG; Smolensky and Legendre 2006; Pater 2009b), with its close ties both to OT and connectionism, has been applied with some success to variation (Coetzee 2009a; Coetzee and Pater forthcoming; Coetzee and Kawahara forthcoming).

Concurrent with these developments in theoretical phonology, it has become easier to obtain laboratory data in recent years. Scobbie (2007b: 19), for instance, attributes the increase in "quantitatively-based arguments" in phonology "to the ready availability of what used to be highly specialized and expensive acoustic analysis hardware and software."

Due to these developments, phonology is now more ready than ever to integrate the apparently disparate approaches of theoretical and laboratory phonology. As

[1] See Cohn (2010) for a review of the history behind the acceptance of laboratory techniques in phonology. Coetzee et al. (2009: footnote 2) provide the following as informal evidence for the change in how traditional phonology valuates laboratory techniques: Of the ten phonology dissertations completed at the University of Massachusetts in the ten years before 1986, none used laboratory techniques. Of the thirteen completed in the ten years before 2008, eight did.

noted above, a hallmark of experimental data is their variability. These data are therefore particularly suited to the study of phonological variation, the focus of the rest of this section. In Section 5.1.2, I show that an adequate study of variation depends on experimental data. In Section 5.1.3, I highlight some of the properties of variation that can be gleaned from experimentally collected data, and for which any adequate theory of variation must account. Finally, in Section 5.1.4, I review some recent models of variation developed in theoretical phonology.

5.1.2 The need for experimentally collected data

Even classical generative phonology acknowledged variation. However, the grammatical architecture assumed in this tradition is not suited to handle quantitative data. Variation was usually handled merely by marking a rule as optional. As a recent example, consider Vaux's (2008) account of variable syllable-final /s/-epenthesis in Dominican Spanish. Examples are given in (1a) (Bradley 2006: 4), with Vaux's rule (2008: 43) in (1b).

(1) a. invi[s]tado ~ invitado 'guest'
yu[s]ca ~ yuca 'yucca'
b. $\emptyset \rightarrow s/\_]_{\sigma}$ (optional)

Though this rule correctly captures that the process is optional, it lacks information about how likely the process is to apply. Such an approach to variability may be adequate if this rule, and variable rules generally, applied randomly. However, it has been documented widely that variable phonological processes do not apply randomly, and that the likelihood of a process applying depends at least partially on grammatical considerations (see below, and Anttila, this chapter, for examples). If we assume that an adequate theory of phonology must account not only for which processes are optional, but also for the likelihood with which optional processes apply, a richer theory of variation is required.

To develop theories that account for the likelihood with which a variable process applies, it is necessary to have data about the likelihood of application and the factors that influence this, and these types of data can be collected most responsibly by carefully planned experiments. There are at least two reasons for this. First, generalizations about the patterning of variable processes can usually only be identified in large data samples. Secondly, there are many factors that influence variation, and drawing reliable conclusions depends either on explicitly controlling for irrelevant variables during data collection, or on collecting large samples and then statistically controlling for the irrelevant variables during data analysis.[2]

[2] Most studies that collect production data through an experiment show how variables are controlled for during data collection. The second way of controlling for irrelevant variables is employed less often. Raymond et al.'s (2006) study of *t/d*-deletion in English serves as an example. They analyze

There are many types of experimentation that can be used to gather the required kind of data (see Kawahara 2011; Chapters 17, 18, and 21 this volume), from speech production experiments that collect acoustic and/or articulatory data, to studies that perform the traditional introspective judgment in a more controlled manner, to studies that perform acoustic analyses of large speech corpora.

5.1.3 Central properties of phonological variation

5.1.3.1 *The ubiquity of variation*

As noted above, variation was not a research focus in early generative phonology. This contrasts with the sociolinguistic variationist tradition that flourished simultaneously with, but independently from, mainstream generative phonology. In this tradition, variation was viewed as a defining property of grammar; as Labov (2004: 6) stated in a recent review, variation is "the central problem of linguistics."

Though research in the variationist tradition has generated a trove of information about variation, this research is not without shortcomings, and there are good reasons to replicate the results of this research in the laboratory. First, data collection is usually done through sociolinguistic interviews. In these settings, it is hard to control for variables not relevant to the research. These data are therefore even noisier than the already noisy laboratory data. A second problem is that most of this research depends on impressionistic transcription, and therefore falls short of the requirements for phonological research. There are many low-level phonetic properties of speech that cannot be transcribed even in narrow phonetic transcription. Additionally, auditory perception research has shown that humans do not accurately perceive acoustic information. Much of the variation in the signal is factored out during perception, so that the final percept can be very different from the actual acoustic signal.[3]

Despite these shortcomings, the basic result of the variationist tradition (that variation is ubiquitous) has been confirmed in the laboratory. Variation has been documented at all levels of phonological grammar, ranging from lexical/morphophonological variation between allomorphs, to variation at the segmental level, to variation at the subsegmental level. Since variation is observed throughout all levels of the phonological grammar, the traditional generative phonology response to variation (placing it outside phonology in the domain of

7,241 examples of t/d in a conversational speech corpus. Due to the conversational nature of the corpus, they could not control for confounding variables during data collection. However, when analyzing their data, they code the data for many different variables, and then perform regression analyses, factoring out the contribution of irrelevant variables.

[3] See, for instance, the research on "compensation for coarticulation" that shows how coarticulatory variation is perceptually factored out (Mann 1980; Mann and Repp 1980; etc.), and the more recent research on auditory illusions (Dupoux et al. 1999, 2001; etc.).

phonetic implementation), is not tenable (see Coetzee and Pater forthcoming for discussion).

Since variation at the lexical/morphophonological level is often reflected in spelling, this type of variation can be studied in textual corpora. In fact, even the internet can serve as a corpus for this purpose (see Anttila, this chapter and Loehr and Van Guilder, this volume). Anttila (1997), for instance, studies the variable realization of the Finnish genitive plural morpheme. This morpheme has a "strong" and "weak" form that are orthographically distinguished (the strong form contains a "d"). Anttila compiled frequency counts of the allomorphs in a corpus of 1.3 million words, documenting that the height of the final root vowel influences the probability of choosing the strong over the weak allomorph. Some examples from Anttila (1997) are given in Table 5.1.1. Studies like this show first that variation is not necessarily random, and that the frequency with which different variants are observed is often influenced by systematic phonological considerations (here vowel height). Secondly, this example shows the need for collecting data in a more quantitative manner than what has been the tradition in generative phonology—the frequency-based interaction of this process with vowel height can only be documented in this way.

Since phonological processes at the segmental level are seldom reflected orthographically, variation at this level can usually not be studied in text corpora. Additionally, as noted above, human perception does not always match the acoustics. Studying variation at the segmental level therefore usually requires more traditional laboratory techniques. As an example, consider Mitterer and Ernestus's 2006 study of word-final t-deletion in Dutch. Dutch words that end in /-Ct/ variably undergo deletion of the final /t/, a process that is not reflected in Dutch spelling. Mitterer and Ernestus also show that Dutch listeners often perceive these word-final /t/s, even when they are in fact absent. Impressionistic transcription alone can therefore not be used to study this phenomenon. Mitterer and Ernestus overcome these problems by performing acoustic analysis on speech corpora. A sample of their results is given in Table 5.1.2. As with the Finnish example above, these results

Table 5.1.1. The interaction of vowel height with the realization of the Finnish genitive plural

Stem		Stem-final vowel height	Allomorph		Frequency (% in parentheses)	
			Weak	Strong	Weak	Strong
lemmikki	'pet'	High	lem.mik.ki-en	lem.mi.kei-den	1131 (83%)	228 (17%)
fyysikko	'physicist'	Mid	fyy.sik.ko-jen	fyy.si.koi-den	352 (43%)	4668 (57%)
sairaala	'hospital'	Low	sai.raa.lo-jen	sai.raa.loi-den	49 (6%)	759 (94%)

Table 5.1.2. Percent word-final t-deletion in Dutch in different contexts

Preceding context	Following context	
	Obstruent gegooid [xəxoit] 'thrown'	Vowel en [ɛn] 'and'
/s/: *kast* [kɑst]∼[kɑs_] 'closet'	26%	10%
/x/: *klacht* [klɑxt]∼[klɑx_] 'complaint'	10%	0%
/n/: *kant* [kɑnt]∼[kɑn_] 'side'	6%	0%

show that the variation is not random, but is influenced by systematic phonological considerations (here the identity of the segment preceding and following the /t/). Only through performing acoustic analysis of actual speech can the existence of this variable process be confirmed, and only through investigating a large enough sample can the frequency patterns in the application of this process be identified.

Finally, variation has also been documented in phonetic implementation, at the subsegmental level. Since subsegmental variation is even less likely to be orthographically reflected and probably even more subject to inaccurate perception, this type of variation requires careful, experimental investigation. The study of the realization of /t/ in Dutch by Warner et al. (2004) is one of many examples. Unlike Mitterer and Ernestus's study above that investigated the variable deletion of /t/, Warner et al. investigated the acoustic properties of /t/s that are actually realized. A sample of their results is represented in Figure 5.1.1, showing that /t/s differ in their fine-phonetic detail depending on the context in which they appear. Both closure and burst durations are longer word-finally (*boot* [bot] 'boat') than word-internally between vowels (*boten* [botən] 'boats'). It is highly unlikely that this type of variation could be detected in any way other than through acoustic investigation.

5.1.3.2 *Gradient and discrete*

Another characteristic of variation that has been established in the laboratory, is that it can occur between two distinct categories (discrete variation), or it can be more continuous, spanning the space between two categories (gradient variation). Again, there are countless examples in the literature, and I give just one example of each kind (see also Ernestus, this volume).

An example of gradient variation comes from Ellis and Hardcastle's (2002) articulatory study of nasal place assimilation in English. Using electropalatography, they

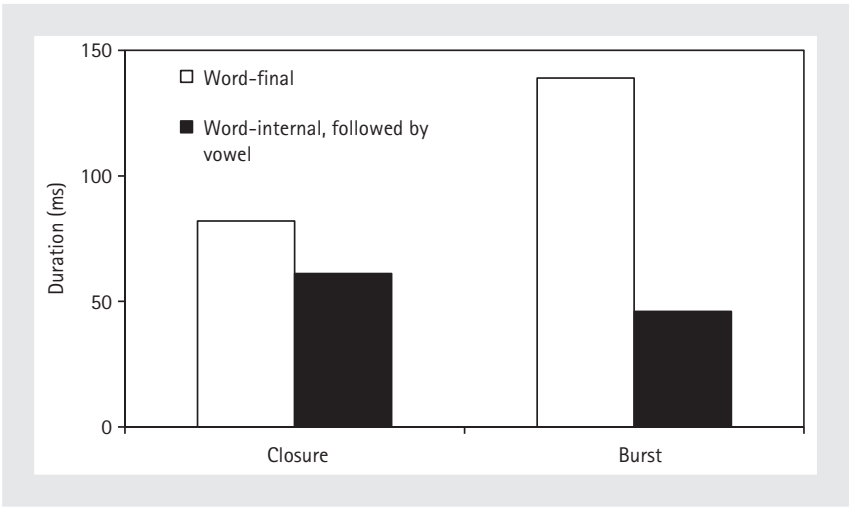

Figure 5.1.1. Closure and burst duration of Dutch /t/ preceded by a long vowel in different contexts.

collect data on the amount and position of tongue–palate contact during the underlined portion of the sentence *It's hard to believe the ban cuts no ice*. Among their ten participants, six showed no variation, pronouncing this sequence consistently either without assimilation as [nk] or with assimilation as [ŋk]. Two participants varied between the endpoint categories, and two used the whole articulatory space between the endpoints. Figure 5.1.2 (based on Ellis and Hardcastle's Figure 7, p. 384) shows three repetitions of the /nk/-sequence for one participant. The top of each image corresponds to the tongue tip, and the bottom to the tongue back. Contact between the tongue body and the roof of the mouth is marked with zeros, and non-contact with periods. Each of rows (a), (b), and (c) represents a different pronunciation of the /nk/-sequence. The progression from left to right in each row represents the time dimension—moving from the part of the utterance corresponding to /n/ into the part corresponding to /k/. Row (a) shows an utterance with a clear [n], indicated by the marked part of the image that shows contact only in the alveolar region. Row (c) shows an utterance where /n/ was realized fully assimilated as [ŋ], indicated by the marked part of the image that has contact only in the velar region. Row (b) shows an utterance where /n/ was realized as intermediate between [n] and [ŋ], with both alveolar and velar contact.

Anttila's study of the Finnish genitive cited above is an example of variation between categories—there are no utterances intermediate between two allomorphs of the genitive plural. Variation between categories is also observed at the segmental level. Coetzee and Pretorius (2010), for instance, report on the process of post-nasal devoicing in Tswana. Traditional grammars of Tswana describe the language as having an invariant process that devoices /b/ to [p] post-nasally (Cole 1955). This

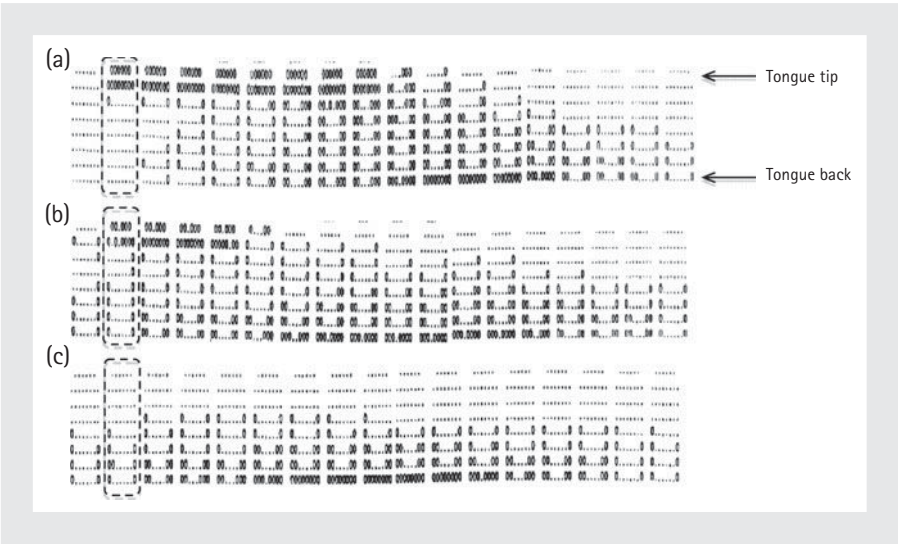

Figure 5.1.2. Contact between the tongue body and palate during pronunciation of an /nk/-sequence. Row (a) represents an unassimilated [n]-pronunciation, row (c) an assimilated [ŋ]-pronunciation, and row (b) a pronunciation intermediate between these options. In each row, part of the image corresponding to /n/ has been marked (reprinted with permission from Ellis and Hardcastle 2002: 384).

gives rise to the alternating realization of a verb like *botsa* [botsa] 'ask'. When it is preceded by the vowel-final first-person plural object clitic, /b/ is realized as [b]: /re+botsa/ → [rebotsa] 'ask us'. However, when preceded by the nasal-final first-person singular object clitic, the /b/ is realized as [p]: /m+botsa/ → [mpotsa] 'ask me'. Though this process is traditionally described as invariant, recent study has confirmed that it is in fact optional (Zsiga et al. 2006). Coetzee and Pretorius (2010) performed an acoustic analysis of Tswana speech collected in the laboratory in order to investigate this process. Their results confirmed that the process is variable, but they also found that this variation is between two discrete categories—between completely voiced [b] and completely voiceless [p]. Figure 5.1.3, based on their Figure 2, shows the distribution of the more than 1,000 tokens in their study in terms of the percent of the plosive's closure that was realized with voicing. These tokens are distributed into two discrete categories, with either complete closure voicing or voicing of about 40 percent of the closure. The space between these categories is only sparsely populated.

Thus while most theoretical phonology accounts of variation have dealt only with discrete variation (see later in this section, and Anttila, this chapter), an adequate theory of variation should be able to account for discrete and gradient variation.

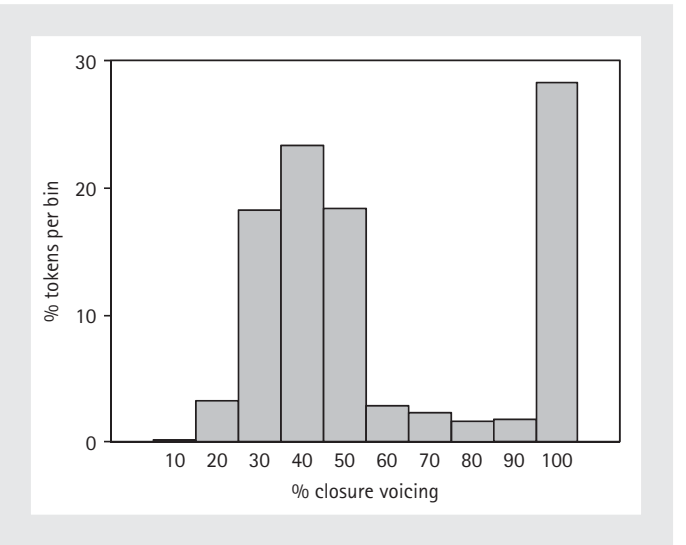

Figure 5.1.3. Distribution of Tswana plosives in terms of percent closure voicing. The distribution shows evidence for two distinct categories: voiced [b] with 100% voicing, and voiceless [p] with voicing of about 40% of the closure (reprinted with permission from Coetzee and Pretorius 2010).

5.1.3.3 *Multiple influences*

Variation can be caused and influenced simultaneously by many different factors, both grammatical and non-grammatical. In fact, Bayley (2002: 118) identifies "the principle of multiple causes" as one of the most important principles that underlie the study of variation. As before, I discuss only a few examples from the literature to illustrate this point. In English, t/d variably deletes from word-final consonant clusters, giving rise to pronunciation of phrases like *west bank* with and without the final /t/ ([wɛst bæŋk] ~ [wɛs bæŋk]). This variable process has been studied extensively in the variationist tradition, and it is influenced by both grammatical and non-grammatical factors. Bayley (1994: 310), for instance, shows that *t/d*-deletion in Tejano English is influenced by grammatical factors such as the following phonological context, and non-grammatical factors such as the age of the speaker. Relevant data are summarized in Table 5.1.3, showing that deletion is more likely before a following consonant than a pause or vowel, and that deletion is more likely in the speech of younger speakers.

Staying with English *t/d*-deletion, Coetzee and Kawahara (forthcoming; Coetzee 2009a) analyze deletion in the Buckeye Corpus (Pitt et al. 2007), and find evidence both of grammatical (the following phonological context) and non-grammatical

Table 5.1.3. Grammatical and non-grammatical influences on *t/d*-deletion in Tejano English

	Following context			Age of speaker	
	Vowel 'best option'	Consonant 'best choice'	Pause 'best'	15–20	26–44
Deletion rate	25%	62%	46%	52%	37%

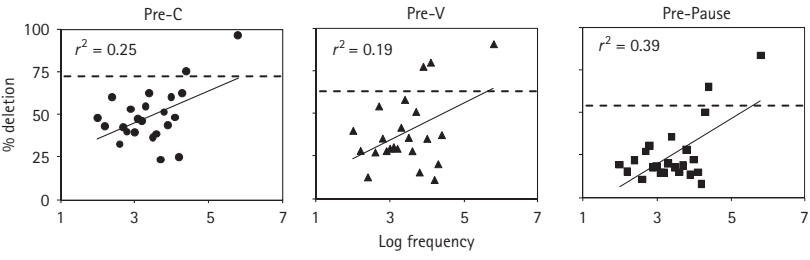

Figure 5.1.4. *t/d*-deletion rate in the Buckeye Corpus.

factors (lexical frequency) influencing the deletion rate—specifically, t/d is more likely to delete from more frequent than less frequent words.[4] Figure 5.1.4 plots the deletion rate in three contexts against the log frequency of tokens, as counted in CELEX (Baayen et al. 1995). Regression lines and their associated r^2-values are also included, showing a positive correlation between frequency and deletion. Horizontal lines represent the mean deletion rates, again showing the influence of the following context. (See also Bybee 2002 for evidence that frequency impacts *t/d*-deletion.)

Results such as these show that phonological variation is often influenced by multiple factors of which grammatical factors are only one type. A model of phonological variation that aims to give a full account of the phenomenon must allow for multiple factors to co-determine how variation is realized in actual speech.

[4] Though most phonologists would agree that speaker age is not a grammatical factor, the same is not true for usage frequency. There are models of phonological grammar in which frequency is directly grammatically encoded (e.g. Bybee 2002; Pierrehumbert 2001a), and hence part of the grammar. At the same time, frequency has no place in classic generative theories of phonology. Whether frequency is a grammatical or non-grammatical factor, the point being made here holds: variation is influenced by multiple factors.

5.1.4 Models of phonological grammar

In this section, I review some of the models that have been developed in theoretical phonology to account for variation. Based on the discussion above, an adequate model should have the following properties: (i) Due to the ubiquity of variation, variability must be part of the central design of the model. (ii) It must be able to account for both discrete and gradient variation. (iii) It must allow for influences from multiple sources, grammatical and non-grammatical, to co-determine how the variation is realized. All the models discussed here address (i) adequately. For a proposal about how to address (ii), see Flemming (2001). I focus here on (iii), reviewing two conceptualizations of the relationship between grammatical and non-grammatical factors.

In what follows, I use the influence of the usage frequency of words on the application of variation as an example of a non-grammatical factor that influences variation. As acknowledged in footnote 4, usage frequency is not considered non-grammatical in all models (see Bybee 2001, 2006, 2007; Pierrehumbert 2001a; Gahl and Yu 2006 eds. etc.). See also later in Section 5.1.4.2 about how other less controversially non-grammatical factors can be handled in the same way as usage frequency.

5.1.4.1 *Equal emphasis on grammatical and non-grammatical factors*

Variable rules, as in the variationist tradition, were originally introduced by Labov (1969). These are ordinary rewrite rules with two augmentations: (i) they are marked as optional, and (ii) their structural descriptions encode how the presence of elements in the context of the rule promotes or inhibits its application. The structural descriptions can contain grammatical and non-grammatical factors. In fact, no formal distinction is made between grammatical and non-grammatical components, and both contribute equally to the application of the rule.

Labov's original variable rules could only account for relative differences in the application of a rule (if A is present in the context, the rule is more/less likely to apply than if B is present). However, variable rules were implemented mathematically during the 1970s (Cedergren and Sankoff 1974; Rousseau and Sankoff 1978). The mathematical implementation that has become the standard, and that is implemented in the widely used VarbRul/Goldvarb software, performs a multivariate stepwise logistic regression over observed token counts (Paolillo 2002: 177; Sankoff et al. 2005; see also Warren and Hay, this volume). Application/non-application is taken as the dependent variable, and different factors hypothesized to influence the probability of application are taken as independent variables. Given a corpus of observed tokens, coded for the dependent and the independent variables, VarbRul/Goldvarb estimates the contribution that each independent variable makes to the probability of rule application. The regression equation used in this model is

Table 5.1.4. Some factor values from the VarbRul analysis of *t/d*-deletion in Tejano English (Bayley 1994: 310)[a]

Factor			
Input	p_0		.469
Following context	p_1	Obstruent	.728
		/l/	.636
		Glide	.479
		/r/	.467
		Pause	.463
		Vowel	.267
Speaker age	p_2	15–20	.528
		26–44	.417

a. Bayley coded the data for more than just these factors, and the factor values listed here are therefore only a part of his results. Also note that Bayley differentiated between different kinds of consonants in the following context. When these results were reported earlier in this section, all the consonants were collapsed into one category.

given in (2). In this formula, p_0 represents the probability of the rule applying independently from any factors $p_1 \ldots p_n$, which influence the probability of application. p expresses the predicted probability that the rule will apply, given the values of the factors $p_1 \ldots p_n$ in the input to the rule.

(2) $$p = \frac{p_0 \times \ldots \times p_n}{[p_0 \times \ldots \times p_n] + [(1 - p_0) \times \ldots \times (1 - p_n)]}$$

The factors, $p_1 \ldots p_n$, can be either grammatical or non-grammatical, without any distinction between these two kinds. A non-grammatical factor can influence the application of a rule just as much as a grammatical factor, so that this model treats all potential influences equally. As an example, consider the results of the Bayley (1994) study of Tejano English mentioned earlier. Table 5.1.4 lists some of the factor values that Bayley found for a grammatical factor (following context) and a non-grammatical factor (speaker age). In order to calculate the probability of t/d-deletion from a word uttered before a vowel by a younger speaker, the factor value for a following vowel (.267) is substituted for p_1 and the value for a younger speaker (.528) for p_2. Solving for p as in (3) gives the result of .265, implying that a deletion rate of 27 percent is expected. Note that the grammatical and the non-grammatical factors are treated alike in this calculation.

(3) Expected deletion rate in pre-vocalic position, by a younger Tejano speaker

$$p = \frac{p_0 \times p_1 \times p_2}{[p_0 \times p_1 \times p_2] + [(1 - p_0) \times (1 - p_1) \times (1 - p_2)]}$$

$$= \frac{.469 \times .267 \times .528}{[.469 \times .267 \times .528] + [(1 - .469) \times (1 - .267) \times (1 - .528)]}$$

$$= .265$$

See Coetzee (2009a) and Coetzee and Pater (forthcoming) for a more detailed discussion and evaluation of this model, and for implementation of the model to several examples of phonological variation.

5.1.4.2 *Grammar-dominant models*

Since the early 1990s, several models of variation have been developed in Optimality Theory (Kiparsky 1993; Reynolds 1994; Anttila 1997; Boersma and Hayes 2001; Coetzee 2006; Anttila, this chapter). These models are all purely grammatical, and do not allow for non-grammatical factors to influence the application of variable processes.[5] More recently, Coetzee and Kawahara (forthcoming; Coetzee 2009a) proposed a model in OT's close relative, Harmonic Grammar (henceforth HG; Smolensky and Legendre 2006; Pater 2009b) that formally incorporates both grammatical and non-grammatical factors. In this regard, it agrees with the variable-rule framework. However, it also differs from it in that it assigns primacy to grammar. Grammar dictates the limits of variation, while non-grammatical factors influence how variation is realized within these limits.

HG works with weighted constraints, rather than OT's ranked constraints. Like OT (Boersma and Hayes 2001), HG has a stochastic implementation known as "noisy HG" (Coetzee and Pater forthcoming). In noisy HG, the weight of each constraint is perturbed by a normally distributed negative or positive value at each evaluation occasion. A well-formedness or harmony score (H) is calculated for each candidate according to the formula in (4), where w_i is the weight of constraint C_i, N_i is the noise associated with C_i, and C_i(cand) is the number of violations that a candidate earns in terms of C_i, expressed in negative whole numbers. That candidate with the highest H is selected as output.

(4) $$H(\text{cand}) = [(w_1 + N_1)(C_1(\text{cand}))] + [(w_2 + N_2)(C_2(\text{cand}))] + \ldots + [(w_n + N_n)(C_n(\text{cand}))]$$

Because of the contribution of noise, constraints with very similar weights can flip around in terms of which contributes more to H between consecutive occasions of using the grammar, and consequently cause variation. This is illustrated in (5). These tableaux show the grammar for the Tejano *t/d*-deletion pattern reported above in Table 5.1.3, as developed by Coetzee and Pater (forthcoming) and Coetzee and Kawahara (forthcoming). The four constraints used are a general markedness constraint *Ct (no word-final clusters that end in t/d), a general anti-deletion constraint Max, and two positional versions of Max that protect against deletion specifically in pre-vocalic and pre-pausal positions. Due to the different noise values

[5] Though see van Oostendorp (1997), Boersma and Hayes (2001: 82–3), and Hammond (2004) for suggestions of how these models may be augmented to allow this.

at the two evaluation occasions, the deletion candidate is selected in (5a) and the non-deletion candidate in (5b).

(5) (a)

‘west bank’	*w*	*N*	*w*	*N*	*w*	*N*	*w*	*N*	
	100.4	–0.2	99.6	0.2	3.0	–0.3	0.8	–0.1	
	*CT		MAX		MAX-PRE-V		MAX-PRE-PAUSE		*H*
west bank	–1								–100.2
☞ wes bank			–1						–99.8

(b)

‘west bank’	*w*	*N*	*w*	*N*	*w*	*N*	*w*	*N*	
	100.4	–0.4	99.6	0.5	3.0	–0.4	0.8	0.2	
	*CT		MAX		MAX-PRE-V		MAX-PRE-PAUSE		*H*
☞ west bank	–1								–100.0
wes bank			–1						–100.1

The model, as implemented above, is still purely grammatical without formal incorporation of non-grammatical influences. Coetzee and Kawahara (forthcoming; Coetzee 2009a, b) propose an augmentation that would incorporate non-grammatical factors. In order to illustrate this augmented HG model of variation, I use usage frequency as an example of a non-grammatical factor. Frequency is chosen here for convenience and not out of principle. Since frequency is quantitative, it is easier to incorporate it into a quantitative model of variation than some other factor, such as speech style, that is not inherently quantitative. However, in principle, even something like style could be expressed in quantitative terms—see Boersma and Hayes (2001: 82–3) for suggestions about how this might be done. Once a factor has been transformed into a quantitative measure, it can be handled exactly as frequency is handled here.

The proposal is that the weight of faithfulness constraints can be scaled up or down by some constant factor determined by extra-grammatical factors. When the faithfulness constraints are scaled up, deletion becomes less likely, and this is therefore associated with non-grammatical contexts that inhibit deletion—given the relation between frequency and deletion in Figure 5.1.4, this could be what happens when a low-frequency word is evaluated. The opposite happens when faithfulness weights are scaled down. With addition of scaling factors, the formula for calculating *H* has to be augmented as in (6), where *S* stands for the scaling factor, *F* for a faithfulness constraint, and *M* for a markedness constraint. Application of the expanded model is illustrated in (7), where the same grammatical settings (weights and noise values) are used under two different scaling conditions. In (7a) weights are scaled up with infrequent *jest*, and non-deletion is selected. In

(7b) weights are scaled down with frequent *just*, and deletion is selected. See Coetzee (2009a) for a more detailed discussion of this model, and for discussion of how the value of the scaling factor S is determined. See also Coetzee and Kawahara (forthcoming) for a different proposal about how to determine the value of the scaling factor.

(6) $$\begin{aligned} H(\text{cand}) =& [(w_{F1} + N_{F1} + S)(F_1(\text{cand}))] + [(w_{F2} + N_{F2} + S)(F_2(\text{cand}))] \\ &+ \cdots + [(w_{Fn} + N_{Fn} + S)(F_n(\text{cand}))] + [(w_{M1} + N_{M1})(M_1(\text{cand}))] \\ &+ [(w_{M2} + N_{M2})(M_2(\text{cand}))] + \cdots + [(w_{Mm} + N_{Mm})(M_m(\text{cand}))] \end{aligned}$$

(7) (a)

'jest'	*w*	*N*	*w*	*N*	*S*	*w*	*N*	*S*	*w*	*N*	*S*	
	100.4	0.4	99.6	0.2	0.5	3.0	–0.3	0.5	0.8	–0.2	0.5	
	*Ct		Max			Max-Pre-V			Max-Pre-Pause			*H*
☞ jest	–1											–100.8
jes			–1						–1			–101.4

(b)

'just'	*w*	*N*	*w*	*N*	*S*	*w*	*N*	*S*	*w*	*N*	*S*	
	100.4	0.4	99.6	0.2	–0.5	3.0	–0.3	–0.5	0.8	–0.2	–0.5	
	*Ct		Max			Max-Pre-V			Max-Pre-Pause			*H*
just	–1											–100.8
☞ jus			–1						–1			–99.4

Although this model allows both grammatical and non-grammatical factors to impact variation, it affords primacy to grammar. Pre-pausal deletion violates a superset of the constraints (Max, Max-Pre-Pause) violated by pre-consonantal deletion (Max). A pre-pausal deletion candidate will therefore always have a lower H than a pre-consonantal deletion candidate, and deletion will consequently always be less likely in pre-pausal position. Weight scaling has no impact on this. Additionally, the weights of all faithfulness constraints are scaled by the same value so that the relative weight differences between these constraints also remain unaffected. In the example here, the weight of Max-Pre-Pause (0.8) is lower than that of Max-Pre-V (3.0), corresponding to the fact that deletion is more likely pre-pausally than pre-vocalically. Since the weights of both constraints are equally scaled, Max-Pre-V will always have a higher weight than Max-Pre-Pause, and there will always be more deletion in pre-pausal position.

Other models that give non-grammatical factors primacy over grammatical factors are possible. The usage-based (Bybee 2001) and exemplar (Pierrehumbert 2001a) models of phonology are examples. In these models, information about the frequency of words, phonemes, etc. (i.e. not a part of grammar in the strictest sense)

occupies a central position, with grammar either emerging from frequency or used only as a last resort when frequency cannot explain some pattern.

There are many, often conflicting, models for incorporating the influence of multiple grammatical and non-grammatical factors on phonological variation. At the moment, data to differentiate between these options are still largely lacking. In fact, this is a prime example of where data collected in the laboratory are needed for phonological theory to make progress.

5.1.5 Concluding remarks

Ohala (1986) argued that data collected through laboratory-based experimentation is the most reliable data that phonologists could use for the testing and development of phonological theories. At that time, they also lamented that theoretically oriented phonologists rarely use such data. In the twenty-five years since, the situation has changed. Primarily because of the increased interest in variation in theoretically oriented generative phonology, laboratory-based research has become common even in these circles. This has led to recent significant overlap between the fields of traditional theoretical phonology and laboratory phonology. The increased use of laboratory data in theoretical phonology has aided in the development of phonological theory. These data have enabled phonologists to test aspects of their theories that could not be tested with the type of data on which theoretical phonologists traditionally relied. But laboratory data also poses new challenges to phonological theory. Since this type of data is characterized by variation and gradience, phonological theory now has to account for more than categorical, obligatory phenomena. This intersection between phonological theory and laboratory phonology is an area of phonological research where much progress is possible and likely in the next several decades.

5.2 Modeling phonological variation*

Arto Anttila

In early generative phonology, the problem of phonological variation was not high on the research agenda. While important progress was being made on the

* This piece was written while I was a fellow at the Stanford Humanities Center in 2008–9. I thank Hideki Zamma, an anonymous reviewer, and the editors for valuable comments. All errors are mine.

quantitative analysis of variation in sociolinguistics (e.g. Cedergren and Sankoff 1974), this work did not have much influence on the mainstream phonological theory of the day. The situation started to change in 1993 when Alan Prince and Paul Smolensky made their influential proposal to replace ordered rules by ranked constraints. Optimality Theory (Prince and Smolensky 1993/2004), as the new theory was called, was not itself a theory of variation, but soon turned out to provide new ways to integrate variable and quantitative patterns into the theory of grammar. The past fifteen years have seen a remarkable growth of activity in this area, both within and outside Optimality Theory. For recent surveys, see e.g. Anttila (2007a) and Coetzee and Pater (forthcoming). These theoretical developments have been fueled with simultaneous methodological developments, including the emergence of phonologically annotated speech and text corpora, computational tools for work on learnability, and developments in experimental psycholinguistics. This has resulted in the broadening of the empirical base of phonological theory and brought generative phonology closer to the concerns of the research community known as laboratory phonologists. In this section, we will briefly review some optimality-theoretic tools that have been used to describe and explain phonological variation and quantitative patterns. Concrete illustrations will be drawn from variable *t/d*-deletion in English and variable word stress in Finnish.

5.2.1 Variation in Optimality Theory

Variation can be defined as a situation where one meaning corresponds to multiple forms. Consider the following examples of *t/d*-deletion in English and *t*-deletion in Finnish:

(1) (a) /kɔst mi:/ = 'cost me' (b) /professori-i-ta/ = 'professor-PL-PAR'

In the English example, /t/ is variably deleted in a complex syllable coda (see e.g. Guy 1980, 1991a, 1991b; Guy and Boberg 1997; Labov 1997). The frequency of deletion is sensitive to a number of factors, including whether the following segment is a consonant or a vowel and whether the deleted segment is the past-tense morpheme /d/. In the Finnish example, a short /t/ is variably deleted between light unstressed syllables (see e.g. Keyser and Kiparsky 1984; Kiparsky 2003) in certain suffixes, such as the partitive.

In what appears to have been the first treatment of variation in Optimality Theory, Kiparsky (1993) proposed an analysis of English *t/d*-deletion. He started by assuming that the stops /t, d/ are preserved when syllabified, else deleted, an instance of STRAY ERASURE (see e.g. Blevins 1995: 223–4). The analysis is set up as

follows: given an input and a set of candidate outputs, the task of the grammar is to find the <input, output> mapping that best satisfies the phonological constraints. For example, assume the input /kɔst mi:/ and three output candidates where the input is syllabified in different ways: [kɔst][mi:] (= no deletion) which has a complex coda; [kɔs]t[mi:] (= deletion) which has no complex margins because /t/ is left unsyllabified; and [kɔs][tmi:] (= resyllabification) which has a complex onset. Kiparsky's (1993) grammar includes the following four constraints hypothesized to be universal:

(2) Constraints (based on Kiparsky 1993):

*Complex	A syllable margin is not complex.
Onset	A syllable has an onset.
Parse-Segment	A segment belongs to a syllable.
Align-Left-Word	A word-initial segment is also syllable-initial.

All these constraints cannot be satisfied simultaneously. For example, Parse-segment can be satisfied by syllabifying all segments, but that violates *Complex. Conversely, *Complex can be satisfied by not syllabifying all segments, but that violates Parse-segment.[6] Optimality Theory resolves such conflicts by ranking the constraints in a language-specific manner. Consider the ranking Parse-Segment >> Onset >> *Complex >> Align-L-W:

(3) Input: *t* followed by a consonant

		Parse-Seg	Onset	*Complex	Align-L-W
kɔst mi:	(a) → [kɔst][mi:]			1	
	(b) [kɔs]t[mi:]	1!			
	(c) [kɔs][tmi:]			1	1!

The candidates are evaluated starting with the highest-ranking (= leftmost) constraint. Parse-Segment is violated once by candidate (b). The fact that some candidates do not violate Parse-Segment makes this violation fatal and eliminates (b) from competition. This is indicated by an exclamation point (!). The irrelevance of the lower-ranking constraints is shown by graying out the remaining cells. The evaluation continues with the remaining candidates (a) and (c). Onset and *Complex do not bring us any closer to the solution because they are unable to distinguish (a) and (c): neither violates Onset and both violate *Complex. The tie is resolved by Align-L-W which is violated by (c), making (a) optimal. The optimal candidate is indicated by an arrow (→). A closer look shows that (c) can never

[6] In addition, a full analysis needs to guarantee that /t/ cannot form a syllable nucleus in English, ruling out [kɔs][t][mi:]; that the cluster cannot be resolved by vowel epenthesis, ruling out [kɔ][sət][mi:]; etc.

be optimal because it incurs a superset of (a)'s violations: both candidates violate *Complex, but (c) violates Align-L-W as well and can thus never win, no matter how the constraints are ranked. Such doomed candidates are called harmonically bounded.

How can variation arise in this model? Under the multiple grammars theory adopted by Kiparsky (1993) variation arises from different constraint rankings. Assume an individual with Kiparsky's four constraints, but no information about their ranking. There are 4! = 24 possible total rankings: 12 predict [kɔst][mi:] (= no deletion); these are the total rankings that conform to the partial ranking Parse-Segment >> *Complex; 12 predict [kɔs]t[mi:] (= deletion); these are the total rankings that conform to the reverse partial ranking *Complex >> Parse-Segment. The rankings are thus evenly split between the two candidates. Assuming that all rankings are equally likely to be selected, this is the quantitative pattern we expect to see in the data.

(4) Predicted output frequencies assuming no rankings

Optimal candidate	Total rankings	Predicted frequency
[kɔst][mi:]	12 rankings	50%
[kɔs]t[mi:]	12 rankings	50%

The situation is different if the *t* is followed by a vowel:

(5) Input: *t* followed by a vowel

		Parse-Seg	Onset	*Complex	Align-L-W
kɔst ʌs	(a) [kɔst][ʌs]		1!	1	
	(b) [kɔs]t[ʌs]	1!	1		
	(c) → [kɔs][tʌs]				1

This time, none of the three candidates is harmonically bounded. The predicted quantitative pattern is also more interesting than in the pre-consonantal case:

(6) Predicted output frequencies assuming no rankings

Optimal candidate	Total rankings	Predicted frequency
[kɔst][ʌs]	5 rankings	21%
[kɔs]t[ʌs]	5 rankings	21%
[kɔs][tʌs]	14 rankings	58%

The multiple grammars theory of variation has served as the foundation for most variationist work in Optimality Theory. The claim is that variation within languages and across languages has the same source: differences in ranking. This makes the prediction that patterns of quantitative variation within individuals should be identical to patterns of typological variation across languages. The generic multiple grammars theory explored by Kiparsky (1993) was further developed in

e.g. Reynolds (1994), Anttila (1997), and Anttila and Cho (1998) who proposed various restrictions on possible ranking relations. A different approach was taken in Stochastic Optimality Theory (StOT, Boersma 1997; Boersma and Hayes 2001) which introduced numerically weighted constraints, but otherwise retained the standard assumptions of Optimality Theory. A concrete illustration will be given shortly. A very different development is the Rank-Ordering Model of EVAL (Coetzee 2004, 2006) which aims at predicting the relative frequencies of variants without addressing their absolute frequencies.

5.2.2 Quantitative typologies

A theory of grammar can be tested by checking how well it predicts the observed linguistic facts. This involves two conceptually distinct tasks. First, one must figure out what kinds of linguistic patterns the theory predicts to be possible and what kinds of linguistic patterns it excludes as impossible. This is purely theoretical work that can be carried out prior to knowing the empirical facts. Second, once the predictions are known, one can check whether they agree with the observations. In the case of a theory of variation, we can check whether these predictions are borne out by the quantitative data.

The *t/d*-deletion grammar fragment predicts that certain quantitative relationships should hold across inputs. For example, a grammar with no rankings predicts that *t/d*-deletion should be more common before consonants than before vowels (50 percent vs. 21 percent), and that resyllabification before vowels should be more common than *t/d*-deletion before consonants (58 percent vs. 50 percent). The question now arises how stable such predictions are under additional rankings. Suppose the speaker acquires evidence for the ranking ALIGN-L-W >> ONSET. How do the quantitative predictions change? The answer is given in (7).

(7) The effect of adding one pairwise ranking into the grammar

	OPTIMAL CANDIDATE	NO RANKINGS	ALIGN-L-W >> ONSET
a.	[kɔst][mi:]	50%	50%
	[kɔs]t[mi:]	50%	50%
b.	[kɔst][ʌs]	21%	42%
	[kɔs]t[ʌs]	21%	42%
	[kɔs][tʌs]	58%	17%

The ranking ALIGN-L-W >> ONSET changes the quantitative predictions selectively: *t/d*-deletion is still more common before consonants than vowels (50 percent vs. 42 percent), but resyllabification before vowels is now less common than *t/d*-deletion before consonants (17 percent vs. 50 percent), reversing the original pattern. This suggests that only some quantitative relationships are stable under additional rankings.

The stable relationships can be found as follows. First, we figure out all the distinct languages that can be obtained by ranking the four constraints in all possible ways. This can be easily done using OTSoft (Hayes et al. 2003). All in all, the twenty-four total rankings yield the four distinct languages shown in (8). This is called the FACTORIAL TYPOLOGY. Optimal candidates with *t/d*-deletion are grayed out.

(8) Factorial typology

	Output #1	Output #2	Output #3	Output #4
/kɔst mi:/:	[kɔst][mi:]	[kɔs]t[mi:]	[kɔst][mi:]	[kɔs]t[mi:]
/kɔst ʌs/:	[kɔst][ʌs]	[kɔs]t[ʌs]	[kɔs][tʌs]	[kɔs][tʌs]

The factorial typology shows that the distribution of *t/d*-deletion is wider before consonants than before vowels. The pattern can be stated as a TYPOLOGICAL ENTAILMENT: if *t/d*-deletion applies before vowels ([kɔs]t[ʌs]) it also applies before consonants ([kɔs]t[mi:]). This statement is true for all four languages.

Typological entailments have important consequences for the quantitative patterns predicted by the multiple grammars theory where a grammar is a set of total rankings drawn from the factorial typology. Each ranking belongs to one of the types 1–4. Suppose we combine four rankings: one of type 1, two of type 2, and one of type 4. The resulting mix will yield 75 percent *t/d*-deletion before consonants, but only 50 percent *t/d*-deletion before vowels.

(9) A variable grammar under the multiple grammars theory

	Before C	Before V
Output #1	[kɔst][mi:]	[kɔst][ʌs]
Output #2	[kɔs]t[mi:]	[kɔs]t[ʌs]
Output #2	[kɔs]t[mi:]	[kɔs]t[ʌs]
Output #4	[kɔs]t[mi:]	[kɔs][tʌs]
Deletion %	75%	50%

The upshot is that it is not possible to construct a grammar that would predict more *t/d*-deletion before vowels than before consonants. This prediction is a QUANTITATIVE UNIVERSAL that holds true no matter how the constraints are ranked. A closer inspection of the factorial typology reveals a second typological entailment: if the syllable and the morpheme are perfectly aligned before vowels [kɔst][ʌs] they are perfectly aligned before consonants [kɔst][mi:]. We call the set of all typological entailments in a grammar a TYPOLOGICAL ORDER, or T-ORDER, described in Figure 5.2.1 as a directed graph.

Finally, resyllabification before vowels <kɔst ʌs, [kɔs][tʌs]> has a special status in the grammar. As can be verified from the factorial typology, it neither entails nor is entailed by any other mapping. This means that the grammar predicts nothing about its quantitative behavior with respect to other mappings. We call such mappings FREE NODES.

Formal work on typological entailments is pursued by Prince (2002a, 2002b, 2007). Particularly relevant is Prince (2006) which discusses typological entailments in terms of ELEMENTARY RANKING CONDITIONS (ERCs) and contains an extended discussion of variable *t/d*-deletion in English based on Kiparsky's (1993) analysis. Empirical studies that make use of typological entailments in the context of quantitative data include Anttila (2008a), Anttila et al. (2008), and Anttila et al. (2010).

5.2.3 Variation in Finnish word stress

5.2.3.1 *Empirical generalizations*

Finnish word stress exhibits variation that is interesting in several ways. First, the variation occurs in relatively "abstract" phonology, namely foot structure, which

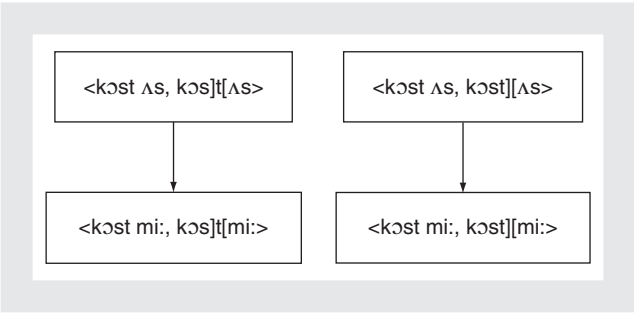

Figure 5.2.1. Typological entailments (T-order) predicted by a syllabification-based grammar of English *t/d*-deletion.

does not always have direct phonetic correlates. Second, the variation pattern is embedded within a categorical pattern, showing that both types of generalizations exist and must be accounted for by the same grammar. Third, several phonological factors interact cumulatively in variation. The small pattern of variation illustrated here is part of a larger pattern studied in Anttila (1997, 2007b, 2008b). Obtaining reliable data on secondary stress can be difficult; see e.g. De Lacy (2007) for discussion. The present study addresses this problem by using segmental alternations as stress diagnostics. The fact that these alternations are represented in the standard Finnish orthography makes it possible to test the analysis on the abundantly available written data, including the Internet (see Loehr and Van Guilder, this volume). The quantitative facts reported here are based on data retrieved from Finnish websites using the Google search engine, approximately nine million word tokens in all.[7]

The basic rule for Finnish word stress is simple: primary stress falls on the initial syllable and secondary stress falls on every other syllable after that. In more theoretical terms, Finnish has trochaic (left-headed) feet, assigned from left to right, with main stress on the leftmost foot (see e.g. Keyser and Kiparsky 1984; Hanson and Kiparsky 1996; Elenbaas 1999; Elenbaas and Kager 1999; Kiparsky 2003; Karvonen 2005; Karttunen 2006).

(10) Finnish word stress: the binary pattern:
  (a) (ká.las)(tè.let) 'you are fishing'
  (b) (ká.las)(tè.le)(mì.nen) 'fishing'
  (c) (íl.moit)(tàu.tu)(mì.nen) 'registering'

The alternating binary pattern is interrupted by occasional ternarity: if the third syllable is light and the fourth heavy, stress falls on the fourth syllable. Syllables of the form (C)V count as light, whereas (C)VC, (C)VV, (C)VVC, and (C)VCC count as heavy.

(11) Finnish word stress: the ternary pattern
  (a) (ká.las.te)(lèm.me) 'we are fishing'
  (b) (íl.moit)(tàu.tu.mi)(sès.ta) 'from registering'

The ternary pattern can be explained by the Weight-to-Stress Principle (WSP): unstressed heavy syllables are avoided (Prince 1990; see also Anttila 1997; Kiparsky 2003; Karvonen 2005). The stress pattern *ká.las.te.lèm.me* (one unstressed heavy) is better than **ká.las.tè.lem.me* (two unstressed heavies), hence the initial dactyl.

In Finnish, stress interacts with various segmental processes. One such process is *t*-deletion (see e.g. Keyser and Kiparsky 1984) which deletes a singleton /t/ between

[7] Instead of typing in word forms one by one we used the "Query Google" software (see e.g. Hayes and Londe 2006) programmed by Timothy Ma of UCLA to query the search engine automatically.

two unstressed light syllables. One possible interpretation of this pattern is that Finnish deletes /t/ outside a foot. This is illustrated in (12): the partitive suffix surfaces as /-ta/ after monosyllabic plural stems, but as /-a/ after disyllabic plural stems.

(12) (a) /maa-i-**ta**/ (mái.**ta**) 'country-PL-PAR'
(b) /talo-i-**ta**/ (tá.lo)**ja** 'house-PL-PAR'

The analysis predicts that four-syllable stems should exhibit invariant *t*-deletion if the third syllable is heavy. This is because secondary stress must fall on the heavy third in order to satisfy the WSP. This prediction is confirmed:

(13) Heavy third syllable, invariant pattern: /margariini-i-ta/ 'margarine-PL-PAR'
(már.ga)(rìi.ne)ja/*(már.ga.rii)(nèi.ta)

What happens if the third syllable is light? There are two ways to satisfy the WSP: an initial trochee with *t*-deletion or an initial dactyl without *t*-deletion. Both are attested:

(14) Light third syllable, variable pattern: /professori-i-ta/ 'professor-PL-PAR'
(pró.fes)(sò.re)ja ~ (pró.fes.so)(rèi.ta)

Examples (13) and (14) illustrate a simple but important general point: phonological variation emerges in contexts where the categorical phonological principles remain silent. Here the categorical principle is the WSP. This shows that in order to understand where variation may and may not occur we must understand the categorical phonology of the language.

The choice between the variants in (14) is not free, but controlled by a more subtle layer of phonology. We observe two preferences:

(15) Preferences:
(a) Light syllables with /a, ä, o, ö/ prefer to be stressed; light syllables with /e, i, u, y/ prefer to be unstressed.[8]
(b) Stress avoids falling next to a heavy syllable.

The first preference seems related to the cross-linguistic observation that low vowels attract stress and high vowels repel stress (Kenstowicz 1996; de Lacy 2002a; Crowhurst and Michael 2005). The fact that the phonetically mid vowels /o, ö/ count as low and the phonetically mid vowel /e/ counts as high seems to reflect

[8] Finnish vowels can be classified in terms of binary features as follows (see e.g. Karlsson 1982: 52):
a = [−high, +low, +back, −round], ä = [−high, +low, −back, −round],
o = [−high, −low, +back, +round], ö = [−high, −low, −back, +round],
e = [−high, −low, −back, −round], i = [+high, −low, −back, −round],
u = [+high, −low, +back, +round], y = [+high, −low, −back, +round].

their morphophonemics: in Finnish, /o, ö/ alternate with the low vowel phonemes /a, ä/ whereas /e/ alternates with the high vowel phoneme /i/. The second preference can be interpreted as a version of clash prohibition: adjacent prominent syllables are avoided.

We now illustrate the effects of vowel sonority and prominence clash in certain four-syllable stems that end in a light syllable. In such stems, stress placement depends on the weight of the second syllable, the weight and sonority of the third syllable, and the sonority of the fourth syllable. Weight is either H (= heavy) or L (= light); vowel sonority is either A (= low vowel, high sonority) or I (= high vowel, low sonority). First, let us examine cases where the third syllable is heavy. The *t*-deletion patterns and the corresponding foot structures are shown in (16). The outcome is a virtually invariant pattern: initial trochee with *t*-deletion. The column labeled GHITS shows the number of Finnish-language websites containing a partitive plural form of the relevant type.

(16) Four-syllable stems: predictions and observations
/edustusto/ 'representation', /termostaatti/ 'thermostat', /affrikaatta/ 'affricate', /margariini/ 'margarine'

	TYPE	EXAMPLE	DELETION%	GHITS
(a)	HHA	(é.dus)(tùs.to)ja	100.0	92,308
(b)	HHI	(tér.mos)(tàat.te)ja	100.0	80,063
(c)	LHA	(áff.ri)(kàat.to)ja	99.7	13,039
(d)	LHI	(már.ga)(rìi.ne)ja	100.0	392,942

The corpus is large enough to contain counterexamples in each category. It is not difficult to think up special explanations for many of them. Consider the following two unexpected examples where *t*-deletion has not applied:

(17)

TYPE	ATTESTED FORM	GHITS	
LHA	politiikoita	30	'politics-PL-PAR'
LHA	mekaniikoita	11	'mechanics-PL-PAR'

In both cases, a special explanation is readily available: the unexpected forms look like production errors triggered by the phonologically and semantically closely related (*pó.lii.ti*)(*kòi.ta*) 'politician-PL-PAR' and (*mé.kaa.ni*)(*kòi.ta*) 'mechanic-PL-PAR' where the second syllable is heavy, the third light, and the initial dactyl without *t*-deletion entirely expected. These forms can thus be plausibly attributed to the analogical influence of related lexical items.

Next, we examine cases where the third syllable is light. The *t*-deletion patterns and the corresponding foot structures are shown in (18). This time, we see variation: initial trochee with *t*-deletion and initial dactyl without *t*-deletion.

(18) Vowel sonority and prominence clash at work

	Type	Example 1	Example 2	Deletion%	Ghits
(a)	LAI	(fí.lo)(sò.fe)ja	(fí.lo.so)(fèi.ta)	90.7%	23,595
(b)	HAI	(pró.fes)(sò.re)ja	(pró.fes.so)(rèi.ta)	84.9%	34,612
(c)	LIA	(gál.le)(rì.o)ja	(gál.le.ri)(òi.ta)	1.0%	91,598
(d)	HIA	(ál.ler)(gì.o)ja	(ál.ler.gi)(òi.ta)	0.3%	190,416

/filosofi/ 'philosopher', /professori/ 'professor', /galleria/ 'gallery', /allergia/ 'allergy'

The examples in (18) illustrate the cumulative interaction of vowel sonority and prominence clash. In (a), both favor *t*-deletion which is observed 90.7% of the time. In (d), both disfavor *t*-deletion which is observed 0.3% of the time. The remaining examples occupy the middle ground: *t*-deletion is favored only by vowel sonority in (b) and only by prominence clash in (c).

5.2.3.2 *Analysis*

We now outline an optimality-theoretic analysis of the Finnish stress pattern. We will only consider candidates that satisfy the following undominated constraints:

(19) Undominated constraints

(a)	Trochee	Feet are left-headed.
(b)	Maxϕ	No deletion within a foot.
(c)	Parse-stem	Stem segments are footed.
(d)	*Unary	No monosyllabic feet.
(e)	*Lapse	Every weak beat must be adjacent to a strong beat or a foot edge. (cf. Elenbaas and Kager 1999: 282)

These constraints reduce the number of foot structures under consideration to four. In particular, Parse-stem guarantees that only suffix segments may remain unfooted and *Unary and *Lapse limit us to binary and ternary feet. We start by checking the predictions for stems with a heavy third syllable. Three dominated constraints are posited: WSP ranks at the top; *t militates against singleton stops; and Max militates against segment deletion. The candidate that best satisfies the WSP is optimal; the ranking of *t and Max is underdetermined by the data. There is no variation. Since stop deletion in stems is blocked by the undominated constraints Maxϕ and Parse-stem, we only mark violations of *t for suffixes.

(20) Constraints

(a)	WSP	Heavy syllables are stressed.
(b)	*t	No singleton stops.
(c)	Max	No deletion.

(21) Invariant stems

HHA	/edustusto-i-ta/	WSP	*T	MAX
	a. (é.dus)(tùs.toi)ta	2!	1	
	b. (é.dus.tus)(tòi.ta)	2!	1	
	c. (é.dus)(tùs.toi.ta)	2!	1	
	d. → (é.dus)(tùs.to)ja	1		1
HHI	/termostaatti-i-ta/	WSP	*T	MAX
	a. (tér.mos)(tàa.tei)ta	2!	1	
	b. (tér.mos.taa)(tèi.ta)	2!	1	
	c. (tér.mos)(tàa.tei.ta)	2!	1	
	d. → (tér.mos)(tàat.te)ja	1		1
LHA	/affrikaatta-i-ta/	WSP	*T	MAX
	a. (áf.fri)(kàa.toi)ta	1!	1	
	b. (áf.fri.kaa)(tòi.ta)	1!	1	
	c. (áf.fri)(kàa.toi.ta)	1!	1	
	d. → (áf.fri)(kàat.to)ja			1
LHI	/margariini-i-ta/	WSP	*T	MAX
	a. (már.ga)(rìi.nei)ta	1!	1	
	b. (már.ga.rii)(nèi.ta)	1!	1	
	c. (már.ga)(rìi.nei.ta)	1!	1	
	d. → (már.ga)(rìi.ne)ja			1

We now turn to stems where the third syllable is light. This time, two candidates satisfy the WSP. In order to account for the weak phonological effects that skew the quantitative pattern, we introduce four more constraints. All these constraints must rank below the WSP, but their mutual ranking remains to be found.

(22) Vowel sonority and prominence clash constraints

(a) *a/ä/o/ö No unstressed light syllable with /a, ä, o, ö/ nucleus.

(b) *é/í/ú/ý No stressed light syllable with an /e, i, u, y/ nucleus.

(c) *H.X No stress next to a heavy syllable.

(d) PARSE-σ Syllables belong to feet.

(23) Variable stems

HAI	/professori-i-ta/	WSP	*T	MAX	P-σ	*H.X	*a/ä/o/ö	*é/í/ú/ý
	(pró.fes)(sò.rei)ta	2!	1		1	3	1	
	(pró.fes)(sò.rei.ta)	2!	1			3	1	
	→ (pró.fes.so)(rèi.ta)	1	1			1	2	
	→ (pró.fes)(sò.re)ja	1		1	1	2	1	
LAI	/filosofi-i-ta/	WSP	*T	MAX	P-σ	*H.X	*a/ä/o/ö	*é/í/ú/ý
	(fí.lo)(sò.fei)ta	1!	1		1	1	2	1
	(fí.lo)(sò.fei.ta)	1!	1			1	2	1
	→ (fí.lo.so)(fèi.ta)		1				3	1
	→ (fí.lo)(sò.fe)ja			1	1		2	1
HIA	/allergia-i-ta/	WSP	*T	MAX	P-σ	*H.X	*a/ä/o/ö	*é/í/ú/ý
	(ál.ler)(gì.oi)ta	2!	1		1	3	1	1
	(ál.ler)(gì.oi.ta)	2!	1			3	1	1
	→ (ál.ler.gi)(òi.ta)	1	1			1	1	
	→ (ál.ler)(gì.o)ja	1		1	1	2	2	1
LIA	/galleria-i-ta/	WSP	*T	MAX	P-σ	*H.X	*a/ä/o/ö	*é/í/ú/ý
	(gál.le)(rì.oi)ta	1!	1		1	1	1	1
	(gál.le)(rì.oi.ta)	1!	1			1	1	1
	→ (gál.le.ri)(òi.ta)		1				1	
	→ (gál.le)(rì.o)ja			1	1		2	1

We now model the quantitative pattern using Stochastic Optimality Theory (StOT, Boersma 1997; Boersma and Hayes 2001). In StOT, each constraint is associated with a real-number RANKING VALUE. Thus, we may have three constraints A, B, C with ranking values like A = 108.956, B = 97.664, and C = 96.650. During candidate evaluation, ranking values are disturbed by "noise" which is a random normally distributed positive or negative value temporarily added to the ranking value. The resulting SELECTION POINTS determine the constraint ranking used in the actual evaluation. The presence of noise causes the selection points to oscillate around the fixed ranking values from evaluation to evaluation. In our example (A = 108.956, B = 97.664, C = 96.650), we will often expect to see the ranking A >> B >> C which corresponds to the ordering of the ranking values, but also A >> C >> B because the ranking values of B and C are close together and easily reversed by the noise. The ranking C >> B >> A is also possible, but much less likely. While StOT uses numerically weighted constraints, it can nevertheless be viewed as an instance of the multiple grammars theory (see e.g. McCarthy 2008: 263): there is a single grammar in the competence (= a set of fixed ranking values), but multiple grammars in performance (= different constraint rankings).

Stochastic Optimality Theory has an associated learning algorithm called the GRADUAL LEARNING ALGORITHM (GLA) which is designed to learn StOT grammars. The algorithm is described by Boersma and Hayes (2001). We took the above constraints and ran the GLA using OTSoft (Hayes et al. 2003) with the goal of learning a StOT grammar that would match the Finnish quantitative pattern. The match was consistently very good. In the test run reported below, the grammar was tested for 2000 cycles, with an average error per candidate of 0.87 percent.

(24) Ranking Values Found

112.000	WSP
108.956	*é/í/ú/ý
102.336	*T
97.664	PARSE-σ
97.664	MAX
96.650	*X.H
96.248	*a/ä/o/ö

(25) Matchup to Input Frequencies

/HAI/	Input Fr.	Gen Fr.	Gen. #
(pro.fes)(so.re)ja	0.849	0.897	1793
(pro.fes.so)(rei.ta)	0.151	0.104	207
/LAI/	Input Fr.	Gen Fr.	Gen. #
(fi.lo)(so.fe)ja	0.907	0.909	1817
(fi.lo.so)(fei.ta)	0.093	0.092	183

/HIA/	Input Fr.	Gen Fr.	Gen. #
(al.ler.gi)(oi.ta)	0.997	0.993	1985
(al.ler)(gi.o)ja	0.003	0.008	15
/LIA/	Input Fr.	Gen Fr.	Gen. #
(gal.le.ri)(oi.ta)	0.990	0.993	1985
(gal.le)(ri.o)ja	0.010	0.008	15
/HHA/	Input Fr.	Gen Fr.	Gen. #
(e.dus)(tus.to)ja	1.000	1.000	2000
(e.dus)(tus.toi)ta	0.000	0.000	
(e.dus.tus)(toi.ta)	0.000	0.000	
(e.dus)(tus.toi.ta)	0.000	0.000	
/HHI/	Input Fr.	Gen Fr.	Gen. #
(ter.mos)(taat.te)ja	1.000	1.000	2000
(ter.mos)(taa.tei)ta	0.000	0.000	
(ter.mos.taa)(tei.ta)	0.000	0.000	
(ter.mos)(taa.tei.ta)	0.000	0.000	
/LHA/	Input Fr.	Gen Fr.	Gen. #
(af.fri)(kaat.to)ja	0.997	1.000	2000
(af.fri)(kaa.toi)ta	0.000	0.000	
(af.fri.kaa)(toi.ta)	0.000	0.000	
(af.fri)(kaa.toi.ta)	0.003	0.000	
/LHI/	Input Fr.	Gen Fr.	Gen. #
(mar.ga)(rii.ne)ja	1.000	1.000	2000
(mar.ga)(rii.nei)ta	0.000	0.000	
(mar.ga.rii)(nei.ta)	0.000	0.000	
(mar.ga)(rii.nei.ta)	0.000	0.000	

5.2.3.3 *Quantitative typology*

What does our theory predict to be possible and what does it exclude as impossible? Such questions are routinely asked in generative linguistics in the context of qualitative data. The same questions can be asked in the context of quantitative data as well. In this section, we will explore these questions in terms of our grammar fragment for Finnish.

We first compute the T-order for our grammar. However, this time the grammar is large enough to render manual methods impractical. For this reason, we used T-order Generator (Anttila and Andrus 2006), a free open-source Python program for computing and visualizing T-orders. All in all, given the eight inputs, the theory derives fifty-four entailments shown in Figure 5.2.2 as a directed graph. The bottom node of the graph collapses four nodes into one box. These are the invariant stems that typologically entail one another: all have an invariant initial trochee

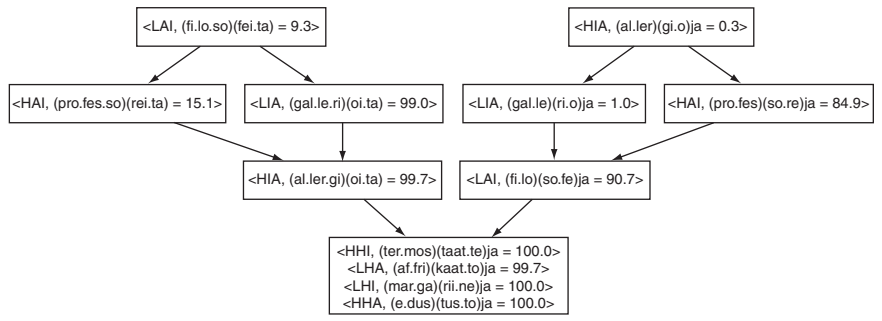

Figure 5.2.2. Typological entailments (T-order) predicted by a stress-based grammar of Finnish *t*-deletion.

with *t*-deletion. The multiple grammars theory correctly predicts that the relative frequencies of variants should increase from top to bottom. Only (*áff.ri*)(*kàat.to*)*ja* in the bottom node is slightly out of line.

Imagine we discovered a new dialect of Finnish where the empirical frequencies of the variants were reversed. In such a dialect, (*fí.lo*)(*sò.fe*)*ja* would occur 9.3 percent of the time and (*fí.lo.so*)(*fèi.ta*) 90.7 percent of the time, the opposite of the pattern in actual Finnish. In such a dialect, the relative frequencies of the variants would decrease from top to bottom in the T-order. Would our theory be able to describe this dialect?

In order to find out the answer, we reversed the frequencies and ran the GLA. In order to make the learning task easier, we only included the variable stems, omitting the invariant stems. The match was consistently very bad. The resulting StOT grammars predicted that all *t*-deletion variants should have the same empirical frequency around 50 percent (say, 53 percent), and all *t*-retention variants should have the same empirical frequency around 50 percent (say, 47 percent). The average error per candidate was large: 21.8 percent. The cause of this descriptive failure is not hard to see: StOT is limited by the factorial typology and hence the T-order. This shows that we do not just have a framework that can replicate any empirical numbers, natural or unnatural, but a theory with empirical content. In other words, Optimality Theory, our constraints, and the multiple grammars theory together impose strict limits on possible and impossible quantitative patterns in natural languages.

5.2.4 Further questions

This brief overview has focused on phonological conditions on phonological variation. In reality, phonological variation often involves multiple factors, including

internal factors such as morphology, syntax, and lexical identity, as well as external factors such as age, gender, style, register, identity, ethnicity, and socioeconomic class. The most straightforward way of dealing with non-phonological factors in an optimality-theoretic analysis is to include them in the grammar as so many independent constraints, as in Kiparsky's (1993) original analysis of English *t/d*-deletion where morphological constraints were assumed to interact with phonological constraints in the same hierarchy. Alternatively, one can imagine a modular solution where the phonological grammar is independent of the rest of the system and the interactions are modeled in some way that does not necessarily involve constraint ranking. Different ways of describing morphological and lexical conditioning in Optimality Theory are discussed by e.g. Itô and Mester (1995, 1999), Inkelas et al. (1997), Kiparsky (2000), Pater (2000, 2009a), Zuraw (2000), Anttila (2002a), and many others. Recent usage-based models of grammar such as Exemplar Theory have emphasized the role of lexical frequency as a factor in phonological variation (see e.g. Fidelholtz 1975; Hooper 1976b; Phillips 1984, 2001, 2006; Johnson 1997b; Myers and Guy 1997; Pierrehumbert 2001a, this volume; Bybee 2001, 2002; Jurafsky et al. 2001; Gahl and Yu 2006 eds., among others). Coetzee (2009b) discusses one possible way of embedding lexical frequency effects into the phonological grammar.

Optimality Theory assumes that grammatical constraints interact in terms of strict ranking. This hypothesis has been challenged on both empirical and learning-theoretic grounds. Alternative proposals usually involve numerically weighted constraints as in Harmonic Grammar (Legendre et al. 1990; Legendre et al. 2006; Coetzee and Pater forthcoming; Pater 2009b) and Maximum Entropy models (Goldwater and Johnson 2003; Jäger and Rosenbach 2006; Jäger 2007; Hayes and Wilson 2008). These numerical models also include the original Variable Rules model (Cedergren and Sankoff 1974; Paolillo 2000, 2002), an early example of the use of logistic regression in the analysis of variation; see e.g. Baayen (2008, this volume) and Johnson (2008) for recent developments in statistical computing for linguistics. The question of how grammatical constraints interact depends on what the grammatical constraints are in the first place and we can expect the debate to continue vigorously as we learn about different types of phonological variation in different languages.

In the current state of the field, there is no lack of theoretical alternatives. The main obstacle appears to be the dearth of linguistically sophisticated empirical resources such as speech and text corpora annotated for phonological variables. For methodologies involving this type of data, see Cole and Hasegawa-Johnson (this volume) and Loehr and Van Guilder (this volume). Such resources are necessary for the detailed evaluation of theories, but developing them is an expensive, labor-intensive, and time-consuming task. Despite recent advances in data-rich linguistics much work remains to be done.

CHAPTER 6

MESSAGE-RELATED VARIATION

SEGMENTAL WITHIN-SPEAKER VARIATION

MIRJAM ERNESTUS

TONAL VARIATION

YIYA CHEN

The authors in this chapter provide parallel discussions of segmental and tonal variations. In each case attention is given to the nature of categorical vs. gradient effects and the question of the degree to which effects are automatic consequences of the production/perception system vs. under speaker control. Ernestus discusses the case of segmental variation, focusing on the rich and complex literature on assimilation and reduction. Chen discusses tonal variation, focusing on coarticulatory effects in languages with lexical tones and global effects of the prosodic encoding of information.

6.1 Segmental within-speaker variation

Mirjam Ernestus

6.1.1 Introduction

It has long been known that a word's pronunciation may be different in connected speech than when carefully produced in isolation. The differences concern the segmental as well as the suprasegmental properties of words. This section focuses on segmental adaptations, while Chen (this chapter) discusses tonal variation.

Segmental differences largely result from the adapation of word-initial and word-final segments to adjacent segments (assimilation) and from reduction (segment lenition and deletion). Within the traditional phonological framework, these types of pronunciation variation were mostly investigated on the basis of data obtained from introspection and impressionistic observation. New technical developments of the last few decades, such as the enormous increase in computer memory and the ability to analyze speech files automatically, have made it easier to study pronunciation variation on the basis of large quantities of real speech (see Cole and Hasegawa-Johnson, this volume). Moreover, the ability to store and search large speech corpora helps researchers find stimuli for psycholinguistic experiments, which facilitates the study of the comprehension process as well.

A strong indication of the importance of these new studies is the finding that simple phonological processes described in the literature, such as assimilation, are not as pervasive as had been thought, while other processes are much more frequent. For instance, Dilley and Pitt (2007) studied place assimilation of alveolar segments to following bilabial or velar segments (e.g. the pronunciation of *green boat* as *gree*[m b]*oat*) in a corpus of American English, a process that has been described as highly productive (e.g. Harris 1994). Contra standard analyses, they found that place assimilation is relatively rare and that deletion, glottalization, or canonical pronunciations of the alveolar consonant are more frequent. Similarly, Ernestus et al. (2006) reported that in a corpus of Dutch, obstruents followed by voiced plosives show regressive voice assimilation (e.g. *we*/t + b/*oek* is pronounced as *we*[db]*oek* 'law book'), as described in many theoretical studies, but they also found that these clusters can show deletion of the first obstruent (*we*[b]*oek*), and most importantly, progressive voice assimilation (*we*[tp]*oek*), a process that has been claimed not to apply to such clusters in Dutch (e.g. Booij 1995).

These new types of empirical studies also show that reduction processes, which had received only limited attention in the phonological literature, are widely attested in informal speech styles. Traditional phonological descriptions mention reduction rules, such as full vowel to schwa reduction and /t/-deletion (see e.g. Booij 1995 for Dutch), but underestimate the frequency and variety of reduction.

Deletion of unstressed vowels in American English, for instance, affects 25 percent of the possible word tokens even in news interviews on television (Dalby 1984), and in casual conversations 25 percent of all word tokens show lenition or deletion of at least one segment (Johnson 2004). Moreover, speakers delete complete syllables, a phenomenon hardly described at all in the phonological literature, resulting in pronunciations like [pʰɛrī] for English *apparently* and [wɛs] for Dutch *wedstrijd* /wɛtstrɛit/ 'contest' (see Johnson 2004 for English; Kohler 1990 for German; and Ernestus 2000 for Dutch examples). These reduction processes lead to a vast number of pronunciation variants for one and the same word, as exemplified by the Dutch word *natuurlijk* [natyrlək] 'of course,' which may be pronounced as [natylək], [ntylək], [ntyk], [tyrlək] (which also has an orthographic representation), [tylək], [tylk], [tyk], [tyg], [dyk], and [dyg], among others (Ernestus 2000).

All these studies provide evidence that pronunciation variation at the segmental level is much more pervasive and above all more complex than previously thought. The detailed characteristics of pronunciation variants and the factors conditioning these variants have important implications for linguistic and psycholinguistic theory. Below, I first discuss the articulatory and acoustic properties of pronunciation variation, focusing on the theoretically important difference between categorical and gradient processes (Section 6.1.2); and how recent findings can be accounted for within generative grammar (Section 6.1.2), Articulatory Phonology, and exemplar-based models (also Section 6.1.3). Then, I discuss different accounts of how listeners process segmental pronunciation variation, including underspecification theory, a perceptual account, and a learning model (Section 6.1.4). Finally, I give an overview of the most important variables conditioning segmental pronunciation variation, which informs phonological theory and psycholinguistic theories of speech production and comprehension (Section 6.1.5).

6.1.2 Categorical versus gradient variation

Traditionally, categorical variation is distinguished from gradient variation. Variation is categorical if it can be well described with the categorical values of phonological features (e.g. [+voice] and [−voice]). Variation is gradient if the acoustic characteristics of the variants reflect values between these categorical values (e.g. partly voiced). The distinction between categorical and gradient variation is theoretically important, since within the generative framework it distinguishes between phonological and phonetic processes: Phonological processes are considered to be categorical while phonetic processes are typically gradient (e.g. Keating 1990b; Cohn 1993). These definitions of phonological and phonetic processes replace earlier definitions that see phonological processes as language-specific and phonetic processes as language-universal and automatically resulting from articulatory

mechanisms (e.g. Kenstowicz and Kisseberth 1979). The distinction between categorical phonological processes and gradient phonetic processes plays a role in both linguistic and psycholinguistic (e.g. Levelt 1989) models.

Within the generative framework (starting with Chomsky and Halle 1968), processes such as assimilation, vowel reduction, and segment deletion are generally assumed to be categorical. Assimilation involves the "spreading" of a phonological feature from one segment to another segment, and this latter segment subsequently cannot be distinguished from segments that have the same feature value in their underlying specifications. For example, [m] has exactly the same surface phonological representation and phonetic characteristics in *a ru*[m p]*icks you up* when the speaker intends *run* or *rum*. Similarly, vowel reduction implies replacement (or deletion) of the phonological features specifying the vowel's quality, and this vowel consequently cannot be distinguished from underlying schwas. Finally, segment deletion implies the complete loss of a segment in the surface phonological representation. Thus, if words differ only in the presence versus absence of a segment in their underlying representations (e.g. English *sport* and *support*, or *miss* and *mist*), deletion of this segment leads to identical pronunciations (see Coetzee, this volume for further discussion of deletion).

Detailed articulatory and acoustic studies have cast serious doubt on the classification of many connected speech processes as categorical. The evidence for gradient rather than categorical variation is especially strong for place assimilation, since the exact location of the closures for plosives can relatively easily be measured by means of electropalatography (EPG, Hardcastle 1972). Several EPG studies have shown that palatalization (as in *hi*/t j/*ou*) may be gradient, with the obstruent becoming more palatal over time, which distinguishes a palatalized obstruent from underlying palatals (e.g. Barry 1992 for Russian; Zsiga 1995 for post-lexical palatalization in American English). The same type of gradience has been documented for place assimilation of /t/ and /d/ in American English (e.g. in *la*/t k/*alls*): These obstruents often start out alveolar and only then gradiently assimilate to the place of articulation of the following consonant (Nolan 1992). Similar results, but showing considerable inter- and intraspeaker variation, have been found for place assimilation of alveolar nasals in American English (e.g. in *gree*/n b/*oat*, Ellis and Hardcastle 2002). All these data suggest that place assimilation, especially assimilation crossing word boundaries, cannot be simply accounted for by the spreading of a phonological place feature from one segment to another.

Several acoustic studies suggest that voice assimilation may be gradient as well. For instance, Ernestus et al. (2006) have shown that Dutch obstruent clusters expected to be subject to regressive voice assimilation (which voices the initial segment) may be produced without any glottal vibration, with glottal vibration throughout the whole cluster, or during only part of the cluster. Even though many acoustic characteristics co-determine the perceptual voiced-voiceless distinction, this result is telling since glottal vibration is considered the most important cue to voicing in Dutch obstruent clusters (van den Berg 1986). In addition, voice

assimilation appears gradient since, instead of affecting all acoustic characteristics cueing the voiced-voiceless distinction, it may affect only some of them. This results in segments that are neither completely voiceless nor completely voiced. For instance, Kuzla et al. (2007) showed that progressive devoicing assimilation in German (e.g. in *ha*/t v/*älder* 'had woods') results in shorter intervals of glottal vibration, while it hardly affects the duration of the fricative, which is the most important cue to the voice (or fortis/lenis) specification of fricatives in German. Similar results were obtained by Ernestus and colleagues (2006), who showed that Dutch obstruent clusters expected to be subject to regressive voice assimilation tend to be shorter (cueing more perceptual voicing), to be produced with longer periods of glottal vibration (also cueing more perceptual voicing), but also with longer release noises (cueing less perceptual voicing) in words of high compared to low lexical frequencies (probably because speakers produce high-frequency words with less articulatory effort). Consequently, in more frequent words, some acoustic characteristics signal more and others less voicing. These data provide additional evidence that assimilation is more complex than the spreading of a phonological feature.

In addition to assimilation, many reduction processes appear gradient. For instance, vowels may show any realization between unreduced pronunciation variants (with formants distinguishing the vowels maximally from the other vowels in the language) and schwas (e.g. Mooshammer and Geng 2008). Simultaneously, vowels may vary in their duration from values typical for accented full vowels to zero, showing all durations in between (e.g. gradient deletion of the first vowel of a sequence of two in Plains Cree, Russell 2008). They may thus have clear, some, or no cues in the acoustic signal. Also, obstruents may show several types of reduced realizations in addition to being fully present or absent (e.g. Mitterer and Ernestus 2006 for /t/ in Dutch). In many cases, consonant deletion thus appears to be the natural endpoint of gradient reduction processes, rather than to result from categorical phonological processes.

Taken together, these studies suggest that most connected speech processes are gradient and thus, according to the definition of the phonological component as containing only categorical processes, they belong to the phonetic component. In other words, the new findings move most of post-lexical phonology from the phonological to the phonetic component in the generative framework. These findings therefore raise the questions of whether the division within generative grammar between phonology and phonetics should be revisited again and how mechanisms responsible for gradient variation should be formalized.

6.1.3 Processing models naturally incorporating gradient variation

The gap between the phonological (i.e. categorical) and physical (i.e. gradient) structure of speech in generative models has stimulated the development of new

models. These alternative models make fundamentally different assumptions and naturally incorporate gradience in pronunciation variation.

One of these models is Articulatory Phonology (see Gafos and Goldstein, this volume), developed by Browman and Goldstein (e.g. 1986, 1992). It assumes that phonological representations consist of abstract articulatory gestures, rather than segments or features. Articulatory Phonology can account for many phenomena that are well explained by non-linear phonology, since the temporal alignment of gestures may be changed, which may result in the overlap of different gestures in time. For instance, regressive place assimilation in *gree*[m b]*oat* may result from the early onset of the bilabial closure, before the realization of the preceding nasal, which then hides the alveolar closure. Importantly, this retiming of gestures may likewise account for gradient assimilation. In addition, it may explain the complete absence of segments in the acoustic signal. For instance, a word-final /t/ may appear absent before bilabial stops (as in *perfec*/t m/*emory*), because speakers close their lips before the /t/ is released, which makes the release of the /t/ inaudible (Browman and Goldstein 1990a). The hypothesis that speakers may produce the articulatory gestures for inaudible segments, as assumed in these accounts, is supported by several X-ray studies (e.g. Browman and Goldstein 1992). Finally, gestures may be reduced in size, which results in the lenited realizations often encountered in casual speech, also a form of gradience. This notion of size reduction, however, has only recently begun to be robustly modelled in Articulatory Phonology (see Gafos and Goldstein, this volume).

The assumption that gestures may overlap in time and be reduced in size, even to zero, makes Articulatory Phonology a very powerful theory. It can account for the absence of any acoustic cue under any condition. Obviously, research is necessary to properly constrain the theory such that it accounts only for those pronunciation variants that really occur. Furthermore, if lexical representations consist of abstract gestures, listeners should extract these gestures from the acoustic signal. Some data suggest that this is indeed what listeners do (e.g. Fowler et al. 2003), but other experiments cast doubt on these results. For instance, Mitterer and Ernestus (2008) found that the speed with which participants shadow words containing the phoneme /r/ is independent of whether participants produce the /r/ with different or with the same articulatory gestures as those used in the words. Furthermore, participants imitate phonetic detail more closely if it is phonologically relevant. These results are unexpected if the basic units of lexical representations are gestures rather than more abstract phonological symbols. More research is necessary also to settle this issue.

Another type of model naturally incorporating the gradience of pronunciation variation is the exemplar-based model (see Chapter 8 this volume). Exemplar models assume that the mental lexicon contains a representation for every pronunciation variant of a word (possibly even one for every token ever heard or

uttered by the language user), with detailed information about all phonetic properties of the variant. Johnson (2004), for instance, following Klatt (1979), proposed that lexical representations can be considered as sequences of spectra with no categorical information at all. The assumption of different lexical representations for pronunciation variants is supported by both production and comprehension data showing that speakers' and listeners' response latencies are affected by the frequency of the given pronunciation variant compared to the frequencies of the other variants for the same word (Ranbom and Connine 2007; Bűrki et al. 2010). These results demonstrate that language users store frequency information about pronunciation variants, which suggests that they store the variants themselves as well. Further specificity within exemplar-based models is necessary to clarify to what extent an actual realization needs to be in line with a corresponding stored exemplar and to what extent its phonetic detail may result from the phonetic implementation of an exemplar. Moreover, future studies have to implement exemplar-based models computationally to test which additional assumptions may be necessary to account for the full range of available data (Ernestus forthcoming).

6.1.4 Comprehension of pronunciation variation

The comprehension of pronunciation variation may be accounted for within the processing models mentioned above (see Nguyen, this volume and Holt, this volume for discussion of the perception of canonical pronunciation variants). Psycholinguistic models based on generative models may assume that the acoustic input is reconstructed to the canonical pronunciation stored in the mental lexicon by means of rules or phonological constraints (e.g. Boersma 1998). This reconstruction may be based, for instance, on the grouping of feature cues distributed over time (Gow 2003). Articulatory Phonology assumes that listeners retrieve the underlying gestures from the gradient acoustic input, while exemplar-based models assume that the mental lexicon contains representations for many pronunciation variants and that an acoustic input is recognized if it is sufficiently similar to one of the stored exemplars. In addition to these models, several other mechanisms have been proposed to account for the comprehension of pronunciation variation.

Underspecification theory (Lahiri and Reetz 2002) assumes, like most models in the tradition of generative grammar, that the mental lexicon contains only one lexical representation for every word (see Lahiri, this volume). In order to explain the recognition of words with assimilated segments (such as *green* in *gree*[m b]*oat*), the theory assumes that phonological features subject to assimilation (e.g. the place feature of alveolar nasals in English) are lexically unspecified and do not contribute

to word recognition. Thus, assimilation does not hinder word recognition, as it does not lead to mismatches with stored phonological representations. Underspecification theory is supported by language acquisition data, which show that young children confuse some words and not others, which is taken as evidence that some phonological segments are lexically underspecified (Fikkert 2005). In contrast, underspecification theory is challenged by perception studies showing that listeners only recognize a pronunciation variant that may result from assimilation if it occurs in the appropriate segmental context (Gaskell and Marslen-Wilson 1996, 1998; Gow 2002). Thus, listeners recognize *gree*[m] as *green* only before bilabial plosives, as in *gree*[m b]*oat*.

Another account for the comprehension of assimilated segments assumes that the human auditory system is not highly sensitive to the differences between assimilated and non-assimilated segments in assimilation contexts (e.g. between *gree*[n] and *gree*[m] if followed by *boat*). As a consequence, assimilation does not pose problems for comprehension. This account explains the role of segmental context in comprehension and receives experimental support from several studies using simple discrimination tasks (e.g. Mitterer, Csépe, and Blomert 2006) and event-related potentials in the brain (e.g. Mitterer and Blomert 2003). Especially convincing is the finding that listeners with native languages that differ in whether they contain the assimilation process under investigation are equally bad in discriminating between assimilated and unassimilated segments in appropriate assimilation contexts (Gow and Im 2004; Mitterer, Csépe, Honbolygo, and Blomert 2006). This account of the comprehension of assimilated segments can be extended to the comprehension of acoustically weak segments. Mitterer et al. (2008) showed that insensitivity of the auditory system can partly explain listeners' ease in the processing of /st/-final words produced with acoustically weak /t/s.

Yet another mechanism that may contribute to the comprehension of pronunciation variation is listeners' learning of language-specific patterns, as advocated by Gaskell and Marslen-Wilson (1998; Gaskell 2003). These researchers showed that listeners are better at inferring the citation form of an assimilated pronunciation variant (i.e. *gree*[n] from *gree*[m b]*oat*) if the word is an existing word in the language (such as *green*) rather than a pseudo word (such as *breen*). Familiarity with language-specific patterns may also explain why Dutch listeners are slightly better than Japanese listeners in discriminating between some pronunciation variants of word-final /st/-clusters: Such clusters are frequent in Dutch, whereas they are phonotactically illegal in Japanese (Mitterer et al. 2008).

In conclusion, the literature contains several proposals, most of them supported by experimental data, to account for the comprehension of pronunciation variation. Probably, the comprehension process results from the interaction of several mechanisms (as also concluded in Mitterer et al. 2008) and further research

should show how these mechanisms interact. Interestingly, some of the relevant mechanisms, including the sensitivity of the human auditory system, are not part of the grammar and are therefore traditionally assumed not to be interesting for linguistic theory. However, in order to define the contributions of linguistic mechanisms to speech behavior, we have to know the contributions of the extragrammatical mechanisms, which together with the grammar will provide us with a complete picture of human speech processing.

6.1.5 Factors conditioning pronunciation variation

All complete models of human language processing should account for the conditions under which pronunciation variation is likely to occur and is best understood by listeners. Data on these conditions are therefore crucial for linguistic and psycholinguistic theories. Whereas data on the comprehension of segmental variation are still relatively scarce, much more is known about conditions favoring the production of non-canonical forms.

Among these conditions, speech style is probably the most important: Non-canonical pronunciation variants tend to be more frequent in less formal speech. For instance, place assimilation of alveolar plosives to velar plosives in English is more common in less formal speech styles (e.g. Kerswill 1985; Barry 1992), and highly reduced pronunciation variants, such as [pʰɛrĩ] for *apparently*, are attested only in truly casual speech. Less formal speech is mostly uttered at relatively high speech rates, which may put speakers under time pressure. Speakers may attenuate this time pressure by deleting segments or by reducing the sizes of articulatory gestures and overlapping them more in time. Speech rate may therefore explain some of the phonetic characteristics of non-canonical forms attested in less formal speech. Importantly, however, a high speech rate does not necessarily lead to non-canonical forms, as documented, among others, by van Son and Pols (1990, 1992). It is therefore speech style rather than speech rate that conditions pronunciation variation, but speech rate may co-determine the type of non-canonical forms occurring in less formal speech.

Another important factor conditioning pronunciation variation is the prosodic structure of the utterance (see Frota, this volume and Turk, this volume). Consonants in the initial position of prosodic domains, such as the intonational phrase or the intermediate prosodic phrase, tend to be longer and to be produced with greater linguopalatal contact (initial strengthening, e.g. Fougeron and Keating 1997; Keating et al.) than consonants in domain-medial or domain-final positions. Domain-initial segments also appear less sensitive to connected speech processes, since initial consonants show less voice assimilation (Kuzla et al. 2007) and vowels in initial syllables show less coarticulation with neighboring vowels (Cho 2004). Among

the non-initial segments, domain-final ones tend to be longer than medial ones, a phenomenon called final lengthening (e.g. Wightman et al. 1992). Several studies suggest that listeners may use these prosodic strengthening cues in comprehension (e.g. Cho et al. 2007).

Less well documented, but probably an equally relevant factor is the word's function in the discourse. Local (2003) reported that the word combination *I think* is more reduced when it occurs in sentence-final position (even though final lengthening would be expected in these positions) and conveys a pragmatic meaning (e.g. in the sentence *they should be here by the time you come out next weekend I think*) than when it is followed by the complementizer *that* and has above all a lexical meaning (e.g. *I think that people have not yet woken up*). Similarly, Plug (2005) reported that the degree of reduction of the Dutch word *eigenlijk* 'actually' depends on whether the word signals contrast with what has been suggested before by the speaker or by the listener. So far, only a few studies have investigated the role of pragmatic function; future studies are needed to determine the exact mechanisms driving the effects and whether their relevance may be restricted to only a few word types.

A final important predictor of a word's pronunciation is its predictability within the context. Words tend to be more reduced when they are more predictable given the preceding or following words. For instance, Scheibman and Bybee (1999) reported that the English word sequence *don't* tends to be produced with a smaller number of segments if preceded by *I*, the word that most frequently precedes *don't*. Likewise, the sequence is more reduced before the words that most often follow *don't* (*know*, *think*, *mean*). In general, function words, like *don't*, are more reduced (in duration and in number of segments) the more predictable they are given the preceding words (e.g. Pluymaekers et al. 2005; Bell et al. 2009). Content words, in contrast, tend to be more reduced if they are more predictable given the following words (e.g. Bell et al. 2009). Thus, the English word *previous* has a higher probability of being reduced when followed by *year* than by *beer*. In addition, words tend to be shorter if they have been mentioned in the conversation before (e.g. Fowler and Housum 1987; Aylett and Turk 2004).

These predictability effects may automatically result from the production process: More predictable words are easier to plan and therefore do not require speakers to slow down their speech rate, which may result in reduction (see e.g. Pluymaekers et al. 2005; Bell et al. 2009). A planning account of the predictability effects has the advantage that it easily explains why reduction degree for content words is correlated especially with the predictability of the following word: Speech rate is determined by the planning of the next words rather than that of the preceding or current words. In addition, this account predicts correctly that words are less reduced if they are followed by hesitations, which indicate planning problems (e.g. Jurafsky et al. 2001).

In contrast to this speaker-driven account are two listener-driven accounts. The first one assumes that speakers would like to reduce their articulation effort as much as possible but adapt their reduction degree to the listeners' needs in order to guarantee smooth communication (in line with the Hyper- and Hypo-articulation theory by Lindblom 1990). More predictable units are easier to understand and speakers would therefore reduce especially highly predictable units. The second listener-driven account states that reduction degree facilitates comprehension as it indicates which information is given or predictable (e.g. Boersma 1998). The hypothesis that speakers adapt their degree of reduction to the listeners' needs is supported by the finding that in Dutch, English, German, and Italian, segmental sequences of medium durations are attested more frequently in corpora of spontaneous speech than sequences of relatively short duration. Probably the short combinations are more difficult to identify and are therefore used less often by the speaker (Kuperman et al. 2008). In contrast, listener-driven accounts are challenged by the observations by Bard and colleagues (Bard et al. 2000) that the second mention of a word tends to be more reduced independently of whether the listener has heard that word before in the conversation. Probably the documented predictability effects are both speaker- and listener-driven and future research has to investigate the relative relevance of the different mechanisms.

In conclusion, factors of very different natures appear to condition pronunciation variation. Speech-processing models can only account for the full range of data if they take the many different aspects of speech (including grammatical form, semantics, pragmatics, and planning) into account.

6.1.6 Conclusions

The last few decades have produced many linguistic studies based on corpus data and psycholinguistic experimentation. These studies have above all shown that speech is much more variable and gradient than has traditionally been assumed: natural speech shows more pronunciation variants than previously assumed, some well-known variants occur less often than expected in favor of others, and many variants show gradient properties. Moreover, speech processing involves mechanisms of very different natures (involving, among others, pragmatic function and speech planning) that appear to interact. So far, no existing model of speech processing can account for all findings. Further studies on the processing of pronunciation variation are necessary to formulate and evaluate comprehensive models of both speech production and comprehension. Since the mechanisms involved appear to be of very different natures, these studies will benefit from the multidisciplinary effort of the laboratory phonology approach.

6.2 Tonal variation*

Yiya Chen

6.2.1 Introduction

In connected speech, various processes occur when speech sounds are grouped into words, phrases, or larger chunks, leading to realizations of sounds that deviate from their forms in isolation. This section complements Ernestus (this chapter) by examining tonal variation, a type of pronunciation variation at the supra-segmental level. Here tone is broadly defined as pitch variation that distinguishes not only lexical meanings, as defined by Zsiga (this volume), but also pragmatic meanings (see Yip 2002; Gussenhoven 2004; and Ladd 2008 for discussions of the various linguistic and paralinguistic functions of pitch variation). In particular, the section focuses on two important sources of tonal variation. At the local level, I will deal with the effect of coarticulation between adjacent lexical tones; at a more global level, I will examine tonal variation due to discourse contexts.

6.2.2 Local tonal variation—coarticulation of adjacent tones

Just as consonants and vowels are sensitive to context and show considerable influence from neighboring segments (see e.g. Hardcastle and Newlett 1999), lexical tones are subject to coarticulatory perturbation.[1]

Instrumental work along this line of research investigates the influence of tones on the phonetic characteristics of tones in adjacent syllables, with primary focus on

* I would like to thank Vincent van Heuven, Jessie Nixon, an anonymous reviewer, and the editors Abby Cohn, Cécile Fougeron, and Marie Huffman, for their questions, comments, and suggestions. Preparation for this contribution was supported by the VIDI grant from the Netherlands Organization for Scientific Research (NWO-VIDI 016084338) and the ERC Starting Independent Researcher Grant from the European Research Council (RPPV-206198).

[1] M. Chen (2000: 27) states that "there is no essential difference between tone sandhi and tonal coarticulation, except that tone sandhi processes are perceptible to the (trained but) unaided ears, and therefore more likely to be reported by fieldworks and integrated to a greater extent into the phonological component of the grammar." I believe that there are different types of tonal alternation processes. It is important to conduct detailed instrumental studies to understand the nature of these various types of contextual tonal variation within and across Chinese dialects so as to understand how exactly they are conditioned by phonetic, phonological, and morphosyntactic environments. For this section, I limit my attention to tonal variations that are conventionally defined as phonetic tonal coarticulation, and refer readers to M. Chen (2000) for cases of tonal variation that are conventionally defined as categorical tonal alternation, commonly known as tone sandhi.

Asian tonal languages (but see Laniran 1992 for tonal coarticulation in Yoruba).[2] Results of earlier studies show that lexical tones coarticulate, but distinctive tonal contours are usually well preserved; they also suggest language-specific differences in the directionality, magnitude, and temporal extent of tonal coarticulation (e.g. Han and Kim 1974 for Vietnamese; Abramson 1979 for Thai; Shih 1987 and Shen 1990 for Mandarin Chinese).

With the development of quantitative procedures in tracking continuous f0 contours as well as adoption of more rigorous experimental control and statistical methods, studies involving larger numbers of subjects have been conducted (e.g. Gandour et al. 1994 and Gandour et al. 1996 on Thai; Peng 1997 on Taiwanese; Xu 1994 and 1997 on Standard Chinese). Results of these studies reveal more similarity than difference in tonal coarticulation across languages. Three general patterns have emerged from this body of literature. First, tonal coarticulation is bidirectional, with both carry-over and anticipatory effects, similar to segmental coarticulation. Second, the influence of carry-over effects is much greater than anticipatory effects, in terms of the number of tonal contexts subject to coarticulation and the temporal extent of their effects. Carry-over effects are observed across all tonal categories while anticipatory effects seem to be present mostly in tones that end with high f0 (at least for Thai in Gandour and his colleagues' work and Standard Chinese in Xu's work). Furthermore, carry-over effects extend well beyond the center of the following syllable while anticipatory effects seem to be more confined to the offset of the preceding syllable. Third, the nature of these two coarticulatory effects differs. Carry-over influence is generally assimilatory. Thus a high-ending tone raises the onset of the following tone. Anticipatory effects tend to be dissimmilatory where a low ending tone raises the preceding tone offset (but see Peng 1997 for anticipatory assimilation in Taiwanese).

It is worth noting that while more recent multi-speaker studies are able to reveal the general patterns of tonal coarticulation, they may have averaged out possible speaker-intrinsic variation. Therefore, in addition to the long-pursued question of cross-linguistic tonal coarticulatory patterns, the new question that arises is to what extent speakers within the same language, either within or across dialects, may vary in the coarticulation of lexical tones.

6.2.2.1 *Factors conditioning tonal coarticulation*

Tonal coarticulation varies as a function of prosodic factors, just like segmental coarticulation (Cho 2004; Ernestus, this chapter; among others). For example,

[2] This is probably due to the fact that Asian languages typically have syllables as the tone-bearing unit and their tones are less mobile than African tones (see section 6.2.1 in Yip 2002 on tonal mobility). Thus, to produce smooth connected speech, Asian tonal language speakers are more likely to coarticulate tones.

when a rising tone in Standard Chinese is in a prosodically weak position (i.e. unstressed at the phrasal level) and embedded in a high-ending and low-starting tonal context, the rising tone is sometimes produced with level or slightly falling f0 contour (Shih and Sproat 1992; Xu 1994). This is different from the rising tone in prosodically strong position where distinctive f0 rising contour typically manifests itself. Prosodic strength also affects the temporal extent of tonal coarticulation. Chen and Xu (2006) examine the effect of lexical tones on neutral tone, which is generally considered to occur in unstressed syllables. They show that influence of lexical tones is not only present in the immediately following neutral tone but also, in cases involving more than one neutral tone, extends into the second and third neutral tone. This, on the surface, is in contrast to results reported for Standard Chinese in Xu (1997) and for Thai in Potisuk et al. (1997) where no significant tonal coarticulation effect was found beyond the immediately adjacent syllables. The lack of a chained, long-distance tonal coarticulatory effect in these studies is likely due to the fact that the lexical tones examined are in prosodically stronger positions (stressed at the lexical or phrasal level) than the neutral-tone syllables and so only the immediately adjacent syllables manifest coarticulatory effects.

We know that the prosodic organization of a language typically includes different levels of grouping, such as prosodic word vs. intonational phrase (see Frota, this volume for details on prosodic structure). Segmental coarticulation is known to vary as a function of the position of target segments within a particular prosodic domain (e.g. initial vs. medial in a phrase) as well as their prominence status (such as stressed vs. unstressed syllable of words or accented vs. unaccented syllable of intonational phrases; in e.g. Cho 2002). Further work is needed to understand how coarticulation of lexical tones is conditioned in different prosodic positions.

Much more work is also needed to understand other factors that condition tonal coarticulation. Segmental studies show languages differ in their coarticulatory patterns (e.g. Öhman 1966). In particular, linguistic contrasts play a role in determining the presence as well as the magnitude of coarticulation (e.g. Manuel 1990, 1999). Interesting patterns of language-specific tonal spaces have been observed. For example, Shi et al. (1987) report that the Gaoba Dong language, a Tai-Kadai language spoken by the Dong ethnic group in China, has five level tones, the acoustic distribution of which, however, cannot be fit linearly into the five-scale pitch system developed by Chao (1930) for pitch patterns of lexical tones. This suggests that lexical tones are not evenly distributed within speakers' pitch range. In Cantonese, Ling and Li (2008) show that the perceptual space of three level tones, commonly transcribed as 55, 33, and 22, consists of different perceptual ranges of f0 values, schematized in Figure 6.2.1. Thus, the question is how intrinsic differences in tonal acoustic and perceptual spaces within and across languages might affect tonal coarticulation.

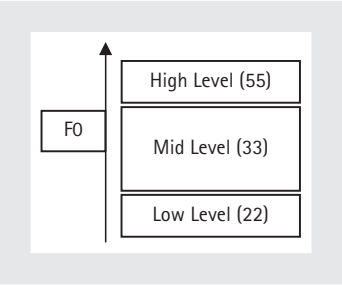

Figure 6.2.1. Schematic representation of the perceptual space of the three level tones (High, Mid, Low) in Cantonese, commonly transcribed as 55, 33, and 22.

6.2.2.2 *Previous accounts of tonal coarticulation*

Different explanations have been offered for tonal coarticulation. One commonly recognized factor is the law of motion and constraints of physiology on articulators in speech production. Pitch variation is produced mainly via laryngeal muscles which place limitations on the maximum speed of pitch changes. Xu and Sun (2002) report that speakers of Standard Chinese need, on average, 142 ms to complete a pitch rise of six semitones. Thus, given a two-syllable Low-High tone sequence, it takes more than half of a syllable for the speakers to finish the transition from the preceding Low to the target High tone. This explains very well the carry-over assimilation effect which is often observed at the beginning of the following tone-bearing syllable.

Carry-over assimilation may also be due to inertia of laryngeal articulators. A case in point is the peak delay phenomenon in Beijing Mandarin. Xu (2001) shows that Rising tones are consistently realized with an f0 peak at the beginning of the following syllable, regardless of speaking rate. This makes it less likely that such peak delay is due to the short duration of a Rising-tone-bearing syllable. Rather, the delayed peak is probably due to the momentum in the implementation of the Rising-tone gesture within the Rising-tone syllable.

Based on neutral tone data from Standard Chinese, Chen and Xu (2006) propose another account for tonal variability in connected speech—strength of articulatory force in tonal implementation. The reduced articulatory force of neutral tone in Standard Chinese (due to its weak prosodic position) means that it takes longer than lexical tones to overcome the influence of the preceding lexical tone, as well as more time for the neutral tone target to be implemented. In other words, the laryngeal muscles need more time to change the speed of vibration so that a

particular pitch range is traversed and a particular pitch target is approximated, which results in great variability in the f0 realization of neutral tones.

Note that the above accounts of coarticulation exclude the possible role of pre-planning in speech production. Whalen (1990) shows that segment coarticulation is, for the most part, planned. For tonal coarticulation, Gandour et al. (1993) have explained the anticipatory dissimilation effect as pre-planned. They propose that tonal dissimilation occurs in order to facilitate pitch transition from a region at one end of the pitch range to the opposite end, similar to a trailer swinging wide to make a sharp turn. Gandour et al. (1996) take a further step and propose, based on data from Thai speakers with unilateral brain damage, that tonal coarticulation in both directions is controlled by long-range planning mechanisms.

Perceptual mechanisms have also been posited as an explanation for tonal coarticulation. Gandour et al. (1993) and Potisuk et al. (1997) speculate that anticipatory dissimilation may in part be motivated by a desire to maximize the perceptual distance between adjacent tones or to generate new acoustic cues for enhanced tonal perception. Peng (1997), with data from Taiwanese, confirms that listeners are indeed able to utilize coarticulatory cues in tonal identification.

Further studies are needed to pin down the exact mechanisms underlying tonal coarticulation. It is possible that both automatic mechanical processes and controlled articulatory planning coexist (for production and/or perceptual reasons). The question is when and how these different processes interact. It is also important to compare tonal and segmental coarticulation. Various models have been proposed to explain segmental coarticulation (see Farnetari and Recasens 1999), none of which, however, has taken into account tonal coarticulatory effects. A model that captures both commonalities and differences in tonal vs. segmental coarticulation has yet to appear.

6.2.2.3 *Asian vs. African tonal variation*

The tonal coarticulatory effects observed in Asian languages bear some similarities to tonal processes reported in African tonal languages. In this section, we discuss a few specific examples to illustrate their similarities and differences, as well as questions that remain for cross-linguistic comparisons.

The anticipatory dissimilation effect, whereby f0 of a high tonal target is raised when preceding a low-starting tone, parallels the High raising effect reported for Yoruba (Connell and Ladd 1990; Akinlabi and Libermann 1995; Laniran and Clements 2003) and other tonal languages (e.g. Rialland and Somé 2000, for Dagara). What differs is that High raising in African languages exhibits more variability than anticipatory dissimilation reported in the aforementioned studies. In Yoruba, Laniran and Clements (2003) show the effect of High raising is relatively small and speaker-specific; High raising also interacts with downstep (i.e. the lowering of High tone after a Low tone) in determining the pitch span of lexical tones. In

Dagara, Rialland and Somé (2000) show that the effect of High raising is robust and that its magnitude is roughly proportional to the number of following downsteps within the utterance, suggesting salient long-distance effects and considerable pre-planning. Furthermore, the effect of High raising in some African languages has been reported to undergo phonologization, giving rise to a new surface-contrastive super-high tone (e.g. Snider 1990, for Krachi; Hyman 1993, for Kirimi; Rialland 2001, for Moba).

Carry-over assimilation is prevalent in African tonal processes. Most assimilation phenomena, however, have been characterized as phonological processes, rather than phonetic coarticulation. For example, in Yoruba, sequences of High and Low tone (in either order) result in the tone of the first syllable being spread onto the second to create a contour tone over the second syllable. Such contour tones are "auditorily comparable to the lexically distinctive contour tones of other West African languages" (Laniran and Clements 2003: 207). In Standard Chinese, for example, comparable tonal sequences (i.e. High-Low or Low-High) are common. A Low tone preceded by a High tone is often realized with falling f0 contour, while a High tone after a Low tone typically surfaces with rising f0 (see Figure 6.2.2a). Despite the clear rising or falling f0 contour, speakers of Standard Chinese have no problem identifying the underlying High or Low tone, especially in context. This is probably due to the fact that the rising or falling contour for High and Low tones are clearly distinguishable from the rising and falling contours for underlying Rising and Falling tones in the same tonal contexts (see Figure 6.2.2b). An interesting line of research is how the phonetics and phonology of tonal systems condition whether some tonal coarticulatory effects may result in phonetic variants

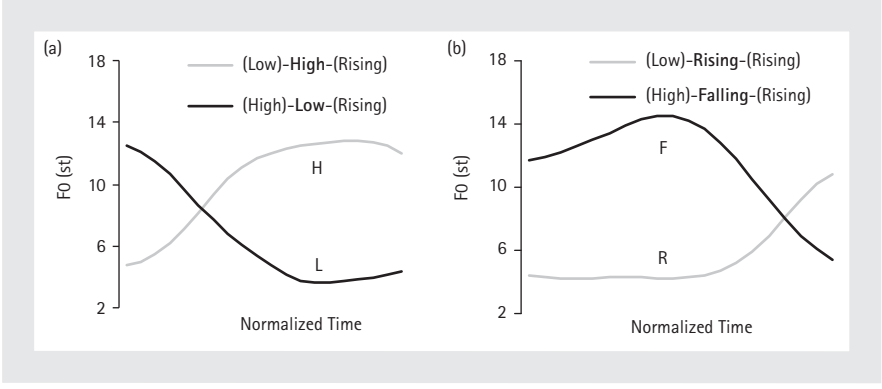

Figure 6.2.2. Mean f0 contours of High and Low tone (a) and Rising and Falling tone (b) in Standard Chinese, averaged across three repetitions of five speakers. The tones were uttered in carrier sentences where they are preceded by a Low or High tone and followed by a Rising tone. [Adapted from Chen and Gussenhoven 2008.]

of underlying tones (as in Standard Chinese) or phonologically distinct new tones (as in Yoruba).

Another common carry-over assimilation rule in African languages is the lowering of High tone to Mid tone in L-H sequences (L-H → L-M) (Hyman 2007). An interesting comparison here is with Cantonese, which has three level tones (L, M, and H). In two similar tonal sequences (i.e. L-H and L-M), the Cantonese H and M tones retain distinctive level tones but their tonal targets manifest mainly in the second half of the tone-bearing syllable, with the beginning part of the syllable making room for the tonal transition (Wang 2006). Thus the difference here lies in that, in African languages, an L-H tonal sequence is typically analyzed as resulting in phonological alternation while in Cantonese, phonetic coarticulation is reported. Without detailed instrumental studies, it is not clear what might be the phonetic difference(s) in these typologically different tonal systems that have led to two different analyses of L-H tonal sequences. Answers to this question should reveal cross-linguistic differences in f0 transition and the role these differences play in determining the phonetic/phonological nature of tonal contextual processes and, possibly, even the course of language change in their tonal systems.

6.2.3 Global tonal variation—encoding of information structure

At the discourse level, tonal variation is employed, along with other prosodic cues (e.g. duration and intensity), to package an utterance so as to integrate it into the information flow of ongoing discourse, a phenomenon known as the prosodic encoding of information structure. Take the English sentence in (1) as an example. In answer to the question of who teaches linguistics, *MARY* is uttered with prominence, signaling that it is new information and focused (indicated with curly brackets and FOC in subscript). Focus here is expressed by the presence of an intonational tone (i.e. salient f0 movement introduced post-lexically to cue the pragmatic function of an utterance; hereafter referred to as pitch accent[3]), which, among other prosodic cues such as duration, gives rise to the perception of prominence.

(1) A: - Who teaches linguistics?
B: - $\{\text{MARY}\}_{\text{FOC}}$ teaches linguistics.

[3] The notion accent has been defined in many ways. Readers are referred to Ladd (1996) for details of the definition of pitch accents within the autosegmental-metrical framework. See also Gussenhoven (2004) and in particular the discussion on p. 47 for various definitions of accents in the literature.

There has been much debate about the basic dimensions of information structure (e.g. theme vs. rheme; given vs. new; focus vs. topic; contrast, among others)[4] and how prosody encodes information structure. An aspect that has been of particular interest within laboratory studies is the prosodic encoding of focus, which provides new information (e.g. *Mary* in 1B as *informational focus*) or highlights contrast/alternatives in the discourse (e.g. *Mary* in 2B as *contrastive focus*), as opposed to given background information (e.g. *Mary* in 3B). Below, I limit my attention to focus-induced tonal variation in two language families with typologically different prosodic systems—West Germanic and Sinitic—with reference to other languages when necessary.

(2) A: - Does Karen teach linguistics?
B: - {MARY}$_{FOC}$ teaches linguistics.

(3) A: - What does Mary teach?
B: - [Mary]$_{Given}$ teaches {LINGUISTICS}$_{FOC}$.

6.2.3.1 *Different approaches to focus-induced tonal variation*

In declaratives, focus in both language families has been reported to boost the f0 contour of on-focus constituents, while for post-focus materials there is typically lowered and/or compressed f0 contour (for West-Germanic languages, see e.g. Cooper et al. 1985; Eady and Cooper 1986; Bartels and Kingston 1994; Sluijter and van Heuven 1996; Xu and Xu 2005; Baumann et al. 2006; Féry and Kügler 2008; Ishihara and Féry 2009; for Sinitic languages, see e.g. Gärding et al. 1983; Selkirk and Shen 1990; Shih 1988; Jin 1996; Xu 1999; Man 2002; Chen 2003; Yuan 2004; Gu and Lee 2007; Chen and Gussenhoven 2008). Pre-focus constituents usually do not show salient prosodic changes. Much less is known about the f0 effect of focus in questions. Eady and Cooper (1985) report f0 raising in English but Liu and Xu (2005) show a lowering effect in Standard Chinese. Haan (2002) found both raising and lowering effects of focus in different question intonations in Dutch.

Traditionally, this general focus-related f0 manipulation has been hypothesized in different languages as due to different linguistic representation or organization. In West-Germanic languages, where f0 changes do not differentiate lexical meanings, a distinction is recognized between the effect of focus and that of givenness on the presence and absence of pitch accents, respectively.[5] With regard to the

[4] For further details and terminological dependencies of these various information structure notions, readers are referred to Lambrecht (1994), and Kruijff-Korbayová and Steedman (2003), Féry et al. (2007), among others.

[5] Note that when the size of a pragmatic/semantic focus domain increases, there can be a mismatch between the focus domain (in curly brackets) and the location of the most salient pitch accent (in capital letters) (e.g. Jackendoff 1972; Ladd 1980), as illustrated in examples (i) and (ii) where pitch accent assignment varies due to the different structures of the answer (i.e. verb + object construction

two types of focus (i.e. contrastive vs. informational), some hold the view that there is no categorical difference between their accent marking, although a gradient paralinguistic difference such as higher and/or delayed f0 peak for contrastive focus, for example, is commonly acknowledged (e.g. Ladd 1996; Gussenhoven 2004). Others propose that their representation should be different in the grammar; only contrastive focus is marked syntactically and this difference is reflected in their different prosodic realizations (e.g. Selkirk 2002, 2007).

In Sinitic tonal languages, f0 changes indicate lexical meanings and speakers do not have the option of inserting or deleting pitch accents according to information status, like in West-Germanic languages.[6] The effect of focus has conventionally been considered to be due to the phonetic modification of pitch range within which lexical tones are realized. Take Standard Chinese as an example. Gärding et al. (1983) were the first, to my knowledge, to make use of the idea of *range grid*, which can be expanded under focus or compressed when out of focus. Most subsequent studies (e.g. Shih 1988; Jin 1996; Xu 1999, the first comprehensive study on focus-tone interaction) have adopted the notion of range expansion and suppression to explain on-focus and post-focus effects on tonal realization respectively. With regard to different types of focus, Chen and Braun (2006) report pilot data showing different f0 adjustment (i.e. more expanded pitch range or f0 maximum for contrastive than information focus).

Although pitch range manipulation often goes hand in hand with prosodic marking of information status, Chen (2003) and Chen and Gussenhoven (2008) show that in Standard Chinese, a number of other f0 adjustments may be prioritized which, as an *ensemble*, ensure that lexical tones are produced with enhanced distinctiveness of their characteristic f0 contours under focus. Furthermore, post-focus tonal realization suggests that f0 range-compression is not the whole story (Chen 2010). As shown in Figure 6.2.3a, when a post-focus Falling tone is preceded by a focused High tone and followed by a Rising tone, it is realized with a more expanded f0 range (indicated by the gray line) than its counterpart in no-focus condition (indicated by the black line). Figure 6.2.3b shows that when a Falling tone is itself focused, in addition to a slightly expanded pitch range, the more salient manifestation of focus is the delayed onset of f0 falling (see the solid line). This makes it very distinct from a focused Low tone (see the dotted line).

vs. verb + prepositional phrase construction). This shows that linguistic structure plays a role in the association between focus and accent (Gussenhoven 1984; Selkirk 1984 and 1995).

(i) A: - What does Mary do?
B: - Mary {teaches LINGUISTICS}$_{FOC}$

(ii) A: - What does Mary do?
B: - Mary {TEACHES in CHINA}$_{FOC}$

[6] In other words, for the purpose of marking focus, there are no additional intonational tones, such as High or Low pitch accent, that are associated with words that already have lexical tones.

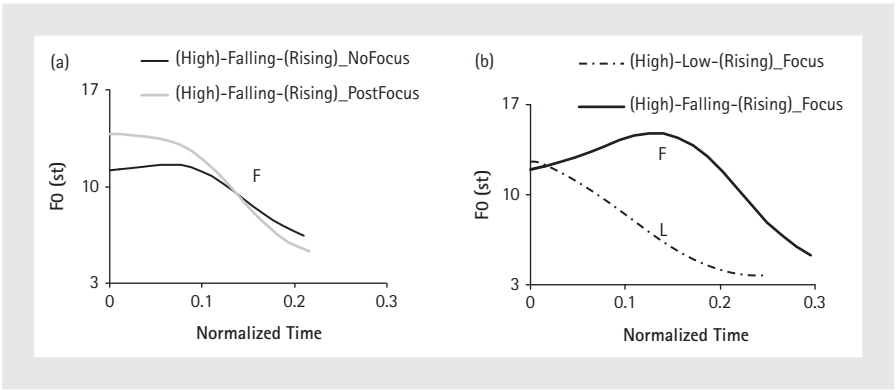

Figure 6.2.3. Mean f0 contours of Falling tones in No-focus vs. Post-focus conditions (a) and Falling vs. Low tone in Focus condition (b) in Standard Chinese, averaged across three repetitions of five speakers. The tones were uttered in carrier sentences where they are preceded by a High tone and followed by a Rising tone. [Part of the data adapted from Chen and Gussenhoven 2008.]

In Shanghai Chinese, Chen (2009) shows that of the five lexical tones, the high Rising tone is the only tone that shows no significant pitch range expansion under focus. This is presumably to ensure its distinction from the low Rising tone because significant range expansion of the high Rising tone would inevitably make the two Rising tones in the language occupy similar pitch range space and therefore be less distinctive. These observations cast doubt on the exclusive and primary role of pitch range expansion in focus encoding.

6.2.3.2 *Unifying approaches to focus-induced tonal variation*

Two different approaches have been taken to providing a unified account for prosodic encoding of focus across the two typologically different language groups. In one approach, focus has been argued to have a direct effect on pitch range manipulation. Xu and Xu (2005) state that "focus realization in English is fundamentally similar to that in Mandarin, i.e. the pitch range of the focused item is expanded, the pitch range of the post-focus items, if any, is compressed and lowered, and the pitch range of the pre-focus items, if any, remains neutral" (Xu and Xu 2005: 193). They deny the existence of pitch accents in English as phonological entities but argue that "Lexical stress and Sentence type jointly determine local pitch targets; and Focus assigns regional pitch ranges" (Xu and Xu 2005: 191). This exemplifies the functionalists' goal of finding direct prosodic correlates of a whole spectrum of communicative functions (e.g. Kohler 2006). In a more recent study, Féry and Kügler (2008) show with data on German that "focus raises tones while givenness

lowers them in pre-nuclear position and cancels them out post-nuclearly," and propose that "[T]hese changes in the values of accents were explained by the influence information structure has on reference lines associated with prosodic domains" (Féry and Kügler 2008: 700). Although different from Xu and Xu (2005) in recognizing pitch accents as a level of representation in intonational phonology, Féry and Kügler share the view that focus primarily modifies the f0 pitch range.

The alternative approach has argued that focus introduces abstract prosodic prominence and thus may indirectly result in pitch range manipulation. Researchers differ slightly in terms of how such focal prominence is defined. Truckenbrodt (1995) and Rooth (1996) argue that a focused element should be the most prominent within a focus domain. Rooth further shows that in certain contexts, such prominence may be cued with segmental lengthening and/or greater amplitude without the additional effect of pitch accents. Ladd (1996) proposes that focus in general is signaled via relative strong-weak metrical strength within the prosodic structure where focused elements are the strongest element of an utterance and associated with the primary pitch accent. Selkirk (2002) holds the view that different types of focus (i.e. contrastive vs. informational) contain metrical prominence at different levels of prosodic phrasing and the focus-prosody relation is mediated by an abstract phonological representation of prominence.

Along a similar vein, Chen (2003) and Chen and Gussenhoven (2008) argue that the phonological reflex of focus does not have to differ between Standard Chinese, a tonal language, and West-Germanic languages such as English and German. In both languages, focus introduces prosodic prominence. The difference between the two types of languages should lie in their different instantiation of focal prominence in terms of their available phonological entities and phonetic cues. For example, English employs intonational tonal events as a salient cue, while Mandarin Chinese manifests focal prominence more in the distinctive realization of lexical tones,[7] since the addition of pitch accents is prohibited.

While more still needs to be understood about the notion of focal prominence and its relation to the general prosodic structure of languages, it should be recognized that an abstract prosodic prominence approach has the advantage of taking full account of the multitude of phonetic and phonological variations reported to mark focus. In addition to pitch changes, cues such as prosodic phrasing (e.g. Kanerva 1989; Downing and Pompino-Marschall 2004) and durational adjustment (e.g. Turk and Sawusch 1997; Chen 2006) are also employed to mark focus. At the segmental level, both consonants and vowels under focus are articulated with

[7] Note that when the focus domain is larger than one syllable, the prosodic manifestation of focal prominence is not evenly distributed across syllables of the focus domain (e.g. Cambier-Langeveld 2000; Chen 2006; van Heuven 1997 for focal lengthening). Further studies are needed to understand the relation between the semantic/pragmatic domain of focus vs. the prosodic domain of focus manifestation.

enhanced distinctiveness (e.g. de Jong 1995; Erickson 2002; Cho and McQueen 2005; Chen 2008). Interestingly, Dohen et al. (2004) show that in French, the degree of lip opening for the vowel /a/ correlates well with the information status of the linguistic constituent (post-focus < pre-focus < on-focus), similar to the kind of asymmetry observed in pitch range manipulation. Among the variety of acoustic and articulatory cues speakers employ to mark information status, Baumann et al. (2007) report that German speakers show considerable variation in the extent to which different strategies are chosen.

6.2.4 Concluding remarks—variation and representation

The very notion of variation implies that "at some level there be invariant, discrete units underlying the variable and continuous activity of speech production," as Kühnert and Nolan (1999: 7) insightfully point out with respect to coarticulation. This short review raises the prominent and recurring question: to what extent should tonal contextual variation be accounted for in terms of phonological processes, and to what extent as phonetic variation? An even more fundamental question is whether indeed there should be invariant and discrete units at some level of representation such as the mental lexicon (see Chapter 8 this volume for more discussion of lexical representation).

While a considerable amount of work has investigated how listeners overcome segmental variability in speech comprehension and how such findings shed light on the nature of the mental lexicon (Ernestus, this chapter), the question of how speakers encode a wide range of variability between the lexicon and the acoustic signal (see Schiller, this volume for an overview of methods for investigating word-form encoding), especially at the prosodic level such as tonal variation, remains a fertile land for future exploration.

CHAPTER 7

SYSTEM-RELATED VARIATION

PHILIP HOOLE,
BARBARA KÜHNERT, AND
MARIANNE POUPLIER

In this chapter, the authors discuss the question of the peripheral motor system as a source of variation in speech production. Through examination of two case studies, one looking at the production of velar stops, the other at intrinsic f0 effects on vowel height and consonant voicing, they argue that motor system effects are deeply integrated into representations driving speech planning, due to speakers' (controlled) enhancement of automatic effects.

7.1 Introduction

In this chapter we consider the topic of system-related variation—patterns of variation caused by, or associated with, the properties of the peripheral motor system of speech production. System-related variation is a critical component for studies in laboratory phonology since at its very heart lies the idea that insight into the nature of phonological units and processes can be gained with the help of experimental phonetic data. A central concern for researchers in this field is to understand—and disentangle—what aspects of observable speech behavior directly reflect the underlying control structures, and what aspects reflect effects of the physics and

physiology of the speech production system. Or, in the words of Munhall et al. in their LabPhon V contribution ". . . in order to determine what contributes to the form of spoken language, we must solve the inverse problem of assigning variance components in the speech output data to different parts of the input planning sequence" (Munhall et al. 2000: 10).

When trying to address this problem, researchers are faced with two primary tasks. On the one hand, it is self-evident that the peripheral motor system can contribute a significant amount of variability to the resulting speech output. Its adequate evaluation and subsequent modeling, therefore, require a detailed knowledge of the complex morphology of the vocal tract, of the myoelectric activity and the resulting movements and response dynamics of the multiple joints and muscles involved, the fluid mechanics of the aerodynamic properties of speech, the articulatory-to-acoustic transformation processes, and so forth. While considerable progress has been made in many of these areas, it is also evident that many crucial parameters "are not available at present and some are difficult if not impossible to acquire with current technology" and "have to be estimated or eliminated through simplifying assumptions" (Munhall et al. 2000: 18).

On the other hand, and more importantly, the research over the last twenty years spanning the laboratory phonology series has increasingly shown that the lower levels of the speech execution system are not merely a passive device for the transmission of linguistic units and processes, but rather that their constraining properties may be an integral part of the internal representation of the more central planning stages. Thus, there has been a major shift in emphasis from the initial key question of "how, in the twin processes of producing and perceiving speech, do the discrete symbolic or cognitive units of the phonological representation of an utterance map into the continuous psychoacoustic and motoric functions of its phonetic representations?" (Beckman and Kingston 1990: 1) to "the insight that the higher levels in the planning process know a great deal about the kinematics and dynamics of the movement" (Munhall et al. 2000: 27). Indeed, the wider point to be made in this chapter is that when seeking insight into the cognitive representation of phonological distinctions, while it is convenient to separate the contributions of variance in terms of a controller and its apparatus, a clear separation between the two components might often not be possible. They are related insofar as they show how the development of phonological distinctions can reflect how speech planning takes advantage of properties of the speech apparatus.

We will not try to consider all issues that might fall under the topic of understanding system-related variation. Rather we try to illustrate for some specific test cases what this teasing apart of controller and apparatus can mean in practice at our present state of knowledge. The case studies, which will be used to illustrate the issue of system-related variability, are:

1. Patterns of tongue movements for velar consonants. There has been considerable discussion as to why they often show strongly elliptical trajectories: For example,

does this reflect basic mechanical properties of the speech production system? Or the influence of air pressure? Or do speakers actively plan movements taking aerodynamic constraints into account, possibly in an effort to maintain or enhance phonological contrast?

2. Patterns of segment-related variation in f0, often referred to as micro-prosody. For example, f0 is typically higher in high than low vowels, and higher after voiceless than voiced consonants. Does this reflect automatic repercussions of the basic articulations involved (tongue position for vowel height, laryngeal movement for voicing control)? Or do speakers actively plan enhancement of the basic contrasts? As Solé (2007) has recently remarked, it can be a difficult matter to tease apart controlled and mechanical properties in speech, precisely because it is quite possible for the status of a phonetic property to change over time, i.e. to shift from a mechanical property, to being adopted into actively controlled behavior. A well-known case in point is the development of tone from the intrinsic pitch differences associated with voiced and voiceless consonants (see e.g. Hombert et al. 1979; Abramson 2004; further references in 7.3.2 below).

Areas which are relevant to the present discussion but will not be further elaborated here include, for instance, which patterns of observable variability would be predicted by models of human speech production. One example comes from studies carried out in the tradition of Stevens's Quantal Theory (1972, 1989) which proposes that non-linearities between articulatory parameters and the resulting acoustic/auditory responses are a key factor (see Iskarous, this volume for a discussion). Thus, the extent of articulatory variability is expected to be higher in vocal tract areas that are relatively insensitive in their acoustic response, but more constrained in areas with critical acoustic impact. For example, the extent of variability of tongue body articulation during the production of the vowels [i, u, a] has been reported to be higher in the direction parallel than perpendicular to the vocal tract outline. This is consistent with the prediction that the formant trajectories of these vowels are relatively insensitive to some variation in exact constriction location, but highly sensitive with respect to variation in the degree of constriction (Perkell and Cohen 1989). Another factor responsible for the extent of observable coarticulatory variation is proposed in Recasens and co-workers' Degree of Articulatory Constraint (DAC) model (e.g. Recasens 2002, 2007), in which different vowels and consonants are associated with specific values of articulatory constraint depending on the articulatory requirements involved in their production. Regarding vowels, for instance, the DAC value should be high (variability therefore low), for the production of [i] and then follow the order [i] > [e] > [ɛ] since forming a palatal constriction highly constrains the raised and fronted tongue dorsum.

Within the field of f0 control, related topics include declination, i.e. to what extent declining f0 is a passive consequence of declining subglottal pressure (see e.g.

Ohala 1990a and, more recently, Demolin 2007). Another example relevant to the general issue, which has been comprehensively covered elsewhere, is the question of whether VOT systematically depends on place of articulation (see Cho and Ladefoged 1999). Useful sources of further discussion of system-related variation in general include the just-mentioned review paper of Solé (2007). This in turn is very much in the tradition of work by Ohala, of whom (among many other papers that could be quoted) we will just cite two review papers (Ohala 1983, 1997: 686–91) which make very clear how sounds can differ in their aerodynamic stability and how the variation this induces can become visible in sound change.

7.2 A case study of supraglottal system-related variation: Velar consonants

Velar consonants provide a particularly instructive example of the difficulties that arise when trying to tease apart those details of the speech movement patterns that are under deliberate control, and those that arise incidentally from the properties of the production system. First, there is the matter of the articulatory "target" for velars. In models of stop consonant articulation the concept of *virtual target* has proved popular. In other words, for lingual consonants the speaker may aim at an unattainable target located beyond the hard palate (see Perrier et al. 2003 for discussion and further references). In a sense this leads to the most basic kind of system-related variation imaginable: For these sounds (but not for vowels, and probably not for consonant categories such as fricatives) the position reached by the main articulator will always be different from the target aimed for. More specifically, the precise position reached might vary depending on the path along which the articulator approaches the target and starts to interact with the impenetrable boundaries of the vocal tract. It is, of course, well-known in many languages that the stop consonants generally referred to as velar can vary considerably in their place of articulation depending on the adjacent vowels (for example, the two classic papers of Öhman 1966, 1967, contain some very pertinent early comments). It is conceivable, in the light of the above, that this surface variability is induced by the interaction of the tongue tissue with the hard palate as it is directed towards some constant but unattainable target. In fact, the modeling studies of Perrier et al. indicate that it is probably more appropriate to assume two separate virtual targets for velars: in front-vowel contexts it is slightly more anterior than in back-vowel contexts.

The situation is even more complex. It has frequently been observed that velar stops, especially in the context of back vowels and much less so in the context of front vowels, may move forwards during the closure phase. Strikingly, this can occur not just when the vowel *preceding* the velar consonant is a back vowel but also when the following one is too, e.g. in sequences like [ugu]; see Figure 7.1. In other words, during the closure the tongue may actually be moving *away* from the following vowel, which runs counter to most ideas as to how coarticulatorily induced variation works. It is instructive to consider the range of explanations that have been offered for such patterns.

The basic movement pattern just described first emerged from Houde's (1968) X-ray study, which used the voiced velar [g]. This led Ohala (1983) to suggest that the movement represented a cavity-enlargement strategy to sustain voicing, and was as such related to the maintenance of phonological contrast. At that time

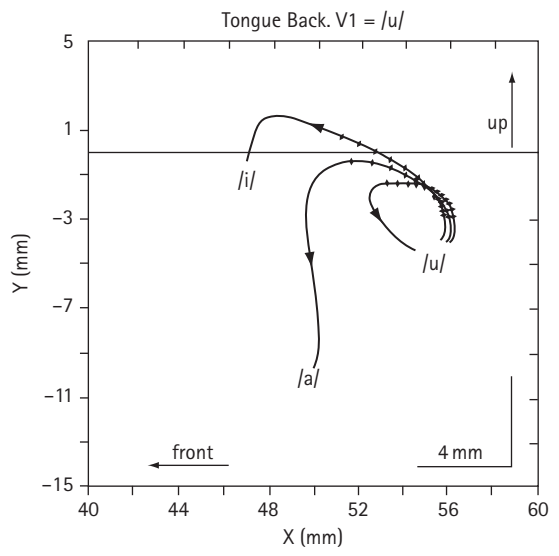

Figure 7.1. Ensemble-averaged trajectories of an EMA sensor located on the tongue dorsum of a German speaker for VCV sequences with V1 = /u/ and V2 = /i, a, u/. Each trajectory extends from the midpoint of V1 to the midpoint of V2. Tick marks (at 8ms intervals) delimit the acoustically defined period of consonantal closure. Note in particular for the /ugu/ sequence that the tongue continues to move away from the target position for V2 even for a short period after the end of consonantal closure.

this suggestion was plausible, because voicing in many consonants constitutes an inherently unstable aerodynamic situation, and speakers undoubtedly have access to articulatory strategies to counteract a potential source of system-related variability. In this case, however, further data of Mooshammer et al. (1995) cast doubt on the explanation since forward movement was larger for voiceless /k/ than for voiced /g/. Subsequently, Hoole et al. (1998) attempted to estimate the contribution that the rise in intra-oral air pressure behind the constriction might be making to the forward movement of the tongue during the closure phase. This followed up on not only the observation of more forward movement for /k/ than /g/ (the former might well have higher intraoral air pressure than the latter) but also an additional finding of the Mooshammer et al. study that movement was very restricted for velar nasals, which presumably have no heightened intraoral pressure at all. The approach followed by Hoole et al. was to compare utterances spoken on an egressive and ingressive airstream. Forward movement was reduced in the ingressive condition, but not completely abolished, indicating that air pressure might be a relevant influence, but not necessarily the only one.

A different approach for explaining these curved tongue-movement paths was put forward by Löfqvist and Gracco (2002) who suggested that they could represent a cost-minimization strategy of the kind often explored in other areas of motor control (e.g. jerk minimization). This might imply that speakers plan complete trajectories following such a principle, rather than basing planning on single point-targets.

In the face of these competing explanations Perrier et al. (2003) investigated the potential contribution of biomechanical effects in a modeling study. Their physiologically based tongue model incorporates realistic biomechanical properties and muscle orientation patterns, making it possible to simulate force development in the tongue tissue over a range of plausible muscle-activation patterns and to study resulting tongue movement while taking tongue-palate interactions into account. The main result was that even a very simple control strategy (a sequence of single target configurations for the consonant and each flanking vowel) was able to generate curved movement paths. In other words—unlike the suggestion of Löfqvist and Gracco—it is not necessary to assume that the curved nature of the paths reflects the planning process of the speakers. Also, contrary to a suggestion in Hoole et al. (1998) it may not even be necessary to assume asynchronies in the activation of the main muscles involved. In addition, Perrier et al.'s study indicated, particularly in back vowel contexts, that the demonstrated biomechanical effects probably outweighed any aerodynamic contribution to the observed movement paths and that the effects of tongue-palate interactions are relatively subtle compared to effects arising in the tongue itself. Moreover, based on systematic manipulation of plausible locations of the virtual target for [k] in front- and back-vowel contexts (marginally more fronted in the former case) they were able to reproduce the empirical observation that the presence and direction of curved movement paths is more stable in the back-vowel contexts.

This example thus serves to illustrate that even very simple sound sequences can present many puzzles when one tries to pin down why the typically observed movement pattern is the preferred one. In the present case there appear to be some basic physical processes that favor a particular pattern of observed behavior (muscle orientation and activation patterns, interaction of tongue tissue with the hard palate, air pressure effects on the tongue tissue). These factors may each actually be quite weak in themselves, but a consistent pattern of movement may nonetheless become established because speakers, where possible, plan their movements to take advantage of the physical bias in the system, particularly when as in the present case the resulting pattern may be efficient for other reasons, such as jerk minimization, and even supporting voicing in voiced plosives. In the following section we endeavor to show that this style of explanation may be fruitfully employed in other areas as well.

7.3 Case studies of segmentally related variation in f0

There has been considerable debate as to the source of readily observable patterns of variability in (1) intrinsic pitch in vowels and (2) effects of consonant voicing on f0. For both of these, researchers have debated whether differences in f0 represent a mechanical consequence of basic articulatory maneuvers required for familiar phonological distinctions such as vowel height and consonant voicing, or whether the differences emerge from a more active enhancement strategy on the part of speakers. We contend that for both areas a more hybrid style of explanation appears promising, and that these areas thus illustrate rather effectively how physical properties of the speech production system interact with the drive towards clear signaling of phonological distinctions. The main difference from the previous section—although the general thrust of the argument remains the same—is that the discussion is now focused much more on how multiple acoustic properties can be involved in a specific phonological opposition. In the previous section we indicated how a single pattern of movement might become strongly entrenched in behavioral patterns for speech, and in that sense enhanced. In the present section the perspective on enhancement will be wider. To illustrate with vowel intrinsic pitch: The basic acoustic correlates of vowel height (above all F1) are generally accompanied by differences in a completely different acoustic property, namely f0. While this additional f0 information for the listener may actually be largely driven by mechanical effects of tongue movement for the vowel, we need to entertain the possibility that this mechanical effect is emulated and thus enhanced

in a completely different articulatory system, namely not in the tongue but in the larynx itself.

7.3.1 Intrinsic pitch in vowels

The tendency for high vowels to show higher f0 than low vowels is an extremely pervasive phenomenon that has been documented in detail by Whalen and Levitt (1995). Indeed, its very universality suggests to these authors that it is an automatic, mechanical consequence of the movements of the articulators required to produce vowels of different heights. This contrasts with the approach of Kingston (e.g. 1992) who sees the f0 differences as an auditorily driven enhancement strategy, given the relevance of the difference between f0 and F1 to the perception of vowel height (e.g. Traunmüller 1981). As pointed out by Whalen et al. (1999), however, there are presumably limits on this effect in running speech because vowel height is not misperceived when a syllable receives f0 prominence.

The most direct evidence for active enhancement would come from vowel-related differences in laryngeal EMG, particularly of the cricothyroid (CT) muscle. In fact, as we will see below, parallel consideration of EMG and f0 patterns equally provides the most direct approach to teasing out possible automatic contributions to f0 variation. Before going into the interpretation of EMG data in more detail we will, firstly, briefly outline the most plausible mechanical substrate of intrinsic f0 (henceforth If0), and, secondly, discuss a more indirect approach to revealing the presence of enhancement strategies ('indirect' means that the point of departure is not the physiology, but rather makes use of the fact that enhancement must reflect speakers' interpretation of the communicative demands of an utterance).

Various proposals have been made over the years for how movements of articulators such as the tongue could influence the tension of the vocal folds (see e.g. Sapir 1989; Dyhr 1990; Honda 2004; Hoole and Honda 2011; Hoole 2006 for reviews). Currently, one of the most plausible accounts (proposed e.g. in Honda and Fujimura 1991) suggests that contraction of the genioglossus posterior to support raising of the tongue exerts a forward pull on the hyoid bone, which in turn causes the thyroid cartilage to rotate forward and lengthen the vocal folds.[1]

In addition to purely anatomical plausibility, another argument in favor of a mechanical account of If0 is that it is present even in languages with a small vowel inventory and where, accordingly, enhancement of contrast might hardly seem necessary. On the other hand, one argument that has been used against the mechanical explanation is that If0 does not appear in all cases where a mechanical explanation predicts that it should. This has been suggested for German tense-lax contrasts where large differences in tongue configuration are not accompanied by

[1] Other lines of explanation have, for example, appealed to coupling effects between f0 and F1. See e.g. Dhyr (1990); Ohala and Eukel (1987); also Hoole (2006) for discussion and further references.

the expected If0 differences (Diehl 1991; Kingston 1992, 2007; extensive discussion in Fischer-Jørgensen 1990; see also Hoole and Mooshammer 2002). Thus If0 in German at first sight appears to be an exception to the pattern: system-related variability is expected, but does not occur. In fact, the EMG evidence to which we now turn suggests that the situation in German may indeed still be consistent with the mechanical account of system-related variation in If0.

An indirect approach to finding evidence for If0 as controlled behavior has recently been looked at in detail in Kingston (2007). The basic technique involves analyzing acoustically how the strength of If0 effects waxes and wanes as the prosodic context of the target vowels is manipulated. This has the great advantage over the EMG experiments, to be discussed below, that purely acoustic recordings are sufficient, potentially allowing a larger number of subjects and a wider range of utterances. The basic argument in Kingston (2007) is that an automatic effect (such as tongue-pull) should not vary with prominence (assuming non-prominent syllables still have full vowel quality), whereas it would be very natural that an active enhancement effect would strongly target prominent vowels with presumably high information load. By this logic, various indications in the literature that If0 is only present in intonationally prominent syllables (e.g. Ladd and Silverman 1984; Steele 1986a) would seem to support an argument against the automaticity of If0.

However, non-prominent syllables often have low f0, and it is certainly not immediately clear that the tongue-pull mechanism necessarily has the same effect on f0 over the full pitch range of the speaker (see Whalen and Levitt 1995, for more discussion; one possibility is that in low tones the activity of the strap muscles may counteract the tongue-pull mechanism). Kingston himself discusses the case of tone languages where it appears that If0 is more evident for high tones than low tones. The crucial test, then, is to compare material where both high and low tones can be both prominent and non-prominent, and determine whether strength of If0 effects follows from prominence more closely than from simple pitch height. Steele (1986a) had found indications that non-prominent high tones might not show If0 effects (see also Reinholt Petersen 1978 for discussion of prominent low tones in Danish). Steele argued that there might be interactions between the strength of tongue-pull effects on If0 and the level of subglottal pressure, thus keeping If0 within the scope of system-related variation. But assessment is hampered here by the lack of relevant physiological data (probably even more acute than in the case of muscle activation data).

The pattern of results in Kingston's experiments was quite complex, but in the more extensive of these experiments (with naive subjects) the results appeared more consistent with the automatic account, i.e. contrary to the original hypothesis. This preliminary conclusion was cross-checked by looking for evidence that speakers were actually using articulatory adjustments to mark prominent syllables. If this were not the case then the experiment may simply have been unsuccessful in eliciting appropriate utterances. In fact, F1 values indicated that prominent syllables

had actually been given articulatory prominence, supporting this conclusion: The speakers simply had not enhanced vowel-related f0 differences on syllables that were indeed prominent, so any f0 differences present had best be assumed to be automatic. Nevertheless, the story may not stop here: Kingston offers as a possible explanation the suggestion that although the speakers were clearly distinguishing prominent and non-prominent syllables, there may not have been a genuine modulation of local information content, since the prosodic manipulations were effectively restricted to the permutation of a small set of proper names (like *Leland*, *Layland*). Thus, the question remains open.

We now turn to the interpretation of direct EMG evidence that speakers actively increase f0 on high vowels. Relevant EMG data is not surprisingly rather limited. Nevertheless, for the single speaker presented in Honda and Fuijmura (1991), for four speakers in Dhyr (1990), and two out of three speakers in Hoole and Honda (2011) activity of the cricothyroid appeared to be higher for the high vowels. This might make it seem tempting to view If0 as an active adjustment. However, such a conclusion would be premature, or at least incomplete. Whalen et al. (1999) failed to find evidence for higher CT activity in higher vowels. Moreover, they argue very cogently that when interpreting positive results it is necessary to look closely at the regression line linking f0 and muscular activity. For example, interpretation becomes a delicate matter if a given change in EMG has a different effect on the f0 of [i] compared to that of [a]. Put another way, if two vowels lie on different regression lines, for example such that at the same CT value a different f0 is predicted, then clearly there remain factors in addition to CT that affect f0. Hoole and Honda looked at this in detail. Indeed, it emerged very consistently from their regression analyses that at comparable EMG levels, f0 was higher in the high vowels. That is, the regression analysis gave the same result in all three speakers (whereas only two speakers showed higher EMG for the high vowels). This suggests that regardless of what pattern of EMG activity speakers adopt, there are still forces at work raising f0 on the high vowels. And this could well be the tongue-pull mechanism outlined above.

Taking all these findings into account, Hoole and Honda argue for a hybrid explanation of the intrinsic pitch phenomenon. The basic driving force is assumed to be rooted in the mechanical contingencies of normal vowel articulation. This output variation emerges from the biomechanical structure of the speech system and consequently is hypothesized to be stably present across speakers. This primary mechanism is then accompanied by a secondary one of reinforcement in some speakers. Speakers may well differ in their sensitivity to fairly subtle differences of this kind, and those speakers for whom the differences appear particularly salient may come to actively reinforce them, adopting them into their repertoire of adjustments for enhancing the clarity of phonological distinctions. Nevertheless, the assumption here is that this is a secondary stage, and not the basic driving force. This scenario is close to that proposed earlier by Honda and Fujimura (1991: 151),

who said in connection with their findings of a process of phonologization: "Thus the cricothyroid activity associated with high vowels 'emulates' the biologically natural f0 rise due to hyoid bone movements."

We will see that this hybrid style of explanation may also be fruitful for the topic to be considered below, namely consonant-related differences in f0. Before turning to that, we return briefly to the apparent puzzle of intrinsic pitch in German vowels. The German data analyzed in Hoole and Honda also allowed comparison of tense and lax vowels. Using very similar regression analyses to those just outlined for high vs. low vowels, some evidence was found for higher cricothyroid activity on the lax vowels, and also when compared at the same EMG level, evidence that f0 is lower for lax vowels. The latter of course is just what the mechanical account of If0 would predict. Thus we have the perhaps slightly curious situation that system-related variation is present but can be masked by active adjustments. (In addition this has interesting implications for our conceptualization of terms like tense vs. lax, but this lies outside the issues being focused on here; for discussion see Hoole and Mooshammer 2002; Hoole 2006; and Hoole and Honda 2011.)

7.3.2 F0 differences related to consonant voicing

F0 differences associated with consonant voicing are an important but subtle test case for bringing into better focus our understanding of how certain readily observable patterns of variation come about. Kingston (1992) has referred in this context to the notion of *co*variation (our emphasis), i.e. what phonetic properties change in parallel when a given linguistic distinction is realized. Is this covariation mechanical and unavoidable given the properties of the human speech apparatus, or does it reveal that speakers can home in on mechanically and physiologically unrelated articulations as a means for enhancing a specific auditory property of the speech signal?

It is not necessary here to review all the literature that has accumulated on the simple fact that there are voicing-related differences in f0. A convenient source of earlier discussion is the book on tone edited by Fromkin (1978), because of the great relevance of the voicing status of consonants for the development of tonal systems. See in particular the chapters by Hombert and Ohala (see also Hombert et al. 1979; further references to the earlier literature can be found in Hoole 2006). Of particular note among more recent acoustic analyses is Hanson (2009). Her carefully balanced corpus supports the view that the f0 differences (at least in English) are probably best viewed as a raising of f0 on voiceless consonants, rather than a lowering on voiced consonants.[2]

[2] We will leave aside here the question of the perceptibility of the consonant-related f0 differences—although this is presumably a prerequisite for tonal developments. We will also not review the somewhat variable results regarding the extent to which f0 may be a relevant cue in the perception

The central issue for present purposes concerns the physiological driving force behind the f0 differences, and—if this is primarily related to consonantal articulation—how it then manifests itself in particular on the post-consonantal rather than the pre-consonantal vowel (this directionality in the effect having clearly emerged from earlier studies). A key early reference for understanding why f0 differences might occur is the seminal paper on laryngeal features of Halle and Stevens (1971). They made clear that increasing the tension of the vocal folds can support the suppression of voicing; increased tension might then also manifest itself as higher f0 in the neighboring vowel. EMG evidence that such an adjustment is not only plausible but actually exists has, once again, been rather slow to accumulate (as remarked by Whalen et al. 1993: 2158). However, the balance of the evidence by now is that the cricothyroid, i.e. the muscle most directly involved in increasing the longitudinal tension in the vocal folds, is more active in voiceless consonants.[3] This emerges in particular from the major study of Löfqvist et al. (1989), and also from the three subjects of Hoole and Honda (2011) on whom findings in the previous section were based (see also Dixit and MacNeilage 1980; Hutters 1985; and for a negative finding, Collier et al. 1979). The acoustic findings of Hanson (2009) appear consistent with the assumption of increased CT activity on voiceless consonants. She compared the consonant-related f0 differences in combination with different intonation contours. In falling intonation the voiced-voiceless differences were neutralized, probably because appropriate realization of the prosodically required contour would be impeded by a local increase in CT activity for the consonant.

Based on a consideration of the timing of CT activity, Löfqvist et al. (1989) are firmly of the opinion that increased CT activity supports devoicing, i.e. it is not planned by speakers primarily to increase f0 in the following vowel, because peak CT activity is located squarely in the consonantal phase, and voiced-voiceless CT differences are weakest around the midpoint of the vowel. Nevertheless, it remains slightly unclear why f0 differences on the vowel are so robust. Below we return to the issue of timing and suggest a physiological mechanism that might account for the propagation of f0 effects onto the following vowel, without calling into question the basic timing pattern identified by Löfqvist et al. Moreover, as pointed out by Kingston and Diehl (1994: 440) there is perhaps also a tension between the assumption that the CT activity helps to suppress voicing—which would mean concentrating the effect at the start of the consonant—and the assumption that it can also account for clear f0 differences after the end of the consonant. In fact, Kingston and Diehl (see especially p. 432) are of the opinion, based on a consideration of the

of the voicing distinction itself. A selection of relevant papers includes Haggard et al. (1970), Abramson and Lisker (1985), Kohler (1985), Schiefer (1986), Silverman (1986), Whalen et al. (1990, 1993).

[3] It might be noted here that the other main muscle responsible for regulating vocal fold tension, namely the vocalis, is unlikely to provide the link between devoicing and high f0 since there is clear evidence that it is suppressed during voiceless consonants (e.g. Hutters 1985; Collier et al. 1979)—probably because it tends to support vocal fold adduction.

voicing contrast in a large number of languages, that f0 differences are not contingent on the specific laryngeal adjustments for the consonant at all, but depend on the consonant's specification with respect to a more abstractly defined [voice] feature. This point of view has recently been reiterated very clearly by Kingston (2007: 174): "f0 is raised next to voiceless unaspirated stops in languages such as French where this is the pronunciation of the [−voice] stops, but lowered next to such stops in languages such as English where this is the pronunciation of [+voice] stops." In short, Kingston (2007) argues that attempts to explain consonant-related effects on f0 as automatic have simply not been successful. In other words, the f0 differences might be better seen as an auditorily driven enhancement of the voicing contrast—enhancing the contrast by increasing the low frequency emphasis in the spectrum in the [+voice] context (any voiced sound will generally have its main concentration of energy in the spectrum lower than for a voiceless sound; this basic pattern can be emphasized by keeping f0 low after a voiced consonant).

The argument that the observable f0 behavior is better predicted by a more abstract phonological specification than by mechanical details of the articulation parallels the argument based on German tense vs. lax vowels against a mechanical explanation of vowel If0. However, we saw above that the latter argument may not be as strong as originally supposed, and, in the present context, we believe it is hazardous to assume that French [−voice] and English [+voice] are completely equivalent in their laryngeal adjustments simply because their surface acoustic manifestations are quite often fairly similar.

We will consider a scenario based on detailed consideration of physiological data which may help to resolve some of the problems raised by Kingston and Diehl (1994). First, however, we should return briefly to Kingston's (2007) experiments (referred to in the previous section as the "indirect" approach to assessing controlled vs. automatic accounts). These applied the same logic of prosodic manipulation as discussed above for vowels to consonantal f0 effects as well. No positive evidence in favor of the controlled hypothesis was in fact found.[4]

Turning now to direct measurements of muscular activity, the main issue concerns the possible discrepancy between the assumption of increased tension to suppress voicing and the concentration of f0 effects on the following vowel. This requires consideration of the timing patterns of the muscular activity, and we will consider these in more detail immediately below. One simple first step towards removing the discrepancy would be to question the relevance of using increased CT activity to assist the suppression of vocal fold vibration at the onset of voiceless consonants. Indeed, it is probably fair to say that the aerodynamically more challenging task for a speaker is not to suppress voicing in a plosive but rather to maintain

[4] But in the most extensive experiment the consonant-related f0 effects were overall very weak, which is also an awkward result for the automatic account (and unexpected in the light of the bulk of the previous literature).

it when this is required by the linguistic code. Given the occlusion of the vocal tract and rising intraoral air pressure, then probably even a small amount of glottal abduction is enough to make voicing stop quickly. There may well be no strong necessity for the speaker to introduce yet a further mechanism to make voicing stop even more quickly. The one or two residual vibrations that often occur as the glottis opens and pressure builds up are unlikely to be communicatively relevant given that offsets are generally less auditorily salient than onsets, and any sound generated is radiated from an already strongly occluded vocal tract.

Even if there is, then, no need for the CT activity to be focused on the consonant onset, what mechanism would nonetheless allow its effect to propagate so strongly from some later location in the consonant onto the vowel? The answer here could be that the delay for the effect of relaxation of CT is longer than that for activation. In other words, after f0 has been increased by CT activation, it may persist at the higher level after relaxation of CT. The delay estimates given in Sawashima et al. (1982) suggest that this is the case, and observations and analyses discussed in Hoole and Honda (2011) are also consistent with such an effect (and this possibility is also mentioned in passing by Löfqvist et al. 1989).

All the same, we have still not converged on a completely coherent picture unless we can make a specific suggestion regarding the timing of increased CT activity for the voiceless consonant: Based on consideration of our own data and perusal of the patterns presented in Löfqvist at al. (1989) we estimate that the time course of CT activity may be quite closely linked to the time course of glottal ab- and adduction. For example, it is known that there are typical differences in the timing of glottal ab- and adduction for aspirated plosives versus fricatives (relative to the relevant supraglottal events; see Hoole 1999). It appears that similar differences in the timing patterns may apply to CT activity. Why could this pattern of coordination be advantageous in the production of voiceless consonants? The basic idea we have proposed in more detail in Hoole and Honda (2011) is that the purpose of the CT activity is to increase the mechanical efficiency of the abductory motion of the arytenoids with respect to the resulting glottal aperture: in other words, the longer the glottis, the greater the increase in area for a given amount of abduction. There may, moreover, be an interesting side-effect of this CT maneuver with respect to vocal fold tension, but not primarily in terms of suppression of vibration but rather in terms of controlling the reinitiation of vibration at the onset of the following vowel, i.e. at a location that is most likely auditorily more salient and communicatively more relevant. The following scenario, which was inspired by illustrations of the articulatory synthesis presented by Hanson and Stevens (2002), seems particularly plausible in the light of the discussion above as to how effects of CT activation can propagate to later points in time.

As one of their examples, Hanson and Stevens discuss the synthesis of voiceless aspirated plosives and concentrate particularly on modulation of the compliance of the vocal folds. They find it useful to modulate the compliance roughly in

parallel with glottal opening, the aim being "to gain more control over the *onset* of voicing" (p. 1173; our emphasis). The aerodynamic background to this is that intraoral pressure is assumed to drop rapidly at release, and the model predicts that under default conditions the vocal folds would be able to vibrate when the glottal area has reduced to about 12mm^2, i.e. before the glottis has completely closed. By reducing the compliance, phonation threshold pressure is raised and the vocal folds are not able to vibrate until the glottal area (in this example) has further reduced to about 5.5mm^2. These adjustments result acoustically in a fast, clear transition from voiceless aspiration noise to modal phonation. Thus the modeling work of Hanson and Stevens shows how a kind of hysteresis effect in the consequences of CT activation could affect the acoustics in a perceptually beneficial way.

Tying together these results, the initial conclusion is that consonantal f0 effects are contingent on articulatory adjustments for the consonant. Accordingly, articulations required for the consonant (e.g. control of glottal abduction and devoicing) provide the basic driving force behind the f0 differences, rather than the speaker's search for an auditorily advantageous enhancement strategy. This basic conclusion parallels closely that reached in the previous section for vowel intrinsic pitch. In fact, the parallels may extend further: In the previous section it was suggested that some speakers may actively reinforce a mechanically (physically) given bias in the system. A small amount of evidence was found in the Hoole and Honda study that this may occur for consonantal f0 as well: CT differences depending on the voicing status of the consonant were found not just on the consonant itself, but in weaker and less consistent but nonetheless sometimes statistically significant form on the following vowel. This is essentially an enhancement effect. This in turn gives empirical support to the basic tonogenesis scenario already mentioned in the introduction above, by which a tonal contrast on vowels remains after a voicing contrast on the previous consonant is lost. As pointed out by Whalen et al. (1993) this scenario logically requires a process of enhancement to have taken place, because if, as contended here, the basic f0 patterns are an automatic consequence of laryngeal articulation for the consonant, then this "automatic" component of f0 would be lost when the voicing contrast is lost.

7.4 Conclusion

System-related variation can be understood as variation emerging from the properties of the peripheral motor system of speech production and the subsequent acoustic transformation process. As we have tried to illustrate in the above case studies, the evaluation of the contribution of system-related variance in the speech

output is a delicate and complex enterprise. This is partly due to the lack of truly comprehensive data on the biomechanics and dynamics of the joints, muscles, and soft tissues involved in speech production. What is more, there is increasing evidence that the higher-level control structures are directly shaped and constrained by the properties and response strategies of the executing system. In the case of velars, the available evidence suggests that in fact phonological factors (maintenance of voicing) and physiologically advantageous articulator movement trajectories converge to produce a stable movement pattern, a scenario reminiscent of Stevens' Quantal Theory. In fact, although Ohala's original interpretation of forward-directed tongue movement during velar closure as a strategy to support voicing was not supported by subsequent data and modeling, this leads to the intriguing speculation that loss of the voiced consonant at this place of articulation might be even more prevalent if velars did not have this pattern of movement. In the case of microprosodic variation, too, the grammar of spoken language seems to exploit the physiological givens of speech production. In practice, teasing apart the system-related, purely phonetic from the intended, explicit phonological variation is often not possible. Additionally, we would like to argue, the assumption that they are separable is undesirable if we wish to gain a deeper understanding of spoken language. Progress in laboratory phonology is therefore likely to benefit from more input of detailed studies of peripheral motor processes as well as from using a combined strategy, in which we "co-develop phonological theory and models of speech motor control" and "in which the elaboration of one depends on that of the other" (Munhall et al. 2000: 26).

PART III

MULTIDIMENSIONAL REPRESENTATIONS OF KNOWLEDGE OF SOUND STRUCTURE

The goals of this part are to examine the substance, access, and acquisition and change over time of representations of speech, with special attention to their variety and richness, the relationships between levels, and the challenges this multiplicity of representations offers to current and future theories of linguistic knowledge.

This page intentionally left blank

CHAPTER 8

LEXICAL REPRESENTATIONS

This chapter presents four different views of the nature of lexical representation. Together the authors paint a picture of "the lexicon" as both abstract and detailed, as a repository of information and a dynamic system at the crossroads of a multi-dimensional system of communication. The contributions also offer a number of ways to investigate the nature of lexical and phonological representations and their relationships to each other, as well as the role and integration of production and perception in structuring this aspect of the linguistic system.

8.1 Probing underlying representations

Adam Albright

8.1.1 Introduction

Lexical representations play many different roles in phonological theory. At the most basic level, the lexicon is assumed to contain all the information needed to recognize and produce words, and therefore most theories of lexical representation encode words or morphemes with at least enough detail to maintain surface contrasts. For the task of lexical recognition, however, representations must be sufficiently schematic (or the matching sufficiently loose, as in exemplar models) to allow listeners to abstract away from surface variation in the phonetic realization of a word due to differences in coarticulation, speech rate, intonation, and speaker-specific properties, as well as random noise. The problem of abstracting away from phonetic variability to arrive at broader categories is discussed by Johnson (1997b, 2007) and Pierrehumbert (2001a, 2001b, this chapter).

If the lexical representations of words were simply their surface realizations with all coarticulatory and speaker/utterance-dependent phonetic details removed, the lexicon might consist of something like broad phonetic transcriptions, with all variable and physically necessary coarticulatory information removed. For example, in the author's dialect of American English, the word *repute* might be represented as /ɹəˈpʰjuːʔt/, omitting details such as the duration of aspiration, the pitch contour, and so on, but retaining properties such as aspiration ([pʰ]), vowel reduction ([ə]), and glottalization of the final [ʔt].

However, phonological analyses frequently posit lexical representations that deviate more radically from surface realizations. For instance, it is often hypothesized that lexical representations lack contextually predictable properties such as the aspiration of pre-tonic [pʰ]. Furthermore, comparison with related forms such as *reputation* [ˌɹɛpjəˈtʰeɪʃn̩] reveals that the glottalized [ʔt] in *repute* is a contextually

predictable variant of [t], while the schwa corresponds to [ɛ] when it bears some degree of stress. Taken together, this yields a representation of *repute* that is different from any surface form: e.g. /rɛpjuːt/.

Thus, in addition to removing non-contrastive surface detail, lexical representations serve two additional roles in phonological analysis: (1) encoding multiple realizations of morphemes in different contexts (ALTERNATIONS, such as [ɹəpʰjuːʔt] ~ [ɹɛpjətʰ]), and (2) constraining the set of alternations that are observed in a given language, independent of the morphemes involved ([ʔt] alternates with [tʰ], but [d] does not). In both cases, lexical representations work together with the phonological grammar to derive the set of possible surface forms.

In this section, I consider experimental evidence concerning several commonly held hypotheses about how underlying representations deviate from surface forms. The discussion rests on several background assumptions. The first is that words are not represented as unanalyzed wholes, but rather, are decomposed into morphemes.[1]

I adopt as a starting point the assumption that each morpheme has a single phonological representation (its UNDERLYING FORM), which is shared by all morphologically complex words involving that morpheme. In addition, consistent with the idea that the lexicon contains enough information to distinguish morphemes from one another, a useful initial hypothesis is that if two sounds occur in the same context, they must be distinguished somehow in their lexical representations—that is, minimally, the lexicon contains information about phonological contrasts; this assumption is revisited in Sections 8.1.3 and 8.1.5 (see also Lahiri, this chapter). Finally, I assume that contrasts are represented using a limited set of discrete entities (e.g. features). Although these assumptions are certainly not universal, versions of them are found in many analytical traditions.

It should also be noted that for many of the theoretical questions raised here, the available experimental evidence is limited or inconclusive. In such cases, preliminary evidence from potentially instructive studies will be considered, as a pointer toward areas where further investigation is needed.

8.1.2 Non-neutralizing allophony

In the simplest cases of allomorphy, a morpheme varies predictably depending on the phonological context, but its segments remain distinct from other contrasting segments. For example in Korean, lax stops are predictably voiced between sonorants within an accentual phrase, and voiceless elsewhere ([tubu] 'tofu' ~

[1] Indirect support for this assumption comes from the literature on morphological priming, which shows that (at least under certain conditions) recognizing a morphologically complex word can facilitate subsequent recognition of other words involving the same morpheme; see McQueen and Cutler (1998) and Feldman (2003) for overviews of evidence and current debates.

[sun **d**ubu] 'pure (soft) tofu') (Jun 1994a), while in English, voiceless stops are predictably aspirated at the beginning of stressed syllables but unaspirated in most other contexts (*ácádemy* [əˈkʰærəmiː] vs. *àcadémic* [ˌækəˈdɛmək]). In many frameworks, including that of *The Sound Pattern of English* (Chomsky and Halle 1968), it is assumed that predictable or redundant features need not be included in lexical representations (UNDERSPECIFICATION) since they are redundant and can be supplied by the grammar.

The question of whether redundant features are removed from lexical entries is especially interesting because it has been called into question by work in Optimality Theory (OT: Prince and Smolensky 2004). OT posits that enforcing surface distributions is entirely the responsibility of the grammar, and that economy of lexical entries is offset by the cost of forcing the grammar to modify strings in order to produce well-formed surface forms. In accordance with the principle of RICHNESS OF THE BASE (Prince and Smolensky 2004: 225), many OT analyses remain agnostic as to whether predictable features are present in underlying representations, while the principle of LEXICON OPTIMIZATION (Prince and Smolensky 2004: 225–30) demands that they be included, barring evidence that a given morpheme alternates.

The arguments for omitting redundant features from lexical representations (underspecification) or including them (lexicon optimization) are primarily theoretical, based on the principles of economy of representation or derivation. However, underspecification of lexical entries has also been the subject of much experimental work, following the influential work of Lahiri and Marslen-Wilson (1991) on underspecification of nasality in English and Bengali. A variety of sources of experimental evidence that have been brought to bear on underspecification are reviewed by Lahiri (this chapter). The logic of these studies is that if a given feature is unspecified in the lexicon, listeners should be insensitive to it when deciding whether an acoustic signal matches a potentially relevant lexical entry.

In the case of underspecification of allophonic features, the question is whether the inappropriate presence or absence of a feature-like aspiration in the acoustic signal is an obstacle to recognizing segments that surface as consistently aspirated (such as the [kʰ] in *cab* [kʰæb]), segments that are consistently unaspirated (such as the [k] in *scab* [skæb]), or segments that alternate (such as in *accuse* [kʰ] ~ *accusation* [k]). To date, few word recognition studies have focused specifically on systematic manipulations of allophonic features. One suggestive finding comes from Gow and Gordon (1995), who presented subjects with phonemically ambiguous target strings such as /pæstɛl/, which could represent either a single word (*pastel*) or a combination of two words (*pass tell*). Due to the presence of the word boundary, the two renditions differed in allophonic aspiration: *pas*[t]*el* vs. *pass* [tʰ]*ell*. These strings were embedded in sentences that were identical up until the ambiguous string (*When the runner(')s* /pæstɛl/. . .). These phonemically ambiguous but phonetically distinct sentences were used in a cross-modal priming task, in which semantic associates of the single-word interpretations (e.g. *color*) were

presented for lexical decision (for discussion of cross-modal semantic priming, see Tabossi 1996). As expected, single-word renditions significantly primed their semantic associates (e.g. *pastel* [pʰæstɛl] primes *color*). Importantly, two-word renditions also significantly primed single-word associates (e.g. *pass tell* [pʰæstʰɛl] also primes *color*), in spite of the fact that aspiration should preclude this interpretation. This result is reminiscent of the results reviewed by Lahiri (this chapter), in which an inappropriate feature does not impede recognition of a lexical entry in which it is unspecified. This leaves open the question, however, of why /t/ is unspecified for aspiration: is it because aspiration is predictable and never specified in English (CONTRASTIVE UNDERSPECIFICATION: Steriade 1987), or because unmarked values (e.g. lack of aspiration) are unspecified (RADICAL UNDERSPECIFICATION: Kiparsky 1982; Mohanan 1991)?

As it turns out, English listeners do not always ignore aspiration. Gow and Gordon (1995) show that only [pʰæstʰɛl] primes semantic associates of the second element *tell*, while [pʰæstɛl] does not. This resembles an asymmetry identified by Lahiri and Marslen-Wilson (1991), in which the percept of an unmarked feature (e.g. coronal) blocks recognition of words with a marked value (labial or dorsal), but not vice versa. In the present case, we might conclude that *tell* is underlyingly /tʰɛl/ with aspiration (in accordance with lexicon optimization), and that the percept of [tɛl] in [pʰæstɛl] causes a mismatch (phonetic [aspiration] ≠ lexical [+aspiration]) that prevents lexical access and priming. By contrast, in accordance with radical underspecification, unmarked /t/ in *pastel* is underspecified, making it compatible with phonetic [t] or [tʰ].

Most studies of priming by inappropriate allophones (including Gow and Gordon 1995; Davis et al. 2002) have focused on pairs like *pastel* vs. *pass tell*, which differ in the presence of a boundary. This introduces a potential confound: perhaps priming is blocked by interference from a word segmentation strategy that is sensitive to boundary-conditioned allophony. Word-spotting studies have consistently shown that listeners are significantly better at recognizing words when the boundary is accompanied by the contextually appropriate allophones (Nakatani and Dukes 1977; Vroomen et al. 1996; Jusczyk, Hohne, and Bauman 1999; Smith and Hawkins 2000; Salverda et al. 2003; Smith 2004; Mattys 2004). Allophonic variation appears to play an important role in word segmentation, perhaps as part of a more general segmentation strategy based on phonotactic probability (McQueen 1998; Hay 2003). In tasks that do not require listeners to segment a continuous speech stream, evidence for allophonic or contextual phonetic effects appears to be weaker and less consistent (Davis et al. 2002). Further research is needed to resolve this issue.

Another source of evidence concerning the lexical status of predictable features comes from acceptability judgments of appropriate vs. inappropriate allophones. Whalen et al. (1997) tested English listeners' sensitivity to aspiration by presenting real and nonce words containing contextually appropriate or inappropriate

Table 8.1.1. Stimuli with appropriate and inappropriate allophones from Whalen et al. (1997)

		Appropriate allophone	Inappropriate allophone
Stressed	Real	*opáque* [oʊˈpʰeɪk]	[oʊˈpeɪk]
	Nonce	*opér* [oʊˈpʰɚ]	[oʊˈpɚ]
Stressless	Real	*Hópi* [ˈhoʊ̆piː]	[ˈhoʊ̆pʰiː]
	Nonce	*óper* [ˈoʊ̆pɚ]	[ˈoʊ̆pʰɚ]

allophones (Table 8.1.1), and asking subjects to rate how native the pronunciations sounded. For real words, subjects judged the contextually appropriate value to be more native. Surprisingly, for nonce words, subjects consistently judged aspirated tokens as more native, even in stressless contexts where the regular English distribution demands unaspirated stops.

Whalen et al. (1997) suggest an interpretation in which, contrary to what we have claimed above, existing words are lexically specified for predictable features. In this account, speakers recognize inappropriate aspiration by comparing perceived tokens against fully specified lexical representations. Another possibility, however, is that listeners are more willing to tolerate inappropriate aspiration for non-words, interpreting it as hyperarticulation of low-frequency words to enhance the contrast between /p/ and /b/ (Wright 2003; Pluymaekers et al. 2005; Zhao and Jurafsky 2009; Scarborough 2010). For familiar, low neighborhood density words such as *Hopi* or *opaque*, such hyperarticulation would be unnecessary, and perhaps even unacceptably odd. In a follow-up study, Jones (2002) tested the acceptability of contextually inappropriate aspiration in /st/ clusters, where the contrast with voiced stops is not an issue. In this context, subjects consistently preferred the contextually appropriate unaspirated realization, both for real and nonce words. This supports the idea that Whalen et al.'s (1997) subjects may have interpreted inappropriate aspiration as hyperarticulation, rather than as an intended featural property.

Taken together, these results confirm that listeners are not impervious to allophonic (or even suballophonic; McMurray et al. 2008) differences in the realization of phonemes. At the same time, they are compatible with the idea that just as with other non-contrastive differences such as speech rate and overall pitch, at least some predictable allophonic feature values are factored out in the course of processing, and do not count as mismatches in accessing lexical representations. The fact that speakers systematically produce allophonic alternations while speaking and extend them accordingly to nonce words can be attributed to the phonological grammar, which fills in predictable features.

8.1.3 Contextual neutralization

Frequently, alternations result in the neutralization of one or more underlying contrasts. For example, the Korean contrast between lax, aspirated, and tense stops is found only before vowels and glides (e.g. [pa**g**-**i**l] 'foil-ACC' vs. [puʌ**k**^{h}-**i**l] 'kitchen-ACC' vs. [pa**k***-**i**l] 'outside-ACC'), while in other positions, only lax (voiceless, unaspirated) stops are found ([pa**k**˺] 'foil', [puʌ**k**˺] 'kitchen', [pa**k**˺] 'outside'). A common analysis is to posit contrasting lexical representations (/pa**k**/ vs. /puʌ**k**^{h}/ vs. /pa**k***/), along with a grammar that derives contextual neutralization.

The Korean example is straightforward because all three underlying values are visible in a single context, and few analysts have questioned the need to posit an underlying contrast in such cases (though we return to this issue below). A more controversial situation arises when a contrasting element is never distinct on the surface, but is neutralized with different elements in different contexts. An example is found in Spanish.

As Table 8.1.2 shows, some Spanish verbs exhibit stress-conditioned alternations between mid vowels and diphthongs ([e], [o] ∼ [je], [we]), while other verbs have non-alternating mid vowels or diphthongs. Following the logic that lexical representations are repositories of unpredictable information, we are forced to conclude that the difference between alternating and non-alternating mid vowels is encoded in their underlying forms. One option is to list both allomorphs in the lexicon without deriving one from the other: {/sent/, /ˈsjent/} (Hooper 1976; see also Mester 1994; Mascaró 1996; Kager 1996; Rubach and Booij 2001). Alternatively, we may encode the fact that stressed and stressless allomorphs tend to differ in precisely the same way by positing a special underlying value for diphthongizing stems; for example, as [−tense], [+long], or more abstractly, [+D] (see J. Harris 1969: 116–18, and Harris 1978 for discussion). This solution requires that the grammar map underlyingly [−tense] (etc.) vowels to diphthongs when they are stressed, but it does not attempt to explain why certain verbs have underlyingly [−tense] vowels. Thus, a claim of this approach is that the lexicon, and not the grammar, is the locus of knowledge about which morphemes alternate.

One reason to question this claim for Spanish is that the difference between alternating and non-alternating segments is at least partially predictable from morphological and phonological factors. Brame and Bordelois (1973: 156–9) observe

Table 8.1.2. A three-way contrast in Spanish

		Stressless		Stressed		Type frequency
a.	Alternating:	sen̲ˈtar	'seat-INF'	ˈsjen̲to	'seat-1SG'	Majority
b.	Mid vowel:	presen̲ˈtar	'present-INF'	preˈsen̲to	'rent-1SG'	Minority
c.	Diphthong:	orjen̲ˈtar	'orient-INF'	oˈrjen̲to	'orient-1SG'	Very rare

that although there are exceptions, diphthongization is typically associated with a specific set of phonological contexts: before sonorants, voiced stops, and stridents. They propose a grammatical analysis in which vowel alternation rules refer to segmental contexts. Although they stop short of eliminating abstract representations of alternating vowels, an extreme version of this approach would rely on segmentally restricted rules to capture the behavior of as many words as possible, leaving the rest as listed exceptions. This grammar-oriented approach makes a testable prediction: if speakers are forced to inflect rare or unknown (nonce) verbs, they should decide the outcome based on the stem's phonological shape.

Several studies have attempted to test this prediction by considering generalization to nonce ('wug') words (Berko 1958). Kernan and Blount (1966) presented speakers with nonce verbs with diphthongs in stressed forms (e.g. *el hombre* [ˈswétʃa] 'the man *sotʃ*-3SG') and elicited stressless forms, forcing subjects to decide whether to apply the alternation. Subjects unanimously chose non-alternating diphthongs (e.g. imperfect [sweˈtʃaba]/*[soˈtʃaba]), even though vanishingly few existing words have non-alternating diphthongs. Similarly, Bybee and Pardo (1981) presented speakers with diphthongized 3SG forms and elicited stressless preterite forms, and found an overwhelming preference for diphthongs in stressless forms. Bybee and Pardo (1981: 940) conclude, in line with the lexicon-oriented approach, that diphthong alternations are "lexically bound."

However, there is reason to think that this conclusion is premature. The preference for non-alternating diphthongs is surprising given their rarity among existing words, but it may nonetheless receive support from the statistical distribution of Spanish verbs. Table 8.1.2 reveals two opposing facts: stressed diphthongs do generally correspond to stressless mid vowels, but overall, most Spanish verbs are non-alternating (since most mid-vowel verbs do not alternate, and verbs with low and high vowels virtually never do). Thus, non-alternation is actually a robust pattern in Spanish, which is perhaps further bolstered by a preference for paradigm uniformity (Burzio 1994; Kenstowicz 1997; Steriade 2000), a reluctance to create new allomorphs (Steriade 1997), or a tendency to treat novel verbs as derived (denominal, onomatopoeic, etc.) and therefore ineligible for diphthongization. These factors may help explain the overall preference for non-alternation. The question still remains, therefore, whether speakers can nonetheless generalize diphthongization according to the segmental context.

To test this, Albright et al. (2001) carried out a statistical study of Spanish first-conjugation verbs, identifying the segmental contexts most strongly associated with diphthongization. The results echoed those of Brame and Bordelois (1973): mid vowels are especially likely to diphthongize when followed by liquids, /st/, or nasal + stop clusters, and they are especially unlikely to diphthongize when followed by voiceless stops and /tʃ/. These correlations were used to construct nonce verbs containing a range of segmental contexts; these were presented to subjects in a stressless form (1PL X-ˈamos), and a stressed form was elicited (1SG ˈX-o). The

results showed that speakers are substantially more likely to apply diphthongization in contexts that favor it in the lexicon. This supports the idea that knowledge of diphthongization goes beyond the lexical specification of individual roots, and is (at least partly) the result of variable or probabilistic rules. This result is parallel to the results of numerous wug tests on lexically restricted alternations in other languages (Zuraw 2000; Ernestus and Baayen 2003; Pierrehumbert 2006b; Becker et al. 2011; Hayes et al. 2009; and many others). In all of these cases, encoding the contrast with lexical specifications alone appears to be insufficient, since this provides no mechanism for generalizing knowledge about regularities in its distribution.

Wug test results on partially predictable contrasts pose an analytical quandary. On the one hand, they show that knowledge of alternations is not limited to the individual morphemes involved. On the other hand, the behavior of individual lexical items is not fully predictable, so a mechanism to encode morpheme-specific knowledge is still required. Becker et al. (2011), following Pater (2006), Becker (2009), and others, propose that learners are biased to posit different grammars, rather than resorting immediately to distinct featural representations. Zuraw (2000), Albright et al. (2001), and Ernestus and Baayen (2003) pursue an account in which the grammar encodes knowledge of overall lexical trends and is used to derive default and nonce forms, while the behavior of exceptional morphemes is encoded by listing their allomorphs or even surface (inflected) forms. What these theories have in common is that the phonological representations of morphemes do not include enough information to fully distinguish all surface patterns. In these accounts, the grammar takes on more of the burden of encoding the behavior of individual morphemes.

An alternative interpretation of such results is that unpredictable alternations are indeed encoded with distinct underlying representations, and that wug tests force subjects to decide on lexical representations via comparison to existing words (Schütze 2005). For example, we might hypothesize that speakers are more likely to infer an underlying [+tense] (or [–D]) vowel in some contexts than in others. The strongest version of this hypothesis is that upon learning a new word, subjects posit a single underlying form unless they receive explicit evidence requiring a more complex representation. This predicts that although different subjects in a wug test may arrive at different conclusions about a given nonce word, each subject should treat that word consistently once it has been learned. There is some evidence that this prediction is too strong. Bybee and Pardo (1981) presented novel verbs to Spanish speakers in two forms: a stressless form with a mid vowel and a stressed form with a diphthong (e.g. INF. [pon'sar], 3SG ['pwensa]). Assuming that alternating vowels are encoded as [−tense], the only underlying form that is compatible with the evidence is /po$_{[-\text{tense}]}$ns/. However, when asked to produce another stressless form, subjects sometimes volunteered diphthongized [pwens-]. This result shows that speakers may apply alternations to a morpheme in some forms but not others. If this is generally true, it indicates that alternations are not encoded at the level of a

single underlying representation (e.g. /po$_{[-\text{tense}]}$ns/ or underspecified /pOns/) that is shared by all forms of the same root. (See also Eddington 1996.) Further work is needed to determine whether such responses are a result of incomplete learning of the alternation, or failure to generalize the alternation to new morphological contexts. If alternations remain bound to the specific morphological forms in which they are presented, this would challenge the idea that all realizations of a morpheme are unified with a single underlying form.

8.1.4 Generalizing across alternations

Another important use of abstract underlying representations is to unify and constrain alternations within a language. For example, in English, the addition of certain derivational affixes triggers a set of vowel changes known as Trisyllabic Laxing, shown in Table 8.1.3. On the surface, these alternations involve diverse features: [aɪ] and [ɪ] differ in (at least) [±tense], [±high], and [±back], while [aʊ] and [ʌ] differ in [±tense], [±low], and [±round]. Chomsky and Halle (1968) posit that all of the vowels in question are underlyingly long, and that the observed alternations are due to a shortening process that is obscured by additional changes: *div*/iː/*ne* → *div*[aɪ]*ne* (with vowel shift of /iː/ → [aɪ]), but *div*/iː/*nity* → *div*[ɪ]*nity* (with shortening and laxing of /iː/ → [ɪ]); see Myers 1987 and Rubach 1996 for literature review and reanalysis. The insight of this analysis is that all of the alternations in Table 8.1.3 can be unified using a small number of rules, provided that speakers are able to arrive at underlying representations such as *div*/iː/*ne*, *extr*/eː/*me*, *t*/æː/*ble*, *verb*/ɔː/*se*, *sch*/oː/*l*, and *prof*/uː/*nd*. The claim is that these alternations form a coherent process of English phonology, which can be contrasted with idiosyncratic alternations that pair the "wrong" vowels, such as *ant*[iː]*que* ~ *ant*[ɪ]*quity* (**ant*[ɛ]*quity*), *abst*[eɪ]*n* ~ *abst*[ɛ]*ntion* (**abst*[æ]*ntion*), and *p*[iː]*ce* ~ *p*[æ]*cify* (* *p*[ɛ]*cify*).

These alternations have been subjected to much experimental investigation, in an attempt to prove (or disprove) both the abstract underlying representations and the accompanying vowel shift rule that changes long vowels like /iː/ and /eː/ into surface

Table 8.1.3. Vowel alternations in English

Tense		Lax	
[aɪ]	*divine, bite*	[ɪ]	*divinity, bit*
[iː]	*extreme, keep*	[ɛ]	*extremity, kept*
[eɪ]	*sane, table*	[æ]	*sanity, tabular*
[oʊ]	*verbose, cone*	[a]	*verbosity, conic*
[uː]	*school, lose*	[a]	*scholar, lost*
[aʊ]	*profound, abound*	[ʌ]	*profundity, abundant*

[aɪ] and [iː] (Moskowitz 1973; Cena 1978; Jaeger 1984; Wang and Derwing 1994). A clear and repeated finding is that speakers are reluctant to apply the alternations in Table 8.1.3 to nonce words in elicitation tasks (Ohala 1974). As with Spanish diphthongization, however, we must be careful to distinguish non-productivity of the alternation from a complete lack of knowledge about the process (Kiparsky 1975). As Jaeger (1984) points out, it may be the shortening rule, rather than the vowel shift rule, that is unproductive; this is confirmed by the diachronic loss of alternations in many morphemes that previously underwent shortening (Myers 1987; Lahiri and Fikkert 2004). This leaves open the possibility that even if some of the components are unproductive, speakers continue to encode knowledge about alternating morphemes with abstract underlying representations, together with the shortening and vowel shift rules.

To avoid the problem of productivity, some studies have presented alternating morphemes and tested whether subjects are better at remembering or grouping alternations that are supported by the hypothesized vowel shift rule. Jaeger (1984) performed a concept formation experiment whose subjects were trained to accept vowel shift alternations (first four rows of Table 8.1.3) and reject similar but idiosyncratic alternations (e.g. *ret*[ɛɪ]*n* ~ *ret*[ɛ]*ntion*, *p*[iː]*ce* ~ *p*[æ]*cify*). Subjects were then tested on two untrained alternations: one additional vowel shift alternation ([aʊ]~[ʌ]), and one alternation that is attested but is not explained by the vowel shift (*red*[uː]*ce* ~ *red*[ʌ]*ction*). The [aʊ]~[ʌ] pairs were systematically rejected, indicating that subjects had not learned a category as general as "vowel pairs whose underlying values are short and long correspondents." This result also echoes a previous finding by Cena (1978) that subjects performed poorly on novel words with [aʊ]~[ʌ] alternations.

This result alone is not conclusive, since it could be that subjects merely preferred to memorize the specific set of trained alternations, even if in principle they might have been able to learn a deeper generalization. The pattern is more interesting when we consider responses to the [uː] ~ [ʌ] pairs: subjects systematically accepted these, in spite of the fact that they had not been included in the training. The puzzle, then, is why subjects reject [aʊ] ~ [ʌ] alternations, choosing instead to treat [uː] ~ [ʌ] as parallel to [aɪ] ~ [ɪ] and [iː] ~ [ɛ]. Jaeger proposes that the answer is orthographic: the first four alternations in Table 8.1.3 represent the "long" and "short" readings of vowel letters (<i> *mite* ~ *mit*, <e> *mete* ~ *met*, <a> *mate* ~ *mat*, etc.), as does the [uː] ~ [ʌ] alternation (<u> *tube* ~ *tub*). The [aʊ] ~ [ʌ] alternation, on the other hand, is always accompanied by a change in spelling. As Jaeger points out, the effect cannot be attributed wholly to orthographic correspondence of "long and short vowels," since subjects were also able to master the trained pair *cycle* [aɪ] ~ *cyclic* [ɪ], even though [aɪ] and [ɪ] are not normally thought of as "long *y*" and "short *y*". Jaeger concludes that knowledge of vowel alternations is orthographically mediated (i.e. [eɪ], [iː], [aɪ], [oʊ], and [uː] are "long vowels," whereas [aʊ] is not), but does not depend on the spelling of individual morphemes involved.

Table 8.1.4. Russian voicing and vowel reduction alternations (after Kenstowicz and Kisseberth 1977: 18–33)

	Nom.Sg.	Nom.Pl.	
a.	'xlep	'xleb-a	'bread'
	'sat	'sad-**i**	'garden'
	'rok	'rog-a	'horn'
b.	'stol	stʌ'l-**i**	'table'
	'vrak	vrʌ'g-i	'enemy'
c.	pi'rok	pirʌ'g-i	'pie'
	sa'pok	səpʌ'g-i	'boot'

The results of Jaeger (1984) and other experiments on English vowel alternations show that speakers can use orthographic criteria in addition to phonological features to group sounds. However, they do not actually disprove the hypothesis that speakers posit abstract phonological representations in order to explain why sounds pattern alike. In order to answer this question, further experimental evidence is needed. As has often been noted in the literature, comparisons between literate and illiterate or normal and dyslexic readers could help shed light on whether orthographic knowledge changes phonological representations, or whether they are merely two different representations that may be recruited depending on the task.

8.1.5 Combining contrastive information from multiple allomorphs

One further use of lexical representations is to integrate information about a morpheme into a single underlying form that can be used to derive all surface allomorphs. Frequently, this requires combining contrastive information from multiple allomorphs; see the *repute* [ɹəpʰjuːʔt] ∼ *reputation* [ɹɛpjətʰ] example in the introduction. Another example comes from Kenstowicz and Kisseberth (1977: 32–3): Russian has a process of final devoicing that neutralizes stem-final voicing (Table 8.1.4).[2]

The forms in (a) show that voicing is neutralized in the nominative singular, where the suffix is Ø; for these words, forms with vowel-initial affixes reveal the voicing of the final obstruent. As seen in (b), some nouns shift their stress onto affixes, causing /o/ and /a/ to reduce to [ʌ]. In some morphemes, shown in (c),

[2] I leave aside for present purposes the question of whether final devoicing is completely neutralizing in Russian, or whether some remnant of the contrast remains; see Chen (1970: 135–7) and Dmitrieva and Jongman (2007) for discussion.

both alternations coincide; for these, the standard analysis posits representations that combine the vowel quality of the unsuffixed allomorph and the voicing of the suffixed allomorph: /pirog/, /sapog/. These representations do not match any one surface form.

This approach can be contrasted with a basic alternant approach, in which learners select a representation from among the set of surface forms (Harris 1951: 308, fn. 14; McCawley 1967; Vennemann 1974; Hooper 1976; Kenstowicz and Kisseberth 1977: 26–8). The criterion for designating a basic alternant could be morphological, perhaps favoring an isolation form (Kenstowicz and Kisseberth 1977: 18–26; Bybee 1985; Kager 1999) or a morphologically informative principal part (Stump and Finkel 2009), or a phonologically informative form (Harris 1951: 308; Vennemann 1974; Albright 2002b).

Limiting lexical representations to one surface variant clearly limits speakers' ability to encode lexical contrasts. Concretely, Russian speakers would be able to encode vowel quality (/pirok/) or final obstruent voicing (/pirʌg/), but not both. As discussed above, removing contrastive features from underlying representations forces speakers to rely on alternative mechanisms to encode information about alternations, such as specific grammatical rules or listed exceptions. This predicts that there should be observable differences between different surface contrasts, depending on whether they are encoded as part of the underlying form or derived by the grammar.

To date, few experimental studies have tested whether speakers treat certain forms in the paradigm as privileged sources of phonological information. One intriguing result comes from the Spanish study of Bybee and Pardo (1981). As described above, speakers in that study volunteered stressless forms with diphthongs surprisingly often, even though they had received overt evidence that the novel verbs had mid vowels in stressless forms (e.g. 3SG [ˈpwens-a], infinitive [p**o**nˈs-ar]). In fact, the rate of diphthong responses varied significantly depending on the particular forms that were presented; subjects were much more likely to take preterite forms as evidence for mid vowels in other preterite forms, and much less likely to take a mid vowel in the infinitive as evidence about preterite forms. Although there are numerous factors that may have led to this result, experiments that investigate in "real time" how speakers learn words could shed light on preferences or constraints concerning how underlying representations are established.

8.1.6 Conclusion

Phonological theory has generally relied on principles such as economy to determine the underlying representation of morphemes. However, as the studies reviewed here show, experimental techniques such as cross-modal priming and wug testing have the potential to distinguish the predictions of different theories of how

speakers actually learn and employ lexical representations. One area that deserves particular attention is the difference between underspecification vs. lexicon optimization (Section 8.1.2), where work is needed to test not only for asymmetries in how inappropriate allophonic values affect recognition, but also for differences depending on whether a particular morpheme is known to alternate. The question of how much of the burden of representing alternations is borne by the lexicon and how much by the grammar (Sections 8.1.3–4) also remains an open issue, though a growing body of evidence points to a larger role for grammar than has traditionally been assumed. All of the experimental results reviewed here are open to multiple interpretations, and many questions remain. It is clear, however, that these studies have identified a number of useful techniques, and laid the groundwork for testing more detailed hypotheses about the principles that speakers use to encode knowledge of words and phonological distributions.

8.2 Asymmetric phonological representations of *words* in the mental lexicon*

Aditi Lahiri

8.2.1 Introduction

Lexical phonological representations typically denote underlying representations of words and their constituent morphemes. In a derivational setting, speaker's rules apply to this idealized representation, providing surface outputs. Within constraint-based approaches, several competing outputs are compared with respect to an idealized input, and the correct ranking of constraints allows the speaker to select the preferred output. For the listener, the acoustic realization of the speaker's output is perceived and must be decoded to access and identify the representation of the intended word. A necessary line of enquiry is whether the lexical representations—"the lexicon"—for the speaker and the listener are the same, and if so, what the phonological criteria are by which these representations are defined. Characteristically, within phonology, underlying phonological representations refer to the lexical

* The research reported here would never have succeeded without the participation of the entire Leibniz group particularly Carsten Eulitz, Verena Felder, Claudia Friedrich, Astrid Kraehenmann, Jonas Obleser, Mathias Scharinger, and Allison Wetterlin. The credit (and blame) for the model is shared with Henning Reetz.

representation relevant for the speaker. The speaker is in control of what she says and after a concept has been created, appropriate sentence structure, grammatical rules, and phonological shapes are fed to the articulators to produce the intended output (cf. Levelt 1989; Levelt et al. 1999). Word comprehension, however, is rather different. The listener must take on the task of decoding the speaker's output. Since no word is ever spoken by the same speaker in an identical fashion, the listener's task of parsing the speech output and identifying the intended word is not straightforward.

Even if we assume that the lexicon is the same for production and perception, the courses of action implicated in *planning and articulating* words as against parsing the speech output to *identify* words are rather different. The listener has to determine the way in which the output matches to the input/underlying representation. (See Nguyen, Schiller, Ernestus, and McMurray and Farris-Trimble, this volume.)

Figure 8.2.1 illustrates place assimilation across affixes and optionally across words. The lefthand box indicates the speaker's output, while the righthand box indicates what the listener has to do. The examples demonstrate CORONAL assimilation, where /ɪn/ or /tɛn/ can be pronounced as [ɪm] when followed by *perfect* or [tɛn]/[tɛŋ] when followed by *cats*, depending on whether the speaker chooses to apply the place assimilation rule. The listener needs to ensure that both [tɛn] and [tɛŋ] are accepted as a variant of *ten*. The point to note is that final [m] of *cream* in *cream cake*, remains unaltered. Such variations are rather common across languages of the world, although there may be language-specific restrictions on domains of application.

Phonology is concerned with patterns of alternations where the phonological shape of a word may vary in diverse contexts. As seen from the examples above, the alternations can be (i) morphophonologically and lexically determined (as in *imperfect*); (ii) purely allophonic (like aspirated [p^h] and unaspirated [p], where [p]*an* vs. [p^h]*an* does not result in a meaning difference); or (iii) a consequence of the phonological context of a neighboring word ([tɛŋ] [kæts] *ten cats*).

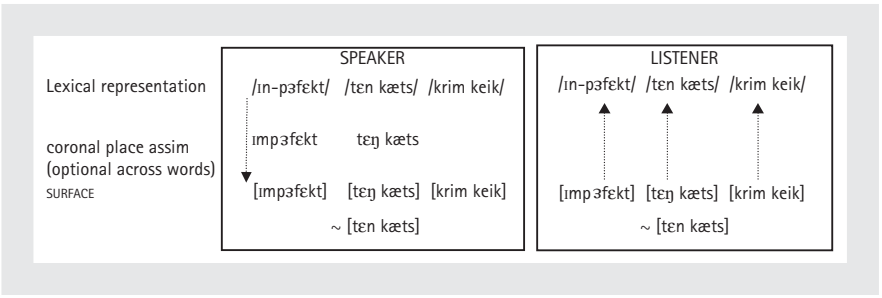

Figure 8.2.1. Speaker's and listener's actions to produce and perceive assimilation across affixes and words.

Despite extensive discussion about the nature of rules and constraints as well as types of contexts and domains relevant for the speaker, very little is said about the listener in most phonological models (though see Hume and Johnson 2001). Mohanan and Mohanan (1986) assumed multiple lexical phonological levels, and concluded that the phonological form which feeds into post-lexical rules (i.e. the phrasal domain) is the appropriate representation for the listener rather than the underlying representation. Language change, however, is an area where the listener has long been accorded importance (e.g. Ohala 1981a; Gussenhoven and Jacobs 1998; Blevins 2004; see Chitoran, this volume), but we are still far from understanding the nature of the listener's representation in language processing.

Psycholinguistic models of language comprehension up until the mid-Eighties rarely dealt with phonological variation. Morphophonological alternations such as *public* ~ *publicity* were discussed not as a phonological problem, but as an obstacle towards morphological decomposition (where there were two rival points of view: decomposition vs. full listing of all morphologically complex words). In the Nineties, psycholinguistic research on morphology began addressing the representational consequences of such alternations. Marslen-Wilson et al. (1994) led with their claim that semantically transparent morphologically complex words with phonological alternations must have unique phonologically abstract (and underspecified) stem morphemes (e.g. *sane* ~ *sanity*).

The possible role of allophonic predictable alternations where non-contrastive features were involved was not addressed until Lahiri and Marslen-Wilson (1991, 1992). They introduced the notion that the nature of phonological representations in the mental lexicon depends in part on phonological alternations and contrastive features of a given language. Investigating the status of vowel nasalization in Bengali (where it contrasts) and English (where it does not), they addressed the question of whether, if the same phonological process leading to identical surface outputs was allophonic in one language and neutralizing in another, the phonological representations of the relevant words would be the same.

Allophonic and morphophonological alternations are often treated in a unified manner phonologically, both being seen as the result of rule application or constraint interaction. Yet it has been argued that some such patterns may be generalizations over stored items in the lexicon (e.g. Bybee 2001). What tends to be overlooked is that the same phonological process may give rise to alternations not only as a consequence of affixation, but also across word boundaries (e.g. Figure 8.2.1). Even if we assume that only post-lexical phrasal alternations are relevant for online processing, such phonological variations are not entirely straightforward. Unlike lexical alternations, phonological variations across words are almost always optional and largely dependent on phrasing. For example, place assimilation is more likely phrase-medially than across two phonological phrases. Furthermore, most phonological systems, like many other aspects of linguistic systems, are asymmetric (cf. Lahiri and Plank 2009). For example, function words tend to cliticize and

attach to the phonological word on the left rather than the one to the right; voiced consonants devoice word-finally rather than initially; coronals tend to assimilate to dorsals more often than the other way around; word-final vowel deletion is more frequent than word-initial deletion, etc. Consequently, it is not unexpected that lexical phonological representations should be asymmetric. The claim that we wish to maintain is that asymmetric representations reflect asymmetry in processing, and in what follows, we provide support for this claim.

In sum, until recently phonological theories as well as psycholinguistics models lacked a well-defined understanding about the nature of the phonological representation of words in the mental lexicon relevant for the listener. Even if some patterns of phonological alternation are assumed to be static generalizations, how does the listener deal with phonological variations that are phrasal and not bound to lexical affixation? We begin by briefly describing the model *Featurally Underspecified Lexicon* (FUL), which defines the nature of the phonological representations assumed for the listener and speaker. This is followed by a description of how FUL operates—the extraction of features as well as the matching process from the signal to the lexicon, ultimately leading to the recognition of words. A variety of experimental paradigms and results support the idea that the processing system tolerates phonological variation to a certain degree. This tolerance is asymmetric as well, reflecting the asymmetry in representation as well as the asymmetry in the matching process.

8.2.2 FUL: Features, representation, and processing

To address the problems of the listener coping with variable shapes of words, we have proposed a view of lexical representation built on the *Featurally Underspecified Lexicon* (Lahiri and Reetz 2002, 2010; see also Reetz, this volume) and subsequent work. The acoustic correlates of features developed gradually in prior work (Lahiri and Blumstein 1984; Lahiri et al. 1984; Evers et al. 1998). The representation-cum-processing approach began in work by Lahiri and Marslen-Wilson (1991, 1992). The full-fledged version of a three-way algorithm assuming *match*, *mismatch*, and *no-mismatch* was developed in Lahiri and Reetz (2002) with a speech recognition algorithm spelt out in Reetz (2003). The idea is that since acoustic variability is infinite and the signal will never yield all details necessary for recognition, storing all variability that has been encountered by the listener cannot be the most efficient solution for word recognition. Instead, the perception mechanism extracts as much information from the signal as it can and transfers it into phonological abstract features based on gross heuristics. Perfect match or mismatch of features to the lexicon is not the only option—the lexicon can also tolerate features if they do not conflict, which is the *no-mismatch* condition. This is the crux of the solution. Lexical tolerance allows for

more options until syntax, semantics, or some other linguistic information helps the listener. A full introduction to the phonological model is not possible here; rather, we highlight the crucial ingredients. (For a fuller discussion, see Lahiri and Reetz 2002, 2010; Scharinger et al. 2009). Inspired by Clements (see Clements and Hume 1995 for a review) and returning to Jakobson et al.'s (1952) view, we take a unified approach, combining all consonantal and vocalic features, as shown in Figure 8.2.2.

According to FUL, every segment in all languages will have one of the major class features—CONSONANTAL or VOCALIC and SONORANT or OBSTRUENT. The members of each pair are conflicting—i.e. CONSONANTAL implies *not* VOCALIC and vice versa. As for PLACE, the idea is that the constriction relevant on the horizontal dimension along the vocal tract is determined by the ARTICULATORS, and on the vertical dimension is characterized by TONGUE HEIGHT. Although these are described articulatorily, they all have acoustic correlates.

Features like HIGH and LOW are mutually exclusive, but not binary. This is because a vowel, for instance, cannot be both HIGH and LOW, but it may be neither. Our feature organization is such that there are no dependencies other than the inherent ones, such as [NASAL] implying [SONORANT]. All segments are distinguished by a combination of these features. A partial list of segment classification is given in (1).

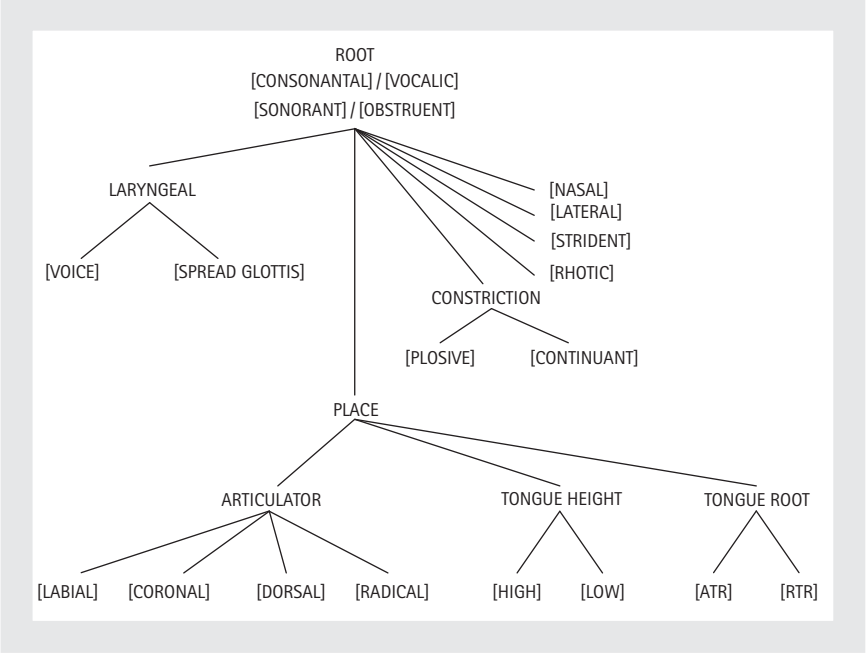

Figure 8.2.2. Feature organization in FUL.

(1) Features and segments

[LABIAL]	labial consonants, rounded vowels
[CORONAL]	front vowels, dental, alveolar palatal, palatoalveolar, retroflex consonants
[DORSAL]	back vowels, velar, uvular consonants
[RADICAL]	pharyngealized vowels, glottal, pharyngeal consonants
[HIGH]	high vowels, palatalized consonants, retroflex, velar, palatal, pharyngeal consonants
[LOW]	low vowels, dental, uvular consonants

8.2.2.1 *Underspecification and representation*

The feature organization leads to feature representation. Underspecification is part of the model, with CORONAL always contrastive and always underspecified. This underspecification is not dependent on context. FUL claims that CORONAL is underspecified in all word positions, initial, medial, and final. Space is too limited to review the history of underspecification, though see Lahiri and Reetz (2010). Instead, we build on Clements (2001), who argues that the feature [coronal] only becomes available when necessary. Coronal transparency can be accounted for either by its absence where it is inactive or "by the nonprojection of [coronal] (in segment classes in which the feature is active but unprojected)" (Clements 2001: 115).

Our approach, however, accepts CORONAL underspecification in all languages irrespective of whether the feature is active or not (also see Dresher 2008). How then are features assigned? For feature assignment, assuming underspecification, we follow Ghini (2001b) and accept the notion "place first" such that a child acquiring a language will assign ARTICULATOR features first. Furthermore, based on Levelt (1995) and Fikkert and Levelt (2008), the assumption is that [LABIAL] is acquired as the specified feature in contrast to [CORONAL], a contrast which is assumed to exist in every language. Thus, the first cut for children would be LABIAL versus CORONAL, and since CORONAL exists always, it can afford to remain underspecified. Only if another distinction is required, DORSAL is assigned. Once the child is settled on the ARTICULATOR distinctions, other distinctions follow (cf. Ghini 2001a; Fikkert and Levelt 2008). LARYNGEAL distinctions come later, and here the choices are easier because the features are either present or absent. Thus, the [CORONAL]/[LABIAL] contrast is established first, with [LABIAL] specified and [CORONAL] always remaining underspecified. The contrasts [CONSONANTAL]/[VOCALIC] and [OBSTRUENT]/[SONORANT] are present in all languages. All other features depend on the phonological systems of individual languages. For other views on this stage of acquisition, see Munson et al. (this volume); Demuth and Song (this volume).[3]

[3] The same lack of full specification holds for tonal contrasts as well (cf. Lahiri et al. 2005 for Swedish and Norwegian) and similar predictions are made for speech perception and matching purposes (Felder et al. 2009).

8.2.2.2 *Specification in representation and presence in the speech output*

Recall that the specification of features does not depend on whether any feature is active in a phonological rule, but only if it is necessary to establish a phonemic contrast. However, the upshot of allowing privative features is that the absence of a feature may be considered to be a form of underspecification. There is, however, a difference. The opposite of a privative feature never plays a role in speech output. For instance, lack of NASAL does not imply that orality is a characteristic of any significance and therefore, no orality can be extracted from the signal to have a bearing on the matching process. That is not the case for CORONAL, which is assumed always to be contrastive but underspecified. Hence, CORONAL can be extracted from the acoustic signal and is relevant for the matching process, as we shall see below. Thus, underspecification of a contrastive feature in FUL only occurs when there is a three-way contrast: PLACE (ARTICULATORS, T-HEIGHT, T-BODY) and CONSTRICTION. Furthermore, all contrastive, and not just specified features, are realized in the speech output and play a role in speech perception. For example, a feature like [VOICE] is specified if the language contrasts a set of voiced and voiceless segments. *Voicelessness* does not exist as a feature and hence consonants which are not voiced will remain unspecified in production and this feature cannot be extracted by the perceptual system. Like [VOICE], features such as [NASAL], [STRIDENT], [LATERAL] etc. are specified if a contrast is established. The absence of these attributes plays no role in production or in perception.

We now turn to the question of how recognition works, which in turn gives us insight into the necessary attributes of lexical representation.

8.2.3 FUL'S *MODUS OPERANDI* for the listener

The key assumption in this model is that the listener requires two steps from the signal to the representation: the extraction of phonological features from the signal (not segments or syllables) and a ternary mapping logic. It is the combination of the ternary matching procedure and the not-fully-specified representations that makes FUL distinctive.

(2) Speech-to-Representation

i. the acoustic signal is parsed using rough holistic acoustic parameters which convert them into phonological features (and not segments)
ii. a mapping process, using a ternary logic of *match*, *mismatch*, and *no-mismatch*, matches up the features extracted from the acoustic signal with those stored in the mental lexicon

The *match* condition is transparent. A *mismatch* occurs when a feature extracted from the signal is in conflict with the feature in the representation. The *no-mismatch*

Table 8.2.1. Matching predictions (German)

CONSONANTS			VOWELS		
Signal	Match	Representation	Signal	Match	Representation
[p,b,m] LAB	*no-mismatch*	/t,d,n/ [ ]	[o,u] LAB	*no-mismatch*	/e,i/ [ ]
[t,d,n] COR	*mismatch*	/p,b,m/ LAB	[i,e,æ] COR	*mismatch*	/u,o,ɑ/ DOR
[k,g,ŋ] DOR	*mismatch*	/p,b,m/ LAB	[y,u] HIGH	*mismatch*	/æ,ɑ/ LOW
[t,d] COR	*mismatch*	/k,g/DOR	[y,u] HIGH	*no-mismatch*	/e,o/ [ ]
[k,g,ŋ] DOR	*no-mismatch*	/t,d,n/ [ ]	[ɑ] LOW	*no-mismatch*	/o/[ ]
[s,z] STRID	*mismatch*	/n/ NASAL	[o,e,ø] *no* TH	*no-mismatch*	/u,i,y/ HIGH
[d,g,b] VOICE	*no-mismatch*	/t,k,p/ [ ]	[o,e] *no* TH	*no-mismatch*	/a,æ/ LOW
[t,p,k] *no* LAR	*no-mismatch*	/d,b,g/ VOICE	[i] COR	*no-mismatch*	/y/ LAB

condition assumes that certain non-perfect matches are tolerated due to underspecified representation. Some matching predictions for German consonants and vowels are given in Table 8.2.1.

As stated above, for unvoiced consonants there is no voicelessness in the representation, nor can a *voiceless* feature be extracted. In the same way, mid vowels lack any TONGUE HEIGHT features and will not extract any such property as "mid." We need to highlight one further point. Although vowels and consonants share same PLACE features, there is one difference. FUL assumes that universally, LABIAL can co-occur with DORSAL and CORONAL for vowels, but not for consonants. Thus, /p/ LABIAL *mismatches* /k/ DORSAL, but /o/ can be both [DORSAL, LABIAL], just as /ø/ can be [CORONAL, LABIAL].

As the features reach the lexicon, following Lahiri and Marslen-Wilson (1991), more than one word is activated in the lexicon. Unlike the traditional cohort model (Zwitserlood and Marslen-Wilson 1989), FUL assumes that the matching procedure involves features and not segments, and the ternary mapping process controls what may or may not be activated.

In Table 8.2.2, we see an illustration of how the word candidates differ after encountering the features from [bu] followed by the respective nasals [m] and [n] of the words *boom* and *boon*. A normal [u] would produce [DORSAL, LABIAL, HIGH, ATR] and activate words with [i, e, o, u] but not mismatching LOW [a, æ]. Many choices are automatically eliminated because there are no words fitting the criterion: e.g. [biz], [bɛp], [bud], [bop], etc. Underspecification increases choices, but the lexicon may limit them. Note that final [m] from *boom* would allow *boon*, but the [CORONAL] extracted from [n] in *boon* would eliminate *boom*.

8.2.3.1 *Underspecification and variation in the signal*

What we have presented so far assumes that the signal-to-feature mapping will always be perfect, which is of course not the case. It is precisely because the speaker's

Table 8.2.2. Feature extraction and word activation for the words *boon* and *boom*. The example reflects lexical activation beginning from the vowel

Extracted from signal of *boon*	Activation of words (subset)
i) *b* [u] DOR, LAB, HI	*b* [u/e/o/i]. . . [*all consonants*] (*boom, boon, booze/babe, bake, bail/bowl, boat/beak, beet, bead*. . .)
ii) *b* [u] DOR, LAB,HI [n] COR, NAS	*b* [u/e/o/i]. . . [COR stops, nasals, liquids] (*boon/bail/bowl, boat/beet, bead*. . .)
Extracted from signal of *boom*	Activation of words (subset)
iii) *b* [u] DOR, LAB, HI	*b* [u/e/o/i]. . . [*all consonants*]. . . (*boom, boon, booze/babe, bake, bail/bowl, boat/beak, beet, bead*. . .)
iv) *b* [u] DOR, LAB, HI [m] LAB, NAS	*b* [u/e/o/i]. . . [LAB/COR stops, nasals, liquids] (*boom, boon/babe, bail/bowl, boat/beet, bead*. . .)

production is never perfect that we assume sparse specification. A concrete example with mid vowels will make the point. The acoustic signal-to-feature mapping is based on general principles where the details may differ from language to language, depending on the number of height contrasts. For instance, the cut-off value for F1 as the (inverse) reflex for height may vary across languages. If a language has a two-height phonological contrast like Turkish, a higher F1 value may be acceptable as HIGH as compared to English. For German, based on the data from the Kiel corpus, we take HIGH = F1 < 350 Hz and LOW = F1 > 600 (Reetz 2000, this volume). Thus, if the F1 of a vowel happens to be below 350Hz, feature HIGH is extracted, else nothing; if it is above 600Hz, it will be assigned LOW, else no height feature. With similar F1 values in English, Figure 8.2.3 illustrates the feature extraction and activation possibilities from the first vowel of an intended utterance of *ape*.

FUL assumes that a mid vowel like [e] may be contextually spoken as a higher or a lower vowel than an idealized intended pronunciation. If it is pronounced as mid-high and the extracted first formant triggers [HIGH], this variant of [e] would be a better match for /i/ than for /e/. Nevertheless, /e/ would still be activated. The low vowel /æ/ would not be activated since the extracted [HIGH] would conflict with the [LOW] of /æ/. Similarly, if [e] was pronounced with a fairly high F1, [LOW] may be extracted and it would be a better match for /æ/, but /e/ would still be activated. Furthermore, if the F1 value happens not to fall within the 350 to 600 Hz range, no height feature will be extracted and there is no mismatch with [HIGH], [LOW], or []. As a consequence, an [e] properly pronounced as a non-high, non-low vowel leads to a no-mismatch with the feature specification of an /e/ in the lexicon irrespective of acoustic variation of F1. This is one of the ways in which variation is handled in FUL. Due to sparse specification in the representation, no matter how variable the initial vowel of *ape* may be, the intended word remains activated.

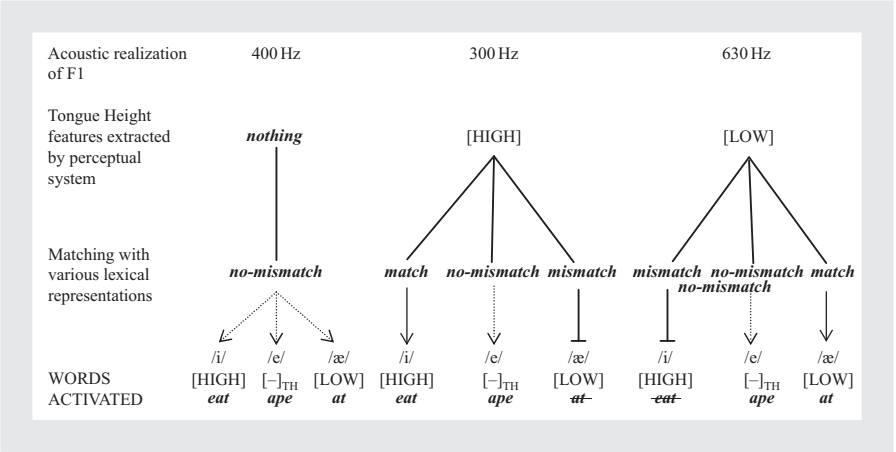

Figure 8.2.3. Variable production of [e] (as the first vowel of *ape*), feature extraction and lexical access. Words which are crossed out (e.g. ~~*at*~~) are not activated.

Thus, FUL assumes that variation in speech may lead to inaccurate and imprecise production and hence incorrect feature extraction. However, due to underspecification in the representation and ternary mapping logic, the inaccuracy will not exclude intended words from being recognized. The lexicon constrains the possible choices and would eventually ascertain the best candidate.

In the next two sections, we review a few experimental results, focusing on effects of CORONAL underspecification, supporting our model. The experimental methodologies include psycholinguistic behavioral tasks measuring reaction time and electrophysiological measurement of brainwave activity known as event-related brain potentials (ERP; see Idsardi and Poeppel, this volume). The combination of techniques offers a highly effective and broader assessment of cognitive and neural phonological processing constraints in adults.

8.2.4 Asymmetry in processing: Word-medial consonant variation

The feature representation in FUL assumes context-free underspecification, in contrast to earlier work in which it was widely assumed to be context-dependent (e.g. Lahiri and Marslen-Wilson 1991, 1992). FUL takes CORONAL underspecification to be universal in lexical representation, and not dependent on whether or not place assimilation occurs in a language. Subsequent to Lahiri and Marslen-Wilson (1991), psycholinguistic and neurolinguistic research on variation and abstract representation has focused on phonological word-shape variation due only to assimilation

(see e.g. Gaskell and Marslen-Wilson 1996, 1998, 2001; Gow 2001, 2002b, 2003; Coenen et al. 2001; Wheeldon and Waksler 2004; Snoeren et al. 2009 and references therein). Although these scholars assumed that an asymmetry in place assimilation existed word-finally (where word-final /n/ could surface as [m] when followed by a labial consonant but a word-final /m/ would never assimilate), they differed in their conclusions as to how to account for this asymmetry. Three main proposals have been advocated in the literature. First, all variants that a listener encounters are stored (Johnson 2006). Consequently since a word like *cream* is never produced as **crean* in any context, the listener will have no need to store it, thereby accounting for the asymmetry. Second, as supported by Gaskell and colleagues, although only one form is stored, there is no necessity for underspecified representations since the context of the assimilated variant provides the cue to the correct lexical item. The assimilatory variant in an incorrect context would not lead to lexical access (i.e. **browm* in the context *browm coat* will not access the word *browm* since [k] of *coat* is an impossible context for the labial [m] in **browm*).[4] Wheeldon and Waksler (2004), however, did not find this context dependency; thus, *browm* successfully accessed *brown* whatever the context. Third, as proposed by Gow, since assimilation is never complete, there is always some acoustic remnant in the signal. Thus traces of coronality of the variant **browm* can activate the real word *brown* (see McMurray and Farris-Trimble, this volume; Nguyen, this volume). The fourth option, supported by FUL, suggests that underspecification is universal in lexical representation, and thus does not depend only on assimilation or any other rule or constraint interaction leading to word shape alternation. Since CORONAL underspecification is context-free, medial as well as initial CORONALS would be underspecified in their representation.[5]

Thus, perception of place in medial position is the strongest test of FULs predictions. The idea is that mispronunciations and misperceptions can occur in all word positions and only the underspecified consonants can be activated by mispronounced variants, not the specified ones. This assumption has been tested in a variety of experiments: Lahiri and Reetz (2002) used a behavioral lexical decision task with semantic priming, where the primes were either words or pseudowords in German. The mispronunciation focused on medial consonants. The assumption was that if the phonological representation of a word is automatically activated, then its semantic field would be activated as well. Consequently, if a CORONAL word is mispronounced as DORSAL or LABIAL, these features would not mismatch with the lexical underspecified representation and by activating the phonological representation, the semantic field and synonyms would also be automatically activated. The opposite would not hold. If a LABIAL word was mispronounced as CORONAL,

[4] It is not clear why **browm* in utterance final or in isolation primes BROWN in a cross-modal task (see Marslen-Wilson et al. 1995). No context is available here.

[5] Lahiri and Reetz (2010) discuss experimental results and other models dealing with assimilation.

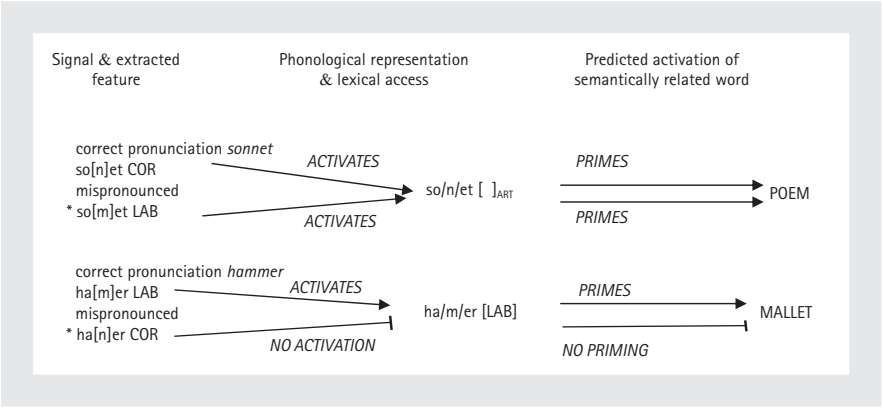

Figure 8.2.4. Activation hypothesis of listener in semantic priming task.

then the extracted coronal feature would mismatch with the lexical representation and would not activate it, in turn failing to activate the concomitant semantic field. An illustration of the activation hypothesis is given in Figure 8.2.4 using English words as illustration.

The experiment had a cross-modal design where the participants heard the prime and saw the target. The task was to decide whether the target was a word or not. The prediction was that [CORONAL] extracted from the medial consonant of mispronounced **hanner* would mismatch with the [LABIAL] of the real word representation *hammer* and would *not* activate semantically related *MALLET*. In contrast, the underspecified /n/ of *sonnet* accepts its non-word variant **sommet* since the [LABIAL] [m] does *not mismatch* with /n/. The activation, then, is reflected in the semantic activation. As expected in a semantic priming task, the real words *sonnet* and *hammer* primed their semantically related targets POEM and MALLET compared to a set of unrelated controls. The real issue was the priming with the non-words. As predicted, **hanner* did *not* prime MALLET, but **sommet* did prime POEM; that is, reaction times to *POEM* were the same irrespective of whether the real word *sonnet* or the non-word **sommet* preceded it. This result could not be attributed to experience, because such mispronunciations are not due to regular phonological rules.

Recall that the listener is not being asked to make a decision on the non-words. Rather, the decision is made on semantically related words. The semantically related POEM was equally fast when preceded by **sommet* or *sonnet*. But MALLET was primed only by the real word *hammer* and **hanner* primed just as badly as the control item. Our hypothesis does not suggest that listeners are unaware of the difference between a word and a non-word. Given time, both **sommet* and **hanner* would be judged as non-existing words. The issue is whether there would be an initial automatic activation of a corresponding real word or not. The semantic acti-

vation of a semantically related word by a set of pseudowords derived by changing coronals to labials or dorsals, supports the view that they must have activated their real phonological counterpart.

Friedrich et al. (2006) tested the same hypothesis about medial consonant CORONAL underspecification in German in an electroencephalogram (EEG) study to examine whether similar asymmetry would be found with a more direct technique measuring brain activity. Again the prediction was that the lexico-semantic memory search processes would be successful when **so[m]et* is presented and activates the corresponding coronal word *so[n]et*, but not when the coronal variant **ha[n]er* is presented for *ha[m]er*, since it would lead to an immediate rejection as a non-existing lexical item. Thus, an asymmetry was expected at least for the initial N400 pseudoword effect, which is most likely related to lexical processing.

The task was "speeded lexical decision" to auditory stimuli. For the behavioral results, the error rates revealed significant differences. Non-coronal pseudowords like **so[m]et (<sonnet)* had significantly more errors than coronal pseudowords like **ha[n]er (<hammer)*, suggesting that subjects more easily recognized **ha[n]er* as a non-word, but had more difficulty in rejecting forms like **som[b]et* as a non-word since it did activate the real word *sonnet*. In the ERP data, the early N400 results showed a clear asymmetry in the earlier activation period of 100–250ms. Mean amplitudes of the coronal non-word variants were significantly more negative than their non-coronal base words. By contrast, ERPs for non-coronal variants did not differ from their base words in this initial part of the N400 non-word effect. Furthermore, a significant difference between both types of non-word variants, but not between both types of words, suggests that this early ERP deflection is related to mismatch detection in the case of coronal non-words.

Thus, medial coronal consonants in contrast to dorsal and labial consonants show a greater tolerance in mispronunciation. The asymmetric pattern was reflected in reaction time data in a semantic priming task as well as in larger N400 amplitudes. Non-coronal mispronounced non-words with labial or dorsal consonants are accepted as variants of the corresponding coronal word, but not vice versa. Unlike word-final consonants, medial consonants do not undergo any assimilatory variance which would have allowed listeners to have become familiar with these non-words. Further, since word frequency was controlled, full specification or specification of phonetic detail cannot account for these results.

8.2.5 Word-initial asymmetries

Word-initial consonants are often assumed to be sacrosanct. Word-final consonants undergo assimilations or deletions while word-initial consonants do not. Nevertheless, word-initial consonants do undergo change, particularly when it comes to prosodic strengthening, like Notker's *Anlautgesetz* (Lahiri and Kraehen-

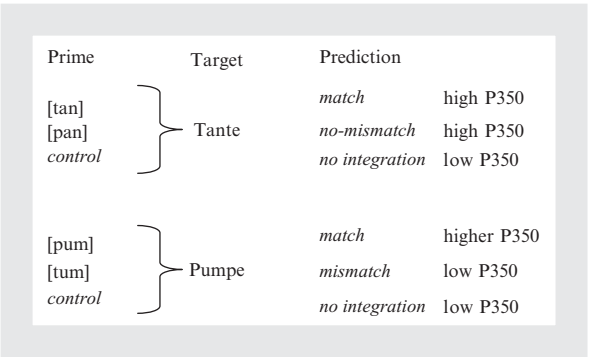

Figure 8.2.5. Predictions for fragment priming in word-initial ERP study.

mann 2004) or the well-known instance of *Raddoppiamento Sintattico* in Italian. Setting aside details, in both instances, word-initial consonants are geminated within certain prosodic phrases depending on the word-final properties of the preceding word. The question here is what happens during normal processing when initial segments are mispronounced. For FUL, since CORONAL underspecification is a must regardless of the position of the word, initial consonants would have the same specifications as final ones, parallel to findings for medial position. Our prediction would be that mispronunciations affecting CORONAL-initial words would be tolerated but not DORSAL or LABIAL-initial words.

Friedrich et al. (2008) tested this claim using a fragment priming task to investigate possible asymmetries (see references therein of earlier studies with manipulated initial consonants). A segmental matching fragment usually gives rise to increased P350 component, argued to show lexical integration (Friedrich 2005). Here, the stimuli consisted of disyllabic German words like *Pumpe* 'pump', *Tante* 'aunt', one set composed of word-initial LABIAL and DORSAL stops and nasals and another set with CORONAL stops and nasals. The task was cross-modal lexical decision. The first syllable of each word and the alternate syllable with the opposing place information (i.e. CORONAL vs. LABIAL/DORSAL) were used as auditory primes: [pum]/[tum] for *Pumpe* and [tan]/[pan] for *Tante*. As before, the expectation was that the initial coronal consonants would be underspecified for place and they would accept variants with [k, g, p, b, m], but not the other way around, as seen in Figure 8.2.5.

The results confirmed the hypothesis. Both *match* conditions showed more negative peak at 350, and our claim was that it supported lexical integration as predicted. The *mismatch* condition was not significantly different from the control and was significantly different from the identity condition. Thus, [tan] and [pan] activated *Tante* equally well as predicted, but although [pum] primed *Pumpe*, [tum] was no different from the control in the activation of *Pumpe*. The ERPs derived

from the control and the non-identity prime were not different. Thus, again the neurolinguistic evidence suggests that word-initial consonants were asymmetrically represented and processed. Similar asymmetric results have been found for isolated vowels in Eulitz and Lahiri (2004), this time using a mismatch-negativity paradigm.

8.2.6 Further implications: The role of frequency

The underspecification that is assumed is especially relevant for three-way contrasts. CORONAL underspecification is universal, as are other contrastive feature specifications. Furthermore, privative features entail underspecification. These assumptions do not exclude the possibility that frequent cliticized elements or idiomatic expressions are stored, nor that frequency and distribution of sound sequences could also have an effect. Nevertheless, the phonological representation of words, based on our feature organization along with the ternary mapping procedure, underlies basic language comprehension. Consequently, homophonous words in two languages with similar feature systems are predicted to show the same activations patterns. English *nine*, *bet*, *gift*, *coal*, *kin* would have the same representations as German *nein, Bett*, *Gift*, *Kohl*, *Kinn* ('no, bed, poison, cabbage, chin'). The representations of the initial consonants would be identical despite the fact that perhaps *coal* ~ *Kohl* have equally high frequencies, but *kin* ~ *Kin* do not.

Our results show asymmetries with initial, medial, and final consonants. Similar asymmetries have been found by many scholars for final consonants, where the goal has been to cope with variation due to assimilation. A recurrent assertion has been that frequencies of occurrence of particular phonemes are responsible for the asymmetries. Although frequency certainly plays a role in day-to-day processing, frequency alone cannot account for the asymmetries particularly if we bear in mind the sound changes that have occurred in time. For instance, voiceless stops which are supposedly rather frequent in the world's languages, *all* became fricatives in word-initial position in Old High German.

Furthermore, word-initial and medial frequencies of individual phonemes are rarely the same. Germanic tends to have more words ending with coronals than other consonants. To clarify this point, we calculated the type and token frequencies of words with initial and final stops and nasals in German. The data, shown in Table 8.2.3, is calculated from the CELEX database. Since German voiced stops undergo devoicing, CELEX calculates word-final voiced stops as voiceless.

These values are informative. Word-finally, CORONALS have a much higher frequency. Word-initially, /n/ and /d/ have lower frequencies than corresponding LABIALS /m, b/ and DORSAL /g/. However, /t/ has a much higher frequency than /p/ but less than /k/. Since both final and initial CORONAL asymmetry was observable in behavioral and EEG results, the frequency explanation alone cannot hold.

Table 8.2.3. Frequencies of consonants (German)

Word-final consonants								
Segment	m	n	ŋ	p	t	k		
Lemma: type count	983	13,584	3,415	288	8,220	1,404		
Lemma: token count	56,309	1,432,443	132,240	21,487	623,954	82,400		
All word forms: item count	22,882	84,175	3,616	5,353	87,125	3,276		
Word-initial consonants								
Segment	m	n	p	t	k	b	d	g
Lemma: type count	2,209	1,409	1,990	2,528	3,039	2,818	1,320	1,859
Lemma: token count	216,718	178,091	71,568	69,461	146,938	217,195	887,571	174,121
All word forms: item count	12,976	9,277	11,287	14,823	19,601	22,685	10,572	15,308

8.2.7 Conclusion

The phonological literature has traditionally privileged the speaker's point of view over the listener's. We have presented evidence for the critical role of the listener in providing evidence for the abstractness of lexical representations. The child learning the language is a better listener than speaker in the beginning. If all variations are to be stored, contexts correctly determined, frequencies established, then language comprehension would become cumbersome right from the start. Our view is that the listener chooses a very simple representation to begin with. The CORONAL feature is there to start out with. Usually these sounds have more high-frequency energy than low-frequency energy (Lahiri et al. 1984; Blumstein and Stevens 1979). These sounds are perceived clearly but all the child needs to know is whether the sound is a consonantal one or a vocalic one. Only when the child needs to make a distinction between [ma] and [na] for instance, then there is a need to add a feature, namely LABIAL. Similarly, unless there is a need to distinguish between [bat] and [pat], there is no necessity for the feature VOICE.

However, *underspecification* is not the complete story—the ternary mapping procedure is at the heart of the model. The entire system depends on the fact that the signal is noisy and what we can extract from the signal may *not* be sufficient to provide a correct match. The system rejects a conflicting sound and rarely finds a good match. It is the *no-mismatch* relationship which allows the system to tolerate variation in a systematic fashion. We have a finite set of words in the mental lexicon and homophones of the same morphological and syntactic category rarely exist. Listeners surely use existing words to help sort out the variation they are faced with every day, as well as any other information available to them, to decode the speaker's intended utterances. In fact, a sparse lexical representation with a mapping process

which allows for tentative decisions helps the listener move toward a final identification of the word. The evidence suggests that the representations are sparse enough to keep words distinct but detailed enough to distinguish them from each other.

8.3 The lexicon: not just elusive, but illusory?

Sarah Hawkins

This section makes the case that a lexicon that comprises a static (or relatively static) list of word- or morpheme-sized units is an analytical convenience, but may have little to do with how speech is actually processed and understood; and that the privileging of one such level skews how we frame theoretical enquiry and is thus empirically limited. This viewpoint encourages us to rethink the nature of a person's linguistic and non-linguistic knowledge, and the relationships between processes in how listeners understand meaning from the spoken signal. Lexical representation is an interesting vehicle for such a discussion, because it is hard for most speakers of Western European languages to regard words as anything other than central to linguistic and even non-linguistic understanding. Yet, as the other contributions to this chapter demonstrate, although we can describe many lexical properties, we do not understand how they function together within a communicative system.

Typically, experimental lexical research assumes the existence of a "mental lexicon" comprising word-like items and then asks how they are identified (word recognition and/or segmentation), or else what properties are represented by the lexicon. The present approach asks what happens if we do not take as axiomatic a mental module that can be uniquely identified as a lexicon. The aim is to raise questions that encourage biologically valid ways of modeling speech and language processing. The roles of context and of individual and shared knowledge in how speech is understood are emphasized; the value of exploring a process-focused account is assessed. Elegant answers are not provided; indeed, much of what is suggested will probably prove to be wrong or distorted. But fear of being wrong should not lessen the value of exploring the viability and implications of a contextually sensitive, process- and function-oriented account of how speech is understood.

8.3.1 Introduction

The lexicon is a concept of linguistic theory which probably applies better in some languages than others, particularly those that are not heavily inflected. Like all

such concepts, its exact definition is problematic, the boundaries between lexical and other types of information are unclear, and these problems are especially obvious when the concept is taken from linguistics and applied to human speech processes.

Adopted by psycholinguistics from theoretical linguistics, the mental lexicon is usually taken to be essentially a dictionary of word-like or morpheme-sized units with phonological form, or patterns, at its core, together with other types of inherent structure that reflect the explanatory scope of the theory adopted (Matthews 1997): the theoretical orientation of the researcher (Rapp and Goldrick 2006) and the structure of particular languages. Theoretical differences engender debate about what these other types of knowledge or representation are, how the lexicon is acquired, and how it is used during speech production and perception.

Thus the term mental lexicon presupposes that word-like units enjoy a privileged status, distinct from other linguistic or meaning units. Setting aside strictly episodic formulations, none of which adequately accounts for how phonetic details are integrated into a linguistic system, it is typically assumed in psycholinguistics that the core phonological representation is citation-form pronunciation, and that lexical knowledge can be treated separately from other knowledge. These assumptions neglect several important issues, including relationships between lexical meaning and contextualized meaning, but they bring other things into sharp focus, such as whether there is a distinction between lexical and pre-lexical representation. (Pre-lexical representations precede and mediate lexical access.) Both lexical and pre-lexical representations are theoretical constructs: there is disagreement about what lexical access entails and hence what is pre-lexical (Frauenfelder and Tyler 1987), and no proof that pre-lexical processes require anything specific to be identified, or comprise a distinct stage of representation.

In sum, defining the lexicon as representations of clearly spoken meaningful units that have reasonably clear referents provides clarity and focus for empirical investigations of how words are produced and recognized, but glosses over many issues central to understanding *interactive talk*. Reassessed in the context of purposeful speech in normal interactive contexts, questions such as the following, which have always been asked, gain new importance. How rich, how tightly bound, and how plastic is information about words and their subcomponents? How are individual lexical items related to meaning and other aspects of language and cognition? How is lexical meaning modified by context (sentence meaning and interactive function)?

In addressing these questions, we need to avoid the circularity that can arise if we unquestioningly impose units of linguistic theory, and especially implications of discrete components, on behavioral data. This point is developed by Pierrehumbert (this chapter) for adults and by Beckman and Edwards (2000a) for children's acquisition of lexical knowledge.

8.3.2 Communicative function and phonetic form

Links between communicative function and phonetic form offer insight into some of these issues. It is now established that detailed phonetic form varies systematically with linguistic category and communicative function (Ernestus et al. 2002; Local 2003, 2007; Kemps et al. 2005; Ogden and Routarinne 2005; Plug 2005; Foulkes and Docherty 2006; Hay and Bresnan 2006). Listeners use such variation to interpret the full meaning of what they hear, and to craft skillful interaction. So, across the full range of normal interaction, form is not fixed, but is affected by communicative process, which in turn is affected by the communicative function of the speech. Skillful speaker-listeners command many functions and hence forms; for people, though functions vary in frequency of use, no single function is more important than many others; hence no single phonological form is privileged.

Function in everyday life usually includes interaction between individuals. Local (2007) shows that words like *so* and *anyway* have distinct acoustic patterns, both inherently and relative to the phonetic properties of the preceding talk, depending on whether their function is to indicate that the speaker will continue the original subject, or is signaling a change in topic or speaker. In such cases, which pepper natural conversational speech, the functional meaning of the word is determined by its pronunciation. Continual adjustment of these processes during interactional talk can affect communication, for example by facilitating continuation, changing the subject, reinforcing a sense of shared values, willfully disrupting the other person's intention, and so on.

In the contexts Local (2007) discusses, words like *so* and *anyway* are spoken as complete intonational phrases, often clearly enough for the word to be identified out of context (though its interactional function might not be understood out of context). But another set of cases that challenges a simple notion of lexicon is utterances pronounced so that the words they represent can only be identified in the complete context in which they are heard. These so-called severely reduced forms carry rich meaning but their phonetic detail, which is often dialect-specific, relates to citation-form phonological structure only by convention. Ernestus's (2002) studies of Dutch show that when common words like *actually* and *naturally* function as discourse markers, they are unintelligible in isolation but highly intelligible—to native speakers—in their original contexts (see also Ernestus, this volume). For English, Local (2003: 327) discusses systematic pronunciation differences that distinguish stronger or more definite versus weaker 'hedging' meanings of the words *I think*. The onset of *think* in confident statements like *I think that's beautiful* typically includes dentality and frication (/θ/), whereas when the same phrase 'hedges' a statement as in *he'll come tomorrow, I think*, dentality and several other features are typically absent (there is no /θ/). Hawkins and Smith (2001) and Hawkins (2003) illustrate how to indicate that one lacks information, with each variant conveying a different type of meaning. The most reduced forms convey rich meaning that

probably governs how the listener responds, yet they are unanalyzable as separate words. The form [ə̃ə̃ə̃] can mean *I don't know* when the person responding is engaged with something other than the question and is not motivated to answer more helpfully. The first speaker must shift to a different conversational gambit if communication is to continue cooperatively. The claim is that the meaning conveyed by this very reduced form is richer than the presumably more common, fluent pronunciation of *I don't know* that is closer to the citation forms of the items. Equally, hyperarticulating *I do not know* with pauses between the words conveys a different rich meaning—normally exasperation.

8.3.3 On representation and processing

These and other observations illustrate two points. First, the complexity of meaning and phonological relations between words entails that mental representations of words are likely to be complex, possibly so complex that they bear little resemblance to what we normally think of as words. Second, whereas we can impose constructs like word and mental lexicon on our data analyses, such constructs do not necessarily reflect any specific representation identifiable in the process of perception or possibly even production, because lexical definitions alone cannot explain the meaning inherent in any meaningful utterance of a particular word. Rather than reflecting biological reality, theoretical constructs like a mental lexicon may be like points on a continuum; directing attention to them allows one to describe certain properties as if they were static, see Rapp and Goldrick (2006). Pierrehumbert (this chapter) and Leach and Samuel (2007) address these points from linguistic and psycholinguistic stances. In short, though lexical representation may seem central to a model of speech processing, we do not know what form lexical representations take, nor what their relationships are, nor how they relate to other aspects of speech processing.

8.3.3.1 *Contextual meaning*

The number and range of meanings a word can have supports the view that there is limited value in equating lexical identification with the identification of phonological pattern in models of natural speech communication. In English and many other languages, many words have more than one meaning. There seem to be no restrictions on how multiple meanings arise: words with multiple meanings may be common or rare, and the meanings related in ways that are obvious (*orange*, *play*) or less obvious (*brace*, *fast*). Metaphor further extends meaning. More pertinently, the exact meaning of many words depends on other words they occur with. One can *play* the violin, the fool, the stock market, with the toy, on words, to the audience, in the park; playwright T. Stoppard can see his own or someone else's play, but

he can only write his own. Finally, the meanings of some words are determined entirely by other words. Consider *quite*. In British English, *quite beautiful* means extremely beautiful, but *quite bad* is not usually extremely bad, and *quite good* means somewhat above average, not extremely good. Some such meanings are dialect-specific and further nuanced by phonetic detail; all are probably learned by association with the words that *quite* modifies, rather than as some property of a generic or abstract word *quite*. Meanings of function words can be even more context-dependent.

Hence, what constitutes the mental representation of a candidate lexical meaning may be situation- and function-specific. Representation of candidate lexical units may likewise be situation-specific. Children learn what the functions are, and apply them as far as their cognitive and linguistic skills allow. Children's one-word utterances have long been recognized as representing different deictic functions: a child who says *car* may mean *look at my car*, *where's my car?*, *listen to the car outside*, etc. Prosody or gestures can indicate the specific meaning. Adults sometimes behave similarly; thus, which lexical items can be used holophrastically, and when, must be part of lexical specification.

8.3.3.2 *Static vs. dynamic representation?*

The above reasoning highlights a problem in trying to draw a boundary between representation and processing. While the distinction may be valid for investigators to make, it may not make sense in terms of brain activity. Moreover, clearer speech and elaborate language are normally needed only when contextual meaning is less clear. So relationships between lexical items, their meaning, and their phonological form can be radically different in different circumstances.

This view of language processing is more abstract than that taken by most psycholinguistic models (e.g. those discussed by Norris and McQueen 2008), yet might provide better insight into the types of neurophysiological behaviors underlying speech-language processing. For example, the vast literature implicating the left inferior frontal gyrus (LIFG) in many behaviors suggests that the LIFG is central to establishing relationships between aspects of sensory stimuli regardless of modality and linguistic status (e.g. using grammar, wrestling, playing tennis, learning to cook, doing mathematics). If this hypothesis is right, then whenever work is required to make sense of complexity or ambiguity, the LIFG will be implicated in what is analyzed variously as phonetic, phonological, morphological, syntactic, and semantic processing. To the brain, however, the activity is one of relating external dynamic sensation to a construction of meaning.

In other words, were models to include a focus on communicative function and how pronunciation varies systematically with function, they might better account for, and thus predict, what a person does in real situations, simultaneously mak-

ing more allowance for differences between, say, genuine talk-in-interaction, and laboratory lexical decision judgments.

8.3.4 Models of processing

What would such a process-focused model look like for speech? The key concepts are memory, prediction, and pattern matching. From these come others, including active combinatorial brain processes whose ephemeral outcomes provide a sense of tangible objectness, and changes in attention that allow task-dependent response plasticity. Short-term variation allows flexible, task-appropriate behavior. Longer-term shifts can be precursors to learning new relationships or categories.

Two lines of evidence are examined: child language acquisition, and how brains process patterns.

8.3.4.1 *Learning words and meanings*

Babies and young children learn words and learn about words, including their phonological structure (e.g. Peters and Menn 1993; Vihman 1996; Waterson 1971). They do this by pattern-matching. Clearer patterns are learned faster. Recent research shows that a clearer pattern is one that is repeated more often (Brent and Siskind 2001; Ota 2006), fits a common pattern, be it prosodic, grammatical, morphological and/or segmental (Jusczyk 1993, 1997; Wauquier-Gravelines 2003; Ota 2006; Rose and Waquier-Gravelines 2007; Gerken and Bollt 2008), and is auditorily clear—e.g. it might be prosodically prominent, or the child might often hear the isolated word (Brent and Siskind 2001; Demuth 2006; Goffman et al. 2007).

Other factors also influence what is learned. Form complexity affects learning rate. For example, two bilingual girls used Hungarian locative expressions months before the Serbo-Croat equivalents. Whereas both languages inflect the noun, Serbo-Croat requires an additional preposition (Mikes 1967 cited in Slobin 1973). Presumably it helps the learning process to have a clear physical or interactional referent (e.g. greetings) that is important to the child. Samoan's rich system for marking affect can involve different words for the same referent. When the affect relates to themselves, Samoan children use affectful grammatical forms (e.g. a pronoun + sympathy, or anger) earlier than phonologically simpler neutral forms (Ochs 1986). Thus early acquisition depends on the child understanding the concept, finding it important, and being able to express it. These examples for production presumably have parallels in perception, where the task can influence what, and how, linguistic units are processed.

In sum, words that children learn early tend to have a clear pattern and a clear deictic or interactional function. This is probably true across the lifespan (Gaskell and Ellis 2009). However, does this constitute proof that single words have a distinct

cognitive status from, say, phrases that have similar deictic functions? In interaction, is *you ought to do it* completely different from *hello*? Perhaps not.

8.3.4.2 *Pattern-matching for meaning*

Hawkins (2010a) explored the hypothesis that speech processing is governed by domain-general perceptual mechanisms which derive meaning from a sensory signal by matching it with memories of similar events or objects in an active, constructive process; specifically, that processes by which expectation, task, and sensory signal are integrated during perception are the same as those that produce illusions.

Part of Hawkins' (2010a) argument is that the brain binds information together to construct auditory objects. Such objects can be thought of as conceptual-linguistic units having detail appropriate for the task, e.g. a phonological feature, a proposition, or something in between. Thus the type of information that the listening brain seeks and attends to depends on values of the parameters x, y, z in the following sentence, which can be thought of as a (usually unconscious) mental orientation to a task: *we are discussing topic x in physical environment y, for reason z; therefore I expect set p sensory patterns, which I can relate to set m meanings, and set w words/phrases, spoken in style s*. Notice the similarity with basic premises in application-specific speech technology.

This view is relatively well accepted for vision and action. Its application to speech and language processing is newer (e.g. Shinn-Cunningham 2008), though rooted in the literature (Halle and Stevens 1962; and phoneme restoration, Warren 1999). The emphasis on interpretation by actively imposing candidate structures on the incoming sensory signal in a matching process seems more convincing when combined with evidence that systematic variation in phonetic detail crucially indicates an utterance's structure and function. It is consistent with models of self-organizing Bayesian processes that rely on old information yet adapt to new information.

Hawkins (2010a) views auditory objects as ephemeral, constructed on the fly during speech processing. (Imagined or remembered forms, lacking current sensory input, normally produce weaker experienced objects.) The construction requires some sort of template, prototype, or group of episodic events (classed as similar) to be held in memory: an object results when current sensation meshes with a memory, as jigsaw pieces fit. Theoretical distinctions between templates, prototypes, and clusters or best-exemplars of episodes are unimportant to this view; they are all abstractions describing properties common to sets of individual experiences, probably bearing only a metaphorical relationship to actual brain processes. What is important to the current argument is that an object results when a "good-enough" match is achieved, that the matching to these abstractions begins when sensation begins, and continues in real time, influenced by feedforward and

feedback. Higher centers influence how new input is interpreted via rich corticofugal feedback, which influences neural responses early in the auditory pathway. Hawkins (2010a) suggests that, for most if not all concepts, the associated templates are embodied, multimodal, and include the limbic system;[6] for speech they are additionally multidimensional and include the cerebellum. Less-distributed neural circuitry might encode familiar and more circumscribed items like faces and names (Bowers 2009).

Principles that seem essential for such networks are: perception grounded in experience, ability to shift attention over time and to different degrees of detail, ability to adjust to current task demands, and matching processes between incoming sensation and memories of past experiences that allow appropriate responses to unexpected patterns. Feedforward and feedback systems will operate over a range of time frames, from a few milliseconds to several seconds and, potentially, several years. These properties should engender self-organizing emergent units comprising richly connected dynamic memory networks, of which words and similar units could be one such set.

Words in such systems will share many parameters. Shared parameters allow rich structure to develop (Moore 2007). A structure that activates a word, then, should include values for specific, broader, and collocational meanings, conceptual categories, grammatical class, phonological form, speaking register, and more. It will hold multiple relationships with other linguistic units (words with similar structural properties and/or associative meanings, and larger and smaller units than itself), and will have non-linguistic associations. Thus a word to be spoken, recognized, or understood may be accessed via many routes.

These claims seem compatible with current understanding of anatomical and functional connectivity in the cerebral cortex, and with the plasticity of memory processes. Neuroimaging studies show that lexical status changes basic phonetic processing rather than affecting so-called post-lexical decisions (Myers and Blumstein 2008) and that phonological patterns akin to templates may register as early as Heschl's gyrus (Jacquemot et al. 2003). Verbs may activate different parts of the motor cortex depending on what part of the body they normally involve (Pulvermüller 1999); yet, crucially, contextual meaning modulates such processing

[6] "Embodied" is a term used in cognitive psychology (and in innovative speech technology and artificial intelligence) that means representation in the brain of an entity (here, a word) by neurons that underpin bodily experience of the physical referent of that word. So, neural circuits for *apple* will include experienced apples' attributes of taste, texture, smell, and looks. Abstract concepts may be built up from basic experiences of simpler things they are related to. For example, *jurisprudence* might contain elements of a child's feelings of well-being when a decision about cake was experienced as fair, and added to later as moral judgments mature. The limbic system is a complex part of the brain concerned with the regulation of emotion and other physical states such as hunger. The basic message is that neural circuits representing even quite simple words may be complex and distributed across wide areas of the brain.

differences (Raposo et al. 2009). Shahin et al. (2009) suggest that auditory objects may be built by feedforward and feedback neural circuits involving, minimally, the temporal, parietal, and frontal lobes (the ventral/dorsal 'what' and 'where' pathways, Rauschecker and Tian 2000; Scott and Johnsrude 2003) and may include illusory completion of sound patterns to match expectations.

To summarize, the proposed neural circuits are complex, self-organizing, richly structured and distributed through many brain regions, in ways that reflect influences of phonology, lexicon, morphology, and so on, but correlate only loosely with each as an independent "module." At times, units corresponding to lexical items will be activated, though via different routes in different circumstances. When a rich superordinate network (e.g. a phrase or complete utterance) is activated in which the lexical item is fully or partially represented, is the lexical item itself independently activated or identifiable? This type of question is unanswerable at present, and may seem perverse. But such questions seem worth considering when the aim is to model how all styles of speech are understood and responded to in natural situations.

8.3.4.3 *Modeling*

The proposed underlying organization of linguistic systems and subsystems has been described in terms of Firthian Prosodic Analysis (FPA, Ogden and Local 1994), which could be computationally modeled as Bayesian functional networks. A start for production was implemented (Ogden et al. 2000), and a framework for perception, Polysp, outlined (Hawkins and Smith 2001; Hawkins 2003). FPA's distinguishing property is its focus on relationships between structures and contrastive subsystems within those structures. FPA is irrevocably context-bound: no linguistic unit is fully describable in isolation from its prosodic and grammatical structural context. Recent formulations also include interactional functions/goals. Such rich structure contrasts with most other models, though see Mattys et al. (2005) and Pierrehumbert (this chapter).

FPA's extreme context sensitivity forces explicit relationships between elements that represent an utterance, which produces extremely accurate descriptions. The cost is complexity; consequently there has been only limited modeling at most analytic levels, including words. Conversely, most models which address lexical items focus on sequential relationships between successive units of one or two types, typically phonemes and words, and on a few influences on those levels, such as frequency and morphological structure. These few parameters produce simple models and broad generalization. Their cost is limited accuracy. For example, Norris and McQueen's (2008) input of diphone misperceptions takes a step towards including detail, but diphones exclude much linguistic structure and knowledge. Ellis and Lambon Ralph's (2000) valuable investigation of frequency and age-of-acquisition effects in written word identification has no superordinate categories

or subcategorization (e.g. abstract vs. concrete nouns which might have different types or numbers of connections), and no enrichment due to differing sensory experiences.

Progress in FPA-based modeling is achieved by describing a subset of language as a slice through situated meaning, semantics, syntax, morphology, phonology and phonetics, plus explicit task goals. Roy's (2005a, 2005b, 2008) perception-action robotics models take a very similar approach. Their abstract control parameters provide the right principles, and are used to model adult-child interaction, but have not yet been applied to normal interactive speech amongst adults. Roy neglects systematic phonetic detail that differentiates non-lexical meaning.

Promising speech models include PRESENCE (Moore 2007), Adaptive Resonance Theory (ART, Grossberg 2003; and see Johnson 2006), DIVA (Guenther 2006), and models based on oscillations and dynamic attractors (reviewed by Engel et al. 2001), and on Hebbian cell assemblies or multilayer perceptrons with reverberation (feedback), e.g. Pulvermüller (2002). However, their focus on phoneme identity or phonological word form limits their approach to clear speech and simple meanings (Hawkins 2003: section 10; Hawkins 2010b).

All these models emphasize plasticity via functional connectivity, emergent units rooted in experience, multiple forms of feedback and feedforward, and prediction. These are generic properties of brain functioning. They contrast starkly with the lean structures that are often claimed to make a good linguistic theory. Though there is much redundancy in language and some models use it (see Pierrehumbert, this chapter), others do not and much interest is focused on what little is needed to distinguish words within a phonologically contrastive system. While theories of linguistic form should inform the types of units a theory of speech behavior accounts for, a narrow set of forms should not dictate the structure that we seek. Function should dictate structure because function defines goals. Formal linguistic contrastive units (phonological, verb phrase, etc.) do indicate functions, but the set is too restricted, yet has been somewhat reified. Complete communicative function should dictate the abstract parameters, and neurobiological functions dictate how they work.

Neurophysiological function is too poorly understood for its application to language to be anything but immensely speculative. Nevertheless, progress demands speculation. Markram's (2006a, 2006b) Blue Brain Project models many neurons, each with unique properties, and combines them into neocortical columns to produce complex responses hoped to mimic intelligent behavior. Markram hypothesizes that dendritic firing in particular brain regions produces three-dimensional electrical objects, each with its particular signature. This type of thinking might help reconcile the paradox between localized and distributed accounts of language processing.

8.3.5 Conclusion

Progress in understanding how people communicate may be facilitated by taking a Wittgenstein-like philosophical position that emphasizes communicative function and the dynamics of spoken communication—with the attention to phonetic detail that this entails. The plasticity and task-oriented use of phonetic detail suggest that lexical specification depends on speech functions and styles. Citation forms may seem stable and fundamental because they have a stable but relatively uninformative context: silence (Hawkins 2004). They therefore demand clear articulation. In other situations, context can reduce the need for clear articulation; here, the intuitive sense of discrete words may be because speech perception follows the principles underlying illusory experiences. When current sensation meshes with past experience (memories), listeners construct subjectively real objects, in this case linguistic ones. However, the physical representation that corresponds with this perception is ephemeral. Word identification emerges from complex networks of multimodal, multidimensional functional connections. These connections, which reflect structured relationships between larger elements and their smaller components, and associations between like elements, change as experience changes.

We do not know what is "purely lexical" in the variety of styles that humans use. Although most speech does not comprise isolated words, presumably phonological-syllabic structure of citation forms should be represented since phonology is a powerful organizing factor and words can be spoken in isolation. But phonology seems unlikely to be most important. Constructed meaning, with grammar, seems prior. These arguments are broadly compatible with the other sections of this chapter, each of which exemplifies some of the properties that specify lexical knowledge.

A common question is whether this account is episodic or abstractionist. It is both; and neither, because to oppose episodic and abstractionist models as mutually incompatible seems misguided. Brains generalize. Therefore many degrees of abstractness are represented. Brains remember details. Therefore details are represented. Generalization requires episodes/details to be grouped or distinguished: such classification is abstraction. Detail, phonetic and otherwise, is needed to construct and maintain the rich knowledge base discussed in this chapter particularly by Pierrehumbert. Detail and word type determine interpretation in many situations (Albright, this chapter). Yet in other situations, a sparse representation may be all that is needed (Lahiri, this chapter)—though the listening brain needs to know which parts of the abstract shape can vary, and in what ways.

The listener's understanding of the task and ambient conditions directs attention to signal properties, thus determining which signal properties are used to understand the message. Thus the challenge is not to choose between the detail and abstraction, but to discover the abstract parameters of the overarching system, and how signal detail maps onto them. The present approach advances the possibility

that the primary parameters are defined in terms of communicative goals and functions; standard linguistic units are secondary.

One implication is that words, or lexical items, can be seen as a possible unit of functional linguistic analysis, without being necessarily fundamental to communication. Lexical identification is no doubt seminal at times, for example in adverse listening conditions, but may not always be essential before longer stretches of speech are understood: higher-order units may be identified before or simultaneously with lower-order units (Hawkins and Smith 2001; Hawkins 2010a, 2010b). By implication, connected speech processing may not necessarily include a discrete stage of lexical access. If this is correct, then psycholinguistic models' so-called prelexical representations could include units longer than words.

These arguments further suggest that a fundamental issue may be whether conceptualization of static linguistic structure should be replaced by function- or process-oriented accounts. Such accounts might not serve some purposes of theoretical linguistics, but might more closely reflect language processing by humans. This conclusion is not restricted to the lexicon, but holds for every level of linguistic analysis: no unit is perfectly definable, boundaries between categories are fuzzy, systems are always in flux. A focus on function, meaning-in-context, and interaction might better elucidate human language processing in general.

8.4 The dynamic lexicon

Janet B. Pierrehumbert

8.4.1 Introduction

The lexicon is the central locus of association between form and meaning. The prior contributions in this chapter focus on the lexicon as it figures in the cognitive systems of individuals. The lexicon can also be viewed at the level of language communities, as shared intellectual property that supports mechanisms of information transmission amongst individuals. This viewpoint is foreshadowed by Hawkins (this chapter), and sketched for linguistic systems in general in Hruschka et al. (2009). Here, I consider the relationship between the lexical systems of individuals and lexical systems at the community level. The dynamics of these systems over time, rooted in their relationship to each other, can inform our understanding of the lexicon, and of the entries and relationships that comprise it. Tackling problems in lexical dynamics, in the light of experimental findings and synchronic statistics,

provides laboratory phonology both with fresh lines of empirical evidence and with fresh arenas for theoretical prediction.

The lexicon is generally assumed to list any associations between form and meaning that are idiosyncratic and must be learned. Thus, it includes not only morphologically simple words, but also irregular or opaque complex words, and collocations. Recently, it has been shown to include morphologically regular words as well (Alegre and Gordon 1999; Baayen et al. 2007). The following discussion emphasizes words (whether morphologically simple or complex), though frequent phrases also appear as a source of new words (Bybee 2001). According to the phonological principle, forms of words (*word forms*) are combinations of basic building blocks, which are characteristic of any individual language, meaningless in themselves, but meaningful in combination. Evidence has recently accumulated that, in addition to this abstract level of characterization, lexical entries also include density distributions over detailed phonetic or socio-indexical properties. I accordingly view word forms as both detailed and abstract (Pierrehumbert 2006a; contributions by Hawkins and Albright, this chapter).

Do people use words? Or, do words use people? At the population level, words inhabit communities of speakers in rather the same way as species inhabit ecological niches (Altmann et al. 2011). Although some words die out, just as some species go extinct, the addition of new words sustains the overall complexity of a lexical system. People both borrow words from other languages, and invent new words by generating names for new concepts (Munat 2007). There is no corpus big enough to include all the words of a language; as a corpus expands to include more topics, more speakers, and longer time periods, new words are always found (Baayen 2002; Manning and Schütze 1999). Even the most stable core vocabulary of Indo-European languages has been replaced at a rate of about twenty words per millenium (Sankoff 1970; Pagel et al. 2007). Berko's wugs paradigm demonstrated the ability of even small children to invent new words through productive use of morphology (Berko 1958), and in adults, this ability is demonstrated both using the same experimental task (Albright and Hayes 2003; Pierrehumbert 2006b), and through the statistics of languages with highly productive morphology (Hankamer 1989). Grammaticalization theory in turn reveals how morphologically complex forms can provide a source of simpler forms on the historical scale (Bybee 2001; Hopper and Traugott 2003).

Like species, lexical innovations compete with pre-existing forms to survive. Words are viable only insofar as they are successfuly replicated. For species, biological reproduction is the mechanism for replication. For words, the mechanism is imitation. Children bring to the task of language acquisition fundamental drives to attend to and imitate speech patterns (Vihman 1996), and to map word forms to word meanings on the basis of phonological and semantical contrast (Clark 1987). Through iterated imitation, linguistic communities converge on shared names for objects and concepts (Steels 1995, 1997) and on shared phonological inventories

(de Boer 2000). This population-dynamic view of the lexicon points to a nexus of cognitive and social factors in determining the long-term dynamics of the lexicon (Komarova and Nowak 2001). I now review some general properties of words and lexicons that are critical for the understanding of this dynamics. I first consider the intrinsic nature of the coding system. Next, I discuss frequency as the reflex of a word's success and as a contributor to lexical dynamics. Finally, I discuss possible mechanisms for new words to overcome the disadvantage of their initially low frequency, and become widespread in the community.

8.4.2 The phonological code

Words are replicated by being learned and then used later. The phonological representations of words supports highly accurate replication—if this were not so, then people would not be able to understand each other as well as they do. But there is also room in the lexicon for new words. These two characteristics can be understood by considering phonological representation as an error-correcting code.

Phonology is a code because it represents the speech stream using sequences of elements from a finite alphabet. A simple illustration of this fact is that blends of two words, such as *celebrademic* from *celebrity* and *academic*, do not average the word forms of the contributing lexemes, but rather sequence components from one lexeme with components from the other, or common to both (Lehrer 2007). In classical linguistic theory, the alphabet was the set of phonemes of the language, defined as minimal units of lexical contrast (Hockett 1961). Though this conceptualization of the phonological code has been updated by autosegmental-metrical theory, the central insight that the code concerns informative contrasts remains. In what follows, I will use the term *segment* as a theoretically neutral term for a phone or phoneme, without commitment to its minimality or abstractness.

From the earliest days of information theory, speech scientists sought to understand the information density of the phonological code, a literature reviewed in Boothroyd and Nittrouer (1988) and Allen (1994). The basic unit of information is the bit, representing a choice or uncertainty between two equally likely alternatives. The smallest number of distinctive features proposed in any phonological theory is 12 (Mielke 2008), which means that English would have an information density of at least 12 bits per segment if all feature values occurred equally and in all combinations. However, mathematical analysis of error patterns for speech perception in noise, with varying amounts of lexical and contextual information, reveals that well-formed English CVC words contain only 10.3 bits of information in total (representing a choice of one word out of 1260 alternatives) or 3.4 bits per segment on the average. Phonotactically well-formed monosyllables (considering words and non-words together) have a greater information density than words, at 4.3 bits per segment. This reflects the existence of accidental gaps in the lexicon,

which provide spaces for new words to be added. Since 4.3 is still far smaller than 12, it also reveals the great redundancy imposed by phonotactics and feature co-occurrence restrictions.

In the theory of information coding and transmission, redundancy is useful for correcting errors. The redundancy in phonological representations reduces the likelihood that a sloppy, erroneous, or poorly heard production will be perceived as an unintended lexical meaning. Word error rates for human speech perception in good listening conditions are neglible, and perception in unfavorable listening conditions is surprisingly robust (Kalikow et al. 1977). Many individual words can be uniquely identified even if one or more segments are missing. This is shown by phoneme restoration experiments, in which people fail to notice that a speech segment has been replaced by noise (Samuel 1981), and by gating experiments, in which people prove able to progressively narrow the set of lexical choices as more and more of the word is provided, often achieving a unique identification before the end of the word (Grosjean 1980). Eye-tracking experiments show that coarticulatory information is used as soon as possible (Dahan, Magnuson, Tanenhaus and Hogan 2001). There is a strong lexical bias in speech perception (Ganong 1980), so that phoneme category boundaries for well-formed non-words (such as *zill* and *woot*) are unconsciously shifted to perceive the most similar real words (*sill* and *wood*). Morphophonological alternations also have a strong tendency to operate within the discrete system of phonological representation (Kiparsky 1985), a behavior that supports error-correcting perception and production for morphologically complex words (and not just simple ones). This functional pressure is so strong that it can cause phonetically conditioned alternations (such as the assimilation of consonants to neighboring vowels) to evolve over time to become more categorical, even at the expense of phonetic naturalness (Anderson 1981).

Redundancy is a reciprocal informational dependency, as discussed in Broe (1993), Steriade (1995), and Frisch et al. (2004). Elements are redundant to the extent that they can be predicted from each other. Predictions can ensue from either positive statistical correlations (known in phonological theory as *harmony* rules or constraints) or negative correlations (known as *OCP* or *Obligatory Contour Principle* constraints). For example, in a language with coronal harmony (such as Chumash), the value of the feature [anterior] for any given strident is largely predictable from any other (Avery and Rice 1989). A strong OCP effect on place of articulation is found in the Arabic verbal roots. The presence of a consonant at some given place in initial position strongly **disfavors** the occurrence of consonants with the same place in second position, **and vice versa**. Frisch and Zawaydeh (2001) demonstrate that such statistics are part of the implicit knowledge of native speakers. Lahiri (this chapter) puts forward some examples of asymmetric informational dependencies relating to the featural make-up of segments. The interest of these examples lies in their contrast with the main thrust of the experiments just reviewed on speech perception in noise, phoneme restoration, gating, eye-tracking, and well-

formedness. Overall, people appear to make optimal use of available statistical information, including the correlations that cause great redundancy in the system. The primary source of informational asymmetry in speech processing is the flow of time in online tasks, which causes some information to be available sooner than other information.

In word phonology, redundancy is found at multiple timescales. At one extreme, consider the avoidance of long words. English has some 43 segment types, whose crossproduct would yield 1849 words with two segments, 79,507 words with three segments, in short 43^n words of length n. But the overall distribution of word lengths is not exponentially increasing. Instead, it is approximately log-normal (Limpert et al. 2001). Relatively few words are extremely short, but past the modal word length of 5 or 6 segments, the likelihood that a given phonological combination exists as a real word becomes vanishing small as length increases. This result can be derived by assuming that a cost function penalizes each additional coding unit (Mitzenmacher 2004). The experiment on wordlikeness judgments by Frisch et al. (2000) establishes the cognitive reality of this basic observation. Feature co-occurrence restrictions within segments provide an example at the shortest time scale. For example, in Indic languages, stops contrast in both breathiness and voicing (2 bits of information), whereas in English these dimensions are conflated (providing only 1 bit taken together). The non-distinction between /r/ and /l/ in Japanese has been particularly well studied. The third formant is the primary cue for this contrast in English. Monolingual Japanese speakers have a poorer neural representation of the third formant than English speakers do, but the neural representation increases if they receive training in the distinction (Zhang Y. et al. 2009). Such results indicate that phonological dimensions (not just phonological categories) are acquired by language learners in a manner that reflects how informative they are in the ambient language.

The nature and interaction of dependencies at different scales provides the motivation for autosegmental-metrical theory as an advance over classic phonemic theory. An autosegmental-metrical constraint amounts to a claim about a statistical dependency at the scale of the constraint. As reviewed in Goldsmith (1990), autosegmental-metrical representations are directed acyclic graphs that encapsulate these dependencies. The leading idea is that dependencies prove to be local if the proper abstract units are defined. Locality is defined in two ways. Metrical units, such as the syllable, the foot, and the prosodic word provide the underpinnings for constraints that involve a head-dependency structure. Tiers provide the underpinning for constraints that pertain to a span without regard to headedness.

The cognitive reality of autosegmental-metrical constraints is demonstrated by a variety of experimental paradigms, including speech segmentation, well-formedness judgments, error patterns, and memory effects. Suomi et al. (1997) show that vowel harmony in Finnish is exploited to segment the speech stream into words. Cutler and Butterfield (1992) show that the typical trochaic stress pattern of

English words is used in the same way. Lee and Goldrick (2008), and Kapatsinski (2009) provide recent best-practice examples of an immense literature on syllable structure. Both bring together multiple strands of evidence to compare the syllable rime and the body (defined as the onset plus nucleus) as cognitively relevant units of prosodic structure.

Accidental gaps in the lexicon are words that do not exist, but are perfectly possible. Autosegmental-metrical theory posits constraints on words in general; these constraints are gradient insofar as the theory is statistically fleshed out. In between the accidental gaps and the general theory lie a set of phenomena that have recently provided critical evidence about the cognitive representations. These are the lexical neighborhood and lexical gang effects.

The lexical neighborhood of a word is the set of words that are minimally different from it (see Frisch, this volume). Though the size of a word's lexical neighborhood is correlated with its overall phonological likelihood, careful experiments have identified dissociations that provide an important argument for a cognitive system with multiple levels of representation, including both an encoding level and a lexical level (Vitevitch and Luce 1998; Luce and Large 2001; Thorn and Frankish 2005). Lexical gangs are sets of words with shared phonological and semantic properties that influence morphological productivity. An example is the set of monosyllabic degree adjectives ending in obstruents that accept the suffix *-en*, such as *black+en*, *white+en*, but not **green+en*, **abstract+en* (Alegre and Gordon 1999). Gang behavior can also be identified for groups of words with shared phonological and semantic components that do not share morphemes in the standard sense, such as *glimmer*, *gleam*, *glint* (Bergen 2004; Tamariz 2008).

Experimental results on lexical gangs and neighborhoods show that subsets of the full lexicon, defined as clusters of words that are particularly similar amongst themselves, have pervasive force. The results support a picture of the lexicon in which words are organized in a network, where the links represent shared phonological and semantic properties (McClelland and Elman 1986; Bybee 2001; Hay and Baayen 2005). The same network is explanatory both for speech processing, and for phonological abstraction and productivity. In processing, activation and inhibition of nodes over time explains perception and production as they unfold in time. Abstractions over groups of words provide the foundation for constraints and for the creation of well-formed new words. Can arbitrary groups of nodes provide the grist for abstraction and generalization? Clearly not. All successful approaches share the insight that the cognitive system forms abstractions from coherent or natural sets of words. A central goal of the network representation is to define the link structure in a way that makes natural groups appear as connected subnetworks of the entire network. Evidence is accumulating that the dimensions of similarity and comparison that define the links are shaped by functional factors at all levels from the perceptual and articulatory periphery to general principles of cognition. For example Lindblom and Maddieson (1988) and Lindblom et al.

(1995) present typological data indicating that the consonant inventories reflect a trade-off of perceptual distinctiveness and articulatory complexity. The results of Albright and Hayes (2003) imply that phonological material temporally adjacent to an affix is more relevant to the productivity of the affix than material in more remote parts of the word. Hudson-Kam and Newport (2009) adduce a cognitive bias towards categorization of frequencies, e.g. interpreting experienced frequencies as more extreme than they really are.

Though these functional factors are reminiscent of innate knowledge in the classic sense of generative phonology, there are also important differences. The differences arise because of the way that functional biases interact with the replication dynamics for the language system. Slight biases can have large effects in structuring the system, because their effects cumulate over time (Reali and Griffiths 2009). Under strong simplifying assumptions, the system is even guaranteed to converge to the prior biases that the learner brings to the learning task (Griffiths and Kalish 2007); but as these authors note, the prior biases may either be innate to the cognitive system, or be rooted in external factors. Under more realistic assumptions, social subgroups can prevent shared norms from emerging (Lu et al. 2009) and oscillations and chaotic variation in the system over time can also arise (Mitchener 2003; Mitchener and Nowak 2004). I return to the challenges raised by these findings in the last section.

8.4.3 Frequency

Statistical learning is central to the picture of lexical dynamics presented thus far. Word types survive to the extent they can replicate themselves through the learner's experience of word tokens (Nowak 2000) and the abstract generalizations that govern lexical productivity are also statistical in nature (Pierrehumbert 2003a, b). Let us therefore consider word frequency more carefully.

Word frequency effects are among the most robust effects known in psycholinguistics. Less-frequent words are recognized more slowly and less reliably than more-frequent words. They are more vulnerable under unfavorable listening conditions (Kalikow et al. 1977). They are also more vulnerable to replacement on historical timescales (Bybee 2001; Lieberman et al. 2007). This last effect arises not only because they are less well learned, but also because they are less likely to be learned at all. A rare word may simply fail to occur by chance in the experience of a learner, and in that case it will not be learned and reproduced for future learners. In the aggregate, statistical sampling considerations mean that the frequencies of individual words are subject to *random walk* effects over generations, and that any word whose frequency happens to become too low will be irretrievably lost. The random walk of frequencies can create morphological gaps (Daland et al. 2007). It entails that the total number of distinct words in the community lexicon

would decrease over time, if new words were not continually added (Fontanari and Perlovsky 2004).

Word frequencies can vary by orders of magnitude across contexts (Altmann et al. 2009), and the context for early word learning—the daily lives of small children—is different from the context for later word learning. Later words are only learned in competition with earlier ones, obeying general principles of contrastiveness in form and meaning (Clark 1987). A new word will be learned only if the powerful error-correcting mechanisms of speech recognition and lexical access do not cause it to be recognized as a pre-existing word. It is initially encoded with the phonological resources that the child commands at that time. Werker and Stager (2000) find that 11 to 12-month olds require multiple points of phonological contrast to successfully map new words onto new referents. A fascinating series of studies by Storkel (2002, 2004) indicates that phonotactics and similarity neighborhoods are dynamically redefined as the lexicon emerges. This dynamics for word learning also predicts individual differences in acceptability of non-words as new words. Frisch et al. (2001) indeed report that individuals with large vocabularies are more accepting of statistically marginal non-words. This might occur because unusual phonological components of the non-words are more likely to already occur in their vocabularies. It might occur because phonological generosity is what permitted them to learn so many words in the first place. These two possibilities can be integrated into a more general and abstract picture, in which a positive feedback loop relating vocabulary size and phonological encoding provides the explanatory dynamics for vocabulary growth; see Munson et al. (this volume) for further discussion.

Frequency effects play a large role in grammaticalization theory, which documents a connection between synchronic statistics on frequency and word length, and typical patterns of historical evolution (Bybee 2001, 2007; Hopper and Traugott 2003). Synchronically, more frequent words tend to be both shorter than less frequent words and less subject to analogical pressure. Diachronically, words and phrases that become more frequent through semantic bleaching (loss of semantic concreteness in connection with usage as grammatical markers) also become shorter. A typical example is the rise of *gonna* as a future from the expression *going to* (Poplack and Tagliamonte 1999; Cacoullos and Walker 2009). Now, frequent words are more expectable than infrequent words. An optimal coding system is obtained if high-frequency words have logarithmically shorter labels than more surprising lower frequency words (Shannon, 1948; van der Helm 2000). Thus, the lexicon is shaped by functional pressures towards uniform information density, a functional pressure that is thought to be relevant for the linguistic system at all levels (Zipf 1949; Goldsmith 2002; Aylett and Turk 2004; Levy and Jaeger 2007; Frank and Jaeger 2008). Shortening words that become frequent is desirable because it helps to optimize the transmission of information. It is possible because frequent words are perceived faster and more reliably even if degraded. It is implemented

through articulatory reduction of word forms that are accessed more easily through their frequency, contextual predictability, and lack of close competitors (Bell et al. 2009).

The loss of internal word boundaries during grammaticalization can further be interpreted within probabilistic models of morphology (reviewed in Hay and Baayen 2005). According to these models, lexical items with meaningful subparts may be accessed either directly as wholes, or indirectly through the subparts. This approach makes nuanced predictions about the decomposibility of words and the productivity of affixes (Hay 2002, 2003). In relation to grammaticalization, the line of prediction is that the complex form will lose word structure as a function of three factors: if its frequency runs ahead of the frequencies of the parts, if the meaning is unpredictable from the parts, and if hypoarticulation induces the loss of phonotactic cues to the boundary. *Gonna* exemplifies this pattern through loss of the motion component of *going*, loss of the velar nasal as cue to a word boundary, and its rise in frequency as it becomes a generic future. Overall, given that a word form rises in frequency, the observed phonological and morphological trajectories documented in grammaticalization theory are predicted.

But what might cause a word's frequency to rise in the first place? Words compete with each other in production, perception, and learning, and the results presented thus far all favor high-frequency competitors over low-frequency competitors. A more frequent form appears more reliably in any finite sample of linguistic experiences used in learning. It is more likely to be learned earlier, interfering with later learning of lower frequency forms. It is more reliably encoded and decoded. The first factor alone already predicts that the lexicon will be simplified over time, and the other factors would only serve to accelerate this trend. To sustain the overall complexity of the lexicon over time, there must be a mechanism for newly invented—and therefore infrequent—words to climb the frequency gradient and come into widespread use.

8.4.4 Heterogeneity

In research on population biology and opinion dynamics, heterogeneity has proved key to understanding innovation and diversity over time. Heterogeneity is the opposite of uniformity. For words, we need to consider both lack of uniformity in the context and lack of uniformity amongst the speakers.

The niche of a word—analogizing to the niche of a species—may be viewed as the thematic and social contexts in which it is used. In population biology, the viability of a species is strongly correlated with the size of its niche (Jablonski 2005; Foote et al. 2008). An analogy can be drawn to the viability of words by considering that a linguistic community explores an abstract conceptual space through its discourse over time, and that a word's viability depends on establishing

a sufficiently large niche (Altmann et al. 2011). Cattuto et al. (2009), analyzing the lexicon of tags on Internet social networking sites, show that the typical growth rate for the number of word types as a function of text length can be derived from a few simple assumptions: Each word has few semantic associates (relative to the total size of the lexicon), and the conceptual exploration by the community takes the form of a random walk. In this picture, global frequency is a chimera and what matters to learning and imitation by individuals is frequency in context. Word types that are very infrequent in general (averaging over time, space, and social context) can be very frequent and predictable in particular contexts (Church and Gale 1995; Altmann et al. 2009), accruing in that context all the advantages of high frequency.

Just as a genetic mutation can create a species with a fitness advantage, a new word can have a fitness advantage deriving from the value and importance of its referent. In studies of opinion dynamics, this type of fitness is called an exogenous factor (in contrast to endogenous factors, which are internal to the system being studied). Studies of recommendation networks for YouTube (Crane and Sornette 2008) and memes (popular phrases) on the Internet (Leskovec et al. 2009) indicate that exogenous factors—such as new inventions, the occurrence of a concert, or the timetable for an election—can cause surges of popularity in the expressions used to discuss them on a scale of weeks or even days. When the value of a product increases with the number of people who have already adopted it, a small minority of users may define a tipping point for universal adoption. Mitchener (2003) develops this line of analysis for language by analyzing the replicator dynamics equations with a fitness function that increases as the number of speakers sharing a given linguistic pattern increases.

Most challenging is the case of endogenous change, in which a new expression gains traction without any real novelty in meaning or functional advantage (as argued for *gonna* in Cacoullos et al. 2009). This case can be analyzed from the point of view of the speakers, as the diffusion of a rare expression through a social network. The links in the network represent social affinity, regions of the network relate to subcommunities of the linguistic community, and adopting a new expression is similar to adopting a new opinion. Mathematical methods similar to those used to study epidemics and catastrophic failures can then be used to explore the likelihood of an *information cascade* (a term introduced in Bikhchandani et al. 1998).

All current models of opinion dynamics that can generate cascades from a small minority of innovators, in the absence of a fitness advantage, depend on heterogeneity in the social network to do so. Baxter et al. (2009) show that a neutral model of social interaction cannot explain convergence to the current New Zealand norm with any realistic choices of parameters. Watts (2002) and Watts and Dodds (2007) generate opinion cascades by positing heterogeneity in the decision thresholds for adopting the new opinion; their *early adopters* can be understood in the present context as people who will use a rare new form because of its association with people that they particularly wish to emulate. Nettle (1999) demonstrated

that linguistic cascading can be obtained by assuming that some highly connected individuals are much more influential than other people. A more sophisticated model by Fagyal et al. (2010) also generates cascading of initially rare innovations by assigning disproportionate importance to input received from speakers who are themselves socially well connected.

Much work remains to be done in this area, because it is far from clear that innovative forms typically originate from or socially close to well-connected high-status people. Indeed, the sociolinguistic literature shows that linguistic change typically originates from lower-status speakers (Labov 1994). However, the models provide clear support for the idea that individual words are associated with indexical information in people's minds. This is necessary because people use words later—sometimes much later—than they last heard them. Preferential adoption of words learned from certain people, or characteristic of certain groups or situations, depends on long-term encoding of these social factors. Indeed, experimental results demonstrate that indexical properties, including speaker identity, are encoded and remembered (Palmeri et al. 1993; Church and Schacter 1994; review in Nygaard 2005). The long-term dynamics of the lexicon provides independent motivation for the conclusions of these studies.

8.4.5 Conclusion

The lexicon is a locus of creativity in language. When invented, novel forms reuse in novel combinations the discrete elements of the system, whether phonological or morphological. To be learned and adopted, novel forms must compete successfully against pre-existing forms in the replicator dynamics, a process of learning and imitation that is generally error-correcting, but also exhibits a systematic bias towards optimal encoding in the relationship of word length to word frequency.

Frequency effects, both in acquisition and in processing, predict the steady attrition of infrequent forms and the steady rise of frequent forms. Research on grammaticalization attests to this trajectory, including the predicted correlation of frequency with shortening. In the absence of additional factors, the lexicon would simplify over time, but the creation of new forms maintains its complexity. A new form must swim against the tide of frequency effects, and it can do so by several mechanisms. It may be intrinsically extremely fit because of exogeneous factors related to its meaning. It may cascade through the population on the strength of social factors. Mathematical models of cascading in related cases of opinion dynamics indicate that cascading of a rare innovative word can occur if the social network is heterogeneous, indexical properties are encoded with words, and these properties play a role in decisions to produce the word.

CHAPTER 9

PHONOLOGICAL ELEMENTS

This chapter discusses three aspects of phonological elements. Mielke reviews recent experimental evidence regarding the role of distinctive features. Zsiga reviews the structure and representation of contrastive tonal elements and implications for implementation. Boersma discusses the ways that phonological category learning can be modeled.

9.1 The nature of distinctive features and the issue of natural classes

Jeff Mielke

9.1.1 Introduction

The sound systems of languages exhibit many properties which linguists recognize as familiar *feature effects*. These include segment inventories that involve contrasts along many of the same phonetic dimensions as inventories of other languages, as well as alternations and distributional restrictions that apply to classes of segments that can be grouped together along some phonetic dimension or combination of dimensions. Feature effects also include the fact that these alternations themselves often involve changes along a small number of phonetic dimensions. These observations have been dealt with in phonological theory by positing that there is a set of distinctive features which form the basis for contrasts, natural classes, and structural changes. Feature theories developed primarily in the second half of the twentieth century sought to identify all of the phonetic dimensions that are relevant for this diverse set of observations, and to place them in a restrictive model which could account for the typology of feature effects.

While distinctive features have been used by phonologists to construct models for all of these observations, it has also been argued that the different observations may have different explanations in phonetics, language change, and general and specific cognitive biases. Laboratory phonology has investigated these mechanisms, which may underlie the observations attributed to features.

9.1.2 Feature theory as explanation

The concept "natural class" has traditionally been defined in terms of both features and patterning, as in (1). This version of the definition is from Mielke (2008: 12–13). See Odden (2005: 156–7) and Hayes (2009: 43) for two recent pedagogical presentations.

(1) Natural class (traditional two-part definition)

a. A group of sounds in an inventory which share one or more distinctive features, to the exclusion of all other sounds in the inventory.
b. A group of sounds in an inventory which may participate in an alternation or static distributional restriction, to the exclusion of all other sounds in the inventory.

This two-part definition is motivated by the observation that sound patterns typically do involve phonetically related classes of sounds, but it is often treated as a truism rather than a hypothesis. The connection in (1) has been accounted for in phonological theory by positing a universal feature set for speech sounds and permitting rules or constraints to apply to classes definable as a conjunction of features (Halle 1959; Chomsky and Halle 1968; Clements 1985; Clements and Hume 1995; Halle 2002). Thus, phonetic similarity (measured in terms of distinctive features) and phonological activity (measured in terms of sound patterns) are treated as two sides of the same coin.

A recent survey (Mielke 2008) shows that recorded sound patterns involve classes which are unnatural featurally as well as phonetically. Out of 6077 phonologically active classes, 4579 (75 percent) are natural according to at least one of the feature systems described in Jakobson et al. (1952), Chomsky and Halle (1968), and Clements and Hume (1995). None of the feature systems draws a clear distinction between classes that are involved in sound patterns and classes that are not, or between classes that are active in multiple languages and classes that are active only sporadically. However, a clear majority of phonologically active classes are phonetically and featurally natural. It is possible to imagine a scenario where features underlie the possible groupings of sounds involved in sound patterns, in which there is a clear distinction between the classes that can be defined with a conjunction of features (which are active in sound patterns) and the classes that cannot be defined in this way. The reality appears to be closer to a more general bias toward phonetically natural classes with nothing ruled out as impossible. Mielke (2008) argues for splitting the traditional definition of natural class into three separate definitions in order to explore whether they are really the same thing:

(2) Phonologically active class (feature theory-independent definition) A group of sounds in an inventory which do at least one of the following, to the exclusion of all other sounds in the inventory:

- undergo a phonological process,
- trigger a phonological process, or
- exemplify a static distributional restriction.

(3) Phonetically natural class
A group of sounds in an inventory which share one or more phonetic properties, to the exclusion of all other sounds in the inventory.

(4) Featurally natural class (feature theory-dependent definition)
A group of sounds in an inventory which share one or more distinctive features within a particular feature theory, to the exclusion of all other sounds in the inventory.

One of the goals of feature theory has been to identify the set of phonetic dimensions which are relevant for phonology. For any non-exhaustive set of phonetically defined features, the set of featurally natural classes is a proper subset of the set of phonetically natural classes. A general observation is that phonetically and featurally natural classes tend to be active in sound patterns (i.e. a connection between (2) and (3)), with many examples of active classes that are not phonetically natural, and many, many examples of phonetically natural classes that are not active in a particular language. Mielke (2008) argues that there is no basis for a stronger connection between featurally natural classes and phonological activity, indicating that the connection is not mediated by any universal feature set. The connection between phonetic naturalness and phonological activity has been attributed to a set of biases. The next sections discuss results from the literature which bear on finding evidence for features, and on how diachronic change and language processing could introduce bias toward phonetically natural classes.

9.1.3 Some experimental results involving features

One of the difficulties with distinctive features as a source of explanation is that phonetically defined features make many of the same predictions as more direct phonetic or historical explanations. Success or failure of feature theory has been determined largely on the basis of accounting for particular synchronic sound patterns, mostly without considering parallel phonetic explanations which are not mediated by an explanatory feature system. A persistent difficulty with finding empirical evidence for distinctive features is that there is considerable overlap between what features are intended to account for and what can be accounted for more or less directly from acoustics and other independently observable information.

Miller and Nicely (1955) showed that the distinctions between different consonants occupy different parts of the speech signal, and can be interfered with through different techniques for signal degradation. This study indicates the distribution of contrasts across phonetic dimensions, but it does not specifically support abstract features. Studdert-Kennedy et al. (1972) show that English-speaking subjects more accurately identify different segments heard in both ears simultaneously if the two segments share phonetic features, regardless of vowel context, which affects the acoustic cues but not abstract features.

Wickelgren (1965, 1966) found that vowel errors were accounted for equally well by conventional articulatory phonetic descriptions and the systematic phonetic level of Chomsky and Halle (1968), but better than the features of Jakobson et al.

(1952) or the phonological level of Chomsky and Halle (1968); while Wickelgren (1966) found that consonant errors were accounted for better by the features used by Miller and Nicely (1955) than by the feature system of Halle (1964), but a feature system based on articulatory descriptions did better than both. Graham and House (1971) found that segments differing by only one SPE feature were overall more likely to be confused by English-speaking girls than other pairs, but otherwise did not predict confusions, and the two most confusable pairs of segments ([f] vs. [θ] and [r] vs. [w]) involve acoustically similar pairs of segments which differ by two and four features, respectively.

Speech errors have been investigated as a source of evidence for features in processing (Fromkin 1973, 1988). Here a recurring confound is that any apparent feature-based error could be reinterpreted as segment substitution (including substitution of a segment differing by only one feature). Shattuck-Hufnagel and Klatt (1979) report that most phonetic speech errors involve manipulating segments rather than features. More recent studies have indicated that subsegmental speech errors are better handled in terms of subfeatural units or gestures (Mowrey and MacKay 1990; Frisch and Wright 2002; Pouplier and Hardcastle 2005; Goldstein et al. 2007).

While some of the experiments described above suggest the presence of particular abstract features, they do not address why particular features are involved. If the features are motivated by the languages of the subjects, a remaining question is why these particular features are involved in these languages in the first place. Some of the experiments discussed further below get at these questions.

Another relatively recent type of investigation into features has been brain-imaging studies which look for evidence of particular features in language processing (see Idsardi and Poeppel, this volume). In a magnetoencephalography (MEG) study, Phillips et al. (2000) report evidence for the feature [voice] in the left-hemisphere auditory cortex of adult English-speaking subjects. Since acoustic similarity is likewise controlled for, an abstract feature can be motivated over acoustic similarity. Other MEG studies report evidence of abstract vowel features (Obleser et al. 2004) and featural underspecification in the mental lexicon (Eulitz and Lahiri 2004). Dehaene-Lambertz and Pena (2001) report electrophysiological evidence that newborns distinguish [pa] and [ta] in a way that they do not distinguish repetition of the same syllable produced by different speakers. Studies with infants have the potential to address more directly the questions about whether these abstract representations are rooted in innate features or in exposure to language data. These studies together do suggest a role for abstract phonological groups, but the connection between these and a specific innate set is weak at best.

See Mielke (2008: ch. 2) for further discussion of these studies. In 9.1.4, we examine some of the essential questions about features that are often addressed; in 9.1.5 we take a closer look at recent results in artificial grammar experiments

9.1.4 What are the features?

It has been observed that certain phonetic dimensions are favored for segmental contrasts and changes, and that phonetically natural classes are favored by sound patterns. With considerable input from phonetics, phonologists have sought to identify the relevant features for defining these classes.

One approach to isolating the relevant features is Quantal Theory (Stevens 1972, 1989), which attributes the phonological oppositions used by languages to the non-linear relationships between articulatory and acoustic parameters (as well as between acoustic and perceptual parameters). Quantal relations provide a natural foundation for binary features with acoustic and articulatory correlates which are common to many languages. This approach has been associated with all-purpose features, used for contrast, changes, and classes, and helps account for the favoritism of particular dimensions in any approach.

A problem with the all-purpose feature approach is that unnatural classes are widely involved in sound patterns, and readily learned, and the preference for certain natural classes and sound patterns involving contextual relevance can often be attributed to phonetically based sound change and more general cognitive biases. Lin and Mielke (2008) identify phonetic dimensions that are easily learned from acoustic and articulatory data, but these are distinct from the dimensions involved in most sound patterns (Mielke, forthcoming), many of which have a basis in particular sound changes. If different feature effects have distinct sources, then it may be unreasonable to expect a single model to account for them. In this case, the approximate success of distinctive feature theory may be an example of model-fitting.

9.1.5 Artificial grammar-learning

The growing literature on artificial grammar-learning experiments seeks evidence of learning biases which may operate independent of or in conjunction with other biases. In the artificial grammar-learning paradigm, subjects (who may be infants or adults) are exposed to data from a hypothetical language. Researchers interpret the relative learnability of different patterns as evidence of cognitive biases that favor the patterns that are learned, or that are learned most easily.

In a recent review of the artificial grammar-learning literature, Peperkamp et al. (2006) suggest three factors which may account for differences in learnability: *phonetic proximity* (changes involving a small number of distinctive features), *contextual relevance* (the target becomes more similar or more different from the trigger), and *markedness reduction* (the result of the alternation is a reduction in markedness). Patterns with these properties may be more readily learned by subjects. Contextual relevance is most closely connected to the issue of features.

9.1.5.1 *Learning natural and arbitrary patterns*

The first group of studies examine the implications for learnability of the difference between well-attested phonetically natural sound patterns and comparable sound patterns that are phonetically arbitrary.

Pycha et al. (2003) exposed adult subjects to words exhibiting a front-back harmony pattern in which two suffix allomorphs contained vowels matching the backness of a context word, a disharmony pattern in which the allomorphs were conditioned by the opposite classes, and a pattern in which suffix allomorphs were conditioned by arbitrary sets of vowels. They found no significant difference between subjects' ability to learn harmony and disharmony, but subjects performed significantly more poorly on the third phonetically arbitrary pattern. Peperkamp et al. (2006) interpret the difference between the learned and unlearned pattern as contextual relevance. In a similar study, Wilson (2003) reports a similar effect, that adult subjects learned assimilatory and dissimilatory patterns more easily than an arbitrary one in which the nasality of a consonant depended on whether or not a root consonant was velar. Both of these studies find that patterns involving a single feature are easier to learn than more complicated patterns. This is another way of viewing contextual relevance: the change and the trigger both involve the same phonetic dimension. However, assimilation and typologically rarer dissimilation are both learned. One reason these could be learned easily is that in both cases the distinction between legal and illegal stimuli (according to the rule exhibited by the experiment stimuli) is whether a particular feature is repeated (assimilation or non-dissimilation) or not (non-assimilation or dissimilation), a distinction that is not available when contextual relevance is not present.

Seidl and Buckley (2005) exposed groups of 9-month-old infants to non-words exhibiting two pairs of phonetically natural and unnatural sound patterns. Both groups of infants (those familiarized to the natural pattern and those familiarized to the unnatural pattern) showed a novelty preference for new words violating the pattern, and there was no difference between the two groups.

Peperkamp and Dupoux (2007) exposed French-speaking adults to two-word phrases accompanied by pictures representing their meaning. Voicing was either phonemic in stops and contextually determined in fricatives, or the other way around. In either case, the allophonically voiced consonants occurred intervocalically and the voiceless counterparts occurred elsewhere. In both cases, the exposure phase did not provide evidence of an alternation among dentals, and subjects did not generalize to dentals when given the opportunity in the testing phase. Subjects were also exposed to phrases exemplifying phonetically arbitrary alternations, which were also learned. In a follow-up study, Peperkamp et al. (2006) found that when the subjects have to name the object rather than just choose which of two pictures matches a phrase, only the phonetically natural allophonic pattern is learned. Still the pattern was not generalized to analogous segments.

In a similar study with similar stimuli but with infant subjects, White et al. (2008) found that 8.5-month-old and 12-month-old infants showed a preference for repetitions of nouns alternating as in the Peperkamp et al. (2006) study as opposed to sequences of phrases with different nouns. When the nouns were played without the conditioning context (vowels, for the voicing alternation), only 12-month-old infants showed a preference for repetition, indicating that only the 12-month-olds were grouping the alternating segments into functional categories.

If all of these results are considered in terms of the factors identified by Peperkamp et al. (2006), the factor that jumps out as being most important is contextual relevance. There is no indication that distinction between assimilation and dissimilation is a significant factor. Both phenomena qualify as having contextual relevance, and they are learned equally well. Since assimilation and dissimilation appear to be equally learnable, the difference in their cross-linguistic frequency may be attributed to the historical development of these patterns as described by Ohala (1983).

Contextual relevance is also related to the phonetic effects from which phonological patterns develop, as coarticulation is often assimilatory and results in changes which are triggered by segments possessing the property that defines the change. This fact about coarticulation does not seem to be related to learnability, so coarticulation and learnability appear to be two independent factors favoring sound patterns with contextual relevance. Phonetic proximity (changes that are analyzed as phonetically simple) also would be an expected consequence of individual sound changes, but so far it has not been shown, on its own, to be a factor in learnability.

9.1.5.2 *Learning natural and unnatural classes*

The previous section looked at studies related to contextual relevance (change and trigger involving the same features) and phonetic proximity (changes involving a small number of features). This section reviews studies that examine the role of the naturalness of sound classes.

Hillenbrand (1983) investigated whether infants can respond to changes between phonetically defined classes of sounds. Following up on studies such as Kuhl (1979) which showed that 6-month-olds detect changes between two segments, Hillenbrand found that 6-month-olds can respond (in a head-turn task) to a change between nasals and voiced oral stops in spite of variation within each category (according to place of articulation and speaker gender), and do not respond to a change between the same stimuli when the nasals and stops are grouped arbitrarily.

Jusczyk et al. (1999) found that 9-month-olds can learn sensitivity to repeated syllables starting with the same consonant or consonant-vowel sequence, but not to syllables ending with the same consonant, or to syllables with the same rime or vowel. In addition to being sensitive to syllables beginning with the same

consonant, the infants were sensitive to syllables beginning with consonants which share manner and voice, but not for initial consonants sharing place of articulation. Jusczyk et al. (1999) observe that this is consistent with the hierarchy proposed by Stevens (1994), in which place features are identified from the speech signal after other features have already been recognized. This is also consistent with the modeling result of Lin and Mielke (2007), that manner features are more easily learned from acoustic data, while place features are more easily learned from articulatory data.

Onishi et al. (2002) show that adult subjects can learn a pattern involving an unnatural class of sounds. Subjects listened to CVC syllables in which the segments [b k m t] were restricted to onset position and [p g n tʃ] were restricted to coda position (or the opposite pattern). In a speeded-repetition task, subjects responded faster to new syllables that obeyed these restrictions. Subjects were also able to learn a more complicated pattern in which the consonant restrictions were reversed according to which vowel was in the syllable, but failed to learn a pattern where the restrictions were reversed according to speaker.

Chambers et al. (2003) conducted a similar experiment with 16.5-month-old infants, with [b k m t f] vs. [p g n tʃ s] involved in the same onset/coda restrictions. In a head-turn preference task, the infants showed a novelty preference for new syllables that violated the phonotactic restrictions.

Subjects in these experiments must have learned independent phonotactic restrictions for each segment rather than a general pattern. This is necessary, since the sets of sounds were deliberately chosen so as not to share any properties other than their phonotactic restrictions. This is analogous to how a learner would acquire a sound pattern involving a phonetically unnatural class.

Whereas Chambers et al. (2003) found that 16.5-month-old infants can learn phonotactic patterns in which the classes [b k m t f] and [p g n tʃ s] each exhibit similar behavior, Saffran and Thiessen (2003) found that younger infants learned generalizations involving natural classes but did not learn generalizations involving unnatural classes. 9-month-olds learned the difference between CV and CVC syllables (showing a familiarity preference), and also for syllables obeying or violating phonotactic restrictions whereby [p t k] or [b d g] was restricted to onset or coda (showing a novelty preference, as the infants tested by Chambers et al. (2003) did for similar phonotactic patterns involving unnatural groupings of sounds). But the 9-month-olds tested by Saffran and Thiessen (2003) did not show a preference with respect to restrictions involving [p d k] or [b t g]. Saffran and Thiessen (2003: 492) speculate that unnatural patterns might require the increased information-processing capability of older children.

Cristià and Seidl (2008) investigated a more subtle version of naturalness. While other studies have compared phonetically robust natural classes with completely unnatural classes, they tested whether 7-month-old infants could learn to identify classes which are phonetically more challenging. Infants who were familiarized with

CVC syllables beginning with nasals and stops, and tested with stops and fricatives, looked longer for the novel fricatives. However, infants familiarized with nasals and fricatives and tested with stops and fricatives showed no preference.

Cristià and Seidl (2008) treat the difference between nasal + stop and nasal + fricative as a difference of naturalness, since this is how they have been treated in feature theory, but as they point out, the "arbitrary" nasal + fricative class is actually involved in more sound patterns. They suggest an alternative interpretation, which is that the acoustic difference between the sibilant fricatives [ʃ z] and the other sounds may make the fricatives harder to group with the nasals. This would be consistent with the finding of Jusczyk et al. (1999) that slightly older infants learned manner classes but not place classes, whose similarity is less acoustically salient.

In summary, the picture is consistent with the suggestion by Saffran and Thiessen (2003) that arbitrary classes are learnable by older infants but not 9-month-olds. The "place" classes used by Jusczyk et al. (1999) are not unnatural from a phonetic or featural standpoint, but they are less clear acoustically. The nasal + fricative class used by Cristià and Seidl (2008) may be unnatural or acoustically difficult. The crucial factor may not be a general notion of naturalness, but rather the acoustic salience of the property shared by the segments grouped together in the stimuli, which is apparently more problematic for younger infants.

9.1.5.3 *Generalization*

One way to tell if language users or experiment subjects have learned a class rather than a set of independent patterns is if they can generalize that pattern to novel segments. Generalization is critical to showing that what has been learned is a class (defined in terms of shared phonetic or grammatical properties), rather than a list of segments. This is the idea behind the "Bach test" (Halle 1978: credited to Lise Menn): If English speakers agree that the plural of "Bach" (which ends with a velar fricative not found in English) ends in [s], they must know a pattern involving the class of voiceless non-sibilant consonants, rather than a pattern involving the arbitrary list of consonants (or words) that take the [-s] plural allomorph.

Generalization also involves an important distinction between phonetically natural and arbitrary classes. Phonetically natural classes can be generalized according to a phonetic property they share, but classes which do not share any properties cannot be. It is possible that a generalization could occur on the basis of a non-phonetic factor, e.g. a phonetically unnatural class that is phonologically active could conceivably form the basis for a generalization. In considering these studies, the important points are whether subjects are able to generalize a pattern they have learned (which indicates the pattern involved a class rather than a list of segments),

and if phonetically or featurally natural classes are generalized more easily than arbitrary classes or each other.

The studies by Peperkamp and colleagues looked specifically for generalization from alternating labial and dorsal pairs to pairs of coronal consonants, but did not find it. Maye (2000), Maye and Gerken (2000), and Maye et al. (2002) found that adults and 6- and 8-month-old infants can learn to discriminate stimuli that are bimodally distributed, whereas subjects exposed to the same stimuli unimodally distributed do not learn to discriminate it (see also Maye, this volume). Maye et al. (2008) found that 8-month-old infants learned a difficult VOT contrast between [t] and [d] on the basis of bimodal distribution, and generalized the contrast to a [k] and [g]. Maye (2000) and Maye and Gerken (2000) found that adults performing a different task do not generalize the contrast, but it is not yet known whether generalization of a distinction learned in this way is unique to infants.

Nielsen (2007) exposed adults to words in which the Voice Onset Time (VOT) of [p] was exaggerated, and found that adults who listen to [p]-initial words with exaggerated VOT exaggerate their VOT in those words, novel words, and [k]-initial words. This suggests that adults can also generalize, but this is not a contrast, and it is not yet known what dimension the subjects are generalizing in (VOT specifically, hyperarticulation generally, or something else).

Wilson (2006) found that adults who are exposed to velar palatalization before [i] do not generalize it to [e] or [a] contexts, but that adults exposed to velar palatalization before [e] do generalize it to [i]. Wilson argues that this asymmetry occurs because [i] is a more natural context for velar palatalization, because [tʃ] and [k] are perceptually more similar there. However, the subjects also generalize to [a], which is a less natural context for velar palatalization.

In summary: the infants tested by Cristià and Seidl (2008) did not generalize when the class was acoustically less robust and did not correspond to a bundle of traditional distinctive features. The adults tested by Maye and Gerken (2000) did not generalize a learned VOT contrast that infants did generalize, but the task was also different. The adults tested by Peperkamp and Dupoux (2007) (also adults, also doing a different task) did not generalize, and the adults tested by Wilson (2006) did not generalize in the condition where the sound pattern (palatalization) was already occurring in the most natural environment, while they did generalize it when it occurred in a less natural environment. The study by Cristià and Seidl (2008) bears most directly on the status of natural classes, because infants generalized only in the case where the resulting class was natural. The asymmetry found by Wilson (2006) also suggests that naturalness plays a role in generalization, but in this case it is the naturalness of the sound pattern that is important, rather than the naturalness of the classes involved.

Mielke (2005, 2008, 2009) suggests that generalization could act as a cognitive bias that results in sound patterns involving phonetically natural classes, e.g. when a sound pattern involves a phonetically natural class that appears to be independent from the phonetic motivation for the sound pattern. It remains to be seen whether a generalization-based account of the phonological activity of some phonetically active classes is necessary, but the above studies have shown that phonetically based generalization is something that humans do.

9.1.6 Conclusion

In contrast to all-purpose feature theory, recent studies of typology and the phonetic and cognitive biases promise a more precise account of why sound patterns are the way they are. Two things should be kept in mind. The first is a clear idea of what aspects of typology or performance are accounted for by a particular factor, and the second is a clear idea of the mechanism by which the factor influences sound patterns.

A good example of the former is the effort by Moreton (2008) to find cognitive biases that account for aspects of typology that are not accounted for by phonetic facts. The development of databases of synchronic patterns and diachronic changes provides an opportunity to identify other recurrent patterns that lack a clear diachronic explanation.

The mechanism of an effect is also important. As has often been noted, if a cognitive bias is to have an effect on the typology of sound patterns, it must be involved in a diachronic change that introduces or eliminates a sound pattern. Possibilities include encouraging or discouraging phonologization of a phonetic effect, and encouraging or discouraging accurate learning of a sound pattern.

Contextual relevance, found to be important by several studies, could play a role in encouraging phonologization, as Moreton suggests. So could the phonetic robustness of the phonetic effect. Hansson (2008) warns that the magnitude of a phonetic effect is not expected to be a direct predictor of phonologization, because listeners typically correct for coarticulation. The potentially complex interactions of these factors need to be considered in figuring out how cognitive biases can influence sound patterns. For the factors to contribute to typology as suggested, they must encourage phonologization, and careful attention must be given to the dynamics of this event. A bias toward phonetically natural classes has the possibility of discouraging the accurate learning of an existing pattern. As Hansson (2008) discusses, this would require that an adult changes the representations of particular lexical items, or that this kind of change (over- and undergeneralization) would only happen for children who are still acquiring these words.

The current study of feature effects in phonology (laboratory and otherwise) is distributed across a wide range of methodologies, frameworks, and assumptions,

but lends itself well to synthesis and promises a satisfying and comprehensive approach to the nature of sound patterns.

9.2 Contrastive tone and its implementation

Elizabeth C. Zsiga

9.2.1 Assumptions and major research questions

Tone is defined as a lexically contrastive pitch pattern (Pike 1948; Yip 2002; Gussenhoven 2004). Not all languages use pitch to create lexical contrasts, but the majority do (Fromkin 1978; Yip 2002). Tone contrasts may consist of different pitch levels (from two to as many as five), or of pitch movements of varying direction, slope, and shape: see Figures 9.2.1 and 9.2.2 for two examples. While pitch changes constantly, and for many reasons, across an utterance, most laboratory phonology approaches to tone assume that there are categorical elements that underlie the constantly changing f0 trace and its articulatory and perceptual correlates. Thus the laboratory phonology approach to tone can be defined as the use of experimental techniques to study the acoustic, articulatory, and perceptual correlates of tone, in order to learn about the underlying categories, and the relation between the underlying representation and its phonetic manifestations, including the coordination between tonal and non-tonal elements. Important questions that arise in the study of contrastive tone include:

1. What is a possible tonal contrast?
2. How should tonal contrasts be represented?
3. How are the contrastive features realized in the articulatory, acoustic, and perceptual domains?
4. How are tones aligned with segments and/or larger prosodic constituents?
5. How do tonal systems arise and change?
6. How are tonal systems acquired or learned?

Section 9.2.2 summarizes some of the answers to these questions that have been proposed in the literature. In each case the issues, major proposals, and phonological evidence will be briefly described. Section 9.2.3 then summarizes and exemplifies different laboratory approaches that have been taken to studying these questions. While the study of tonal alternations is important in addressing these questions, tone rules will not be a focus of this section: see Yip (2002, 2007), Gussenhoven (2004), Chen (2000).

9.2.2 Issues and proposals in the phonological literature

The question "What is a possible tonal contrast?" may be rephrased as "What are the necessary and sufficient universal tone features?" (Fromkin 1978: 1; see also other contributions in that volume). Research on tone features has focused on establishing which aspects of the pitch pattern are significant–what speakers are paying attention to and systematically manipulating.

Currently, most phonologists agree that the representation that best accounts for cross-linguistic patterns of tonal contrast and alternation consists of H (high) and L (low) autosegments associated to a tone-bearing unit (TBU) such as the mora or syllable (see e.g. Yip 1989, 1995, 2002; Duanmu 1994; Zhang 2002; Gussenhoven 2004; following Leben 1973, 1978; Gandour 1974; Goldsmith 1976; Anderson 1978). Languages where the tonal contrasts consist of distinct (relative) pitch levels (*register tone languages* in the terminology of Pike 1948), contrast in the presence of H vs. L (or H vs. L vs. Ø) for every tone-bearing unit (TBU), while *contour tone languages* allow multiple associations for each TBU, so that a rising pitch pattern, for example, is represented by LH. More complex systems may require additional intermediate class nodes: Yip (1995), for example, argues for the addition of a register node [+/− upper] in order to account for systems with more than three pitch levels. (See also Odden 1995; Yip 2002, 2007; Gussenhoven 2004 for further discussion of tone feature geometry.) The decomposition of contours accounts for cases such as *tonal melodies* that spread over the required number of syllables, and tonal alternations which treat the parts of a contour tone separately (Goldsmith 1976). In addition, an autosegmental representation allows for a consistent formalism to be used for register tone, contour tone, pitch accent, and intonation, with varying systems differing only in the sparseness of the tonal representation (McCawley 1978; Pierrehumbert 1980).

A drawback of the autosegmental approach, however, is that the correspondence between autosegments and the parameters of perception and production is not always straightforward. Acoustically, the complex shapes of contour tones do not necessarily consist of an obvious sequence of H followed by L. In the perceptual domain, a number of studies such as those by Gandour (1978), have shown that listeners judge similarities between tones based on shape rather than endpoints. Thus, some researchers (e.g. Sapir 1921; Pike 1948; Gandour 1978; Abramson 1978; Clark 1990; Xu 1998, 2004; Roengpitya 2007; Barrie 2007) have argued that a representation of tone based on movement rather than high or low points is more phonetically accurate and psychologically plausible. These argue, following Pike, that some languages encode pitch differences in terms of levels (static targets), others in terms of movements (dynamic targets) and that "for a dynamic target, the movement itself is the goal." (Xu 2004: 13).

Both the autosegmental and non-compositional approaches to tone features take acoustic or perceptual targets, either movements or endpoints, as basic. Another

recent approach (Gao 2006) has suggested using articulatory gestures as the units of tonal contrast. A strength of this approach is that it incorporates tone into the theory and practice of Articulatory Phonology (Browman and Goldstein 1992), which had previously been implemented primarily for segmental phenomena. The articulatory approach to tone has had success in modeling some complex patterns with simple underlying gestures. A weakness is that it is still largely untested, and much further work will be required to test it against a range of cross-linguistic patterns.

Each approach to tone features must also address the question of the coordination of the tonal melody with other speech events. One approach to this problem focuses on the identity of the TBU: are there universal constraints governing the association of tones to prosodic units, or will the units and principles of association vary from language to language? (See the discussions in Clements 1984, 1986; Odden 1989, 1995; Clark 1990; Pulleyblank 1994; Duanmu 1994; Yip 1995, 2002.) Another approach focuses on the alignment of tonal specifications to segmental or syllable-level landmarks. Studies of alignment from the dynamic targets perspective include Xu (1998, 1999) and Roengpitya (2007): these studies find evidence for treating tone slopes as indivisible entities. From the autosegmental perspective, studies including Myers (1996), Morén and Zsiga (2006), and Zsiga and Nitisaroj (2007) argue that H and L targets align independently. Other studies of the interaction of tone and prosodic structure have examined the relationship between vowel length and the distribution of tonal contours (Ohala and Ewan 1972; Blicher et al. 1990; Zhang 2002; Yu 2006), the mutual effects of tonal and intonational specifications (Downing 1989; Inkelas and Leben 1990; Myers 1996), and the attraction of pitch peaks to prominent TBUs (Bickmore 1995; Yip 2001; de Lacy 2002b).

A further question related to tone features is whether "tonal" contrasts sometimes involve dimensions other than pitch, particularly voice quality. In a number of languages, voice quality conveys lexical contrast in a manner similar to tone, or tone and voice quality vary together: for example, a high-toned syllable may always be realized with breathy voice, or a low-toned syllable with creaky voice. Such "mixed systems" are common in Southeast Asia (e.g. Vietnamese: Brunelle 2009) and in the Americas (e.g. Yucatec Maya: Gussenhoven 2004). To some degree, the treatment of mixed systems is a matter of definition: should the definition of "tone" be revised to include laryngeal contrasts other than pitch? One solution is to adopt a feature system that encompasses all laryngeal contrasts within a single system, such as [+/− stiff] or [+/− slack] vocal folds as proposed by Halle and Stevens (1971). To the extent, however, that both voice quality and tone are controlled systematically and independently, as is the case in many languages, a cross-classifying set of features is needed (see Yip 1992; Andruski and Ratliff 2000; Brunelle 2005; Keating and Esposito 2007). The interaction of tone and voice remains an active area of research. Acoustic and articulatory studies document the co-occurrence of voice

and tone parameters, while perceptual studies address the question of whether one or the other aspect has precedence as a perceptual cue.

Finally, questions of how tonal systems arise and change must be addressed, both for the system as a whole and for the individual. Segmental and prosodic effects are particularly important in addressing *tonogenesis* (the process by which pitch becomes lexically contrastive in a language) and diachronic change (Hyman 1978). Numerous researchers (e.g. Connell 2002; Kingston 2005; Svantesson and House 2006; Abramson et al. 2007) have supported the hypothesis of Hombert et al. (1979) that tonogenesis comes about when pitch differences that are the unintended result of a particular laryngeal configuration are reinterpreted as intended. Concerning individual change, studies of how an individual acquires a tonal contrast in a first language (L1) or learns a tonal contrast in a second language (L2) have lagged behind studies that address segmental acquisition and learning. Crucial questions in both L1 and L2 thus often focus on how the acquisition of tone may be the same as or different from the acquisition of segmental contrasts (e.g. Li and Thompson 1978; Demuth 1993, 1995a, 2003; Tsukada et al. 2004; Hao and de Jong 2007). The question of how systems of tone and intonation interact in adult learners has also recently become an important area of interest (e.g. Wayland 1997; Wayland and Guion 2004; Francis et al. 2008; Nguyen and Macken 2008).

There is no one-to-one relationship between the questions listed above and different laboratory approaches. Multiple questions might be addressed, and approaches used, in a single study. The goal of the next section is to briefly review representative laboratory phonology studies of contrastive tone, organized by type of data examined.

9.2.3 Laboratory approaches

9.2.3.1 *F0 measurement*

The most common laboratory approach to studies of tone is acoustic measurement of f0 patterns, using pitch-tracking algorithms such as autocorrelation. The most basic example of this type of study is documentation of f0 patterns and contrast in citation form or in an invariant frame. Long lists of descriptive work on languages in every part of the world could be cited, beginning with Bradley (1911); two recent representative examples are shown in Figures 9.2.1 and 9.2.2. Figure 9.2.1 (Picanço 2005) documents three contrastive tones in Munduruku citation forms. Figure 9.2.2 (Nitisaroj 2006) documents five contrastive tones in Thai in sentence-initial position. Note that the data in Figure 9.2.1 presents actual pitch traces of multiple repetitions by a single speaker, while the data in Figure 9.2.2 averages over multiple productions by different speakers, normalized in both pitch range (by transformation to z-score) and duration (% of syllable duration). Both types of presentation are common.

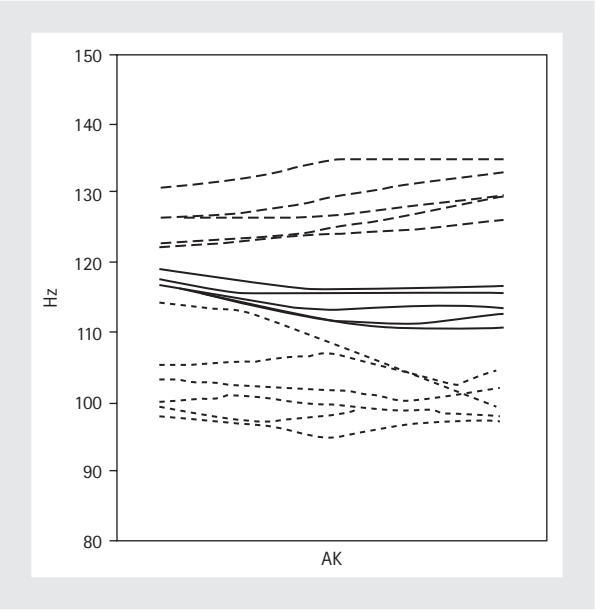

Figure 9.2.1. Three contrastive tones in Munduruku (Picanço 2005: 46, with permission).

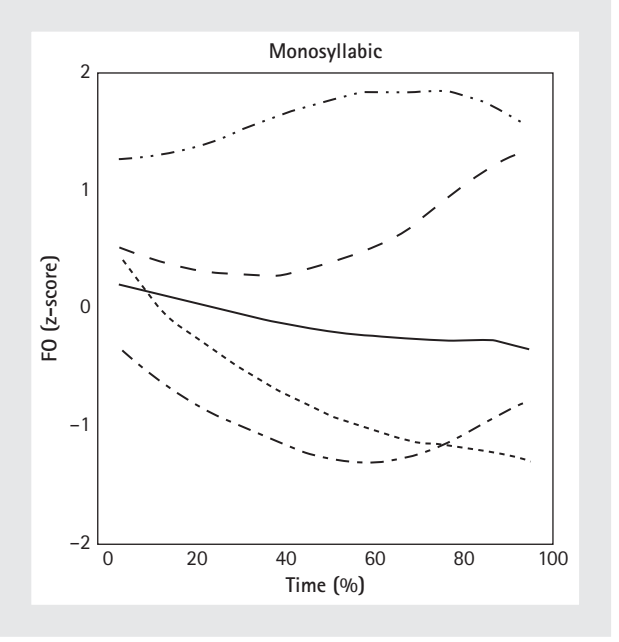

Figure 9.2.2. Five contrastive tones in Thai (Nitisaroj 2006, with permission).

While studies using citation forms or invariant frames are an important first step, they do not necessarily provide the data needed to uncover the underlying tonal features. A next step is to vary the context: position in the utterance, surrounding tones, or discourse context (such as focus, or statement vs. question). The basic assumption of contextual studies is that aspects of the tonal shape that remain constant reveal underlying features, while changes illuminate the causes of variation (see also Chen, this volume on this topic).

Contextual studies may address the influence of tones upon one another (Shen 1990; Gandour et al. 1994; Xu 1997; Potisuk et al. 1997; Agwuele 2007; Daly and Hyman 2007); downstep and downdrift in sequences of tones (Hyman 1979; Connell and Ladd 1990; Snider 1998); phrasal and prosodic influences including the interaction of tone and prominence (Yip 2001; Gussenhoven 2004; Morén and Zsiga 2006; Roengpitya 2007), and the interaction of tone and intonation (Downing 1989; Inkeless and Leben 1990; Myers 1996; Herman 1996; Kallayanamit 2004; Yuan 2004; Hyman 2008). Data from such studies may determine whether contextual changes are more categorical, characteristic of what many would call phonological, or whether they are better characterized as gradient, within-category variation. Data from studies of contextual variation are often used to argue for different models of featural representation. Daly and Hyman (2007), for example, argue that the mid tone in Peñoles Mixtec must be phonologically unspecified, based on varying contextual realizations. Gussenhoven (2004) documents a tonal contrast in Yucatec Maya that is realized with glottalization in phrase-final position, but as a falling contour in phrase-medial position, and argues for a phonological association between glottalization and high tone. Inkelas and Leben (1990) demonstrate a number of phrasal and intonational effects on tone realization in Yoruba, and use the data to argue for the necessity of a register node in phonological representation. Figure 9.2.3 illustrates the difference between high and low register in Yoruba tone in a statement and yes-no question.

Another type of variation is speech rate. Xu (1998), for example, compares the realization of Mandarin tonal contours on syllables of different lengths pronounced at different speech rates, and evaluates changes in the shape and alignment of the contours. He concludes that rising and falling contours move as a unit, rather than peaks and valleys aligning independently, and from this draws support for the hypothesis that contours are integral dynamic units rather than being composed of sequential H and L. Roengpitya (2007) reaches a similar conclusion for Thai based on contour realizations over syllables of different lengths. On the other hand, Nitisaroj (2006) finds that H and L points in the contours of Thai align independently under changes in speech rate. Myers (1996) finds a difference only in peak alignment in Chichewa.

Acoustic analysis is also used to study the interaction of tones with vowels and consonants. Early acoustic studies (Peterson and Barney 1952; Lehiste and Peterson

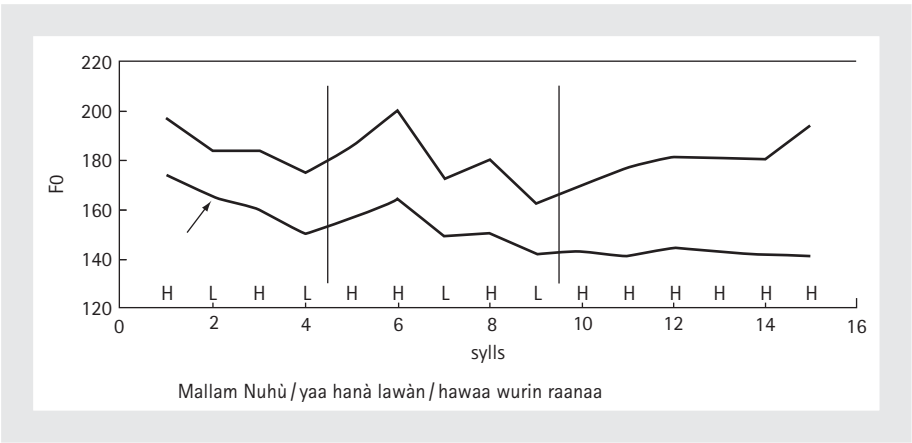

Figure 9.2.3. A statement and yes–no question in Yoruba (Inkelas and Leben 1990: 18, with permission).

1961; Lehiste 1970) established that high vowels, probably due to the interconnectedness of tongue muscles and the structures of the larynx, have slightly higher intrinsic pitch than non-high vowels (see also Ohala and Eukel 1987 for discussion). Whalen and Levitt (1995) and Connell (2002) confirm an intrinsic pitch effect in tone languages, and find that the effect of vowel quality on f0 is greater for high tones than for low tones. Other early studies, such as Hombert et al. (1979), conclude that voiced consonants lower f0 and voiceless consonants raise f0. These studies provided acoustic evidence for voicing distinctions in consonants as a source of tonogenesis, a hypothesis previously based on written sources alone (e.g. Haudricourt 1954 on Vietnamese). Teeranan (2007) and Hyslop (2009) provide recent examples of such tonogenesis in progress in, respectively, a dialect of Malay and the Tibeto-Burman language Kurtop. Kingston (2005) focuses on glottalization in Athabaskan languages, finding that in some cases glottalization is associated with raised pitch due to increased vocal fold tension, while in other cases it is associated with lowered pitch due to irregular vocal fold vibration. Picanço (2005) uses acoustic analysis of vowels to determine that f0 is a more reliable correlate of lexical contrast in Munduruku than is voice quality. Other acoustic studies of mixed systems (e.g. Svantesson and House 2006 and Abramson et al. 2007 on Khmu, and Brunelle 2009 on Cham) make diachronic applications: the data show these systems evolving from a voice quality contrast to a mixed system to a pure tone system. Dialectal differences are often evident, with different dialects at different stages. Studies of mixed systems often include both acoustic and perceptual components: acoustic studies to document what voice qualities and tones occur together, and then perceptual studies to determine how varying the combinations changes listeners' judgments.

9.2.3.2 *Perception studies*

The simplest form of tone perception study is lexical identification: native speakers of a language listen to tokens of natural speech and name the word they hear. A lexical identification task can be used to check that the linguist's understanding of the system is correct—listeners can indeed distinguish the tones that the linguist believes are contrastive—and can serve as a baseline for further studies. Studies based primarily on natural-speech lexical identification include Roux (1995), Peng (1997), Connell (2000), Andruski (2006), and Khouw and Ciocca (2007). Svantesson and House (2006) and Brunelle (2009) introduce dialectal variation, and mechanisms of diachronic change. Svantesson and House find that some dialects of Kammu use f0 for lexical contrast and some do not, hypothesizing that tonogenesis is underway in this language. Brunelle concludes that Northern speakers of Vietnamese use voice quality distinctions that Southern speakers have lost.

To further probe the cues that are necessary and sufficient for a particular contrast, researchers often digitally alter speech tokens for perception studies: resynthesizing pitch contours (Vance 1977; Abramson 1978; Gårding et al. 1986; Lin and Repp 1989; Repp and Lin 1990; Zsiga and Nitisaroj 2007; Abramson et al. 2007); filtering to remove f0 information (Liu and Samuel 2004); truncating syllables (Lee 2001); or combining pitch and other dimensions such as voice quality or duration in different ways (Blicher et al. 1990; Yu 2004; Brunelle 2009). By independently varying the parameters that occur together in natural speech, or requiring listeners to respond to unnatural contours that contain hypothesized cues, perceptual studies with digitally altered speech can tease apart the effects of different cues that are inseparable in natural speech. For example, Brunelle (2009) shows that perceptual judgments do not correspond to generally accepted tone features for Vietnamese, and argues in favor of a new system.

Abramson (1978) tested the degree of slope that was necessary for Thai speakers to identify a tone as "rising." On the other hand, Zsiga and Nitisaroj (2007) tested various synthetic contours, and conclude that peak alignment, not slope, is the main perceptual cue to tonal distinctions in Thai. Figure 9.2.4 (Zsiga and Nitisaroj 2007: 377) shows that lexical identifications switched from "falling" (filled diamonds) to "high" (asterisks) as the pitch peak was moved later, with the crossover point occurring about three quarters of the way through the syllable (220 ms). Peaks in the first half of the syllable caused ambiguity and confusion, consistent both with the hypotheses that peaks are aligned to the *right* edges of moras, and with the findings of House (1990) that, cross-linguistically, tones are better perceived later in the syllable, after the spectral changes associated with syllable onsets have subsided.

Another type of perception study involves asking listeners for similarity judgments rather than lexical identifications, using a same-different (AX) task, or similarity (AXB) task. An advantage of such studies is that listeners do not have to be native speakers of the language(s) under study, or trained in making

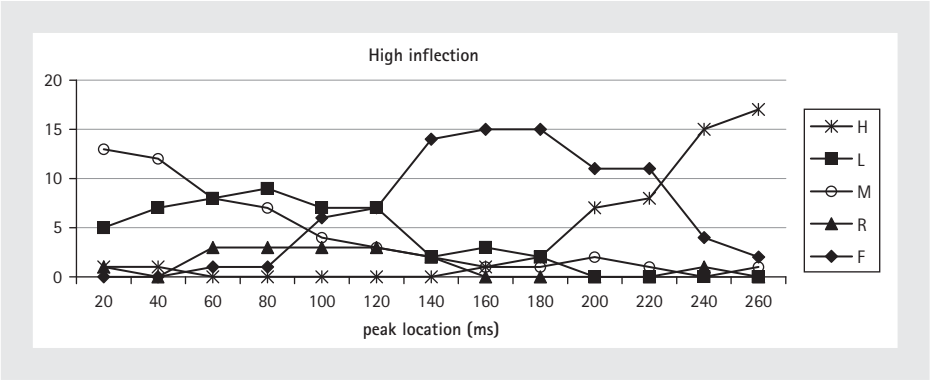

Figure 9.2.4. Tone identifications as a function of peak alignment on digitally-altered f0 contours in Thai (Zsiga and Nitisaroj 2007, with permission).

categorizations. A drawback is that it is not clear whether listeners are using the same strategies for similarity judgments as for lexical judgments (see discussion in Zsiga and Nitisaroj 2007). Gandour and colleagues (Gandour 1978, 1981, 1983; Gandour and Harshman 1978) have used this technique, along with the statistical analysis of multidimensional scaling (see Clopper, this volume), to test whether speakers of different languages use the same dimensions to group pitch contours in the perceptual space. Gandour argues that listeners with different language backgrounds use the same five dimensions, but weight their importance differently: speakers of tone languages give more weight to pitch slope than speakers of intonational languages do, for example. He thus concludes that perceptual parameters including direction and slope of pitch change must be included as part of the universal feature set. In a recent version of similarity judgment tasks, researchers use neuroimaging studies, relying on known brain responses to within-category and across-category stimuli, to measure directly whether two sounds are perceived as the same or different (Gandour et al. 2000; Li et al. 2008).

A final perceptual issue is how tone normalization works. It is clear that tonal contrasts are relative: a "high" tone does not refer to an actual pitch level, but to a tone realized in a certain part of the speaker's range. Studies investigating how listeners normalize for pitch differences between speakers (Leather 1983; Moore and Jongman 1997) present syllables with identical f0 patterns in different contexts, or vary pitch with other segment-internal information, to determine which changes influence listener decisions.

9.2.3.3 *Articulatory studies*

The earliest form of articulatory study was autopsy. Ohala (1978: 10) credits Vesalius (1543) for providing detailed descriptions of laryngeal anatomy based on autopsy.

Ohala also cites experiments conducted by Müller (1851) "done using freshly excised human larynges, sometimes with most of the rest of the vocal tract attached," which demonstrated that pitch could be changed both by altering vocal fold tension and by regulating subglottal air pressure. Figure 9.2.5 illustrates the laboratory set-up.

Modern articulatory investigations of tonal contrast use airflow measures, laryngoscopy, electroglottography (EGG), and electromyography (EMG) (see Hanson, this volume). Studies of airflow focus on the effect that oral constrictions have on transglottal airflow and thus (potentially) on pitch. Guion and Wayland (2004), for example, use airflow data to argue that the aerodynamic requirements of an apical trill condition a falling f0 contour, with implications for tonogenesis. Edmonson and Esling (2006) use laryngoscopy to investigate laryngeal mechanisms in the interaction of tone, register, and stress. EGG uses electrodes placed on the skin of the throat to measure the impedance of a low-level current passed across the larynx. Because impedance is proportional to glottal opening, EGG directly but non-invasively measures both f0, a correlate of pitch, and open quotient, which correlates with breathy and creaky voice quality. Brunelle (2005, 2008) uses EGG,

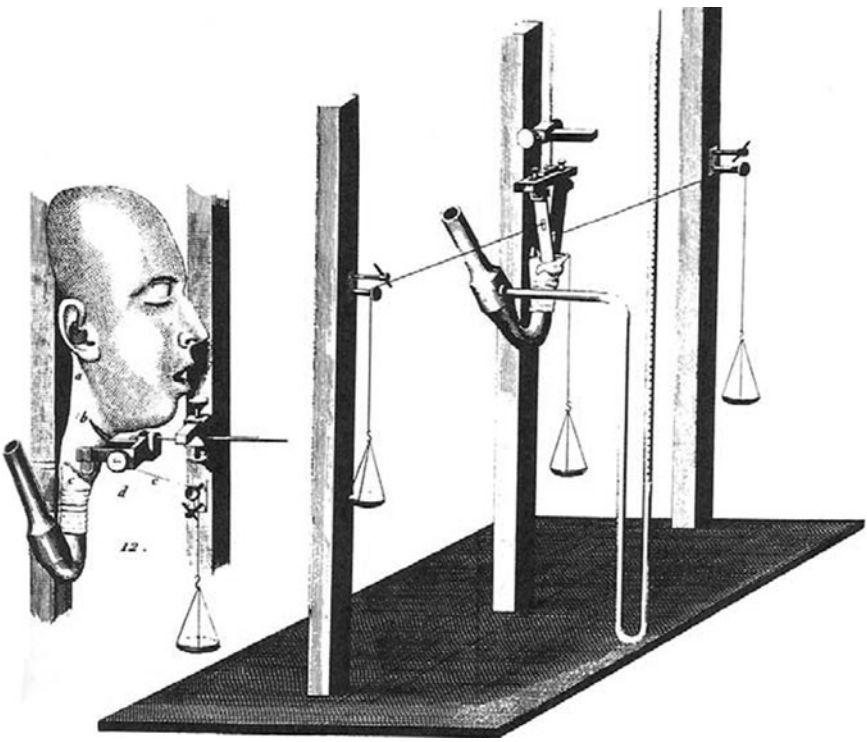

Figure 9.2.5. An early laboratory phonology experiment: Müller (1851), reprinted in Ohala (1978, reprinted with permission. © Elsevier).

among other techniques, to investigate the interaction of voice quality and tone in Cham, and argues for the separation of laryngeal and tonal features.

EMG is a difficult methodology to use for laryngeal studies, because the muscles of the larynx are small, intertwined, and not easily accessible, and because the insertion of the hooked-wire electrodes is not pain-free. The technique is therefore most often used in the study of voice disorders (e.g. Heman-Ackah and Barr 2006). Nonetheless, evidence from EMG studies can be useful in establishing whether there is an active articulatory gesture (and thus phonological target) associated with a particular pitch pattern. EMG studies in the 1960s and 1970s established that the cricothyroid muscle is the primary muscle responsible for pitch-raising, whereas the sternohyoid and sternothyroid are most active in pitch-lowering (Ohala 1978). The technique can be used to investigate whether similar pitch patterns in different languages are brought about by similar articulatory actions, and can thus be useful in defining cross-linguistic features and natural classes (Erickson 1976, 1994; Halle 1994).

9.2.3.4 *Modeling*

Recently, computer modeling has begun to be used in the study of tonal contrast. Computer models and "analysis by synthesis" test whether the right shapes can be derived from the proposed primitives and a given understanding of contextual influences. Languages for which computational models of tonal implementation have been proposed include Mandarin (Shih and Kochanski 2000; Yuan et al. 2002; Yuan 2004; Xu 2004), Thai (Mixdorff et al. 2002; Roengpitya 2007), Vietnamese (Mixdorff et al. 2003), and Yoruba (Agwuele 2007). Fujisaki et al. (2007) apply their model to Thai, Vietnamese, Mandarin, Cantonese, and other Chinese languages. Gao (2006) synthesizes trajectories for Mandarin based on gestural scores. Other models (e.g. Cao et al. 2004; Zhang and Hirose 2004) are implemented in speech recognition systems for tone languages.

9.2.3.5 *Studies with special populations*

A final set of experimental approaches to consider is studies with special populations. These studies allow researchers to examine tone systems in development, decline, and conflict in order to provide new evidence for cognitive representations. Studies of L1 acquisition in children (Tse 1978; Demuth 1993; So and Dodd 1995; Tsukada et al. 2004) use perception and production techniques to address the chronology of tone acquisition, differences between the acquisition of tone and segments, and cross-language differences and similarities. Parallel questions can be asked concerning adult L2 learners. In addition, studies of L2 learners can explore ways in which the L1 and L2 systems interfere with one another, and what sorts of interventions may be most useful. L2 studies of tone may focus either on speakers

learning a new tone language (Wang et al. 1999; Wayland and Guion 2004; Hao and de Jong 2007; Francis et al. 2008; Wayland and Li 2008) or on speakers of tone languages learning a non-tone language (Wayland et al. 2006; Nguyen and Macken 2008). Finally, studies of atypical populations such as patients who have suffered aphasia or stroke (Gandour et al. 1996, 1997; Gandour 1998; Becker and Reinvang 2007) can assess the ways production and processing of tone may change in the damaged system.

9.2.4 Conclusion: Consensus and directions for further research

None of the questions raised in Section 9.2.1 have been fully answered. Research continues on the question of defining tonal systems. Undocumented languages remain to be described. The interaction of voice quality and tone, involving the synchronic and diachronic study of mixed systems, is an especially active research area. Regarding tonal features, the current consensus among phonologists is that tonal representations are autosegmental, but much work remains to be done in determining how these autosegments are produced and perceived, and a number of researchers who pay close attention to perception and to phonetic implemention, especially of contour tones, remain unconvinced. The question of how tones are aligned to other speech events remains an active area of research from all theoretical perspectives. Research on change, acquisition, and learning of tonal systems has in the past lagged behind segmental studies, but is currently increasing. All of the laboratory techniques described in this contribution—acoustic, perceptual, articulatory, and computational—will continue to contribute to our increased understanding of contrastive tone.

9.3 Modeling phonological category learning

Paul Boersma

9.3.1 What is category learning?

The term "phonological categories" refers to the discrete elements that make up a phonological representation, i.e. elements of its temporal organization (e.g. the foot, syllable, mora, segment, or autosegment) and elements of its internal content

(e.g. phonemes such as /p/ and /n/ or feature values such as [+nasal] and the high tone H). From the title of this contribution, the reader can already infer that the present author assumes that at least some of these phonological categories can be learned. This assumption is opposite to the assumption held by most generative phonologists, which is that all phonological categories are innately given to the human infant. Thus, Chomsky and Halle (1968: 4) state that "phonetic features" belong to the "substantive universals", which are a subgroup of linguistic universals "available to the child . . . as an a priori, innate endowment." Likewise, Prince and Smolensky (1993 [2004: 2–3]) state that Universal Grammar, i.e. the innate language endowment, "consists largely of a set of constraints on representational well-formedness"; in their examples, such innate constraints often refer to substantive phonological elements, which therefore have to be innate *a fortiori*.

As has been pointed out for syntax and semantics by Braine (1992), Slobin (1997: 289–96), and Tomasello (2003: 183–5), the generative assumption of innate categories comes with a learnability problem, namely the *linking problem*. In the phonology and phonetics domain this means that even if all phonological categories were innately given, the language-acquiring child would still have to connect at least some of these innate categories to auditory events available in the incoming speech data, and this is a problem because the mappings between some phonological categories and auditory events vary cross-linguistically and cannot therefore be innate. After all, the hypothesis of innate categories presupposes the universal existence of, for example, the phoneme category /u/ (or of the feature values [+back], [+high], and [+round]); but since a phonological element representable as /u/ (or as the corresponding feature bundle) is typically pronounced slightly differently in every language, the mapping between this phonological category and auditory events must be language-specific and cannot therefore be innately given.

The reasoning in the previous paragraph may not convince nativist phonologists. After all, one could object that an innate category /u/ could correspond to a *region* of auditory events, like a cloud of F1–F2 pairs, and that different languages select different parts of this cloud. This objection fails if one realizes that the perceptual *boundary* between vowels like /u/ and /o/ is also language-specific, so that two different languages should be able to both have the innate categories /u/ and /o/, but there will be some sounds that are perceived as /u/ by listeners of one language and as /o/ by listeners of the other language. This abundantly established fact (for direct cross-linguistic comparisons, see Savela 2009 for vowels or Hamann et al. 2010 for the [f]–[v]–[ʋ] continuum) proves that at least some categories cannot be innately connected to specific sounds (or that the categories themselves are different feature bundles in all these languages, so that the innate feature set must be very large, as Hale et al. 2007: 662 ultimately have to admit).

It might still be possible to hold the innatist viewpoint here and devise a learning algorithm that starts in some default sound-to-category-connection state and subsequently shifts the category boundaries on the basis of incoming speech

data, analogously to ideas known from the syntax-semantics interface such as Grimshaw's (1981) innate "canonical structural realizations" or Pinker's (1984, 1989) "bootstrapping" with innate linking rules. To my knowledge, however, no such algorithm has been explicitly proposed in the phonological-phonetic literature. For this reason, I will in the remainder of this contribution assume the *emergentist* viewpoint of category learning, which holds that the language-acquiring child begins without any innate phonological categories and subsequently *creates* her categories on the basis of incoming speech data.

9.3.2 Where do categories emerge?

Assuming, then, that phonological categories emerge in the language-acquiring child, the question is in what location (representation in the brain) these categories emerge. *A priori*, it would be good to have phonological categories in the phonological lexicon, which is the location where humans typically have to store enough sound information to make thousands of morphemes pronounceable and perceivable: Categories are discrete internal representations of raw continuous data in the outer world, and can thus provide a helpful reduction of the required data storage. For this reason, most psycholinguists and phonologists agree that the phonological lexicon consists of discrete categories (for the opposite standpoint of exemplar theory, which is less concerned about lexical economy, see below in Section 9.3.7, and Chapter 8 this volume).

But is the lexicon the *only* location where these categories exist? There are three sources of evidence that phonological categories exist even outside the lexicon. The first source of evidence comes from psycholinguistic experiments; psycholinguists with quite diverging convictions on the details of phonological comprehension (McClelland and Elman 1986; Samuel 1996; Norris et al. 2000) can agree that the speech comprehension process goes through a *pre-lexical representation* that consists of the same kinds of phonemes (and other phonological elements) as the lexical representation; the basic idea is that human beings in the lab can readily identify phonemes in tasks that do not involve access to the lexicon, such as those involving short non-word syllables. The second kind of evidence comes from phonological theory, where it is widely agreed (e.g. Prince and Smolensky 1993) that the speech production process goes through a *phonological surface structure* consisting of discrete phonological elements such as feet, syllables, segments, and features; the basic idea is that especially the larger metrical structures (feet, syllables) cannot be specified in the lexicon (the *underlying form*), because the domain of their assignment is often the phrase rather than the word (i.e. these structures tend to span across word boundaries). The third kind of evidence comes from infant studies, which find that the perception of children of 8 to 10 months of age is already adapting to the phonological categories of their ambient language

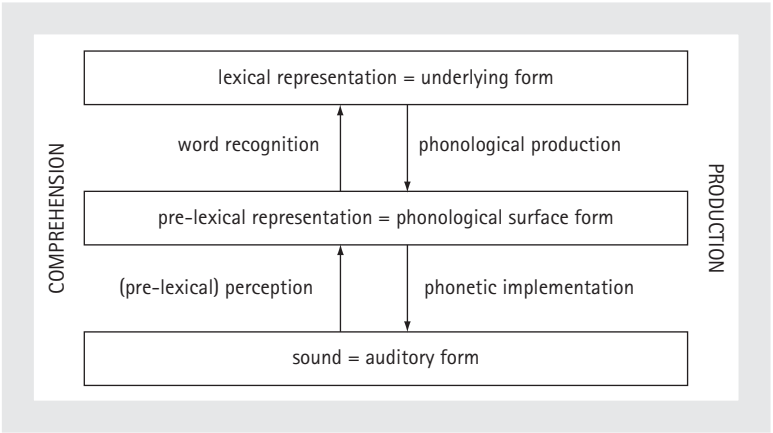

Figure 9.3.1. The simplest model of speech comprehension and production compatible with the evidence from psycholinguistic experiments, phonological theory, and infant studies. Categories emerge in the intermediate level, as a result of the acquisition of perception.

environment (Kuhl et al. 1991; Polka and Werker 1994; Jusczyk 1997); the basic idea is that although these infants have no words in their lexicons yet, they have already increased their ability to distinguish between sounds that belong to different phonological categories and have decreased ability to distinguish between sounds that belong to the same phonological category.

When we combine these three sources of evidence, and assume that humans of any age have the same levels of representation, the simplest hypothesis must be that categories emerge in the intermediate level (the pre-lexical representation or phonological surface structure), and that this happens in the infant's comprehension process, and more specifically in the infant's acquisition of her (pre-lexical) perception. This is shown in Figure 9.3.1.

9.3.3 What do categories emerge from?

In Figure 9.3.1, the phonetic correlate of a phonological category is auditory. Although this is in line with the acquisitional evidence discussed in 9.3.2 (infants perceive contrasts before they articulate them), the possibility that the phonetic correlate of a phonological category is instead articulatory, especially in production, cannot be ruled out. In linguistics, the auditory view is shared by Saussure (1916) with his *image acoustique* and by Jakobson et al. (1963) and, from a nativist camp, Anderson and Ewen (1987). Since the phonetic implementation process must also somehow feed into articulation, this view has to entail that articulation happens in

the service of audition. Thus, Harris and Lindsey (1995) argue for the primacy of audition from biteblock experiments (Lindblom et al. 1979), in which speakers maintain auditory forms by modifying their articulations (for a slightly different view see Folkins and Zimmermann 1981). Likewise, in an explicit model of phonological-phonetic production, Boersma (1998) places the articulatory form below the auditory form in Figure 9.3.1, arguing that for implementing the phoneme /s/ the auditory correlate of loud high-frequency noise is primary whereas the articulatory correlate of alveolar constriction is secondary. The idea is that in order to articulate a legitimate /s/ you also have to make sure that your lungs contract, your glottis is wide, your velum is up, and your lips are open, with everything in service of producing auditory loud high-frequency noise.

Many authors (Chomsky and Halle 1968; Clements 1985; Browman and Goldstein 1989; Keyser and Stevens 1994; Hale et al. 2007), and therefore probably many readers of the present contribution, do not share the auditory view of phonetic implementation: they assume instead that the phonetic correlate of phonological categories is articulatory in nature. In Saussure (1916: 98), for instance, Bally and Sechehaye found it necessary to include a footnote explaining Saussure's standpoint against the articulatory bias of those days, and Ramus et al. (2010) mention Boersma's model but deviate from it (without argument) by positing in their boxes-and-arrows model the articulatory rather than the auditory form as the direct output of phonetic implementation. Ramus et al. do not provide an explicit, let alone computational, account of how production or comprehension could proceed; I predict that attempts to devise an explicit account of the production of /s/ would fail in the case of their model. I stress here that boxes-and-arrows graphs can be verified or falsified only by explicit, preferably computational modeling, something that very few psycholinguistic accounts presented at laboratory phonology conferences provide. By contrast, phonological accounts by linguists do tend to be fully explicit (e.g. with ordered rule sets or with ranked constraint sets), and therefore have the desirable level of explicitness. Of course, I do agree with Ramus et al.'s point that linguists should address not just what Figure 9.3.1 calls phonological production (as e.g. Prince and Smolensky 1993 do), but also word recognition (as Smolensky 1996 does), pre-lexical perception (modeled explicitly by Boersma 1998 *et seq.*; Pater 2004; Berent et al. 2009), and phonetic implementation (Boersma 2007, 2009; Boersma and Hamann 2009).

The traditional bias in favor of articulatory correlates in production has been extended to comprehension. The hypothesis of *direct realism* (Fowler 1986; Best 1995), for instance, maintains that listeners directly perceive the speaker's articulatory gestures, and the *motor theory of speech perception* (Liberman and Mattingly 1985) holds that listeners access their phonological forms only after activating their *own* articulatory gestures. In these two models, then, even the left side of Figure 9.3.1 would have to be extended with an articulatory level (either the speaker's or the listener's) between the auditory and surface forms. While such extensions are

imaginable, most explicit models of category creation only consider the lower two levels of Figure 9.3.1, and it is those models that I discuss here.

Another issue relevant to the relationship between Figure 9.3.1 and category creation is whether the arrows on the left and on the right represent separate modules or not. According to Ramus et al. (2010), for instance, the arrows "word recognition" and "phonological production" must be separate, because in foreign-language perception, Japanese listeners insert vowels but in their phonology they do not (Polivanov 1931; Dupoux et al. 1999; Jacquemot et al. 2003). Two things are wrong with this reasoning. First, in Smolensky's (1996) explicit (namely, Optimality-Theoretic) bidirectional model, where word recognition and phonological production employ the same ranked relations, insertion in comprehension corresponds to deletion in production (again, we see a dramatic example of why the common practice of translating the results of psycholinguistic experiments to boxes-and-arrows plots must fail without an explicit model of what the boxes and arrows mean). Second, the psycholinguistic evidence shows that Japanese perceptual vowel insertion takes place in the module of "pre-lexical perception," i.e. at a different *level* (not a different *direction*) than phonological production (for an explicit Optimality-Theoretic account of such cases, see Boersma 2009 for Japanese and Boersma and Hamann 2009 for Korean); on the right side of Figure 9.3.1, this perceptual capability of inserting vowels corresponds to the capability of Japanese speakers to delete vowels in "phonetic implementation," which is an uncontroversial aspect of Japanese pronunciation (Akamatsu 1997). There thus does not seem to be any strong evidence against the bidirectionality proposed by Smolensky (1996) for the top two arrows and by Boersma (2007) for the bottom two arrows in Figure 9.3.1; if this bidirectionality is true, categories created on the basis of correct pre-lexical perception can be employed immediately in phonetic implementation, with correct auditory targets (i.e. potentially hampered only by articulatory effort).

The last issue with Figure 9.3.1 is whether or not the two arrows at the left or right represent sequential modules. Interactive (top-down) influences of the lexicon on phonological categorization in comprehension would make at least the "word recognition" arrow bidirectional (for an explicit model see McClelland and Elman 1986), and interactive (bottom-up) influences of phonetic considerations such as articulatory effort and the quality of auditory cues on phonological production would make at least the "phonetic implementation" arrow bidirectional (for an explicit model see Boersma 2007, 2008).

9.3.4 How do categories emerge?

If categories emerge in the phonological surface form (the intermediate level in Figure 9.3.1), then one or more other representations must play a role in this

process. The simplest computer simulations of phonological category creation (e.g. Guenther and Gjaja 1996; Boersma et al. 2003) indeed assume that the discrete phonological categories emerge in the surface form from continuous auditory representations such as formants, pitch, duration, noise, silence, and their combinations and sequences (which are in the lowest level in Figure 9.3.1). Artificial-language-learning studies have shown that this modeled procedure is realistic (Maye and Gerken 2000; Maye et al. 2002). We can conclude that bottom-up processing in speech comprehension plays a major role in category creation.

What also might play a role in category creation are all the representations *above* the phonological surface form, not only the underlying form in Figure 9.3.1, but perhaps also the syntactic and semantic representations, which must be located even further up. Whether the lexicon plays an active role in determining a perceived category in online *comprehension* is a matter of vigorous debate (e.g. Norris et al. 2000; Samuel 1996), but it is more widely accepted that the lexicon (perhaps via higher-level representations) can act afterwards as a correcting supervisor telling the listener what she should have perceived, because this kind of top-down processing in perceptual *learning* has been observed in the laboratory (Eisner 2006; Eisner and McQueen 2006). Many explicit models of perceptual learning, e.g. the TRACE model by McClelland and Elman (1986) and an Optimality-Theoretic model by Boersma (1998), therefore include such a supervising mechanism. However, such supervision can only occur once the categories exist, and it is possible that top-down processing plays no role whatsoever in the *creation* of categories.

9.3.5 Requirements for a model of category emergence

Despite the fact that 8-month-olds can profit little from higher representations when creating their first phonological categories, the ultimate comprehensive model of category creation will probably have to be embedded in a larger model that can handle not only the creation of phonological categories and the acquisition of the connections of those categories to auditory cues, but also the acquisition of their connections to higher representations. Such a larger model therefore should not just do category learning but also exhibit many "effects" known from the literature on psycholinguistics, phonological theory, and infant studies, such as perhaps the Ganong effect (Ganong 1980), the McGurk effect (McGurk and MacDonald 1976), the prototype effect in best-token experiments (Johnson, Flemming, and Wright 1993), the perceptual magnet effect (Kuhl et al. 1991), the relation between phonological activity and frequency (what phonologists call "markedness"), auditory dispersion (Liljencrants and Lindblom 1972), licensing by cue (Steriade 2001), and so on. After all, all these phenomena appear in the same language-processing

brain, and we should not have to create a separate model for every observed phenomenon. Hence, all these phenomena should ultimately be viewed in relation to each other.

If the ultimate larger model is as emergentist as the category creation model must be, this causes a problem for the hypotheses of direct realism and motor theory discussed in Section 9.3.2, because the fact that infants can categorize before they can speak may require those models to assume an innate connection between sound and articulation. In the following, I therefore assume the simpler model of Figure 9.3.1, and also assume that all parts of it are emergent.

9.3.6 Existing models of emergence (but not of categories)

Some comprehensive emergentist models exist already. The neurobiologically inspired TRACE model (McClelland and Elman 1986) considers the three levels of Figure 9.3.1 and derives several effects, including the Ganong effect. The present author's linguistically inspired Optimality-Theoretic model of bidirectional parallel multilevel constraint competition (for an overview, see Boersma forthcoming) brings together the seven effects mentioned in Section 9.3.5 under one umbrella: the Ganong effect results from parallel multilevel evaluation; the McGurk effect from Optimality-Theoretic interactions between auditory and visual inputs; the prototype effect and auditory dispersion from the idea that constraint rankings optimized for perception are reused in production; and markedness effects and licensing by cue from a bidirectional multilevel learning algorithm. It has to be remarked here that the model does *not* handle category creation, nor its developmental precursor, the perceptual magnet effect.

9.3.7 Existing models of category creation (but not of phonology)

Several existing models can handle category creation, although these have rarely been applied to the learning of phonological categories, let alone been embedded within a larger model of language processing. Adaptive Resonance Theory (Grossberg 1976, 1980, 1987; Carpenter and Grossberg 1987) proposes that a new category is created at a certain level of representation (e.g. the phonological surface form in Figure 9.3.1) as soon as the brain detects a mismatch between bottom-up information to that level (e.g. from the auditory form in Figure 9.3.1) and top-down expectations (e.g. from the lexical representation in Figure 9.3.1). It would be interesting to see how these complicated models perform within a large linguistic model.

Connectionist models also hold a promise of providing mechanisms for category creation. McClelland and Rumelhart (1986) show that if categories are not represented as unitary symbols, but as distributed representations in a neural network, categoryhood must be a gradient concept, so that categories can be created in a gradual manner. Connecting these ideas to the representations of phonology would be an interesting enterprise for the future. A connectionist model that does address phonological issues (Soderstrom et al. 2006) unfortunately works with innate constraints (specified in the genome), and therefore, *a fortiori*, with innate categories (because the constraints refer to phonological categories such as codas); this model therefore cannot handle category emergence.

A separate strand of research involves the modeling of the *perceptual magnet effect* (Kuhl et al. 1991), in which listeners discriminate two sounds more easily if they belong to different phonological categories than if they belong to the same phonological category; it is as if the auditory properties of two sounds within the same category are nearer to each other than one would expect on the basis of their acoustic distance. Guenther and Gjaja (1996) show with computer simulations that such *perceptual warping* can emerge as the result of the formation of an auditory map in a neural network model. The inputs to the network are auditory values encoded directly as neural activities. For instance, there is one pair of neurons whose activities reflect the second formant (for the first neuron, low activity means low F2, high activity means high F2; for the second neuron, low activity means high F2, high activity means low F2), one pair of neurons whose activity encodes F3, and so on. The model also has a "neural map" consisting of, say, 500 neurons, all of which inhibit each other and all of which are connected to each of the four input neurons. The model is then fed auditory events (F2–F3 pairs) drawn from language-specific distributions. Thus, an English language environment is simulated as a Gaussian distribution centered around an F2 of 1000 mel and an F3 of 2075 mel, reflecting the phoneme /l/, plus a Gaussian distribution centered around an F2 of 1000 mel and an F3 of 1200 mel, reflecting the phoneme /ɹ/. As auditory events come in, a standard learning rule that tries to increase the correlation between pre-synaptic activity and connection weight for every active cell (a continuous variant of Hebbian learning) causes most cells in the map to become "tuned" to the most frequent combinations of formant values. After learning, a combination of input formants F2–F3 will then generally lead to a different *perceived* combination of formants F2′–F3′, if the latter is defined as an average over the "best" tuning frequencies of all active neurons in the map (weighted by their activities); the learning rule will have made sure that the perceived F2′–F3′ tends to be close to a frequent combination of input formants, even if F2–F3 are not. This is illustrated in Figure 9.3.2.

Whereas Guenther and Gjaja used unrealistically low values for the standard deviations of F2 and F3 (40 and 60 mels, respectively), so that there was essentially no overlap between the formant clouds for /l/ and /ɹ/, Figure 9.3.2 was produced

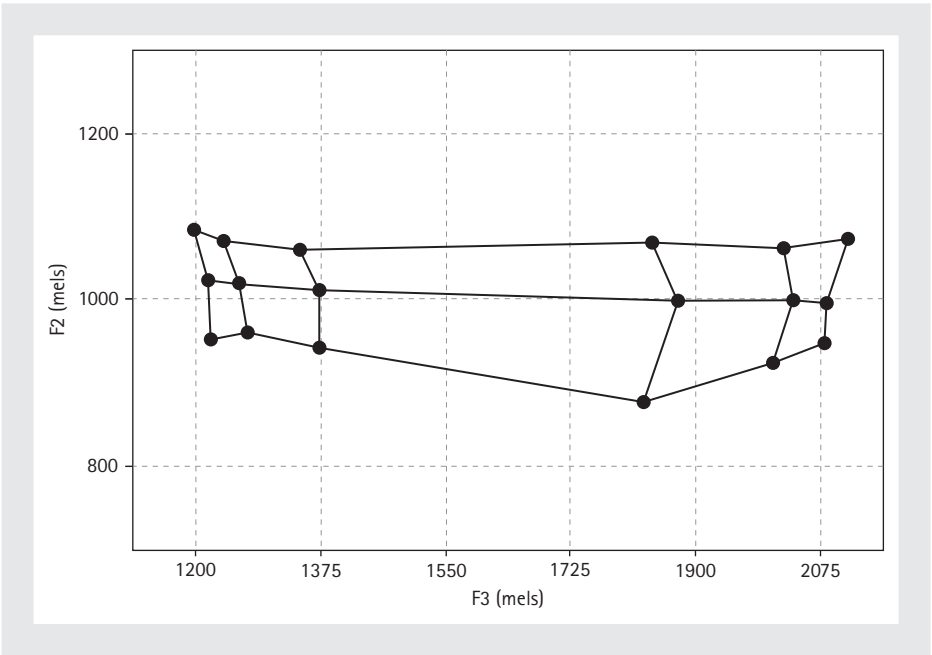

Figure 9.3.2. Perceptual warping, based on Guenther and Gjaja's (1996) model after 100 pieces of English-distributed input data for liquids. The input formant combinations are the 18 crossings of the dotted lines. The "perceived" formant combinations are the 18 dots.

with realistic standard deviations (100 and 200 mels, respectively), which required raising the size of a map cell's "neighborhood" from 35 to 150 cells (Wanrooij 2009). We can see that for equidistant input formant combinations (the crossings of the dotted lines) the perceived formant combinations (the dots) are no longer equidistant but instead cluster around the centers of the English distributions (F3 = 1200 and 2075 mels; F2 = 1000 mels). If the distance between any two dots in the figure is a measure of how well the two sounds can be discriminated, the perceptual magnet effect is explained; for instance, the perceived distance between an input F3 of 1375 and an input F3 of 1550 Hz is reduced to approximately 100 Hz (the distance between the second and third columns of dots in Figure 9.3.2), which presumably makes for poor discrimination ("acquired similarity" in terms of Liberman 1957), whereas the perceived distance between an input F3 of 1550 Hz and an input F3 of 1725 Hz is raised to approximately 500 Hz (the distance between the third and fourth columns of dots), which presumably makes for good discrimination ("acquired distinctiveness", in Liberman's terms).

A similar result was obtained by Boersma et al. (2003) with computer simulations of an Optimality-Theoretic learning algorithm: perceptual warping emerged

through the use of constraints in favor of perceiving all input F2 and F3 values, constraints *against* perceived F2′ and F3′ values, and constraints *against* perceptual warping. Although both Guenther and Gjaja's and Boersma et al.'s simulations are meant to be a part of a larger linguistic model, they would have to rely on a discrete event (a "category creation day") to turn the warped perceptions into discrete symbolic categories suitable for inclusion in a linguistic model. These models would become more principled if combined with gradual category creation, such as is promised by the distributed connectionist models discussed above.

Finally, there is the promise of *exemplar theory* (Nosofsky 1988), which has been applied to phonological storage by Pierrehumbert (2001) and Wedel (2004, 2006, 2007). This family of theories holds that the lexicon consists of a massive number of stored phonetic (or auditory) events, with or without category labels. Those subtheories that touch on category creation can do so because they include no category labels, but subtheories that make interesting linguistic generalizations (e.g. on auditory dispersion: Wedel 2006) do require the presence of category labels. Thus, although exemplar theory has the potential of becoming a comprehensive theory of language at some point, it cannot yet combine category creation with linguistic theorizing. For instance, exemplar theory cannot yet handle the simplest examples of sentence phonology, such as nasal place assimilation, because it cannot distinguish between underlying forms in the lexicon and surface forms. One could make a version of exemplar theory that includes both surface and underlying forms (Wedel 2004: ch. 4), but even such a version cannot handle sentence phonology, because it is incapable of singling out unambiguous underlying forms. It seems that in order to begin accounting for basic phonological phenomena such as nasal place assimilation, exemplar theory would have to be extended with stored relations between morphemes and underlying forms, and with relations between underlying and surface forms, thus becoming very much like the model of Figure 9.3.1.

A problem shared by all the above models (apart from having trouble linking to phonology) is that they rely on the existence of neural mechanisms that do computations with auditory distance. These mechanisms are Guenther and Gjaja's weighted summation over formant values; Boersma et al.'s distance-dependent anti-warping constraints, and exemplar theory's nearest exemplars in perception and neighborhood averaging in production, the underlying networks that should provide such distance-processing mechanisms are not specified. By contrast, models of associative memory (Kohonen 1984) can derive auditory-distance effects without representing auditory distance anywhere underlyingly. Likewise, there exist Optimality-Theoretic models in which auditory-distance effects emerge without represented auditory distance (for auditory dispersion: Boersma and Hamann 2008), but they do not handle category creation. There still seems to be a divide between models of category creation and models of linguistic processing in many respects.

9.3.8 Conclusion

The conclusion must be that as of yet, no model combines category creation with other emergent properties of language processing, but that some partial answers have been given, so that we may well find a comprehensive model in the future. Such a model may include the linked representations of Figure 9.3.1 (plus an articulatory form, as the speaker's output representation), and represent categories gradiently as distributed across a neurobiologically inspired network, preferably without representing auditory distance explicitly.

CHAPTER 10

ORGANIZATION OF PHONOLOGICAL ELEMENTS

Contributions to this chapter present a myriad of issues relating to the organizing structures of speech. Gafos and Goldstein review advances in explanation made possible by treating articulatory gestures as theoretical primitives. Côté reviews our current understanding of the evidence for syllables and their status as linguistic units. Turk discusses how higher level prosodic structure is reflected in the timing patterns of speech.

10.1 Articulatory representation and organization

Adamantios Gafos and Louis Goldstein*

10.1.1 Introduction

As a scientific enterprise, phonology is a theory of speakers' knowledge of the abstract patterns of speech events that characterize and differentiate languages. Work has largely proceeded from the assumption that the observables used to build the theory are transcriptions of speech as a sequence of segmental units. Not surprisingly, therefore, the internal representation of speech that theories of phonology have traditionally postulated is likewise a sequence of segmental units, not much different in kind from the observables themselves. With the advent of practical acoustic analysis in the 1950s, it became possible to consider using acoustic patterns as the observables, rather than relying on transcription. However, the acoustics revealed by spectrographic analysis appeared so unlike the transcriptions of speech, and so apparently incompatible with it (Harris 1953; Hockett 1955), that it was rejected as a primary observable. It was assumed that somehow a listener must reconstruct a segmental transcription when listening to speech (through something like categorical perception, e.g. Liberman et al. 1967), so neither the basic observables, nor the nature of the hypothesized representations were much changed.

* Research supported by NIH NIDCD DC008780 and NSF 0922437 grants to Haskins Laboratories and Advanced ERC grant 249440.

Two independent developments later together spawned novel work that challenged the standard view of phonological representation and its reliance on transcription for observables. These were autosegmental phonology (Goldsmith 1976) and the beginnings of availability of techniques (x-ray, magnetometer, ultrasound) for obtaining quantitative records of articulator motion during speech. Autosegmental theory hypothesized that the phonological representation was composed of multiple feature sequences whose boundaries were not necessarily aligned in time in a manner that would be implied by a single sequence of segments. While many autosegmental theorists continued to employ segmental transcriptions as the observables, others saw that the temporal structure of these representations was in many cases isomorphic with the (newly) observable structure of articulatory events (e.g. Fujimura 1981; Browman and Goldstein 1986). Thus, it became possible to use articulatory events and their timing as observables informing autosegmental-type representations.

While phonological representations are no longer seen as autosegmental, the working assumption that phonological representations are isomorphic with speech production events has produced a new, deeper, and more general understanding of several phonological phenomena. This is so partly because the isomorphism makes it possible to test specific hypotheses about representations and processes using (quantitative) articulatory observables. This has proven particularly revealing in circumstances in which the nature of the articulatory-acoustic mapping obscures an articulatory event in the acoustic (and transcription) record. The hypothesis of such "hidden" events can afford a novel, simple description of particular phonological processes, and can be tested in laboratory experiments.

10.1.2 Units of articulatory organization and units of phonological encoding

One of the major stumbling blocks to being able to systematically relate the observables of speech to phonological representations and processes is the apparent incompatibility in the nature of the entities involved at the two different levels of description. Combinatorial phonological units are discrete, qualitative, and context-invariant, while speech is continuous (in space and time) and highly context-dependent. The same issue of the relation between the qualitative and the continuous is met in (diachronic) sound change (see Harrington, this volume).

Advances in understanding of the coordination and control of action, beginning with Bernstein (1967) and Turvey (1977), have provided a principled way of unifying these descriptions. This approach was first applied to speech in the work of Fowler (1980) and Fowler et al. (1980), and was made explicit in the concept of speech *gesture*, as developed in the Articulatory Phonology of Browman and Goldstein

(1986, 1989, 1990a), and the Task Dynamics model of Saltzman (1986), Saltzman and Munhall (1989), and Kelso et al. (1986).

A gesture is a functional unit of action that achieves a specified task (most often in the case of speech, a vocal tract constriction task). Two properties of gestures are key to bridging the conceptual divides between qualitative and continuous, and between context-invariant and context-dependent. The first is the notion of *coordinative structure*. The constriction tasks of gestures are defined abstractly in the space of constriction variable goals. For example, reducing to 0 (or less) the distance between the upper and lower lips (or Lip Aperture) results in producing a lip closure gesture. The many articulatory (e.g. upper lip, lower lip, jaw) and muscular (orbicularis oris, anterior belly of the digastric, risorius, etc.) components that can contribute to this task form a coordinative structure (or functional synergy) within which they cooperate flexibly to achieve the task. This flexibility means that the task will be achieved with quantitatively different movements of the articulatory components, depending on the context in which the gesture is produced. The articulator motions are context-dependent, but the task description guiding them is invariant. The motions are not themselves gestures, but are guided by the current active gesture(s). A similar approach to contextual flexibility is also found in Guenther's (1995) neurally inspired model of speech production.

The second relevant property of gestures is that the continuous motion of a controlled task (its kinematics) is modeled as the output of a dynamical system, i.e. a system of differential equations. The signature property of such systems is that while the state (position and velocity of the object or, here, constriction) is changing continuously, the equations that give rise to the time-varying state are fixed during the lifetime of the gesture and constitute an underlying law governing this surface variability (e.g. Saltzman 1995). Most speech gestures have been hypothesized to be governed by *point-attractor* dynamics: all possible trajectories converge on a single state over time, as specified by the target, or equilibrium position parameter of the system. The hypothesized dynamics will give rise to quite different trajectories depending on the initial condition (as determined by context). Dynamical laws defining gestures thus stand at the same level of abstraction as the invariant context-independent units of representation in a purely symbolic view of phonological units. Crucially, however, it would be misleading to view the kinematics as *implementing* these (dynamical) symbols because no additional formal construct is needed to go from the dynamical law defining gestures to the continuity and context specificity of their kinematic patterns.

While point-attractor dynamics (and the related Equilibrium Point Hypothesis, Perrier et al. 1996) provide a good first approximation to a lawful description of speech kinematics, other findings suggest that speech gestures do not have targets that are single points but rather ranges of values. The arguments for this approach have been developed most explicitly in the work of Keating (1990b), wherein "windows" prescribe ranges of variability within individual articulatory

dimensions, and by Guenther (1995), wherein targets are defined as convex regions in a multidimensional space of orosensory parameters (such as tongue body height with respect to jaw, tongue tip pressure receptor, and so on). Within the dynamical systems model of speech gestures, for a proposal on defining targets as ranges see Saltzman and Byrd (2000) and for a different proposal on specifying targets using "activation fields" see Gafos and Kirov (2010).

Research on speech errors has also shown that the choice of observables can strongly influence conclusions about the minimal units of phonological encoding in speech production. Levelt, Roelofs, and Meyer's (1999) theory of phonological encoding in speech production hypothesizes that these are wholistic, symbolic units, and one of the major sources of evidence presented for this view is the nature of (segmental) speech errors. Analyses of transcriptions of speech error corpora (Fromkin 1971; Shattuck-Hufnagel and Klatt 1979; Shattuck-Hufnagel 1983) have been argued to show that errors result from the insertion of an intended phonological segment in the wrong *slot* within a prosodic *frame* for an utterance. Apart from this misplacement of a unit (or units), an errorful utterance is assumed to be both phonotactically and phonetically well-formed. Fromkin (1971) originally argued for the abstractness of the unit involved in errors by uncovering errors in which a phonological segment is phonetically accommodated to its new position, e.g. *slumber* [p^{h}]*arty* → *lumber s*[p]*arty*.

The observables that were used to develop these theories of speech errors are segmental transcriptions. However, recent work reveals a very different picture when quantitative measures of speech articulation (Mowrey and MacKay 1990; Pouplier 2003, 2007, 2008; Pouplier and Goldstein 2005; Goldstein et al. 2007) and/or acoustics (Frisch and Wright 2002; Goldrick and Blumstein 2006) during error production are examined. The key result is that the measured properties of a segment when it appears as a substitution in an error are usually not identical to those of the same segment when it is produced in a non-errorful utterance. In fact, the substitution combines properties of the intended and substituted segments. In the most extreme case of this (Goldstein et al. 2007), errors appear to involve simultaneous production of the intended constriction gesture and the substituted gesture. For example, when talkers repeat the phrase "cop top" or "top cop," they produce *gestural intrusion* errors, in which the tongue dorsum gesture for /k/ and the tongue tip gesture for /t/ are coproduced (Goldstein et al. 2007). Such intrusions are the most frequent type of error observed, both in their repetition task and in a SLIP task (Pouplier 2007) in which there is no overt repetition. These errors (and their frequency) call into question the classic arguments for segments as units of phonological encoding. First, they show that many errors are not in fact phonetically well-formed (coproduced tongue tip and tongue dorsum constrictions are not licensed in English). Second, the occurrence of gestural intrusions can be explained as qualitative shifts in systems of coupled oscillators to a dynamically more stable state (Goldstein et al. 2007), which relates the errors to a wider class

of non-linear phenomena. Finally, for multigestural segments such as nasals, either of the component gestures (oral constriction, velum lowering) can intrude without the other intruding (Goldstein et al. 2007). This argues that gestures function as units of encoding.

The fact that gestural intrusion errors are the most frequently observed error does not, of course, rule out the possibility that segments as well as gestures are units of phonological encoding. It merely shows that the classic arguments are inadequate, because they are based on purely transcriptional observables. Richer experimentation will be required to determine if there exist certain types of errors that provide unambiguous support for segments as units of phonological encoding.

10.1.3 Articulatory events and phonological processes

Another insight of the gestural approach is the idea that phonological units and processes may be realized as patterns of gestural coordination among the gestures that constitute these units and that interact in phonological processes.

We will illustrate this with examples of allophonic variation, assimilation and harmony. First consider the difference between "clear" and "dark" allophones of English /l/, as in *lip*, *late*, *lie* versus *pill*, *feel*, *cool* ([l] versus [ɫ]). In atemporal models of phonology, the difference would be expressed by saying that the basic allophone is the clear /l/ and in syllable-final position this /l/ changes to a "dark" or velarized version by a feature-change rule adding the feature [+back]. Looking closely at this variation with the X-ray microbeam system, Sproat and Fujimura (1993) discovered that English /l/ is composed of two gestures, a tongue tip "consonantal" gesture and a tongue dorsum "vocalic" gesture, and that the relative timing of these varies as a function of syllable position and adjoining prosodic boundary. In syllable-initial position, the two gestures show a synchronous pattern of relative timing, with tongue tip and tongue dorsum attaining their goals at the same time. In syllable-final position, the tongue dorsum gesture significantly precedes the tongue tip gesture, with the tongue dorsum attaining its target at the onset of the tongue tip gesture. In syllable-final position, then, the acoustic portion of the syllable corresponding to the vowel is significantly more overlapped with the tongue dorsum gesture. The acoustic consequence of this difference in overlap is what gives rise to the distinction between the "clear" and "dark" variants of /l/ (see also Browman and Goldstein 1995). Krakow (1989, 1999) finds a strikingly similar pattern of timing in English nasals between the component gestures of velic lowering and oral closing, and shows how the timing differences can be used to explain the allophonic variation between oral and nasalized vowels as in *meat* versus *team* (see also Byrd et al. 2009 for a recent replication of these results using real-time MRI). The insight of expressing phonological processes such as allophony as patterns of gestures and their coordination has inspired the development of grammar models based on gestural

representations. In a study of the phonological system of Moroccan Colloquial Arabic, Gafos (2002) argues that phonological knowledge can make reference to the temporal dimension of linguistic form. This proposal adopts Optimality Theory (Prince and Smolensky 1993/2004) by expressing language-particular patterns as the result of optimization under a set of violable constraints, some of which must crucially refer to temporal relations among gestures. Angermeyer (2003), Benus, Smorodinsky, and Gafos (2004), Bradley (2002), Davidson (2003, 2006c), and Hall (2003) also pursue a model of grammar based on gestural representations and Optimality Theory in analyzing independent phenomena in other languages.

Another area of focus has been assimilation. A sample inventory of experimental studies on local assimilation includes: Bulgarian [t d] palatalization, Wood (1996); Chaga nasal-stop sequences, Browman and Goldstein (1986); English /s/ palatalization, Zsiga (1995); English /s/ to [ʃ] assimilation, Holst and Nolan (1995), Nolan et al. (1996); English camper, camber, Browman and Goldstein (1986); English casual speech, Browman and Goldstein (1989, 1990a); Castillian Spanish nasal place assimilation, Honorof (1999); German CC clusters, Kohler (1990), Kröger (1993); Igbo vowel assimilation, Zsiga (1997); Greek vowel hiatus altenations, Baltazani (2006b); Italian CC clusters, Farnetari and Busà (1994); Russian palatalization, Keating (1988); Russian coronal-dorsal sequences, Barry (1991). For a discussion of assimilation and experimental data on lexical access, see Ernestus (this volume).

Here we review two examples. Zsiga (1995, 1997) compared [s + j] sequences as in *confess your*, whose acoustic consequences resemble [ʃ], especially under fast speaking rates, to other [ʃ]s as in *fresh* and *confession*. In *fresh* the [ʃ] is part of the mental lexicon entry. In *confession*, the [ʃ] is assumed to be derived by a lexical phonological rule of palatalization changing [s] to [ʃ] when an [s]-final verb combines with the Latinate suffix *-ion* to form its deverbal noun. Using electropalatography, Zsiga found that the tongue-palate contact pattern during the acoustic interval corresponding to the [ʃ] in *confession* is indistinguishable from that of the [ʃ] in *fresh*. However, in *confess your*, tongue-palate contact patterns during the underlined portion of the utterance change in a way that reveals the bisegmental make-up of such sequences. Across word boundaries, therefore, an [ʃ]-like acoustic output arises via coarticulation; that is, as the by-product of the temporal overlap between [s] and [j]. Thus, coarticulatory overlap and the result of the presumed phonological rule of palatalization may have similar acoustic consequences, but the two can be teased apart by examining how articulation unfolds in time.

In his work on Castillian Spanish nasal place assimilation, Honorof (1999) finds that the alveolar nasal [n] assimilates completely to the place of the following labial or dorsal obstruent, e.g. in /digan # paxa . . ./ → [diɣampaxa. . .] 'say (form. pl) straw', the alveolar /n/ assimilates completely to a labial nasal. This subset of the data is therefore fully consistent with standard phonological treatments of assimilation. According to these, the place specification for the nasal is replaced by a copy of the place specification of the following obstruent (Chomsky and Halle 1968), or

in an autosegmental view the domain of the place specification of the obstruent extends via spreading to also encompass the nasal with concomitant delinking of the nasal's specification (McCarthy 1988). However, when the obstruent trigger of nasal place assimilation was the dental [t], Honorof's data showed that the result of the assimilation is not a dental [n]. Rather, [n] and [t] blended variably with a constriction location intermediate between an alveolar and a dental. The blending seen in the /n/ plus coronal sequences is a notable result that speaks to the issue of underspecification. In particular, the radical underspecification theory of representations has promoted the idea that unmarked segments are not specified for certain features, which marked segments are (Archangeli 1988; Stemberger 1991), and since coronals are considered to be the prototypical unmarked segments, coronals should lack a specification for their place of articulation. Such unmarked segments receive fully specified representations by the action of default rules that fill in the missing values or by assimilation rules that spread the values from nearby segments to the underspecified targets (see Steriade 1995 for a review). The Castillian Spanish blending facts indicate that, if /n/ is considered to be a target of assimilation, then it cannot be said to be underspecified.

So-called *long-distance* assimilations such as vowel and consonant harmony have also been investigated using laboratory techniques. Since Gay (1977, 1978) it has been known that a non-contiguous sequence of identical vowels such as [u-u] in [kutup] is produced by speakers of English with a discontinuity both in the articulatory and the electromyographic measures of lip rounding (see also Boyce 1988, 1990). For example, in the electromyographic signal there is a trough coincident with the production of the intervening consonant. The cessation of muscle activity during the consonant is consistent with the hypothesis that the linguistic representation underlying the production of lip-rounding schedules the rounding of the two identical vowels as two independent events, $[u]^{\text{Round}}C[u]^{\text{Round}}$ where C is a variable for any permissible intervocalic consonant or consonant cluster. A number of other studies have documented the same trough pattern in the production of non-contiguous, identical vowels in Spanish, French (Perkell 1986), and Swedish (McAllister 1978; Engstrand 1981). In contrast to these cases, Boyce (1988, 1990) found a plateau of continuous activity in Turkish for [uCu] utterances both in muscle excitation patterns (of the orbicularic oris) and in lower-lip protrusion kinematics. This pattern of results, the English trough versus the Turkish plateau, seems to reflect the fact that Turkish but not English has vowel (rounding) harmony.

Cohn (1990, 1993b) studies a case of nasal harmony in Sundanese, in which nasality spreads rightward from a nasal consonant until it encounters a supralaryngeal consonant, e.g. [ɲãũr] 'say', but [ɲãtur] 'arrange'. But the laryngeals /h ʔ/ can intervene in the domain of nasal spread as if they were skipped by the spreading, e.g. [mĩhãk] 'take sides' and [nũʔũs] 'dry.' Using oral/nasal airflow traces, Cohn presents evidence that these "transparent" consonants are in fact nasalized. This result is consistent with the standard autosegmental treatment which sees harmony as

an extension of the domain of the assimilating property. Gerfen (1999) studies nasal harmony processes in Coatzospan Mixtec using airflow recordings, and Walker (1999) is an acoustic study of nasal harmony in Guarani. For vowel harmony, using a combination of electromagnetic articulometry and ultrasound methods, Benus and colleagues studied transparent vowels in Hungarian vowel harmony (Gafos and Benus 2003; Benus and Goldstein 2004; Benus 2005; Benus and Gafos 2007). Their results indicate that transparent vowels are affected by the harmonic context. Gick et al. (2006) used ultrasound to study the transparency of the low vowel [a] in Kinande tongue root harmony. Walker et al. (2008) studied transparency in the consonant harmony of Kinyarwanda using electromagnetic articulometry. Both of the above studies provide evidence that transparent segments are participants in the domain of harmony. From the perspective of the typological richness and specificity of harmony systems across languages much remains to be done, both in terms of charting the phonetic data in a more rigorous way than with transcriptions and in terms of integrating that data with phonological theory (for a review see Archangeli and Pulleyblank 2007).

We highlight a critical outstanding issue in relating experimental data on harmony to phonological theories. Focusing on an apparently simple case, we can ask what relation can be established between phonological theory and, for example, the continuous activation of lip rounding observed in Turkish [uCu] sequences. Two hypotheses suggest themselves: the continuum is an extended unitary rounding gesture, or the continuum is the aggregate by-product of overlap of separate shorter rounding gestures. According to the former view, in Turkish, rounding would extend over a domain encompassing both vowels in [uCu] and this is what gives rise to the plateau seen in Boyce's study. In the latter view, the plateau is the result of two separate rounding instructions, each with its own temporal domain, and it is the juxtaposition of these two rounding domains which results in a rounding plateau across the entire [uCu] sequence. The choice between the two hypotheses corresponds to a fundamental issue in phonological theory. This is the issue of assimilation and harmony as feature spreading (Goldsmith 1976, 1985, 1990; Clements 1976, 1977, 1985; Kiparsky 1981; Hayes 1986; Sagey 1986) versus feature change (Chomsky and Halle 1968). Although the former view is widely assumed, it has never been subjected to systematic investigation across languages and across assimilating parameters. Deciding between these two views is not an easy matter. It is well known that due to coarticulation the shape of the vocal tract at any time is an aggregate of multiple gestures associated with different segments. Aggregation has been observed for gestures that involve different constriction variables and for gestures that involve the same constriction variables. For different constriction variables, Hardcastle (1985) and Marchal (1988), using electropalatography, find that the gestures of two successive consonants, such as those in /kt/ (one with a tongue dorsum constriction goal, the other with a tongue tip constriction goal), show different degrees of overlap, and that the amount of overlap increases with

speaking rate (Hardcastle 1985). Similar results are reported when overlapping gestures are specified for the same constriction variable. Munhall and Löfqvist (1992), for example, study the effects of speaking rate on two successive laryngeal abduction movements in *kiss Ted*, where the two units with laryngeal abduction gestures correspond to *ss* and *T*. The basic finding is that the distance between the two glottal peak openings decreases as rate increases. At slow rates, two opening movements occur and the glottis is closed between these two openings. At fast rates, a single movement is found with similar durations for the abduction and adduction phases (see also Boyce et al. 1990 for similar results that relate the occurrence of a one- or two-movement pattern for the velum to the rate of speech). Munhall and Löfqvist (1992) find that the *shape* of the observed trajectories could be reasonably well modeled by adding two underlying gestures at different degrees of overlap. When the gestures do not overlap, summation produces two clear peaks in the shape of the simulated trajectory. As overlap increases, the simulated trajectory resembles in shape the blends or single movement patterns observed in the actual trajectories. However, there were inconsistencies between the amplitude of the simulated trajectories and that of the actual trajectories, especially at intermediate to large amounts of overlap. These inconsistencies derive from the assumption that the aggregation function can be estimated by simple algebraic summation or linear superposition.

An alternative is to hypothesize that the dynamical parameter values (target and dynamical stiffness, or time constant) of overlapping gestures of a constriction variable are averaged, rather than added (see Saltzman and Munhall 1989). In the case of a partially overlapping sequence of identical gestures, as might represent certain types of geminate consonants, this would mean that the same dynamical regime would be involved in single vs. geminate consonants, with the only difference being the span of time over which the regime is active. This representational difference could account in a simple way for Löfqvist's (2005) findings on the kinematic properties of geminates in Japanese and Swedish. In addition to geminates being of course longer, they are produced with greater articulatory displacements and result in tighter constrictions. Yet they lack the increase in peak velocity that is usually associated with an increase in displacement (cf. Beckman et al. 1992; Cho 2006). This combination of greater displacement without a corresponding increase in peak velocity could result from effective undershoot in the case of the single consonant. If we hypothesize that the relation between gestural time constant (or stiffness) and the activation duration of single consonants is such that single consonants do not have sufficient time to reach the target value, then they will exhibit undershoot. The longer activation time of the geminate would allow the target to be reached.

Other results suggest that in some cases the aggregation function must be more complex than either adding or averaging. Nolan et al. (1996) investigated the properties of s-ʃ overlap in English examples like *this shop*. Using electropalatography and acoustics, they find that for modest degrees of overlap, the results are consistent with the predictions of parameter averaging. However, for extreme degrees of

overlap, the palatographic and acoustic characteristics of the maximum constriction are not significantly different from those of [ʃ] by itself; i.e. there appears to be no influence of [s] at all on those characteristics. Yet, the overall constriction duration is longer than that for a single [ʃ], suggesting that the [s] gesture is still somehow contributing to the observed movements.

The key unresolved issue thus can be summarized by asking: how do the planning or execution systems combine multiple inputs for a given constriction variable? This is a critical question for the study of the relation between linguistic representation and articulatory organization and it is a question we can only ask if coordination of gestures is a fundamental part of our model.

10.1.4 Syllable organization

Laboratory phonology work over the last twenty years has developed both theoretical models and empirical methods that pursue the consequences of defining syllable structure as patterns or modes of temporal coordination among phonetic primitives. This approach is possible when the primitives are articulatory units that have observable, dynamic temporal properties among which abstract coordination relations can be defined (Browman and Goldstein 1988, 1995; Gafos 2002). Thus, the organization of compositional primitives into syllables, and the structural relations among units within a syllable (onset, rime, nucleus, coda) are implicit in the same representation (coordination topology) as required to adequately model the temporal regularities of speech. The consequences of this view have been investigated for a wide range of phenomena from syllable-structure-sensitive allophony (see Section 10.1.3 above) to universal preferences (markedness) of syllable structures (Nam et al. 2009). For a broader discussion of syllables examined with laboratory phonology approaches see Côté (this chapter). For higher prosodic structure and rhythm, see Turk (this chapter).

One specific theory of syllable structure developed in this framework is based on coupled oscillators (Goldstein et al. 2006; Nam et al. 2009). The theory attempts to account for why CV syllables are preferred to VC in several ways: they are more frequent cross-linguistically (and may be the only universally attested syllable type), they are acquired earlier than VC syllables, and they afford relatively freer internal combination (between onsets and nuclei) than do VC (between nuclei and coda). That theory attempts to relate these generalizations in a principled way to the fact that C and V gestures are triggered relatively synchronously in onsets (Löfqvist and Gracco 1999), but not in codas. In this theory, stable temporal coordination among articulatory units during speech production is achieved by associating each unit with a clock responsible for triggering that articulatory action, and by coupling the clocks for a given syllable's gestures to one another in a plan or *coupling graph* (a specific model of coordination topology). The coupling relations

within the graph are hypothesized to leverage the intrinsically available modes of coupling oscillatory motions (Haken et al. 1985; Turvey 1990), in-phase and anti-phase. Much work summarized in those papers shows that the in-phase mode is more *accessible* and more *stable* than the anti-phase mode. Thus if a syllable is to be composed of a consonant unit and a vowel unit, there are only two ways of coordinating them using these intrinsically available modes: in-phase, in which C and V are triggered synchronously is hypothesized for the onset-nucleus (CV) relation, and anti-phase (sequential triggering) is hypothesized for the nucleus-coda (VC) relation. Given the independently motivated properties of in-phase and anti-phase modes, the differences between CV and VC syllables can be explained.

A promising implication of the coordination topology model of syllable structure is that it opens the possibility of using temporal properties of articulatory events to infer syllabification. Whereas in English strings such as /kru/ *crew* or /gli/ *glee* are parsed into a single syllable with a complex two-consonant cluster as its onset, in Moroccan Arabic similar strings are claimed to be parsed into two syllables, e.g. /kra/ → [k.ra] 'rent', /skru/ → [sk.ru] 'they got drunk,' /glih/ → [g.lih] 'he grilled' (<.> marks syllabic divisions; Dell and Elmedlaoui 2002). In terms of coordination topology, the consonants composing the onset in English should all share the same (onset) coordination in relation to the vowel, while they are coordinated sequentially with respect to one another (Browman and Goldstein 2000). However, in Arabic, only the single (simplex) onset consonant bears the onset relation to the vowel. The different topologies should be associated with distinct temporal patternings of articulatory intervals. Pursuing this prediction, articulatory studies of syllable structure have examined the variability of structurally relevant intervals. Two distinct patterns of stability have emerged, each characteristic of a particular qualitative syllabic organization. In languages that admit complex onsets, the most stable interval across CVC, CCVC, and CCCVC utterances (where C is any consonant and V is any vowel) is an interval defined by the center of the pre-vocalic consonantal string and the end of the hypothesized syllable (Browman and Goldstein 1988; Honorof and Browman 1995; Byrd 1995). The stability of this interval is predicted by models in Browman and Goldstein (2000) and Gafos (2002) as the result of optimization in systems of competing C-V and C-C constraints on coordination, and also by the coupled oscillator model (Saltzman et al. 2008) as the result of a loop in the coupling graph in which all onset Cs are coupled in-phase with the V and anti-phase with one another. In contrast, in languages that do not admit complex onsets such as Arabic, the most stable interval across CVC, CCVC, and CCCVC utterances is defined by the immediately pre-vocalic consonant and the end of the hypothesized syllable (Shaw et al. 2009). See Côté (this chapter) for a summary from other languages.

Shaw et al. (2009) introduce computational and analytical methods in the study of the relation between syllable structure and experimental data. Given a

hypothesized coordination topology, their models generate simulated temporal structure via a probabilistic version of a theory of temporal coordination constraints (Gafos 2002). The simulated data are then compared to the experimental speech movement data for their goodness of fit. Using this method, Shaw and Gafos (2010) show that for a CCV string in a language that does not admit complex onsets like Moroccan Arabic, the simplex onset topology provides a better fit to the experimental data from that language than the complex onset topology. The situation is reversed for English data. Shaw et al. (2009) also show that variability in the experimental data can influence the behavior of stability indices projected from an underlying qualitative syllabic organization. As variability across the lexical sample over which stability measures are assessed is increased moderately, the stability indices corresponding to the qualitative organization of a simplex onset parse remain in the quantitative region characteristic of simplex onsets. But as variability increases further, a tipping point can be seen beyond which the stability pattern turns to a state characteristic of complex onsets. The stability pattern can therefore change, thus exposing the range of validity of earlier heuristic methods (discussed above) that do not employ explicit stochastic modeling. Overall, instead of ignoring variability or treating it as a nuisance, Shaw et al. (2009) develop methods which harness variability as a tool for elucidating the relation between mental organization, in the form of syllable structure, and its complex behavioral instantiations.

10.1.5 Concluding remarks

We have reviewed how research in articulation has informed phonological inquiry across a wide range of domains. In each case, we have presented the key results obtained and noted areas of convergence or lack thereof with phonological theorizing. A shared notion in the research reviewed is the construct of the speech gesture, a dynamic event with both spatial and temporal properties. This notion, and the model in which it is embedded, have sustained research in laboratory phonology by keeping in perspective both theoretical developments in phonological theory and novel methods of acquiring experimental data. Because the model is formally fleshed out, it can be used to derive explicit predictions. These predictions have been pursued in various studies using a wide range of experimental methods. In turn, the studies pursuing these predictions have produced new data patterns which present opportunities for sharpening the theory, the model, or the relation between the two. We have discussed examples of this interleaving of theory and experiment around the issues of, most notably, the nature of speech errors, the formal mechanism of assimilation (spreading versus feature change), the notion of transparency in harmony systems, the relation between phonological plan and surface-produced output, and finally syllable structure.

10.2 The role of the syllable in the organization and realization of sound systems

Marie-Hélène Côté

10.2.1 Introduction

Theories of phonology include an inventory of basic elements, such as segments, tones, stress, features, and gestures, and a model that accounts for their distribution and realization. The phonological context is taken to be organized into higher-order prosodic constituents, which structure and constrain the realization of phonological elements. Moving from elements to their distribution, this section addresses issues pertaining to the syllable as an organizing constituent of phonological units, bridging between the coordination of articulatory events (Gafos and Goldstein, this chapter) and prosodic structure (Turk, this chapter, and Chapter 11).

A large range of approaches to the syllable have been offered, varying in their level of abstractness and supported by an impressive variety of experimental results. Despite the richness and sophistication of this body of research and the intuitive attractiveness of the syllable, the syllabic domain remains difficult to define and characterize, physically and formally. Questions surrounding the syllable concern its nature (1–4), its role in the organization and production of sound systems (5–6), and its cross-linguistic variability (7).

1. Is the syllable an abstract primitive, governed by universal organizational principles such as those requiring syllable onsets or prohibiting syllable codas, or an epiphenomenal category emerging from linguistic experience (e.g. Ohala 1992b, 2008; Bertinetto 2001)? Does it have a substantive status in the analysis of sound systems, or is it merely a descriptive concept, a surface and loosely defined segmental grouping?
2. What is the internal structure of the syllable? One can distinguish between a rhythmic organization involving the mora as a unit of segment quantity and syllable weight, and a segmental organization relying on subsyllabic constituents (see Zec 2007 for a recent discussion). These include the nucleus, the onset (the preceding consonants), the coda (the following consonants), the rime (nucleus+coda), and the body (onset+nucleus).
3. What is the basis (phonetic or other) of syllable structure and syllabification procedures?
4. How does the syllable relate to the segmental content and to other prosodic constituents?

5. What categories of processes and generalizations, if any, are characterized in terms of syllable structure? Are they sensitive to the position of syllable edges and/or the syllabic affiliation of segments?
6. At what levels, if any, does the syllable intervene in mechanisms of speech perception and production?
7. What aspects of the syllable are universal or subject to cross-linguistic variation?

Several reviews of the syllable from a formal phonological perspective are available, which establish the syllable as an organizing unit between segments and higher prosodic structure (e.g. Blevins 1995; van der Hulst and Ritter 1999; Rubach 1999; Zec 2007). I focus here on the contribution of laboratory phonology to some of the issues listed above, discussing a number of recent experimental results that speak to the activity, existence, and physical correlates of the syllable or particular subsyllabic constituents. The interpretation of the results is subject to much ambiguity, however, a point also emphasized by Pierrehumbert and Nair (1995) and Shattuck-Hufnagel (2011). Shattuck-Hufnagel, for instance, argues that speech errors, regularly cited as supporting the syllable, provide basically no unambiguous evidence for it, at least in English. Discussions may fail to properly distinguish between word and syllable boundaries, between evidence for some subsyllabic grouping and evidence for the whole syllable. Results interpreted as supporting the syllable may also be compatible with linear characterizations: in certain linguistic or experimental contexts, "onset" corresponds to "pre-vocalic," the number of syllables to the number of vowels.

I concentrate on two broad categories of findings. First, those that directly address the role of the syllable in phonological generalizations (Section 10.2.2.1) and in speech perception and production (Section 10.2.2.2). I suggest that the evidence for the syllable is not as clear as generally admitted. In Section 10.2.3, I turn to the internal organization of the syllable or subsyllabic relationships, considered in relation to statistical tendencies in the lexicon, acoustic/perceptual factors, and articulatory organization. Focus is on the asymmetry, central to the syllable, between CV and VC sequences. Section 10.2.4 offers additional comments and concluding remarks. The discussion straddles phonetics, phonology, and psycholinguistics; while specialists in each domain might be left unsatisfied, it is hoped that this multidisciplinary perspective will end up being more than the sum of its parts.

10.2.2 The role of the syllable

10.2.2.1 *In phonological generalizations*

The syllable has enjoyed considerable success in the analysis of phonological patterns, segmental and suprasegmental. Syllable-based accounts refer either to the

syllabic affiliation of segments (e.g. lenition or neutralization of coda consonants, vowel epenthesis in complex codas), or to syllable shape (e.g. vowel laxing or shortening in closed syllables, stress attraction to heavy syllables). I will not expand on this traditional type of evidence, focusing instead on some recent challenges to the syllabic formulation of phonological generalizations.

Descriptively, the syllable is not necessary since phonological processes that are expressed with reference to the syllable can always be reformulated in sequential terms. Conceptual economy has motivated the exclusion of the syllable from the set of basic phonological units, yet the perceived explanatory power of the syllable and the simplicity of syllable-based accounts have secured its place at the center of phonological theory. More recently, however, the syllabic basis of phonological generalizations has been questioned on empirical grounds: syllabic analyses do not necessarily make the correct predictions when a closer look at the data is taken (e.g. Steriade 1999b, 2001; Côté 2000; Blevins 2003). For example, McCrary (2004) provides experimental evidence against the role of the syllable in different aspects of the phonology of Italian, which have standardly been understood with reference to the syllable: the allomorphy of the masculine definite article (*il*∼*lo*) and segment duration. In particular, no evidence for vowel lengthening in open stressed syllables is found; instead, vowel duration is inversely correlated with the duration of the following consonantal sequence, irrespective of its syllabic organization.

In mixed typological and experimental studies, Ahn (2000) and Zhang (2004) argue that stress attraction and contour tone distribution, classically formulated in terms of syllable weight, are sensitive to duration, not syllable structure. Only long vowels, not closed syllables, truly attract stress; CVC syllables may only fail to repel stress, as opposed to CV syllables, in languages with vowel-length distinctions. The distribution of contour tones is determined by the phonetic duration of the sonorous portion of the rime, which is affected by elements of the wider context such as word length and phrasal position. Such proposals put into question the correspondence between the realization of phonological elements and syllabic structure, but they still require a distinction between rimal and onset consonants. In other words, it is the activity, not the existence, of syllabic categories that is at stake here.

In the domain of segmental phonotactics, a well-articulated non-syllabic alternative is Steriade's (1999a, 2001) "Licensing by Cue," as opposed to "Licensing by Prosody." The core idea is that the occurrence of a feature or segment in a given context is determined by its relative perceptibility in that context rather than its syllabic position. Voicing contrasts, for instance, tend to be maintained before a sonorant, where the cues to voicing are rich enough, and neutralized elsewhere. This usually corresponds to the onset position, hence the traditional syllabic formulation of voice licensing (e.g. Lombardi 1999), but Steriade uncovers crucial cases where the licensing-onset and neutralization-coda correspondences break up. This cue-based approach has also been applied to the resolution of consonant

clusters (Côté 2000) and, with experimental support, to palatalization (Kochetov 1999, 2006b) and retroflexion (Hamann 2003); see also Kirchner and Varelas (2002). In response, licensing by cue has been argued to be either insufficient or restricted to the diachronic domain. Critics have often specifically supported the syllabic alternative (e.g. Gerfen 2001; Howe and Pulleyblank 2001; Flack 2005; Wheeler 2005; Kaplan 2006; Moreton et al. 2008). But (a pure version of) licensing by cue may also be rejected without advocating a role for the syllable (e.g. Wagner 2002; Hansson 2003; Yu 2004).

10.2.2.2 *In speech perception and production*

Supplementing conclusions based on phonological generalizations and offering a different perspective on the role of the syllable, numerous experimental studies have investigated speaker behavior in speech perception and production. Mehler et al.'s (1981) classic study indicates that French listeners detect a target sequence in a lexical form faster when the sequence corresponds to a syllable in the form; for example, the sequence [ba] is detected faster in [ba.lɑ̃s] 'balance' than in [bal.kɔ̃] 'balcony,' [bal] faster in [bal.kɔ̃] than in [ba.lɑ̃s] (periods indicate syllable boundaries). This suggests that the syllable constitutes a processing unit in speech perception. Similar syllabic effects have been observed in production experiments using syllable priming (e.g. Ferrand et al. 1996), and in other studies summarized in Cutler (1997), Perret (2007), and Schiller (2008). However, subsequent (and better controlled) experiments failed to replicate these effects (e.g. Content et al. 2001; Schiller et al. 2002; Schiller and Costa 2006; Perret 2007; see Schiller 2008 and Cholin et al. 2006). Instead of a syllabic effect, what is generally observed is a segmental overlap effect, with CVC primes associated with faster production latencies than CV primes, irrespective of the syllabic position of the post-vocalic consonant in target words. This has led to the conclusion that syllables are not activated in lexical retrieval and not present in lexical representations, in accordance with the standard view in phonology that syllable structure is excluded from the lexicon, due to its predictable nature.

If correct, this conclusion calls for a reinterpretation of the "syllabic effects" observed in earlier studies. A phonetic account has been put forward as follows (Altmann 1997; Content et al. 2001; Segui and Ferrand 2002). The sequence [al] is phonetically distinct in [ba.lɑ̃s] and [bal.kɔ̃], due to the different quality of the liquid and degree of coarticulation with the preceding vowel. If speakers use this phonetic information in lexical access, the sequence [bal] will initially activate words like [bal.kɔ̃] but not [ba.lɑ̃s], and vice versa for the prime [ba]. Content et al. (2001) observed a "syllabic" effect with liquid post-vocalic consonants but not with obstruents. This contrast is consistent with liquids being more coarticulated with the preceding vowel than obstruents, less so with a general syllable-based mechanism. The appeal to subphonemic phonetic distinctions speaks to current

debates on the content of lexical representations, between minimally specified and phonetically detailed forms (see Chapter 8 this volume). It may also explain Cholin et al.'s (2004) finding, interpreted in terms of syllable structure preparation, that sets of words with identical initial syllables (e.g. Dutch *spui.en*, *spui.de*, *spui.er*, *spui.end*) are easier to produce than sets in which one of the words has a different syllabic segmentation (e.g. *hui.len*, *hui.ler*, *hui.lend*, but *huil.de*).

More promising evidence for the syllable may come from the effect of syllable frequency on speech production. Words composed of more frequent syllables are produced faster than words composed of less frequent syllables (with adequate control of phoneme and word frequency) (Cholin et al. 2006; Cholin and Levelt 2009). This result highlights the possible implication of the "mental syllabary," a repository of articulatory routines or pre-compiled motor instructions corresponding to syllable-size sequences. Less frequent syllables are either less easily accessed in the syllabary, or not stored in it and computed online. Frequency effects, however, do not in themselves support the existence of a suprasegmental syllabic level. There is no indication that the stored units are anything other than frequent segmental sequences, and Cholin and Levelt (2009) do not exclude that the mental "syllabary" also contains chunks that are smaller or larger than a syllable.

A different line of enquiry into the role of the syllable in speech perception and the organization of sound sequences exploits the phenomenon of perceptual epenthesis. Dupoux et al. (1999) show that Japanese speakers perceive an epenthetic vowel in sequences of consonants that are not legal in their native language; *ebzo*, for example, is heard as *ebuzo*. The form of the phonotactic constraint triggering perceptual epenthesis is unclear, however. At least two options can be entertained: a linear one, which bans the sequence [bz], and a syllabic one, which excludes [b] in coda position.

Kabak and Idsardi (2007) seek to disentangle this issue by contrasting two types of impossible biconsonantal clusters in VC_1C_2V context in Korean: clusters excluded by a syllabic constraint against C_1 in coda position (e.g. [cm]) and clusters banned by a sequential restriction against C_1C_2 (e.g. [km], C_1 being a possible coda before consonants other than C_2). Perceptual epenthesis is observed only in sequences of the first type, suggesting that perception is modulated by a syllabic organization. However, consonants excluded from the coda position in Korean (e.g. [c]) only appear before vowels or diphthongs, while other consonants (e.g. [k]) are familiar in pre-consonantal position. Perceptual epenthesis could exploit this linear distinction rather than syllabic constraints. A similar ambiguity arises in the interpretation of the results of Coetzee (2011, this volume), which indicate that American listeners hear an epenthetic vowel after [s] significantly more often in nonce forms like [st^{h}ápi] than [lust^{h}ápi]. This finding shows that listeners attend to allophonic cues in treating incoming sequences and make use of a grammatical constraint banning aspirated stops following a word-initial [s], as in [st^{h}ápi].

But different formulations of this constraint remain available, syllable-based (e.g. aspirated stops only occur syllable-initially) or not (e.g. stops are not aspirated after a tautomorphemic fricative): [lust^{h}ápi] may be licensed either by the insertion of a syllable boundary after [s], as suggested by Coetzee, or by parsing it as two separate words [lus#t^{h}ápi], with a normally aspirated word-initial [t] (like in *this tablet*).

10.2.3 Subsyllabic asymmetrical relationships

If support for the syllable, as reviewed above, may be considered mixed or ambiguous, evidence for asymmetrical segmental relationships within the syllable appears stronger. The dominant model of segmental association is the onset-rime one, which expresses a closer relationship between the nucleus and the coda than between the nucleus and the onset; the competing body-coda structure groups the nucleus and the onset. In terms of weight contribution, the mora establishes a contrast between pre-vocalic and post-vocalic consonants, as only the latter, like vowels, may bear a mora.

Three lines of investigation, reviewed in 10.2.3.1–10.2.3.3, can be identified in documenting and explaining the basic asymmetry between onset and coda consonants. Durational correlates have also been uncovered, such as compensatory mechanisms inside the rime (closed-syllable vowel shortening; Maddieson 1985) and the correlation between segmental duration and the moraic status of coda consonants (Broselow et al. 1997).

10.2.3.1 *Psycholinguistic and statistical evidence*

In discussing the internal constituency of the syllable, classic phonological data are complemented with a variety of psycholinguistic results, relying in particular on novel word games and phonotactic distributions in the lexicon. In English, the onset-rime model is supported by experiments indicating that speakers tend to maintain the integrity of onsets and rimes and break monosyllabic words at the onset-rime boundary (see Treiman 1989 and Treiman and Kessler 1995 for reviews). Statistical analyses of the lexicon reinforce the idea of a closer relationship of vowels with following codas than preceding onsets. Kessler and Treiman (1997) show that in uninflected CVC words in English, vowels more strongly interact with codas: certain vowel-coda combinations are more frequent than expected by chance, but no comparable tendency is observed between vowels and onsets (see also Frisch, this volume).

The onset-rime model has dominated the literature on the syllable, but its universality has been challenged. Yoon and Derwing (2001) and Derwing (2007)

present a series of experimental results supporting the body-coda structure in Korean and Minnan Chinese, with a stronger onset-vowel association than vowel-coda. This is consistent with a parameterized approach to syllable structure, each language adopting one among a small number of possible options (e.g. onset rime in English, body coda in Korean).

Recent work by Lee and Goldrick (2008) argues against such a categorical approach to the difference between English and Korean. The authors confirm the link between psycholinguistic results and statistical analyses of the lexicon: both support a stronger vowel-coda association in English and a stronger onset-vowel association in Korean. In a list-recall task in which participants repeat non-word CVC syllables, Korean and English speakers are more likely to recall correctly the CV and VC portions, respectively, consistent with the stronger associations observed in the lexicon. However, this general pattern is reversed under particular conditions. English participants do better on the CV portion when the CVC syllables contain CV sequences chosen among those that show a strong statistical association in the lexicon; the same applies to the VC portion in Korean. In other words, the speakers' behavior follows closely the statistical tendencies of the lexicon: a global preference for rime or body sequences but opposite local preferences when they are statistically favored. Since speakers do diverge from the "default" rime or body association of their language under certain conditions, the results are not immediately compatible with an abstract and invariant syllable architecture in each language.

This type of result is consistent with others that challenge the stability of syllable structure within a language. Italian leans toward an onset-rime model but the evidence appears much weaker than in English (Bertinetto 1999). Syllable weight is often not uniform language-internally but is determined by process-specific criteria; the categorization of syllables between light and heavy may be different, for example, for stress and tone (Gordon 2004).

10.2.3.2 *Gestural organization*

One flourishing research direction explores the links between intersegmental coordination patterns and prosodic organization. For example, investigating the traditional classification of languages as stress-timed, syllable-timed, or mora-timed (see Turk, this chapter for discussion), Smith (1995) suggests that in mora-timed languages, exemplified by Japanese, vowels in CVCV sequences are primarily coordinated with adjacent consonants, whereas syllable or stress-timing, as in Italian, is associated with a pattern of vowel-to-vowel coordination across intervening consonants.

Looking inside the syllable, much work has attempted to characterize the syllabic organization in terms of magnitude and relative timing of gestures (e.g. Browman and Goldstein 1988, 1995, 2000; Byrd 1995, 1996b; Krakow 1999; de Jong 2003). At least for American English, this research provides consistent results. In a

CVC sequence, there is more precise timing of articulatory movements in onset than in coda consonants, resulting in increased coarticulation in VC compared to CV sequences. For example, velic lowering in [m] is synchronized with labial constriction in onset position, but it precedes it in coda position (see Gafos and Goldstein, this chapter for a discussion of this aspect of articulatory organization and further examples). Concerning the magnitude of gestures, pre-vocalic consonants tend to be produced with a tighter constriction (a more extreme consonantal articulation) than post-vocalic ones. For example, post-vocalic nasals are associated with a lower velic position and longer low velic plateau, increasing the amount of nasal airflow; this makes post-vocalic nasals more sonorant-like or less obstruent-like than pre-vocalic ones. Likewise, post-vocalic laterals show a weaker tongue tip constriction.

Onset and coda clusters also display distinct timing characteristics. Onsets are characterized by the c-center effect, which corresponds to a relatively stable interval between the vowel and the center of the preceding consonantal sequence, irrespective of the number of consonants. For example, a stable interval is maintained between the vowel in *sayed*, *spayed*, and *splayed* and the center of [s], [sp], and [spl]. The c-center effect is not characteristic of coda clusters, which are produced more sequentially and also display more variability of intergestural timing between consonants. This has been formalized in terms of in-phase and anti-phase coupling modes (see Gafos and Goldstein, this chapter).

These generalizations are largely based on American English, and a growing body of work has begun to investigate their applicability to other languages and the range of cross-linguistic variation in coordination patterns; see Kochetov (2006a) for a comparison with Russian and Gick, Campbell, Oh, and Tamburri-Watt (2006) for a cross-linguistic study of the production of liquids. A number of recent papers have looked at initial clusters, contrasting languages that display the c-center effect (English, Marin and Pouplier 2008; French, Kühnert et al. 2006; Italian, Hermes et al. 2008; Georgian, Goldstein, Chitoran, and Selkirk 2007) with languages in which the vowel aligns not with the center but with the right edge of initial consonant sequences (Tashlhiyt Berber, Goldstein, Chitoran, and Selkirk 2007; Moroccan Arabic, Shaw et al. 2009). The two different alignment patterns have been interpreted as corresponding to different syllable structures: center alignment with complex onsets and right alignment with simple onsets (in this case additional initial consonants are extrasyllabic or belong to a separate syllable). Different clusters may also display different coordination patterns in the same language: Hermes et al. (2008) indicate a c-center effect for stop+liquid clusters but not /s/+obstruent in Italian; see Marin and Pouplier (2008) for coda clusters in American English. More research is needed to understand the language-specific and cluster-specific nature of articulatory coordination patterns, the factors that determine those patterns (e.g. perceptual recoverability, segmental contrasts), and the relationship between coordination patterns and syllable structure.

10.2.3.3 *Acoustic and perceptual factors*

CV and VC sequences also differ acoustically and perceptually, with several experiments pointing to the privileged status of CV transitions and the enhanced perceptibility of pre-vocalic consonants. Onset (pre-vocalic) consonants are more accurately perceived than coda (post-vocalic) consonants (Redford and Diehl 1999; Kochetov 2004). In a VCV sequence, the following vowel contributes to the perceptibility of the consonant more than the preceding vowel (Fujimura et al. 1978; Ohala 1990b; Wright 2001): when faced with contradictory transitions from the preceding and following vowels, listeners mainly rely on the CV transition to identify the consonant.

This perceptual asymmetry results from a number of factors. Acoustically, onset consonants and CV formant transitions display greater distinctness and spectral differentiation between different consonants than coda consonants and VC transitions (Öhman 1966; Kawasaki 1982; Redford and Diehl 1999). This is likely related to the articulatory factors mentioned above. Better synchronization of articulatory gestures and tighter constriction in onset position enhance the contrast between the consonant and the following vowel and positively affect perceptibility, since salience is partly determined by the degree of modulation in the acoustic signal (e.g. Kawasaki 1982; Ohala and Kawasaki 1984; Ohala 1992b; Boersma 1998). Pre-vocalic consonants also benefit from additional cues, notably stop bursts, which are not reliably audible in non-pre-vocalic position. Perceptually, the response of the auditory system confers increased salience to the onset of an acoustic signal (e.g. frication noise, formant structure, release burst), which gives rise to a marked burst of activity of the auditory nerve fiber (Bladon 1986; Delgutte 1997; Wright 2004). This provides a perceptual advantage to CV cues: the onset of formants (at the CV juncture) is amplified in a way that their offset (at the VC juncture) is not. The CV boost is optimal with stops and less pronounced with sonorants, which display less syntagmatic contrast with following vowels (Wright 2004).

The CV advantage, like articulatory patterns, appears to vary cross-linguistically. Tabain et al. (2004) indicate that CV transitions are acoustically less variable and more controlled than VC ones in English, but both transitions are equally stable in Arrernte, Yanyuwa, and Yindjibarndi (three Australian languages). Arrernte has been analyzed exceptionally as a VC language, banning onsets and requiring codas (Breen and Pensalfini 1999). This is consistent with the reduced advantage of CV in this language. However, there is no indication that Yanyuwa and Yindjibarndi also have a VC structure. One factor that is common to all three languages is the presence of an exceptionally large number of place distinctions, including subcoronal contrasts. Tabain et al. speculate that greater control of the VC transition (and limited gestural weakening in coda; see Kang 2000) is required to maximize the cues available to distinguish between places of articulation. So phonemic inventory contributes to determining a language's articulatory and acoustic structure—and possibly also its syllabic organization.

10.2.4 Discussion and conclusions

This review has brought forward data from a number of separate domains: phonological processes, production and perception studies, statistical analyses of the lexicon. Globally, the evidence for the syllable as a stable and active constituent is mixed or ambiguous, but position-specific articulatory or perceptual properties and patterns of intersegmental cohesion appear more strongly supported. One additional body of experimental work not discussed so far is that dealing with speakers' syllabification judgments, in particular the syllabification of intervocalic consonants and clusters (see Côté and Kharlamov 2011, for references). Such judgments do not directly address the syllabic basis of the organization of sound systems, but they contribute to defining the syllable by identifying the factors implicated in speakers' string division judgments: nature of the consonants, onset dominance, stress position, vowel quantity, morphological structure, word-edge phonotactics.

A general consistency emerges between phonological processes characterized as syllable-based, asymmetries between CV and VC sequences, and tendencies observed in syllabification judgments. Perceptual and articulatory factors—greater salience and tighter constriction in CV, greater coarticulation in VC, all modulated by the nature of the consonants—could largely contribute to explaining "syllabic" effects such as the increased vulnerability of codas with respect to deletion, neutralization, and assimilation, the tendency for onsets to be of low sonority and codas of high sonority, and the tendency for vowels and codas (especially sonorant ones) to act as a unit. The same factors may also be reflected in basic tendencies observed in syllabification judgments. For instance, the general tendency to syllabify VCV sequences as V.CV rather than VC.V could have a perceptual origin: the consonant receives better cues from the following vowel and may be said to be perceptually more strongly associated with it. Likewise, sonorants are less consistently syllabified as onsets than obstruents in identical contexts (e.g. Fallows 1981; Barry et al. 1999; Zamuner and Ohala 1999; Content et al. 2001; Ishikawa 2002), reflecting articulatory and perceptual distinctions between sonorants and obstruents. One may hypothesize that, in string division tasks, speakers tend to group together segments with relatively stronger perceptual and articulatory dependency, and insert divisions in locations of weaker interaction. This proposal obviously needs to undergo specific testing and further development.

One key question is whether the impact of articulatory, perceptual, or other factors on sound patterns and syllabification judgments is direct or mediated by a level of syllabic constituency. The perceived stability and categorical nature of many phonological processes might be interpreted as supporting an intermediate syllabic level. But the concept of a stable and invariant syllable is at odds with the observed variability in syllabification judgments. In VCV sequences, for example, while consonants are generally syllabified as onsets, they can also be codas or ambisyllabic

segments (e.g. Content et al. 2001 for French; Côté and Kharlamov 2011, for Russian), contrasting with the invariant onset syllabification predicted by phonological theory. Côté and Kharlamov also show that syllabification judgments may differ significantly depending on the task speakers are asked to perform. The gap between categorical phonological processes and variable judgments on string division needs to be addressed and might be interpreted as evidence against a rigidly defined syllabic constituent; alternatively, such experimentally obtained judgments could be argued not to reflect the syllable, at least the phonologically relevant one.

Issues regarding the nature and role of the syllable in sound systems will have to be addressed at multiple levels, but much progress can be expected in the parallel study of three empirical domains: intersegmental relationships, phonological processes, and string division judgments. Recent work on perceptual and articulatory asymmetries in segment sequences has highlighted the role of intersegmental (horizontal) patterns in our understanding of syllabic effects, which complements the more traditional perspective centered on prosodic (vertical) constituency. A key question is to what extent the factors underlying asymmetrical intersegmental relationships—coarticulation, perceptual dependency, and others—are implicated in different phonological processes and in syllabification judgments.

10.3 The temporal implementation of prosodic structure

Alice Turk

10.3.1 Introduction

There is general agreement that speakers organize phonetic material into a hierarchy of prosodic constituents and signal several degrees of relative prominence of syllables and words (cf. Frota, this volume for a fuller discussion). The main functions of prosodic constituency are twofold: (1) to demarcate words and phrases from adjacent material, and (2) to group smaller units into larger ones (e.g. syllables in words, words in phrases, etc.). The main function of prosodic prominence is to highlight particular elements within the speech stream, e.g. to stress syllables within words, and to focus attention on words within phrases. In this section, I outline aspects of what we currently know about the temporal implementation of prosodic structure. I discuss three issues: (1) ways in which speakers use duration to signal prosodic structure, (2) the stretches of speech whose durations are affected

by prosodic structure, and (3) whether and how speakers avoid ambiguity when using duration for prosodic as well as other purposes. The first issue relates to available mechanisms for using duration to highlight and demarcate constituents. The second issue is important from an organizational point of view, because the units that define these stretches must form part of the structures that speakers use during phonetic implementation. The third issue is relevant to modeling the interacting effects of multiple factors on duration, as well as the interaction of duration with other phonetic correlates of prosodic structure.

10.3.2 Temporal strategies for highlighting, grouping, and demarcating

Although a variety of phonetic parameters are used to implement prosodic structure, e.g. f0, intensity, spectral properties, glottalization, segmental sandhi phenomena, etc., duration is one of the primary correlates. Duration is a common correlate of prominence, although its use for signaling word-level stress can vary cross-linguistically (van Heuven and Sluijter 1996).

Durational effects used to implement prosodic structure are traditionally described as lengthenings and shortenings (e.g. initial lengthening, final lengthening, polysyllabic shortening), as if to imply default phrase-medial monosyllabic durations that speakers adjust. An alternative view is that surface durational patterns are the result of the phonetic implementation of abstract, symbolic, phonological representations. Because the representations that relate to speech timing are symbolic, there are no underlying durations to be lengthened or shortened. On this view, output durations of e.g. consonant constriction and vocalic intervals, are the result of a set of interacting factors, i.e. prosodic factors, intrinsic segmental duration factors, rate, etc. (cf. Klatt 1976; van Santen 1992).

Available evidence does not distinguish between these alternatives, but does suggest that speakers commonly use duration to signal prominence and constituent boundaries, in several different ways: (1) prominent elements are often longer than the same elements when non-prominent (variously referred to as stress-related lengthening, prominence-related lengthening, or accentual lengthening); (2) segments that begin constituents are often longer than the same segments when constituent-medial (initial lengthening); (3) segments that end constituents are often longer than the same segments when constituent-medial (final lengthening); and (4) pauses often signal boundaries between constituents. The magnitudes of these effects tend to vary with boundary strength, or level in the prosodic hierarchy (e.g. Grosjean and Collins 1979, Wightman et al. 1992, Keating et al. 2003, and Sugahara and Turk 2009, for subtle effects at a very low level in the hierarchy; Fletcher 2010 for a review). For example, initial- and final-lengthening and pause tend to be greater in magnitude at intonational phrase edges than at the edges of minor

phrases that are intonational phrase-medial, and a syllable that bears phrasal stress as well as lexical stress is likely to be longer than a syllable that bears only lexical stress. These types of effects are very common, and are attested for most, if not all, languages studied. Details of their use can vary cross-linguistically, however, both in terms of magnitudes of observed effects, and in terms of levels in the hierarchy that show the effects (van Heuven and Sluijter 1996; Fletcher 2010).

Other types of mechanisms for signalling constituency have also been proposed. Syllable ratio equalization was proposed by Abercrombie (1965) to account for differences in relative syllable duration within cross-word Abercrombian Foot constituents, depending on morphosyntactic affiliation. For example, in phrases such as *Take Greater London* and *Take Grey to London* spoken in a non-rhotic variety of British English, the syllables /greɪ/ and /tə/ are more equal in *Greater* than they are in *Grey to*, both of which form Abercrombian Feet. Polysegmental shortening (including Closed-Syllable Vowel Shortening: Abercrombie 1967; Jones 1950, cited in Maddieson 1985; Lehiste 1960; Waals 1999) and polysyllabic shortening (Lehiste 1972 for English; Lindblom 1968 for Swedish; Nooteboom 1972 for Dutch) are mechanisms that suggest shorter segments or syllables when more occur in a larger unit. For example, Lehiste (1972) reported that the syllable *stick* in *stickiness* is shorter than the monosyllabic word *stick*. These three proposed mechanisms are conceptually very different from the edge-marking mechanisms described as initial and final lengthening. Syllable-ratio equalization proposes that syllable durations are planned relative to the duration of adjacent syllables within larger constituents, and that durational differences are intended to be proportional. Polysegmental and polysyllabic shortening proposals suggest that speakers plan the duration of elements on the basis of the number of elements within a larger constituent. In contrast, edge-marking proposals such as initial and final lengthening are ambiguous about what speakers might use as a reference duration (e.g. default phrase-medial duration, adjacent syllable duration, etc.). They do not require any reference to the number of elements in a larger constituent, but do require information about the strength of the boundary whose edge is being signaled.

As discussed in Turk and Shattuck-Hufnagel (2000), Sugahara and Turk (2009), White (2002), White and Turk (2010), and van Santen and Shih (2002), it is often difficult to distinguish syllable ratio equalization, polysegmental shortening, and polysyllabic shortening from edge-marking mechanisms, since these mechanisms often make similar predictions. For example, longer *stick* in *stick* as compared to *sticky* may be the result of final lengthening on *stick*, as opposed to polysyllabic shortening or syllable ratio equalization on *sticky*. Nevertheless, polysyllabic shortening may be required to account for longer syllables in disyllabic as compared to longer units, e.g. for longer *-mend* in *commend* vs. *recommend* that White and Turk (2010) observed in phrasally stressed contexts, since *mend* is final in both cases.[1]

[1] Another possibility mentioned in White and Turk (2010) is that observed durational differences on *mend* in *commend* vs. *recommend* in phrasally stressed contexts might be due to the implementation

Unambiguous evidence for syllable ratio equalization is even harder to find, but Turk and Shattuck-Hufnagel (2000) present one possible argument in its favor.

Assuming mechanisms like polysegmental and polysyllabic shortening exist, there is a great deal of uncertainty about the units within which they occur. For polysyllabic shortening, Turk and Shattuck-Hufnagel (2000), White (2002), and Kim (2006) document word-level effects, e.g. *-un-* in *tuna choir* is shorter than *-un* in *tune acquire*. However, durational differences consistent with polysyllabic shortening have also been observed in cross-word, inter-stress intervals containing more syllables, even when the number of syllables in the word is controlled (Kim 2006). It therefore appears that polysyllabic shortening can apply within units larger than the word, but it is unclear how best to characterize these units. Possibilities include content + function word groups, groups of multiple content words, as well as inter-stress (non-word-based, rhythmic) intervals, such as Abercrombian Feet. Shattuck-Hufnagel and Turk's (2009) study of limerick (poetic) speech was designed to distinguish between these possibilities. Their preliminary findings suggested that word-based polysyllabic shortening within words (*bak-* in *baking* shorter than *bak-* in *bake avocados*) and clitic groups (*bake* in *bake us apples* shorter than *bake* in *bake elixirs*) is more common than inter-stress interval polysyllabic shortening, even in poetic contexts, where inter-stress interval rhythmicity would be likely to surface. Only one of the three speakers in Shattuck-Hufnagel and Turk's study showed inter-stress interval polysyllabic shortening (e.g. *bake* in [*bake avo*]*cados*, shorter than *bake* in [*bake*] *apples*, where brackets indicate cross-word foot boundaries.

While edge-marking mechanisms seem to be reliable and nearly always present in studies that look for them, polysyllabic shortening is not always observed. Turk and Shattuck-Hufnagel (2000) and White and Turk (2010) have shown that the effect of polysyllabic shortening can be much reduced and even absent in non-phrasally stressed contexts. For example, White and Turk found no difference in duration between e.g. *mend* in *mend* and *mend* in *commend* or *recommend* when these were not phrasally stressed, and found that the difference between e.g. [mes] in *mace* vs. *mason* and *masonry* was greater when the words were phrasally stressed than when they were not. In addition, Suomi et al. (2008) report the absence of polysyllabic shortening in Finnish, even in prominent contexts. To date, it is unclear to what extent polysegmental shortening occurs in phrasally prominent vs. non-prominent contexts; to my knowledge, studies of polysegmental shortening have not systematically varied phrasal prominence.

With edge-marking mechanisms at their disposal, one might ask why speakers might choose to use mechanisms such as polysegmental and polysyllabic shortening

of accentual lengthening on words of different lengths. On this view accentual lengthening would be spread over three syllables in *recommend*, but only over two in *commend*, yielding a shorter *mend* in *recommend* as compared to *commend*. Note that this proposed mechanism is similar to polysyllabic shortening in that it too requires speakers to take into account the number of syllables within a larger constituent when planning syllable duration.

at all. One possibility might be a drive towards rhythmicity, where the term *rhythmicity* refers to the periodic recurrence of events: shortening elements within units that contain more of them would lead to a tendency towards isochronous units. However, neither polysegmental nor polysyllabic shortening yields isochronous units. For example, Munhall et al. (1992) compared the kinematics of jaw movement in words with coda clusters vs. coda singletons (e.g. *baps* vs. *bap*). Acoustic measures showed that coda clusters were on average 97 ms longer than coda singletons, and vowels in words with coda clusters were on average 13 ms shorter than vowels in words with coda singletons. Compensation for longer codas therefore occurred, but was very weak, and did not result in comparable durations. Similarly, Shattuck-Hufnagel and Turk (2009) found that polysyllabic shortening did not compensate completely an additional syllable within a word. Their preliminary study of three speakers' productions of ten pairs such as *bake apples* and *baking apples* found that although stressed syllable rime durations were on average 27 ms shorter in e.g. *baking* than in *bake*, disyllabic words (e.g. *baking*) were on average 119 ms longer than monosyllabic words (e.g. *bake*). It is possible, and even likely, that polysyllabic shortening is one of the mechanisms that speakers use to achieve rhythmicity when asked to speak to an external rhythmic stimulus, e.g. a metronome (Cummins 2003). However, they fail to fully achieve rhythmicity in non-musical, non-poetic contexts (Uldall 1971). One explanation of this failure is that rhythmicity is an underlying principle of speech production, but that factors other than rhythmicity are more heavily weighted in the speech production process. One of the main factors is likely to be the positioning of articulators to produce segments with appropriate targets for their featural or gestural specifications (see also Saltzman and Byrd 2000, and Saltzman et al. 2008 for a coupled oscillator model in which rhythmicity at multiple levels competes to give surface rhythmic tendencies).

Another possibility is that compression effects are the result of purely non-rhythmic factors. For example, stressed syllable duration might be computed on the basis of the identity and number of segments in the syllable, along with the number of syllables in the word. On this view, rhythmicity would not be a principle that guides the planning process, although signaling the number of syllables in a word would be.

Whether or not rhythmicity is intended, or used as a mechanism that guides speech motor control, it is clear that available speech production mechanisms are used to signal appropriate prominence and grouping relationships, as encoded in prosodic constituent and prominence structure. Edge-marking techniques like initial-lengthening, final-lengthening, and pause insertion are common. In addition, mechanisms like polysegmental and polysyllabic shortening may be used in normal speech production to signal the number of segments in syllables or words, and the number of syllables in words or larger units (White and Turk 2010). Shattuck-Hufnagel and Turk's (2009) preliminary results suggest that even in poetic contexts, where the rhythmicity of inter-stress intervals would be expected, speakers

appear to use their timing skills primarily for linguistic purposes, i.e. signaling the number of syllables within words or clitic groups.

The fact that speakers are more likely to use mechanisms like polysyllabic shortening in phrasally stressed contexts also fits with the view that speakers use durational mechanisms to signal linguistically relevant information. The Smooth Signal Redundancy perspective (Aylett 2000; Aylett and Turk 2004; Turk 2010) proposes that prosodic structure mediates between predictability and the phonetic characteristics of utterances. On this view, prosodic structure has evolved to complement predictability (language redundancy), and directly controls acoustic saliency. Where language redundancy is high and words are highly predictable, there is no need to use prosodic structure to highlight or demarcate words, either via phrasal stress, or by signaling their boundaries. When words are unpredictable, speakers may use phrasal stress as well as word-boundary signaling mechanisms to make words stand out. On this view, speakers would be more likely to use mechanisms like polysyllabic shortening and initial- and final-lengthening to signal word boundaries in phrasally stressed contexts, than in non-phrasally stressed contexts, since it is in phrasally stressed contexts that words are least predictable. This view also provides a potential explanation for the lack of polysyllabic shortening in Finnish. In this language, there may be no need to use polysyllabic shortening to demarcate words, since word edges are predictable from the location of fixed initial word stress.

10.3.3 Which stretches of speech are affected?

The durational correlates of prosodic structure can affect multiple segments and even syllables (e.g. Nooteboom 1972, Sluijter and van Heuven 1995, Turk and White 1999, Heldner and Strangert 2001, Suomi et al. 2003 for prominence-related lengthening; and Klatt 1975, Kohler 1983, Silverman 1990, Wightman et al. 1992, Berkovits 1994, Cambier-Langeveld 2000, Turk and Shattuck-Hufnagel 2007, for final lengthening). On the other hand, initial lengthening appears to be largely restricted to the first segment of a word or phrase (Oller 1973; White 2002; Cho et al. 2002). How are we to define and explain the temporal nature of these effects?

Structural theories such as Klatt (1975) propose the sublexical constituents of the prosodic hierarchy (syllables, onsets, rimes, words) as a set of candidate structures that define the segments affected by durational processes. On such an account, whether a segment is affected by final lengthening or not depends entirely on its structural affiliation. That is, because e.g. /d/ in both *Ida* and *Fridays* have the same structural specification as final syllable onsets, they will be equally likely or unlikely to be affected by phrase-final lengthening.

Within the Articulatory Phonology framework (Browman and Goldstein 1992; Gafos and Goldstein, this chapter), Byrd and Saltzman (2003) propose that final

lengthening can be accomplished via a Pi-gesture (Pi is mnemonic for *Prosodic*) which has a fixed temporal extent as specified in abstract durational units, is anchored at its temporal midpoint to a phrase boundary, and lengthens the segmental gestures with which it overlaps. It is the Pi-gesture which therefore defines the stretch of speech that is lengthened. The amount of lengthening (or stretching) of the overlapped gestures is specified by the height of the Pi-gesture, which at any moment in time slows the ticking of an internal clock by the amount specified by the Pi-gesture height. To derive appropriate surface articulator movement duration, this type of model must stipulate appropriate intrinsic gestural durations, as well as appropriate Pi-gesture heights.

As discussed in Byrd and Saltzman (2003) and Turk and Shattuck-Hufnagel (2007), Pi-gesture theory predicts that only the gestures overlapped by the Pi-gestures will be lengthened, and the number of gestures overlapped by the Pi-gesture will vary depending on their number and intrinsic duration. Therefore, assuming a fixed duration Pi-gesture anchored at its temporal midpoint to a boundary, if there is only one segment following /d/, e.g. the schwa vowel in the word *Ida*, the Pi-gesture is more likely to overlap the penultimate segment (e.g. /d/), than if there are two segments following /d/ as in e.g. *Fridays*. Therefore, [d] in *Ida* is more likely to show final lengthening than [d] in *Fridays*. Figure 10.3.1 illustrates Pi-gesture overlap for final vowels of different intrinsic durations.

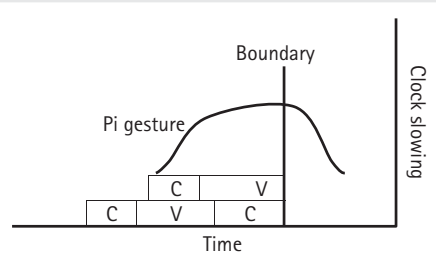

Figure 10.3.1. Schematic representation of a Pi-gesture and the degree to which it overlaps two different types of segmental sequences. The post-onset-C interval in the top CV sequence this short in intrinsic duration, whereas this interval is longer in the lower CVC sequence. As a result, more of the consonant in the top CV sequence will be overlapped by the Pi-gesture and as a consequence will show more final lengthening.

Cambier-Langeveld (1997) makes a third type of proposal in which structural aspects of the final syllable as well as properties of the final segment determine the likelihood of lengthening on preceding segments. She proposes that phrase-finality requires a certain amount of lengthening that occurs preferentially on the phrase-final syllable rime; whether or not final lengthening also affects earlier segments depends on the expandability of that final syllable rime. On her view, lengthening might be more likely on [d] of *Ida* than [d] of *Fridays*, since schwa may have constraints on expandability.

Available studies on this topic are limited. Cambier-Langeveld's (1997) study supports her expandability view as well as the Pi-gesture view: final lengthening on penultimate syllables occurred in her Dutch data only when the final syllable contained schwa. In contrast, results reported for English are consistent with a structural view (e.g. Oller 1973; Klatt 1975; Wightman et al. 1992; Fougeron and Keating 1997; White 2002; Cho et al. 2002; Turk and Shattuck-Hufnagel 2007) in which initial lengthening effects are concentrated on the initial segment in the word, and the greatest magnitude of final lengthening occurs on the final syllable rime. However, materials in these studies could not unambiguously distinguish structural views from Pi-gesture interpretations. Studies of polysyllabic shortening within words (e.g. *mend* longest in *mend*, shorter in *commend*, shortest in *recommend* when phrasally stressed) point to the stressed syllable nucleus as a potential target (White and Turk 2010), consistent with a structural view, but again, the materials were not designed to distinguish the Pi-gesture, expandability, and structural views.

Several studies have shown that final lengthening and prominence-related lengthening processes can target multiple syllables. For example, Turk and Shattuck-Hufnagel (2007) showed that in words like *Madison*, the rimes of both the primary stressed syllable and the final syllable showed significant lengthening, although the largest effects were observed on the final syllable rime. Because lengthening did not affect the middle syllable in words like *Madison*, they concluded that final lengthening did not target a single, multisyllabic stretch, but instead targets multiple, potentially distinct, lengthening sites.

Studies of the temporal extent of prominence suggest that phrasal prominence can also affect multiple syllables within prominent words (Chen 2006 for Chinese; Suomi 2007 for Finnish; Heldner and Strangert 2001 for Swedish; Bouchhioua 2008 for Tunisian Arabic; Nooteboom 1972, Eefting 1991, Sluijter and van Heuven 1995, Sluijter 1995, Cambier-Langeveld and Turk 1999 for Dutch; Sluijter 1995, Turk and Sawusch 1997, Turk and White 1999 for English). Studies of disyllabic words are consistent with the view that phrasal stress targets a single, potentially multisyllabic domain that is constrained by constituent boundaries, e.g. word boundaries (see Turk and Sawusch 1997; Turk and White 1999; and Chen 2006 for details). However, a study of longer, four-syllable words (Turk and Dimitrova 2007) suggests that phrasal-stress-related lengthening in English, like final lengthening, actually targets multiple potentially distinct sites. In English, these sites are roughly

syllable-sized, and can include the secondary-stressed syllable (if there is one), the primary-stressed syllable, and the final syllable. For example, in the word *presidency*, Turk and Dimitrova found that contrastive stress targeted the first and last syllable. In addition, as observed elsewhere, lengthening optionally spilled over from a primary stressed syllable onto a following unstressed syllable (see Chen 2006 for evidence that anticipatory lengthening can also occur on preceding unstressed syllables). Results were consistent with structural definitions of lengthening sites to the extent that observed patterns were not detectably different for different test words. However, it may be that the range of variation in intrinsic segment duration was too limited to provide an adequate test of Pi-gesture-type theory.

Why should final- and prominence-related lengthening affect both word-final and lexically stressed syllables? One possibility for final lengthening is that the lengthening effect on primary stressed syllables preserves the syntagmatic durational relationship among syllables in the final word. That is, lengthening makes the stressed syllable more recognizable as being stressed in the face of the very long final syllable. Syntagmatic motivations for final lengthening patterns have also been observed for Finnish (Nakai et al. 2009). However, a syntagmatic explanation is not available for the fact that prominence-related lengthening affects both the initial and final syllables of e.g. *presidency*. Turk (2010) proposes the following account: Phrasal stress highlights words in two ways: (1) by highlighting its prominent syllable or syllables, and (2) by demarcating it. On this view, lengthening on the final syllable (and possibly also on the initial syllable onset) can be seen as a way of demarcating the phrasally stressed word from adjacent words.

10.3.4 How do speakers avoid ambiguity when duration is used for multiple purposes?

Speakers use duration for many different linguistic and paralinguistic purposes, e.g. signaling phonemic length distinctions, prosodic purposes, and overall rate. How do they do this without causing ambiguity?

One way is to target different structural units for different functions. For example, although phrasal-stress-related lengthening and final lengthening can both affect multiple syllables in polysyllabic words (Turk and Sawusch 1997; Turk and Shattuck-Hufnagel 2007), final lengthening appears primarily to affect syllable rimes, whereas phrasal-stress-related lengthening affects syllable onsets as well as rimes (often with less lengthening on codas). In contrast, initial lengthening appears to target onset segments only. In addition, magnitudes of lengthening can differ for different functional purposes and can be distributed in different ways. In English, final lengthening is greatest on phrase-final syllable rimes (Wightman et al. 1992; Byrd and Saltzman 2003). This lengthening can be much greater than

the lengthening that occurs on phrasally stressed syllables in contrastively stressed words (White and Turk 1999; Turk and Shattuck-Hufnagel 2007).

Languages with phonemic length distinctions (quantity languages) present interesting cases that might result in the ambiguous use of duration for phonemic vs. prosodic purposes. Berinstein (1979) went so far as to hypothesize that the use of duration for segmental contrasts would preclude its use for signaling stress. However, Berinstein's own results and other studies suggest that this functional load hypothesis may be too strong. For example, final lengthening has been observed in several quantity languages: Muskogee Creek, Swedish, Dinka, Estonian, and Hungarian (Lindblom 1968; Krull 1997; Hockey and Fagyal 1999; Johnson and Martin 2001; Remijsen and Gilley 2008; White and Mády 2008), and Finnish, a language with few non-durational correlates of quantity (Nakai et al. 2009). However, both Remijsen and Gilley's (2008) and Nakai et al.'s (2009) results suggest that the proportional magnitude of final lengthening is regulated in a way that preserves the quantity system. The proportional magnitude of final lengthening that Remijsen and Gilley (2008) observed for Dinka, a language with three levels of vowel length, was half or less than that found by Turk and Shattuck-Hufnagel (2007) for English. These findings support a weaker version of the functional load hypothesis, in which the prosodic use of duration is constrained, but not precluded, by its use for segmental contrasts. In addition, it may be the case that the number of quantity distinctions in languages of the world is limited because of the use of duration for prosodic purposes: three-way quantity systems are rare (Remijsen and Gilley 2008).

Yet another way speakers distinguish the different functions of duration is through the use of different articulatory strategies to produce segments of different lengths (Summers 1987; Beckman and Edwards 1992). Speakers can adjust durations by slowing down movements towards and/or away from segmental targets, and/or by lengthening the time articulators spend in target regions. For example, Beckman and Edwards (1992) and Edwards et al. (1991) suggest different strategies for the duration differences associated with rate, lexical stress, phrasal stress, and final lengthening. They found that English lexically unstressed (reduced) vowels (e.g. the final vowel in *poppa* [pɑpə]) were produced with smaller jaw movement distances, and with faster speed/distance ratios than lexically stressed (full) vowels. That is, among other things, speakers adjusted the speed of the movement towards the vowel target. This strategy for distinguishing unstressed vs. stressed vowels was similar to that used for fast vs. slow productions of full vowels. In contrast, phrasal-stress-related lengthening on *pop* was achieved by a different strategy: longer time spent in the target region. Beckman and colleagues found less overlap of the nucleus vowel by a coda consonant (resulting in longer quasi-steady states), greater movement amplitudes, but with no change in the peak speed/distance relationship, suggesting that movement speeds towards and away from these longer vowels were faster than for non-phrasally stressed vowels. Final lengthening of full vowels was

achieved by a combination of speaker- and rate-dependent strategies: a change in the timing of the onset of the final closing gesture and/or a decrease in the peak speed/distance relationship of the final closing gesture. The equivalence of different articulatory strategies in the production of final lengthening suggests the possibility that speakers are aiming for a target interval duration that can be achieved several different ways (Edwards et al. 1991; Ferreira 1993).

And finally, the potential ambiguity of duration is reduced via the use of duration along with sets of other phonetic correlates, where sets of correlates differ according to function. For example, in Finnish, final lengthening often co-occurs with boundary tones, pause, and non-modal phonation; pause and non-modal phonation are not correlates of phonemic vowel length (Suomi et al. 2008). Initial lengthening often co-occurs with initial strengthening (Fougeron and Keating 1997; Cho 2005); prominence-related lengthening often co-occurs with spectral correlates.

10.3.5 Conclusion and discussion

To summarize, available evidence suggests that speakers commonly use duration to signal prominence and constituency. Prominence is signaled by lengthening stressed syllables or (parts of) prominent words. For signaling constituency, edge-marking mechanisms such as pause, initial lengthening, and final lengthening are widely attested, but mechanisms for signaling the number of segments or syllables in larger units are also available. In implementing final lengthening and prominence-related lengthening, speakers often target multiple segments and even syllables, but it is still unclear which units best predict the stretch of segments whose durations are affected. In spite of the fact that speakers use duration for multiple prosodic, segmental, and other purposes (e.g. overall rate of speech), they are able to distinguish separate functions by controlling the magnitudes of effects, the stretches of speech that are targeted, as well as the other phonetic parameters with which durational effects co-occur. It is still unclear how speech movements are controlled to produce these effects.

The structures of Articulatory Phonology, task dynamic theory, and their recent developments (Browman and Goldstein 1992; Saltzman and Byrd 2000; Byrd and Saltzman 2003; Saltzman et al. 2008; Gafos and Goldstein, this chapter) provide one set of candidate structures and mechanisms that account for durational variability associated with prosodic organization. This type of theory is attractive because it provides ways of modeling lengthenings and shortenings associated with boundaries and prominences, as well as interactions among levels in the boundary and prominence hierarchies. In addition, it can model things not discussed here, e.g. interactions with global speech rate, and intergestural timing.

More traditional, symbolic theories of phonology also account for some aspects of variation, e.g. the positive correlation of hierarchy-derived constituent boundary

strength with duration, and provide a set of candidate structures that define the stretches of speech whose durations are affected by context. However, although Guenther (1995) and Bullock and Grossberg (1988) provide possible motor control mechanisms that could implement a symbolic theory, these theories do not provide mechanisms to predict actual segment durations in different contexts.

To be able to fully predict segment durations in the full range of linguistic and paralinguistic contexts, both types of theory will need to meet several challenges. One challenge is to provide a principled account for intrinsic segmental duration differences (see Fitts 1954; Schmidt et al. 1979; Bullock and Grossberg 1988; Harris and Wolpert 1998; Saltzman et al. 2000). At the same time, these models must account for the interactions of segmental factors with other factors such as prosody and rate. More generally, there is still much to be learned about how interacting prosodic, segmental, and other factors are controlled to produce resulting surface durations.

CHAPTER 11

PROSODIC REPRESENTATIONS

PROSODIC STRUCTURE, CONSTITUENTS, AND THEIR IMPLEMENTATION

SÓNIA FROTA

SEGMENT-TO-TONE ASSOCIATION

AMALIA ARVANITI

TONAL ALIGNMENT

MARIAPAOLA D'IMPERIO

Contributions in this chapter address the nature of prosodic representations and their implementation as illuminated through experimental work. Frota addresses the nature of prosodic structure(s) and the implementation of prosodic constituents. Arvaniti deals with segment-to-tone association, that is how intonational elements relate to the segmental string. D'Imperio focuses more specifically on tonal target alignment, addressing the question of its stability and its variability.

11.1 Prosodic structure, constituents and their implementation*

Sónia Frota

11.1.1 Introduction

After three decades of literature focusing on prosodic structure, the view that prosodic structure has a role to play as the organizing framework of speech is well established. This structure consists of the grouping of chunks of speech into prosodic constituents arranged according to a hierarchy, delimited by prosodic boundaries or edges and with prominences or heads at the various levels. Prominence strength and boundary strength reflect the hierarchy. Prosodic domains are marked by constellations of cues, which stand as the major empirical evidence for prosodic structure and the constituents it comprises. These cues have been shown to be used in lexical processing, in the disambiguation of syntax, or in the identification of morphosyntactic units (as in bootstrapping).

Laboratory phonology approaches to the study of speech have been instrumental in the discovery and discussion of cues to prosodic structure, in the shaping of the essential questions that need to be accounted for and of the challenges for future research which will sharpen and deepen our understanding of prosodic constituency across languages. In this section we provide an overview of the contribution of work in laboratory phonology to the present knowledge of prosodic structure, prosodic constituents, and their implementation. Illustration of the already considerable amount of research is provided on the basis of selected examples. In sections 11.1.2 and 11.1.3 we deal with two essential questions on the nature of prosodic structure, namely whether there are different kinds of such structures or instead a unique representation, and how a prosodic constituent is defined and whether levels of constituency and of phrasing are equivalent. In section 11.1.4 we examine the implementation of prosodic structure across languages. In section 11.1.5 we highlight recent developments and explorations in research on prosodic structure.

* I am grateful to Amalia Arvaniti, Pilar Prieto, Marina Vigário, and to the editors Abigail Cohn, Cécile Fougeron, and Marie Huffman, as well as an anonymous reviewer, for their comments and suggestions. Preparation of this contribution was partially funded by grant PTDC/LIN/66202/2006, FCT-Portugal.

11.1.2 On the nature of prosodic structure: Is one enough?

Different views of prosodic structure have been proposed both in the general literature on prosody and in the laboratory phonology literature. Research on phonological rules has successfully shown that morphosyntactic structure influences prosodic structure, so that syntactic constraints, together with phonological constraints, yield the constituent structure that accounts for contextual segmental rules (Selkirk 1984, 1986, 2000, 2005; Nespor and Vogel 1986; Truckenbrodt 1999; among others). Parallel to rule-based structure, an intonation-based structure has been posited to describe the intonation of several languages (Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988; Jun 2005a, among others). A prominence-based structure has also been proposed, where levels of constituency correspond to levels of prominence seen as stress and/or accent manifestations with no direct relation to "other" prosodic constituency (e.g. Beckman and Edwards 1990, 1994). The different views cannot be reduced to grid versus tree-based models of representation of prosodic structure. Instead, they seem to emerge from independent research traditions and/or angles of approaching prosodic structure: phonological rules, intonation, and prominence. Some researchers have assumed an integrated view and set out to empirically test the hypothesis that phrasal rules, intonation, and prominence phenomena all refer to the same structure of prosodic constituents (Hayes and Lahiri 1991; Frota 2000; Hellmuth 2007). In this section we briefly review and compare these approaches.

The hierarchies given in Table 11.1.1 capture the main aspects of the different approaches.[1] They all share three basic observations about prosodic structure: prosodic constituency is non-isomorphic to morphosyntactic constituency and thus is properly phonological; prosodic constituents are metrical constituents of some sort that are hierarchically structured; the limits of higher constituents are also the limits of lower-level constituents.

However, the approaches differ in that they propose different principles of prosodic organization for the structure above the word level and, to some extent, different types of constituents. For (a) in Table 11.1.1, prosodic constituency partially results from the interface of phonology with other components of grammar, and thus it bears some systematic relation to morphosyntax. For example, Phonological Phrases (PhPs) relate to syntactic phrases (XPs) and Intonational Phrases (IPs) to syntactic clauses, but crucially it is not the case that all syntactic boundaries of a certain type must correspond to prosodic boundaries of a given type and vice versa (Nespor and Vogel 1986; Selkirk 1986, 2000; Truckenbrodt 1999). The principles of

[1] The prosodic hierarchies in Table 11.1.1 are organized for comparison purposes and only the most well-established and/or discussed levels are included. The relevance of a given constituent may depend on the language, as is the case in (b) for the mora in Japanese, and the foot in English (see Pierrehumbert and Beckman 1988; Grice 1995a). Structure at the word level and above is the focus of interest in this section.

Table 11.1.1. Different prosodic hierarchies proposed in the literature

a. Rule-based	b. Intonation-based	c. Prominence-based
Intonational Phrase (IP)	IP	Nuclear accent
Phonological / Major Phrase	Intermediate Phrase	
Clitic Group/Minor Phrase/Prosodic Word Group	Accentual Phrase	Accent
Prosodic Word (PW)	PW	Stress
Foot	Foot	Full vowel
Syllable	Syllable	Syllable
Mora	Mora	

syntax-to-phonology mapping are themselves partially responsible for this non-isomorphism, by promoting the alignment of a specific pair of syntactic/prosodic constituent edges (left or right) or enforcing the inclusion of a morphosyntactic phrase within a prosodic phrase. The combination of syntactic constraints on prosodic structure with well-formedness conditions on the size and eurhythmicity of prosodic constituents is a further factor behind the autonomy of prosodic structure (e.g. Ghini 1993; Selkirk 2000; Prieto 2005; Elordieta et al. 2005; Frota and Vigário 2007). For approaches (b) and (c), by contrast, prosodic structure is intonation-defined or prominence-defined in the sense that prosodic constituents are posited (and labeled) with reference to the phenomena that characterize them (rather than to the morphosyntactic constituents they relate to). Thus, for example, the presence of a nuclear accent and a boundary tone defines the IP, whereas the presence of a nuclear accent and a phrase accent defines the Intermediate Phrase (Beckman and Pierrehumbert 1986, and much subsequent work). In these approaches the highly variable character of prosodic structure is usually highlighted as a consequence of factors such as speaking rate, style, discourse structure, or rhythm.

Despite the clear differences in the underlying principles and definitions of the prosodic constituents, a closer inspection of the hierarchies in Table 11.1.1 reveals some striking similarities. In all cases, syllables, prosodic words (PWs), and IPs are constituents of the prosodic structure. Also in all cases, there seems to be variation in the number and/or type of constituents between the PW and the IP. Sources for this variation have been suggested in the literature: it may well be the case that certain levels of structure are language-specific (e.g. Selkirk 1990); it may also be the case that some of the variation is a side effect of the specific approach to prosodic structure, and that, for example, PhPs and Intermediate Phrases represent the same type of constituent (an early suggestion in this direction can be found in Pierrehumbert and Beckman 1988), in the same way as the Clitic Group, the

Minor Phrase, the Accentual Phrase, and more recently the Prosodic Word Group (PWG) can be seen as essentially equivalent (Selkirk et al. 2004; Vigário 2009). The question thus arises whether the hierarchies in Table 11.1.1 are not fundamentally versions of the same prosodic structure. This question has been empirically addressed in work where morphosyntactic and phonological constraints on prosodic structure are assessed by more than one of the possible correlates of prosodic constituency, namely phonological rules together with intonational phenomena (and sometimes also relative prominence, rhythmic and lengthening phenomena). This research is crucially laboratory-based, in that it requires the development of experimental paradigms that control and provide a test for the properties of prosodic structure. Illustrative examples of this line of research are described below.

In what is probably the first systematic empirical test of an integrated view of prosodic structure, Hayes and Lahiri (1991) have shown that in Bengali the distribution of boundary tones and the application of segmental phrasal rules of assimilation refer to the same prosodic hierarchy, which is defined on the basis of syntax-to-phonology mapping principles plus phonological constraints. In Frota (2000) a set of production experiments was designed to examine whether there is a match in European Portuguese among the phonological structures required to account for phrasal rules, the domains of rhythmic phenomena such as stress clashes, the distribution of intonational events, and the facts of boundary-related lengthening (for production studies in the analysis of prosody, see also Prieto, this volume). The findings show convergent results pointing to the same prosodic structure established on the basis of both syntactic and phonological conditions (namely on the size of prosodic phrases). In detailed production studies of prosodic phrasing in Egyptian Arabic, Hellmuth (2004, 2007) has inspected a range of post-lexical tonal phenomena, including pitch accent distribution, together with a syllable repair rule of epenthesis, and pause distribution. The phrasing patterns that emerge are, again, consistent with a prosodic structure established by the interaction between syntax-phonology interface principles and phonological well-formedness conditions on the size of prosodic constituents.

In a related line of research, experimental procedures have been used to examine in a systematic fashion the import of the different syntactic and phonological factors affecting prosodic phrasing, such as alignment to syntactic edges, syntactic complexity, prosodic complexity, or phonological length (e.g. Jun 2003 for Korean; D'Imperio et al. 2005, Elordieta et al. 2005, and Prieto 2005 for several Romance languages; Shaked 2007 for Hebrew). The findings have shown an important role for language specificity in the relative weight of those factors, at the same time as they strengthen the view of prosodic structure as the result of the combined action of syntactic and phonological conditions.

Table 11.1.2. Levels of constituency and levels of phrasing in prosodic representations

a.	b.	c.
( ) A	( ) A	( ) A
( ) B	( ) B	( ) ( ) A
( ) ( ) C	( ) C	( ) ( ) B
( ) ( ) D	( ) C	( ) ( ) C
( ) ( ) ( ) E	( ) D	( ) ( ) ( ) D
	( ) ( ) ( ) E	( ) ( ) ( ) E

In short, the question of whether there are different kinds of prosodic structure organized on independent principles or whether one prosodic structure is enough is ultimately an empirical one.

11.1.3 Defining prosodic constituents and levels of phrasing

In most approaches to prosodic structure, whether rule-based or intonation/prominence-based, this structure is considered to be fundamentally different from morphosyntactic structure in that it is crucially flatter. This observation has been embodied in the Strict Layer Hypothesis, which determines a fixed, layered organization of prosodic structure that contrasts with the indefinite depth of syntax (Selkirk 1984; Nespor and Vogel 1986). Under this view, prosodic structure consists of a fixed number of possible constituents and thus levels of constituency strictly correspond to levels of phrasing, as in Table 11.1.2 (a).[2] However, work on prosodic phrasing in various languages has led to the relaxation of this strong view. For example, accounts of the prosodification of clitics have shown that it is not necessarily the case that a given level of the hierarchy consists exclusively of constituents of the next lower level, and proposals of recursive prosodic words and recursive prosodic phrases have also been put forward (Selkirk 1996; Booij 1996; Peperkamp 1997; Vigário 2003; Gussenhoven 2004; *inter alia*). Thus, prosodic representations as in Table 11.1.2 (b) have been argued for. These structures raise an important question about the depth of prosodic structure and its essential difference from syntax. The proposal of compound prosodic structure (Ladd 1996, 2008; Frota 2000) addresses this question by constraining recursiveness to compound structures, as in Table 11.1.2 (c).

Unlike in Table 11.1.2 (a), in the structures in (b–c) the levels of prosodic constituency do not necessarily correspond to the levels of phrasing, and thus the

[2] The labels A, B, to E represent constituent types, where A is higher in the hierarchy than B, and B than C, and C than E.

question arises as to how a given prosodic constituent is defined, both within and across languages. Experimental approaches to prosodic structure have been instrumental in providing evidence for prosodic constituents and levels of phrasing. Indeed, they have been crucial to evaluate the empirical basis of proposals such as (a) and have motivated new proposals like (b) or (c). However, the issue of how levels of constituency and levels of phrasing are defined is clearly not settled yet, as shown by the research cases described below.[3]

Jun's experimental work on prosodic phrasing in Korean, simultaneously based on the tonal patterns and the application of phonological rules, established two levels of constituency above the PW, namely the accentual phrase and the intonational phrase (Jun 1996, 2005b). The former is characterized by the underlying tonal pattern THLH (where T is either H or L depending on the laryngeal properties of the phrase-initial segment), and is usually signaled by a phrase-final LH pattern, and is also the domain of three different phrasal rules; the latter is signaled by a final boundary tone, and by limiting the application of two other phonological rules. In Jun (2007), however, experiments on prosodic phrasing and relative clause attachment prompted a revision of Korean prosodic structure. An additional constituent is proposed, the intermediate phrase, on the basis of juncture strength differences: a stronger accentual phrase boundary, phonetically denoted by a higher tone and/or by a following higher pitch range, is interpreted as the boundary of a different and higher constituent. Most strikingly, the phonetic and phonological definitions of the accentual phrase and the IP are clearly independent of each other, whereas the definition of the intermediate phrase seems to be dependent on the properties of the accentual phrase, of which it only provides a stronger version.

Work on the prosodic phrasing of European Portuguese (EP) has provided similar data to that reported for Korean. EP has been shown to have a phonological phrase level and an intonational phrase level. The PhP plays an important role in the account of rhythmic and prominence-related phenomena: for example, the clash between two adjacent stressed syllables is solved by lengthening of the first of these syllables if both of them belong to the same PhP, but not across a PhP boundary; and the deletion of a word-final vowel when followed by a word-initial vowel is blocked if the second word involved is the head of a PhP, but not otherwise. The intonational phrase level in EP is the domain of many phonological processes, the domain of final lengthening, as well as of the minimal tune (only the IP head must be pitch-accented and only the right edge of the IP requires tonal boundary marking in the language—Frota 2000, forthcoming). When an IP is short, however, it is found to group with an adjacent IP. This grouping is signaled by weaker

[3] See also Ladd (1996/2008) for an extended discussion on the empirical adequacy of proposals of phonological structure. Other work directly bearing on the issue is D'Imperio and Gili Fivela (2003) for levels of phrasing above the word in Italian; Hellmuth (2004) and Chahal and Hellmuth (forthcoming) for the discussion of the presence/absence of a minor phrase in Egytian Arabic; and Arvaniti and Baltazani (2005), Baltazani (2006b), and Kainada (2009) for levels of prosodic phrasing in Greek.

boundaries of the short IP expressed by less final lengthening and a narrower boundary rise, at the same time as the domain span rules may apply across the weaker boundary. These facts are interpreted as pointing to a recursive intonational phrase compound structure, where the difference between the inner and outer edges of the compound phrase is merely a gradient one and rule application across the inner edge simply follows from the span character of the rules within the IP.

In recent work on the prosodic phrasing of word-like structures in various languages (especially compounds), a prosodic constituent different from the PW and the PhP has been proposed, rather than a recursive PW structure (Kabak and Revithiadou 2006; Vigário 2009). The key argument for the PWG (as Vigário 2009 calls it) is that it functions as a domain for phonological processes distinct from those that apply with reference to PWs or PhPs.

The syntactic grounding of the prosodic hierarchy view proposes that prosodic constituent types relate to morphosyntactic constituents, and experimental work on prosodic phrasing has provided ample phonetic and phonological evidence for levels of constituency and/or levels of phrasing. Taken together, they offer an important empirical insight to be explored in further research: a prosodic constituent involves some kind of morphosyntax-to-prosody mapping and an array of phonological properties, including size and prominence, acting as the domain for phonological and phonetic phenomena (segmental, tonal, temporal), and cues to boundary marking; the morphosyntactic constituent it relates to and at least a subset of the phonetic and phonological properties it shows should be different in type from those defining the other prosodic constituents. In contrast, recursion and compounding refer to forms of grouping of instances of a given prosodic category, yielding levels of phrasing that are reflected only by gradient differences in the strength of the same phonetic properties.

11.1.4 Cues to prosodic structure across languages and language varieties

The detailed study of the implementation of prosodic structure across languages has been perhaps the most fruitful research program within laboratory approaches to prosodic structure. In this section we review the types of cues that have been reported and present illustrative examples of cue variation and language specificity of phonetic cues.

11.1.4.1 *Types of cues*

Phrasal phonological processes in many languages have been among the cues to prosodic phrasing since the early proposals in prosodic phonology. These have included assimilations, lenitions, fortitions, deletions, insertions, and so on (see Nespor and Vogel 1986, Selkirk 1986, Jun 1996, Frota 2000, Hellmuth 2004, Baltazani

2006b for exemplification in various languages). Whether these processes are best described as categorical or gradient changes, they have been instrumental in the signaling of prosodic structure across languages (see Ladd and Scobbie 2003 and Zsiga 1995 for detailed analysis and arguments in either direction; see also Ernestus, this volume and Coetzee, this volume).

Another area where the realization of segments and tones has been shown to be affected by the implementation of prosodic structure is constituent-initial strengthening, a set of phenomena to the study of which laboratory phonology approaches have strongly contributed (see also Ernestus, this volume). Both acoustic and articulatory studies have shown that initial strengthening is highly correlated with constituency (or phrasal) level, although all the levels posited in the various studies are not necessarily distinguished either within or across languages (Pierrehumbert and Talkin 1992; Jun 1995, 1996; Byrd et al. 2000; Keating et al. 2003; Pan 2007). Similarly, final lengthening has been investigated as a result of the phonetic implementation of prosodic representations. Detailed acoustic and articulatory studies using controlled laboratory materials have examined the presence of final lengthening, its correlation with prosodic boundary level, and its temporal scope and distribution (Beckman and Edwards 1990; Beckman, Edwards, and Fletcher 1992; Wightman et al. 1992; Cambier-Langeveld 2000; Frota 2000; Turk and Shattuck-Hufnagel 2000, 2007; Byrd et al. 2006; see also Turk, this volume). While the presence of final lengthening at the IP level seems well established in many languages, empirical findings are less clear with regard to lower phrasal levels and the PW, both within and across languages.

Rhythmic phenomena like stress clash resolution strategies are also sensitive to prosodic structure. The phonology and phonetics of these phenomena, especially the rhythm rule, a phenomenon whereby the major prominence within a word is moved to an early vowel when the stress of the following word is adjacent, has been studied in detail. There are two main accounts for the rhythm rule: the stress-shift account, according to which the main stress moves leftwards to avoid the stress clash, and the early accent account, that sees the change in prominence as a reflection of early pitch accent placement within the word (e.g. Nespor and Vogel 1989; Horne 1990; Grabe and Warren 1995; Vogel et al. 1995; Shattuck-Hufnagel 1995, 2000). However, in both accounts there is agreement that the rhythm rule applies within a prosodic domain, is blocked across a phrase boundary, and is constrained by prosodic conditions related to the rhythmic organization of prosodic word- and phrase-level prominences.

The realization of tonal targets has been shown to rely on the implementation of prosodic structure, as is illustrated by work on pitch scaling and final lowering phenomena (see Ladd 1996, 2008). For example, empirical studies of scaling in German show that different prosodic constituents define different phonetic reference lines that establish the relative height of a tone (Truckenbrodt 2002, 2007a). In the same vein, experimental evidence suggests that the lowering of the final peak in

a series of peaks is the phonetic manifestation of a grammaticalized pitch range relation determined by prosodic constituency, at least in some languages (Arvaniti and Godjevac 2003; Arvaniti 2007c).

Distributional properties such as those established by edge tones, pitch accents, patterns of occurrence of pre-nuclear and nuclear accents or of nuclear accents and edge tones and their relative frequencies, have been shown to reflect prosodic structure. Specific patterns of combinations of pitch accents may be informative of their pre-nuclear/nuclear position in a prosodic phrase (Dainora 2006). Languages may exhibit a dense or sparse distribution of pitch accents, depending on the level of prosodic constituency that serves as the domain for accentuation. The lower the prosodic domain relevant for pitch accent distribution, the more dense pitch accentuation is; the reverse obtains if a higher-level domain regulates pitch accent distribution. Illustrative examples are: Egyptian Arabic, with the PW as the relevant domain and a dense distribution of pitch accents (i.e. every PW is accented); Standard European Portuguese, with the IP as the domain for accentuation and thus a sparse pitch accent distribution (i.e. only IP heads must be accented); and Northern European Portuguese, with a lower phrase as the relevant domain and therefore a richer distribution of pitch accents than in the standard variety (Vigário and Frota 2003; Hellmuth 2007; Frota forthcoming).

11.1.4.2 *Variation of cues*

The cues mentioned above may show variation across languages in their presence/absence, in the level of constituency or phrasing they signal, or in the specific ways they are implemented. We illustrate this variation with three examples.

In a comparison of cues to phrasing across Bantu languages, Zerbian (2007) shows that similar patterns of phrasing are found across some languages but with considerable diversity in the phonetic cues that implement them: different cues can be used to signal the same level of phrasing, like blocking of high tones at phrase boundaries in Northern Sotho, but deletion of high tones within the same phrase in Kinyambo; and the same cues can indicate boundaries of different levels, as in the case of penultimate lengthening (i.e. the lengthening of vowels in the penultimate syllable of a prosodic domain) that signals the PhP in Chichewa, but marks the IP in Northern Sotho.

In their comparative study of intonational phrasing in Romance, Frota et al. (2007) show that while a high boundary tone is the main cue across languages, both nuclear pitch accent choices and the detailed phonetics of intonational boundaries vary in consistent ways and group the languages in two sets: the Catalan-Spanish group and the Italian-European Portuguese group. In the former, rising accents are the dominant choice, and the scaling of the boundary tone is correlated with the scaling of the first peak in the phrase, while there is no impact of phrase length on the height of the tonal boundary. In the latter, by contrast, both rising and falling

accents are common, and the length of the phrase and not the height of the first peak, crucially affects the scaling of the boundary tone.

Final lengthening (especially at the IP level) has been shown to be present in many languages, but references to its absence are also found (as in Chimwiini, Estonian, Finnish). Myers and Hansen (2007), based on the results of a series of production and perception experiments, have shown the presence of both final lengthening and final devoicing in Finnish, and argue that final devoiced vowels tend to be identified as short vowels by native speakers. Employing highly controlled laboratory materials, Nakai et al. (2009) have also shown that a quantity language like Finnish exhibits final lengthening, but its implementation is regulated to preserve the language-specific quantity system, namely the contrast between single or short vowels and double or long vowels. This important empirical finding raises the question whether final lengthening, and perhaps also other prosodic cues, is a universal cue to phrasing that is implemented in language-particular ways. If so, cue variation may be the result of the conspiracy of specific phonologies against universal tendencies in language, and experimental approaches are crucial for disentangling the two factors.

11.1.5 Recent developments and explorations

There are at least three areas in which experimental research in prosodic structure is developing rapidly and holds the promise to provide new insights into the nature of prosodic phrasing and its implementation: sign languages, language processing, and language acquisition.

Work on the prosody of sign languages has shown a similar chunking into prosodic constituents, which are signaled by sets of cues, as in spoken languages (e.g. Sandler 2006; Sandler and Lillo-Martin 2006). Although very different articulators are used, sign languages also exhibit sandhi rules (like the spreading of the non-dominant hand) and intonation (facial expressions), as markers of prosodic constituency.

Using both behavioral methods and ERP measures (cf. Prieto, this volume), the investigation of the processing of prosodic structure has shown that adult listeners are sensitive to different levels of constituency, and that prosodic boundaries play an important role in lexical access and syntactic disambiguation (Christophe et al. 2004; Millotte et al. 2007; Li and Yang 2009; Frota et al. 2009). Strikingly, lexical processing is not only affected by local boundary cues, but also by distant prosodic properties such as pitch and rhythm patterns (Dilley and McAuley 2008). Infant listeners seem to show a similar sensitivity to prosodic structure, and they are able to use it both for lexical segmentation and syntactic analysis (Gout et al. 2004; Homae et al. 2007; Christophe et al. 2008).

Clearly, the task for the future is twofold: the cross-linguistic exploration of these recent lines of research, and the development of experimental studies that integrate the simultaneous assessment of the multiple cues to prosodic structure (i.e. intonation, boundary strength, prominence, rhythm, and the realization of segments and lexical tones in connected speech). Together, these two movements promise to significantly push the frontiers of our understanding of prosodic structure in language(s).

11.2 Segment-to-tone association

Amalia Arvaniti

11.2.1 Introduction

Intonation is the linguistically structured and pragmatically meaningful modulation of pitch. A key theoretical question surrounding intonation is that of the formal ways in which intonational elements are said to relate to the segmental string. In order to be able to provide a satisfactory answer to this question, it is important to know what the structure of intonation might be. It is of course possible to conclude that such a structure (in the sense of a combination of discrete elements) does not exist: for instance, in their model OXIGEN, Grabe, Kochanski, and Coleman (2007) treat tunes as gestalts, as did some earlier researchers, such as Cooper and Sorensen (1981).

This kind of approach to intonational structure is problematic for two reasons. First, if melodies are gestalts, their meaning should be unique and relatively constant across utterances. However, it has been repeatedly noted that the same melody can be used with different meanings (for discussions and relevant data, see Pike 1945; Lehiste 1970: 95 ff. and references therein; Baltazani 2006a; Arvaniti 2007a; Ladd 2008: ch. 1). For this reason, viewing melodies as composites of smaller and meaningful elements is more likely to be successful in accounting for intonational meaning (cf. Pierrehumbert and Hirschberg 1990).

In addition, seeing melodies as gestalts is problematic from the point of view of form, since melodies do not simply shrink or stretch to fit the duration of the utterance with which they co-occur. Rather, parts of the melody appear to coordinate independently with parts of the segmental string (Arvaniti 2007a, b; Ladd 2008: ch. 2). Recent experimental evidence for this view is provided in Arvaniti, Ladd, and Mennen (2006a), and Arvaniti and Ladd (2009), who examine the intonation of Greek polar and wh-questions respectively, and show that the shape of pitch

contours is radically affected by the position of stressed syllables, the location of the focal element, and the length of the utterance. These results cannot be accounted for if contour shape is seen as a primitive, but are compatible with the idea of discrete tones that can vary in their phonological association, and consequently in their phonetic realization, depending on the parameters mentioned above.

It is thus clear that melodies must be composed of some kind of primitives, the nature of which has preoccupied many researchers. Answers have varied extensively, from level tones (e.g. Pike 1945; Trager and Smith 1951), to configurations such as rises and falls (e.g. the IPO model, presented in detail in 't Hart et al. 1990), to elements that can span f0 stretches of arbitrary length, such as the *head*, *pre-head*, *nucleus*, and *tail* of the British school of intonation (e.g. O'Connor and Arnold 1973; Halliday 1967, 1970). Choosing the right answer, however, is neither trivial nor a matter of taste, as the answer has important (and often empirically testable) consequences for our understanding of the relationship between the text and the tune. As a discussion of intonational primitives is beyond the scope of this section, in the remainder I proceed under the working assumption that these primitives are static tones and combinations thereof (for supporting experimental evidence see Pierrehumbert and Beckman 1988, and Arvaniti et al. 1998; for a review of the issue of tonal primitives, see Arvaniti 2011).

The remainder of the section reviews several crucial issues in the study of intonation. Section 11.2.2 briefly reviews the experimental results that led to the development of the *autosegmental-metrical model of intonational phonology* (henceforth AM), the main principles of which are presented in Section 11.2.3. Finally, Section 11.2.4 reviews experimental research that addresses several issues arising from the central tenet of AM that intonation involves the phonological *association* of tones with constituents of the prosodic hierarchy (phonological association gives rise to phonetic *alignment*, the temporal coordination between segments and tones, reviewed in D'Imperio, this chapter). Section 11.2.5 briefly concludes.

11.2.2 The relationship between text and tune: Early empirical evidence

The relationship between text and tune did not feature prominently in intonational models until researchers started examining acoustic data and the close connection between text and tune was uncovered. For example, researchers like 't Hart et al. (1990), in the IPO tradition, noted that some of the f0 movements that constitute the primitives in their model showed tight temporal coordination with stressed syllables (*prominence-lending* pitch movements), while others tended to spread over several syllables and did not co-occur with either prominent words or stressed syllables (*non-prominence-lending* pitch movements).

A significant influence on our current understanding of the relationship between text and tune has been Bruce's study of the Swedish lexical pitch accents, known as Accent I and Accent II (Bruce 1977). Bruce showed that these accentual patterns can be successfully accounted for if both accents are seen as falls that differ in terms of their timing with respect to the accented syllable (see also D'Imperio, this chapter, Section 11.3.1.1 and Figure 11.3.1). By doing so, Bruce essentially distinguished the *phonological* connection between tonal elements and segmental structure from the coordination between the two in real time.[4]

Following Ladd (1983), these two properties have become known as *association* and *alignment* respectively. Simplifying somewhat, we could say that in Swedish both Accent I and Accent II associate with the same syllable but align differently with it (for an analysis along these lines, see Bruce 1987).

In addition, Bruce showed that the second pitch peak seen in words with Accent II was not part of the pitch accent per se but the reflex of phrasal tones. Thus, his study highlighted the distinction between tonal elements that co-occur with prominent syllables and tonal elements that co-occur with phrasal boundaries. It further showed that despite their different origin in the grammar, phrasal tones and lexical tones are part of the same representation and realized in a similar fashion, rather than forming distinct layers of tonal structure.

11.2.3 The autosegmental-metrical model of intonational phonology

Although the research discussed above provided empirical evidence for the relationship between text and tune, models like that of the IPO or Bruce did not formalize their findings in terms of phonological representations. Such formalizations appeared in Leben (1973), Goldsmith (1976), and Liberman (1978). The combined insight of these early approaches culminated in Pierrehumbert's thesis (Pierrehumbert 1980), which gave rise to the AM model. Pierrehumbert's dissertation is an early example of laboratory phonology as it combines a formal phonological analysis with instrumental and quantitative data.

Pierrehumbert proposed that English tunes are composed of high (H) and low (L) tones which are linearly ordered on an autosegmental tier and associated to strong nodes and edges of metrical trees. Thus, similarly to Bruce's model (and unlike early autosegmental accounts of English intonation), these H and L tones do not exhaustively represent the course of f0. Phonetically, the reflexes of L and H tones are *tonal targets* (typically, though not necessarily, f0 minima and maxima respectively), with the pitch between them being generated by interpolation. Thus,

[4] Strictly speaking, Bruce saw association as a property of the systematic phonetic level, not of phonology, which does not feature in his model.

both at the phonological and at the phonetic level melodies are underspecified. A corollary of underspecification—in combination with the association properties of tones discussed in more detail immediately below—is that the number of tones and the number of *tone-bearing units* (henceforth TBUs) need not match: in some instances, several TBUs may not be associated to tones, while in others, several tones may associate with the same TBU, giving rise to *tonal crowding* (Bruce 1977).

Experimental evidence for underspecification was first provided in Pierrehumbert and Beckman (1988) who showed that the f0 of Tokyo Japanese accentual phrases with unaccented words can be captured by positing just a H phrasal tone co-occurring with the second mora of the phrase and a L% boundary tone co-occurring with the right phrasal boundary. In their data these two landmarks were separated by increasingly more segmental material, leading to an increasingly shallower f0 slope between H and L%. This change of slope cannot be explained if all the syllables between the H and L% are specified for tone, as in earlier models of Japanese, but is compatible with the idea that the melody is underspecified both phonologically and phonetically. Similarly, a host of laboratory studies have demonstrated that tonal crowding results in controlled variability in the realization of tones, including tone truncation (Bruce 1977; Grice 1995a; Arvaniti 1998; Grabe 1998; Grabe et al. 2000; Arvaniti and Ladd 2009), tonal undershoot (Bruce 1977; Prieto 1998; Arvaniti et al. 2000, 2006a,b; Arvaniti and Ladd 2009), and the temporal realignment of tones (Silverman and Pierrehumbert 1990; Arvaniti and Ladd 2009). None of these effects should be observed if melodies were gestalts or if all syllables were tonally specified, as such views of intonation predict uniform expansion and compression of their primitives.

As noted, Pierrehumbert's system also incorporated the distinction between tones that associate with prominent syllables, that is strong nodes in the metrical tree, and tones that associate with utterance edges. The former, known as *pitch accents*, are notated with an asterisk, e.g. H*; the latter, known as *boundary tones*, are notated with a percent, e.g. H%. Pierrehumbert also noted that English tunes included a stretch between the last pitch accent (or *nuclear accent*) and the following boundary tone, where f0 was clearly not a simple interpolation between these two tones. She analyzed these stretches as the reflex of *phrase accents*, unassociated (floating) tones with variable realization. Pierrehumbert also proposed that in bitonal accents,[5] such as L*+H, the starred tone is metrically strong and phonologically associated to the accented syllable, while the unstarred tone is a floating tone that precedes or follows the starred tone by a fixed amount of time.[6] In

[5] Bitonal accents have been notated in various ways; to give an example, L*+H−, L*+H, L*H−, and L*H represent essentially the same entity.

[6] In Pierrehumbert (1980) and in Beckman and Pierrehumbert (1986) the unstarred tones of some bitonal accents are not realized, but are used as a means of triggering various scaling changes, including downstep. The formal treatment of downstep is beyond the scope of this paper, but see Ladd (2008: ch. 3) for a discussion.

Pierrehumbert (1980) this analysis is supported by quantitative data showing that the H of L*+H is located approximately 200 ms after the L* and its location does not correlate with the segmental structure of the accented syllable.

Phonological association is formalized in more detail in Beckman and Pierrehumbert (1986), and in Pierrehumbert and Beckman (1988) where new formalisms are introduced on the basis of experimental data from Japanese. Specifically, in Pierrehumbert and Beckman (1988) metrical trees are replaced by prosodic trees which represent both prominence relationships and constituency. These prosodic trees differ from those proposed by Selkirk (1984) and Nespor and Vogel (1986) by permitting limited extrametricality and a language-specific number of prosodic levels (for a comparison of the different prosodic structures proposed in the literature see Frota, this chapter).[7]

Crucially, according to Pierrehumbert and Beckman (1988) tones may associate with any node in the tree, including phrasal nodes, and not just with specific TBUs (or with prosodic boundaries as in Hayes and Lahiri 1991). This new formalization had several advantages. First, it provided a stricter formalization of the association of phrase accents, which are now seen as phrasal tones that associate with the smaller of two phrasal constituents posited for English, the *intermediate phrase* or *ip* (the other being the *intonational phrase* or IP). Second, the phonological association of tones to prosodic nodes of different levels is also said to determine their scaling (i.e. differences in pitch level). Simplifying considerably, Pierrehumbert and Beckman (1988) show that the scaling of Hs and Ls in Japanese is determined by their association, with tones associated to lower-level constituents in the prosodic tree exhibiting less extreme scaling than those associated to higher-level constituents. In addition, downstep (*catathesis* in the terminology of Pierrehumbert and Beckman) is shown to apply in Japanese within but not across intermediate phrases. Formalizing tonal association along these lines provides a way to account for both local and long-distance effects on tonal scaling while retaining a linear representation of tones, that is without resorting to hierarchical representations or the notion of registers (for an alternative view of hierarchical representations in intonation, see Ladd 2008: ch. 8 and references therein). In addition to Japanese, the formalization of scaling relations proposed by Pierrehumbert and Beckman (1988) has been successfully used by Truckenbrodt (2002) to account for peak scaling effects in Southern varieties of German.

As noted by Pierrehumbert and Beckman (1988), one drawback of allowing tones to associate with prosodic nodes of various levels is that in some instances phonological representations include no ordering among tonal elements, such as lexical and phrasal tones, thus providing no guidance as to how these elements are to be

[7] Prosodic structure is taken to be independent of intonation (Ladd 2008: ch. 7) and thus it is not discussed at length here, though laboratory research has uncovered interesting interactions between the two (see e.g. Shattuck-Hufnagel et al. 1994; Beckman and Edwards 1994; Campbell and Beckman 1997; Harrington, Fletcher, and Beckman 2000; Baltazani 2006b).

produced in time. To address this issue, Pierrehumbert and Beckman formulate a constraint which is informally stated as follows: "a substantive element that is associated to a node in the prosodic tree must either also be associated to the center [head constituent at some lower level of the tree] of the constituent or be realized somewhere at its left or right periphery" (Pierrehumbert and Beckman 1988: 131). This distinction between *central* and *peripheral* tones regulates the temporal order of tones in phonetic realization.

11.2.4 Empirical evidence and the notion of association

In the early autosegmental work, a crucial assumption that was largely accepted without comment was that the relationship between phonological association and phonetic alignment is straightforward: elements that associate in phonology co-occur in time (e.g. Goldsmith 1976: chs. 1 and 3). This idea was challenged by a series of empirical results and has led to new proposals regarding association that I briefly review here.

11.2.4.1 *The internal structure of pitch accents*

One of the first studies showing that the relationship between phonological association and phonetic alignment is not simple was that of Silverman and Pierrehumbert (1990), who examined the phonetic realization of pre-nuclear H* accents in American English. Their results showed that the pitch peaks of these accents do not always co-occur with the accented vowel but may show *peak delay*, the extent of which is determined by the segmental make-up of the accented syllable's rime and its distance from the following accent. Empirical evidence for peak delay and its contrastive potential within a linguistic system, as well as its susceptibility to dialectal variation, have since been documented in many typologically unrelated languages with widely different prosodic systems (for a review, see D'Imperio, this chapter).

Arvaniti et al. (1998, 2000) address a complication that arises from peak delay with respect to the temporal patterning of bitonal pitch accents. Specifically, they examined Greek rising accents that can be plausibly analyzed as bitonal L+H (Arvaniti and Ladd 1995) and showed that neither is it the case that one of the tones aligns with respect to the other (as the analyses of Pierrehumbert 1980, or Beckman and Pierrehumbert 1986, would predict) nor that they both align with respect to the same segmental landmark (as the association of both tones with the same TBU would predict). Rather, the L tone occurs slightly before the onset of the accented syllable and the H tone occurs early in the first post-accentual vowel. Thus, the distance between the two tones is variable and positively correlates with the duration of the accented syllable. The finding that tones may align stably with segmental landmarks gave rise to the "segmental anchoring" hypothesis (see D'Imperio, this chapter).

In addition, these results served to consolidate the distinction between discrete phonological association with a specific structural position and gradient phonetic alignment with segmental material (Arvaniti et al. 2000). Finally, they showed that it is possible for grouped tones to be aligned independently of each other.

The independent alignment of grouped tones is formalized in Grice (1995b), who proposed two distinct representations for bitonal accents: accents with leading tones, e.g. L+H*, are represented as clusters in which both tones independently associate with the same TBU, while accents with trailing tones, e.g. L*+H, are represented as contours that associate with a given TBU as a group. Although Grice's proposal that this difference in association is tied to the presence of a trailing vs. a leading tone is not empirically supported cross-linguistically (e.g. Arvaniti et al. 1998, for Greek; Ladd and Schepman 2003, and Arvaniti and Garding 2007, for English), the notion of two association options with concomitant effects on phonetic alignment is certainly valid and has been supported by experimental data from Portuguese (Frota 2002). Taken all together, the results briefly reviewed here suggest that tones may align in different ways with the segmental string and with each other both within and across languages. In turn, these different modes of alignment can be formally represented as different types of phonological association.

11.2.4.2 *Primary and secondary association*

The complex nature of phonological association is addressed in detail in Pierrehumbert and Beckman (1988) who introduced the notion of *secondary association* to account for a series of data from Japanese. As mentioned, in Japanese each accentual phrase is associated with two tones, the phrasal H and the boundary L%. Typically, phrasal tones are realized at the boundary of the constituent with which they are associated, but the Japanese phrasal H co-occurs with the second sonorant mora of its accentual phrase if the first syllable is short (otherwise it co-occurs with the first sonorant mora), while the boundary L% co-occurs with the first mora of the *following* phrase if its first syllable is short and unaccented (otherwise it is realized at the right edge of its own accentual phrase). Pierrehumbert and Beckman (1988: 126 ff.) account for this variability by positing that in addition to their primary associations with the accentual phrase node, the phrasal H and boundary L% have secondary associations to the second sonorant mora and the first mora of the following accentual phrase respectively; these secondary associations are realized when certain conditions are met.

The notion of secondary association was taken up by Grice et al. (2000) who examined a variety of tunes in English, German, Dutch, Cypriot and Standard Greek, and in Standard and Transylvanian Hungarian and Romanian. They showed that the variability in the phonetic realization of these tunes can be accounted for if we assume the existence of a phrase accent with both a primary and a secondary association. Phrase accents have a primary association to a phrasal constituent

(the exact nature of which is left unspecified in Grice et al.) but also a secondary association to a specific TBU, such as the last metrically strong syllable of their phrase or the first syllable of the last word (independently of metrical strength). When this TBU is associated with another tone, the phrase accent is realized at the phrasal boundary, i.e. by means of its primary association; but when the TBU is available, the secondary association takes over instead. Quantitative support for this view of phrase accents has been reported for Cypriot Greek (Arvaniti 1998), Standard Greek (Arvaniti et al. 2006a; Arvaniti and Ladd 2009), French (Welby 2004), Dutch (Lickley et al. 2005), and English (Barnes et al. 2006).

A different use of secondary association is developed in Prieto et al. (2005). They present data from Catalan and Italian involving bitonal accents which are phonologically contrastive but phonetically show only small differences in peak alignment. They propose that these differences be represented by means of secondary association of one of the tones to a syllabic or segmental edge, with peak delay being the default. Face and Prieto (2007) further apply this idea to a three-way contrast of peak alignment in Spanish.

This use of secondary association differs from the use of the mechanism by Pierrehumbert and Beckman (1988), Grice et al. (2000), and others, where secondary association accounts for a discrete alternating pattern that does not affect meaning: in Prieto et al. (2005) secondary association is used *in conjunction* with the primary association, not in place of it, and thus does not involve an alternation in the realization of a tone. Arvaniti et al. (2006b) argue that this use is essentially a formalization of phonetic alignment rather than of association proper and propose alternative ways of representing the patterns uncovered by Prieto et al. that do not require the use of secondary association.

The notion of secondary association is also explored within OT by Gussenhoven (2000a,b) in his analysis of Roermond Dutch, a variety with a lexical tonal contrast. Simplifying somewhat, in Roermond Dutch phrasal tones show controlled variability in spreading that is determined by the location of the focal word (which carries a *post-lexical* pitch accent) and (for some phrasal tones) the possible presence of a *lexical* H tone on the focal word. Gussenhoven accounts for these differences in the spreading of phrasal tones by drawing a distinction between *alignment* (in the OT sense) and *association*: phrasal tones are aligned with the right edge of their intonational phrase and also show alignment to the right of a preceding tone. This formalization of alignment does not entail association with a particular TBU, though phrasal tones may also have such an association; e.g. in Gussenhoven's analysis the L_i of Roermond Dutch declaratives associates with the focal word's second mora if the word is bimoraic and its second mora is not associated with a lexical H. As noted by Gussenhoven (2000a, b), this analysis accounts for tonal spreading and more generally for the durational aspects of tonal implementation without the need of positing additional targets in phonetic realization (but at the expense of additional alignment constraints).

11.2.4.3 *Association and the temporal ordering of tones*

In addition to the issues they raise about the role of secondary association, the Roermond data question a standard assumption about the realization of boundary tones, namely that they occur as close as possible to the boundary with which they are phonologically associated. This idea, which is largely supported by both qualitative and experimental results, is connected to the well-established notion of the linear ordering of autosegments and is an instance of the prohibition of crossed association lines (Goldsmith 1976; Sagey 1988), the aim of which is to ensure the same left-to-right sequence among associated elements in each tier. Maintaining temporal order is also the motivation behind the distinction of Pierrehumbert and Beckman (1988) between central and peripheral tones, discussed in Section 11.2.3.

Gussenhoven's Roermond data provide rare evidence that f0 contours do not always reflect the expected order of tones. Specifically, in this variety boundary tones are realized *before* a lexical H tone, if this tone occupies the last mora of the intonational phrase. In order to account for this unusual pattern, Gussenhoven posits that AlignT$_i$Right, the constraint for the right alignment of Roermond Dutch boundary tones, is outranked by a similar constraint for the lexical tone, AlignLexRt, and argues that other AM models cannot account for this pattern (for further discussion see also Gussenhoven 2004: ch. 7).

Although Gussenhoven's contention appears to be prima facie correct—or at the very least involves a pattern difficult to formalize in AM except in ad hoc ways—ultimately the issues raised by the Roermond Dutch boundary tones and more generally by the timing patterns discussed in Gussenhoven (2000a,b) and Prieto et al. (2005) have to do with the interface between phonology and phonetics. If it is accepted that the phonological representation of intonation need not be faithful to surface facts any more than other phonological representations are, it is possible to represent tonal contrasts making sparing use of formalisms such as secondary association and alignment constraints, and to view differences in tonal timing, spread, and duration as the realm of phonetic realization, as advocated by, among many, Pierrehumbert and Beckman (1988), Silverman and Pierrehumbert (1990), Elordieta and Calleja (2005), Arvaniti et al. (2006b), Arvaniti (2007a), and Ladd (2008: ch. 5).

11.2.4.4 *Alternative views*

Most of the experimental research reviewed above has been couched in terms of the AM model of intonational phonology. Results have led, as is inevitable, to revisions of Pierrehumbert's original model, and also to the development of alternatives, such as ToDI (Transcription of Dutch Intonation; Gussenhoven et al. 2003), and IViE (Intonational Variation in English; Grabe 2001). These particular models deviate in some respects from the original AM model—e.g. both ToDi and IViE dispense

with the phrase accent and make boundary tones optional—yet they retain the main assumptions of AM, that is, the notion that melodies are composed on tones, that tunes remain largely underspecified, and that phonological structure mediates between f0 and meaning.

A radically different approach is taken by Xu and colleagues in the Parallel Encoding and Target Approximation (PENTA) model (e.g. Xu 2005; Xu and Xu 2005, among many). PENTA rests on the idea that f0 directly encodes "communicative functions" (such as statement vs. question or the location of focus). Crucially, every syllable in an utterance is specified for f0, and f0 events are synchronized with syllables. Although PENTA is based almost exclusively on Mandarin, Xu and Xu (2005) have also applied it with some success to English declaratives with narrow focus. It is important to note, however, that their results are *not* incompatible with existing AM accounts, while results from several other studies are incompatible with main PENTA assumptions (for relevant data and discussion see, among others, Arvaniti et al. 2006a; Arvaniti and Ladd 2009; Chen 2010). Thus, the experimental evidence so far argues against the full specification of f0 contours and for a principled distinction between phonetic realization and an abstract phonological level of representation that involves tones and their association with the segmental string.

11.2.5 Conclusion

In conclusion, experimental evidence from a variety of typologically distinct languages supports the main tenets of AM, namely, that pitch contours consist of strings of tonal targets which temporally coordinate with segmental events of various sorts. For intonation, these are typically metrically strong or boundary-adjacent syllables. In turn, these tonal targets are considered the reflexes of underlying tones.

Although this general conception appears to be fundamentally correct, several issues remain that require further investigation. For example, it is still unclear how many different types of tonal association are possible cross-linguistically. Recent research that examines the connection between f0 excursions and articulatory gestures (e.g. Mücke, Grice, Becker, and Hermes 2009) may shed light on this issue, as it has already uncovered parallels between different types of articulatory phasing relations and existing formalizations of tonal association, such as those proposed by Grice (1995b; see section 11.2.4.1). Similarly, experimental research should help elucidate the mechanisms that determine the choice of melody and tonal association at a higher level, i.e. the association of accents to specific words in an utterance (*accentuation*). Though existing research provides no clear answers at present (cf. the accentability of particular words explored by German et al. 2006, vs. the metrical explanation of accentuation advocated by Calhoun 2010), it is clear that more research along these lines—using both experimental paradigms

and natural data—will be necessary, if we are to understand the phonological representation of intonation, its phonetic realization and its function within a given linguistic system.

11.3 Tonal alignment

Mariapaola D'Imperio

11.3.1 The notion of tonal alignment

In the last twenty to thirty years, there have been numerous arguments from cognition, linguistics, and technology for assuming that the words, on the one hand, and the tonal specification of an utterance, on the other hand, are stored and created independently of each other (Liberman 1978; Goldsmith 1976; Pierrehumbert 1980; see Ladd 2008 for a synthesis of the autosegmental-metrical approach to intonation, or AM theory). But this independence comes at a price: tune and words have to be synchronized with each other in time. *Tonal alignment* is concerned with how target tones are temporally coordinated, or synchronized, with prosodic units (e.g. syllables) and their constituents (the segments that make up syllables). A large body of alignment studies in laboratory phonology has therefore explicitly tested the basic tenet of the AM theory, i.e. that underlying tonal structure is reflected in the signal through the presence of well-defined fundamental frequency (f0) targets, which are specified both in terms of tonal alignment and according to some well-defined melodic value (*scaling*). Among the two "coordinates," alignment with the segmental string (i.e. the temporal distance from edges of segments or syllables) has been the most studied aspect so far, including acoustic, production, and perception studies, all of which will be reviewed below.

Indeed, the question of the alignment of intonation contours and their impact on perception was first investigated in *contour* approaches to intonation. For instance, the IPO school, while developing a rule-based generative approach to intonation, extensively investigated the alignment of rises and falls and found that an *early* vs. *late* alignment with the segmental string could induce different interpretations of the contour, hence different meanings (cf. Section 11.3.3 below). Here I more closely concentrate on the insights of AM theory in stimulating laboratory investigations of tonal alignment. Earlier alignment studies set out to question the controversial nature of intonational primitives: finding evidence for the existence of static tonal targets characterized by stable temporal and melodic coordinates calls into question the validity of the primacy of intonational movements (traditionally represented by

the British School approaches and the IPO system (cf. O'Connor and Arnold 1961; 't Hart et al. 1990; see also Arvaniti, this chapter).

Despite the growing attention of the prosody community to alignment studies, there is still a great deal of controversy about the forces as well as the exact mechanism driving the alignment of tonal targets with segmental gestures, and many issues remain unsettled, as will be shown in this section. Among the most significant points in question are how best to tease apart the phonetics and the phonology of alignment, and how to identify universally motivated tendencies (both from a production and a perception standpoint) in both intonational and tonal languages. It should be noted that the choice of concentrating upon tonal alignment regularities relative to specific segmental landmarks or regions is generally not theoretically (acoustically, perceptually, and/or articulatorily) driven, making the study of alignment mostly an exploratory enterprise at this point, though certainly not devoid of interest.

In this section, the notion of tonal target alignment is reviewed both from a general perspective and according to more specific claims related to the autosegmental-metrical theory of intonation. The notions of tonal stability and alignment variability (as a consequence of various phonetic and phonological factors) are also discussed in the light of acoustic studies. The final portions of this section aim at reviewing more recent empirical research on the perception and articulatory production of tonal target alignment.

11.3.1.1 *Accounting for variability in tonal target alignment: Early acoustic studies*

Bruce's seminal work on Stockholm Swedish (Bruce 1977; see also Arvaniti, this chapter) paved the way for viewing tonal elements as abstract (phonological) targets, mapping onto concrete phonetic targets. It is those targets that are responsible for the varied, surface tonal representations that occur in different phrasal and prosodic environments. Bruce's work is thus a classic study integrating phonetics and phonology, and the first to show how an abstract, underlying, tonal form can be revealed by careful phonetic and contextual manipulations. By separating the different contribution of accentual and phrasal melody Bruce discovered that the two basic accent types of Swedish, Accent I and II, are composed of the same HL tonal sequence (i.e. they are both falls), but contrast through the specification of a *peak alignment* feature. While in Accent I the H target precedes the stressed syllable (while the L target falls within it), in Accent II this H target is timed to occur at the onset of the stressed syllable, with the consequence that the entire falling gesture is globally later, spanning the entire stressed syllable. This difference can be seen in Figure 11.3.1 for the Accent I word *längre* (solid line) 'longer' and the Accent II word *långa* 'long' (dashed line) in pre-focal position, with the stressed syllable being initial in both words. Similar meaningful differences in accentual fall

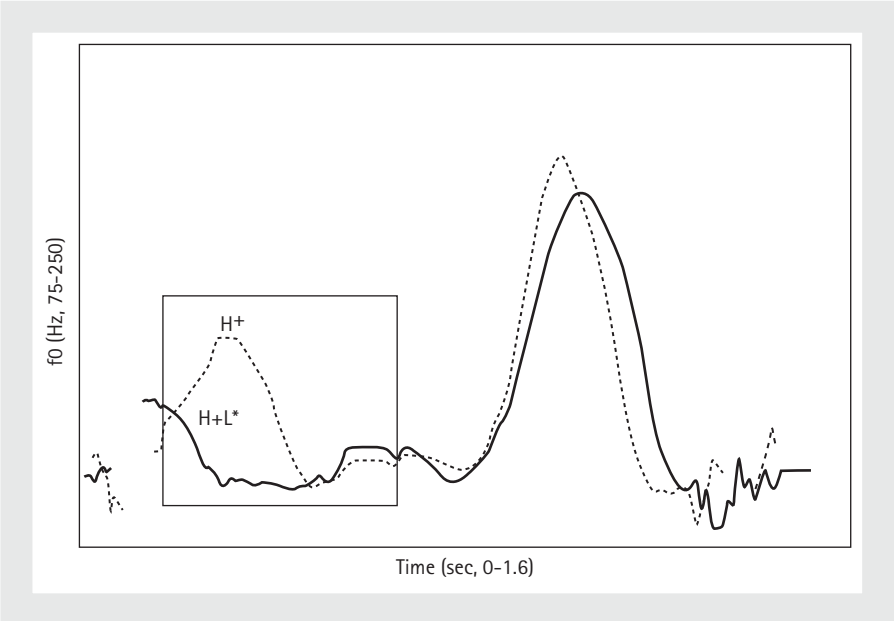

Figure 11.3.1. Schematic representation of the f0 curves for the Swedish utteranced *Jag sa långa NUMMER* 'I said long number' and *Jag sa längre NUMMER* 'I said longer number', with focus on *Nummer*. The alignment contrast between Accent I (H+L*) word *längre* 'longer' and Accent II (H*+L) word *långa* 'long' is shown within the box, delimiting the word boundaries.

alignment have been later pointed out in several languages, such as in Palermo Italian and European Portuguese (see Grice 1995a and Frota 2000 for the nuclear H+L*/H*+L distinction).

Note however that the shape of the same word accents in other sentential and focus contexts can be subject to systematic variability due to, among other things, "time-dependent" adjustments. For instance, the simultaneous presence of a word accent and a sentence accent on the same syllable (e.g. when the syllable is phrase-final) leads to a situation of tonal crowding that is resolved either through temporal readjustments of peak location or through melodic readjustments of the f0 level reached by the tonal target. For instance, the fall in Accent II is anticipated when immediately followed by the sentence accent, suggesting, according to Bruce, that the sentence accent command has interfered with the word accent command.

Similar contextual dependencies on tonal alignment have also been found for intonational languages, where tonal contrasts carry a pragmatic or semantic meaning. The idea that intonational languages are characterized by sparse tonal distribution is by now widely accepted within intonational phonology. If an intonation contour is abstractly viewed as a sequence of a limited series of tonal targets,

connected by (linear or non-linear) interpolation, then the question is how the temporal location and the pitch height of such targets is controlled. Inspired by Bruce's work, the grammar of intonation built by Pierrehumbert (1980) embraced the idea of phonologically contrastive alignment by postulating the existence of differently associated bitonal pitch accents (see also Arvaniti, this chapter). This fundamental notion emerges more clearly in the theory proposed by Pierrehumbert and Beckman (1988), postulating that the coordination between tones and segments is mediated through prosodic structure, i.e. both the tonal and the segmental tiers (the "substantive" tiers) are linked to specific prosodic domains, at different levels of the prosodic hierarchy (mora, syllable, accentual phrase, etc.). The authors speculate that tones associated to a specific prosodic unit are realized at the "phonetic boundary" of this unit, unless differently specified, so that initial and final tones at a prosodic level would be realized at the same time as the initial and final phonemes belonging to the same structural unit. For instance, a phrase accent is usually realized at the phonetic periphery of the intermediate phrase. However, in cases when the alignment of peripheral tones appears to be different from this general stipulation, this might be accounted for by secondary association with a lower level of the prosodic hierarchy. For example, in Japanese an initial phrasal H associated with the left edge of an accentual phrase can be realized on the second instead of the first mora of the phrase, which is taken to be evidence for secondary association to this structural position. This mechanism of secondary association has been recently employed to explain some alignment phenomena relative to phrase accents and edge tones in several different languages (see Frota 2003 for phrasal H tones in European Portuguese and Grice et al. 2000 for Eastern European languages, *inter alia*).

Centrally associated tones, such as pitch accents (which are associated to heads of prosodic domains) can also be differently aligned relative to the segments composing the *tone-bearing unit* (TBU). In this case, language-specific phonetic implementation rules can stipulate whether the exact alignment of centrally associated tones is either "late" or "early" relative to the TBU, such as the stressed syllable. However, the authors mention that a mechanism of secondary association could also be envisaged for centrally associated tones.[8]

This is what has been proposed by Prieto et al. (2005) to account for the different alignment patterns found for the H tone of LH pitch accents in Catalan and Italian. Controlled experiments are hence required to establish, within and across languages, how phonological association is translated into exact alignment patterns, and whether there are universal tendencies due to either perception or production constraints.

[8] The authors also observe that languages might display certain alignment tendencies. For instance, in a language such as English, late alignment of starred tones seems to be preferred (i.e. H* targets tending to occur towards the right edge of the stressed syllable, or even later).

Some early alignment studies in English served also to evaluate the claim that the inventory of pitch accents is the same in pre-nuclear and nuclear position, a point of dispute between the British and American approaches to intonational analysis. Support for the "same inventory" hypothesis can be found in studies conducted by Steele (1986b) and Silverman and Pierrehumbert (1990), who investigated the timing of the target f0 peak for pre-nuclear H* accents as a consequence of prosodic context. The authors carried out a carefully controlled study in which the presence of an upcoming word boundary and/or an upcoming (nuclear) accent as well as global speech rate resulted in significant modifications of H alignment, which was explicitly accounted for through linear, additive models. It was here that the notion of "peak delay" was first introduced in the tonal alignment literature. The notion of a delay stems from the observation that H* pre-nuclear peaks are generally temporally aligned beyond the boundaries of their TBU (i.e. the stressed syllable for English and Italian). In this study, peak delay refers more simply to an objective measure of the latency of the tonal target relative to stressed vowel onset.[9]

It was eventually a proportional measure of target alignment, relative to stressed rime duration, and not absolute peak delay, which yielded a better fit in the regression models. Smaller proportional alignment was interpreted as a consequence of "tonal repulsion" due to the upcoming tonal gesture. Thus an earlier alignment of nuclear H* peaks would be caused by the proximity of an L- phrase accent. An important outcome of this study was that the alignment differences observed were continuous rather than discrete, thus supporting the idea that the pattern of tune-text association is an abstract, coarse-grained description of how the melodic targets and the segmental structure are to be timed relative to each other in running speech, while exact phonetic alignment stems from language-specific phonetic implementation.

Subsequent studies on tonal alignment in the 1990s extensively tested the impact of contextual prosodic factors on the temporal location of both H and L targets in a number of languages. Of these factors, the most studied are proximity to upcoming accentual and phrasal tones or upcoming word boundaries, as well as rate effects. Other studies concentrated on the effect of segmental composition of the TBU and intrasyllabic duration on the placement of pre-nuclear accent peaks. Among them, van Santen and Hirschberg (1994) showed that the timing of pre-nuclear H* peaks in American English can be modeled as the result of weighted effects of intrasyllabic segmental composition (consonant manner and voicing of onset and coda) and duration. The authors predicted the timing of "anchor points" (located at specific proportions of an entire accent gesture) and then warped the accentual contour in a non-linear way, so that the timing of earlier anchor points would depend more on onset duration, while the timing of later anchor points would largely depend on rime duration.

[9] Later, this notion was reanalyzed by a number of researchers to indicate the delay of peak alignment in pre-nuclear H* accents, especially in work on tonal alignment in Mandarin conducted by Xu and colleagues (Xu 2001).

A similar approach was adopted by Prieto et al. (1995) who investigated peak alignment for the pre-nuclear H* in Mexican Spanish. Among the relevant findings, both onset and stressed vowel duration contributed in successfully predicting peak delay, while the effects of upcoming prosodic boundaries vs. position in the word were not entirely homogeneous. Alignment alterations due to tonal crowding have also been observed for Neapolitan Italian (D'Imperio 1995, 1997, 2001, 2002a). In sum, the temporal alignment of f0 targets seems to be affected by a number of fine phonetic and phonological variables that need to be taken into consideration while studying the inventory of tonal events for a given language.

11.3.1.2 *Stability of tonal alignment and the "segmental anchoring hypothesis"*

Despite its variability, tonal alignment can also be strikingly systematic. Work by Ladd and colleagues on tonal alignment in Greek, Dutch, English, and German has uncovered a certain number of stability effects (see Ladd 2008 for a review; see also Arvaniti, this chapter). When right-hand prosodic effects are excluded (i.e. when the tonal features under investigation are not in the vicinity of pitch accents or boundary tones), the alignment of f0 targets appears to be consistently governed by segmental anchoring (*segmental anchoring hypothesis*, SAH henceforth). According to this view, the beginning and the end of a tonal movement would be anchored relative to some specific segment within the stressed syllable (such as its onset consonant, its vowel nucleus, or its coda) or even the post-accentual syllable, and this alignment would be resistant to modifications in syllabic composition, segmental structure, and speech rate (Ladd et al. 1999, 2000; Ladd 2006, *inter alia*). On the other hand, holistic features of the contour, such as rise or fall time and speed (hence slope), would vary as a consequence of the specific tone-to-segment alignment pattern.

Ladd et al. (1999) showed also that the L target preceding the peak of the English pre-nuclear H* accent is consistently aligned with stressed syllable onset, despite rate and segmental differences. This fact was interpreted in terms of a stability of low target alignment due to the *anchoring* of the L tone to stressed syllable onset (see also van Santen and Hirschberg 1994 for American English; Prieto et al. 1995 for Spanish; Caspers and van Heuven 1993 for Dutch; Xu 1998 for Mandarin; and Igarashi 2004 for Russian).

Strict segmental anchoring of tonal targets has been taken to be strong evidence for the AM approach to intonation. Yet, the strict version of the SAH is contradicted by results indicating that syllable structure detail, segmental composition, and even speech rate can crucially affect target alignment. First, there seems to be an asymmetry in the behavior of L and H tones with regard to anchoring. While L targets in LH accents appear to be consistently anchored to the onset of the accented syllable, H targets of rising accents are found to be much more variable. For instance, H peak

alignment measures reported in Ladd et al. (1999) are less supportive of a strict segmental anchoring hypothesis since an effect of speech rate was found even when alignment was measured relative to the closest segmental anchor (i.e. the onset of the post-accentual syllable).[10] Several other studies have found a significant effect of speech rate on peak alignment (Xu 1998; Ishihara 2003; D'Imperio, Espesser, Loevenbruck, Menezes, Nguyen, and Welby 2007; Prieto et al. 2007).

Among other sources of evidence for systematic variability of tonal alignment, some studies have underlined syllable structure effects. D'Imperio (2000; see also D'Imperio, Petrone, and Nguyen 2007) found that the peak of the Neapolitan Italian nuclear L+H* was located closer to the vowel offset in closed syllables, while the peak of L*+H (typical of yes/no questions) would shift from being aligned within the stressed vowel in open syllables to being aligned with the coda in closed syllables (see Figure 11.3.2). Interestingly, alignment was stable across syllable structures when measured relative to the onset of the syllable (or the vowel), and contrasted instead when a landmark on the right-hand side of the syllable was taken as a reference point. The choice of the putative anchors for tonal alignment can hence be of paramount importance when it comes to uncovering significant or non-significant effects. However, no principled selection of one such choice over the others has been found yet. Moreover, the alignment of L tones is much more variable than previously thought, especially when automatically and not manually calculated. For instance, L tones are timed later within the associated syllable for question L*+H than for statement L+H* rises in Neapolitan (D'Imperio 2000, 2002). Note that in this study L timing was calculated through an automatic technique,[11] i.e. a two-line regression fit (see Figure 11.3.3) already employed by Pierrehumbert and Beckman (1988), and which has since been employed in the study of various other languages (e.g. Welby 2006 for French; Frota 2002 for European Portuguese).

Prieto and Torreira (2007) investigated the alignment of Spanish LH* pre-nuclear peaks with segmental landmarks as a function of syllable structure type (open, closed), segmental composition, and speaking rate. Alignment was indeed affected by both syllable structure and speech rate: in open syllables the peak was located around the end of the accented vowel, while in closed syllables it was located within the sonorant coda and slow speaking rate caused peaks to align relatively earlier. For similar effects, see also Gili Fivela and Savino (2003) for Bari Italian and Hellmuth (2005b) for Egyptian Arabic.

[10] This effect was partially removed only when alignment was calculated as a proportion of a segmental region, corresponding to the post-accentual onset consonant duration.

[11] D'Imperio (2000: 93): "[A]n automatic procedure was employed by which two straight lines were fitted to the f0 segment [stretching between two references points, such as i and j or p and q in Figure 4b]. The parameters of the two linear models were estimated by means of conventional linear least-squares methods. To estimate the elbow position, i.e. the intersection of the two fitted lines, two linear regressions were computed for each possible elbow location. The location eventually selected as the 'elbow' was the one leading to the smallest total modeling error."

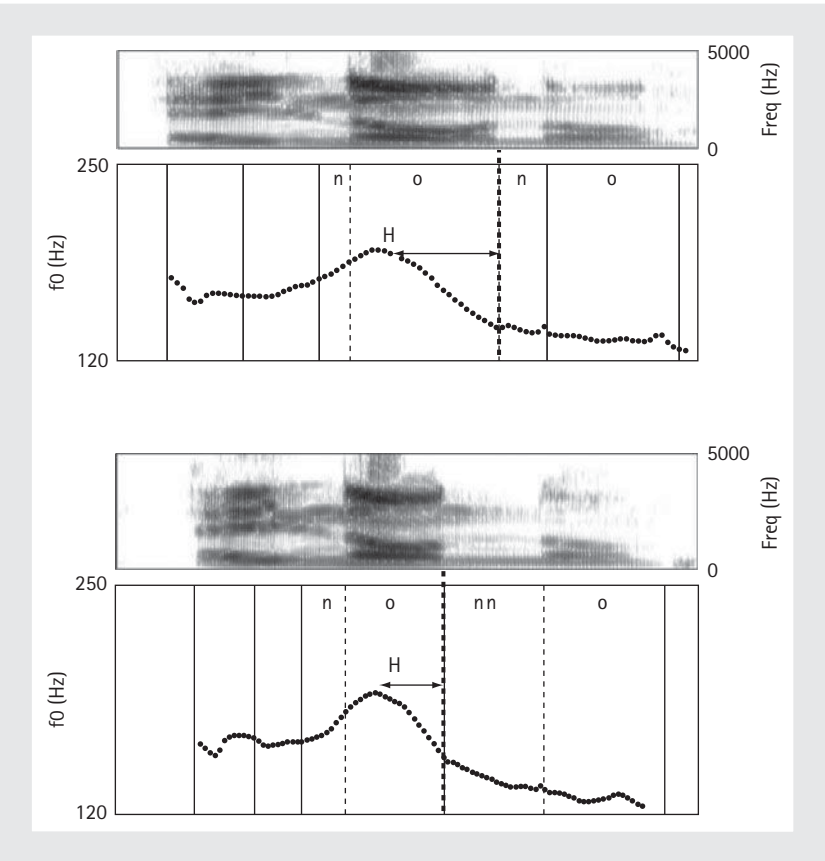

Figure 11.3.2. F0 curve and spectrogram for the sentences *E' il nono* 'It is the ninth' (upper) and *E' il nonno* 'It is grandpa' (lower), uttered as a narrow focus statement, with a L+H* nuclear accent. Note the different relative alignment of the peak in the two cases. The dashed vertical line indicates stressed vowel offset.

The findings for Dutch reported in Ladd et al. (2000) also show that a phonological length contrast in vowels seems to affect peak alignment in Dutch prenuclear rises, independent of actual duration values. In other words, H peaks of rises would be aligned later in phonologically long than in phonologically short vowels, even when this contrast does not translate immediately into an acoustic duration difference. More recent contributions represent a challenge for the SAH, for instance by positing a notion of "anchorage" to account for H target alignment of French LH* rises (Welby and Loevenbruck 2006). The "anchorage" notion would translate into a more or less extended temporal region within which a tonal target can anchor. For French late rise peaks, this region stretches from just before the end of the vowel of the last full syllable of the accentual phrase to the end of the phrase.

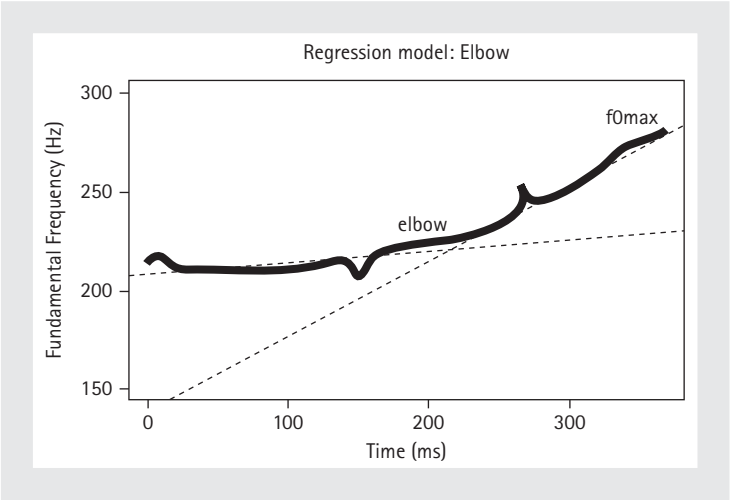

Figure 11.3.3. Schematic representation of the "two-line" regression methodology. (Taken from D'Imperio 2000.)

Very recent research on alignment variability across dialects of the same language has also revealed a quite complex picture, which is difficult to reconcile with the idea of strong segmental anchoring. For instance, Atterer and Ladd (2004) report great variability in the alignment of syllable-initial L as well as H targets between Northern and Southern German dialects, which has been more recently replicated in a study by Mücke, Grice, Becker, and Hermes (2009). Similar kinds of effects have been reported for dialects of American English (Garding and Arvaniti 2007). Note also that in both varieties of German described, both L and H targets were aligned later than in English. It remains to be seen whether the difference between the striking regularity of L target alignment as opposed to a higher sensitivity to rate, syllable structure, and segmental composition of H targets in rising pitch accents might be due to articulatory or perceptual constraints on its location.

11.3.2 Tonal alignment and articulatory 'anchors'

A hypothesis tested in some recent studies is that laryngeal gestures might be timed to co-occur with some specific supra-laryngeal gesture (e.g. minima and maxima of consonantal trajectories or peak velocity within the onset consonantal gesture) for specific pitch accents. Hence, there may exist anchor points around which tonal targets are located, governed by principles of synchrony and stability (D'Imperio et al. 2003, 2007). Specifically, D'Imperio, Espesser, Loevenbruck, Menezes, Nguyen, and Welby (2007) suggest that the H targets of the nuclear declarative accents of Neapolitan Italian (and to a certain extent the L of the initial LH rise of French)

might be synchronous with maximum velocity for the primary constrictor trajectory (such as lower lip for labials and tongue tip for coronals) within the stressed syllable onset (see similar findings for Catalan in Prieto et al. 2007). Comparably, L and H targets in two varieties of German (Mücke, Grice, Becker, and Hermes 2009) have been shown to align consistently with articulatory anchors: in prenuclear rises, they are aligned with vocalic gestures (measured as transvocalic minima of lip/tongue tip constriction), whereas nuclear L and H are more likely to align with consonantal gestures (here consonantal maxima of tip/lip constriction).

On the other hand, a position close to that of strict segmental anchoring, which would actually be better characterized as *syllabic anchoring*, is the one maintained by Xu (2002; Xu and Wang 2001), claiming that the phasing of tones and syllables is constrained by the speed at which speech is produced, as well as by some kind of human-to-human coordination system. In other words, the observations that, in rising contours, L tones tend to be timed at syllable onsets while H targets are timed at syllable offsets might be due to a preference for an in-phase relationship (or phase-locking) between tonal and syllabic gesture, so that the two might start and end together. Tonal sparsity is a potential problem for this model, though, since it may be that synchronization between the laryngeal and supralaryngeal systems, and the pressure to achieve the kind of phase-locking proposed, is more true of tonal languages such as Mandarin than intonation languages such as English, French, or Italian. In this latter group of languages, co-production of tonal and syllabic cycles might be less needed, which would account for the greater tonal alignment variability observed.

11.3.3 Perception of tonal alignment contrast

On the one hand, then, there is much about tonal alignment that is predictably influenced by the phonetic context. On the other, though, small differences of alignment can apparently also create clearly perceptible differences of meaning, both phrasal and lexical, as well as help listeners to identify word boundaries. Among the first studies on the impact of tonal alignment on meaning, Pierrehumbert and Steele (1989) investigated whether the timing of the "rise-fall-rise" (scooped) pattern in English could be contrastive and hence justify the presence of both an L+H* and an L*+H rise in English. The two pitch accents L*+H and L+H* appeared to be related to either an "incredulous exclamation" or a "forceful assertion" meaning (Ward and Hirschberg 1985). They hence created a set of rise-fall stimuli over the utterance "Only a millionaire" by means of LPC resynthesis, in which they varied only the relative timing of the LH targets, and then asked subjects to imitate the contour they heard. Despite the continuous timing manipulation, subject productions showed a discrete and bimodal alignment pattern. A related set of experiments conducted by Kohler (1987), through the categorical perception

paradigm (see Prieto, this volume and Iverson, this volume) also showed that a three-way contrast for the timing of an HL fall (early, medial, and late) can induce three different interpretations in German, i.e. "established," "new," and "emphatic."

More recently, D'Imperio and House (1997) showed that in Neapolitan Italian, the intonation patterns for narrow-focus statements and questions are distinguished primarily by the alignment of the nuclear peak of a rise-fall pattern. Through f0 target timing manipulations alone, the authors were able to induce a clear question and a statement category, with a region of ambiguity at the center of the continuum. The results were replicated with stimuli controlled for peak shape (flat or sharp peak) in D'Imperio (2000), which are schematically shown in Figure 11.3.4. The materials consisted of a series of stimuli in which the L and H target of a rising-falling configuration were shifted forward within the stressed syllable ("peak-shift" manipulation) from a typical alignment for statement to that of a yes-no question. A similar manipulation was performed to obtain stimuli containing a 30 ms plateau ("plateau-shift"). Thirty Neapolitan subjects listened to the stimuli and identified each as a question or a statement. The results are shown in Figure 11.3.5, where percent of question responses is plotted for both the peak and the plateau series. Note that shifting the peak to later timing continua in both peak manipulations caused an increase in question responses that appeared to be categorical. Moreover, peak shape affected the perceived alignment of the target tone, in that flat peak stimuli received a greater number of question responses already at early locations within the continuum. This result was interpreted in terms of the notion of a "perceived" target, whose location can be identified roughly with the end of the plateau. More recent work also questions a direct link between timing of f0 extrema and perception, showing that peak shape can affect pitch accent identity also in German (Niebuhr 2003) and in other varieties of Italian (see Gili Fivela and D'Imperio 2008; D'Imperio et al. 2010).

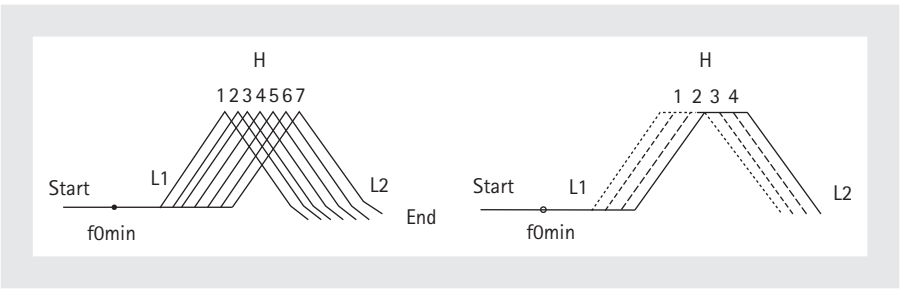

Figure 11.3.4. Schematic representation of the timing manipulation for the "peak" series (left) and the "plateau" series (right) for the experiment in D'Imperio (2000). (Taken from D'Imperio 2000.)

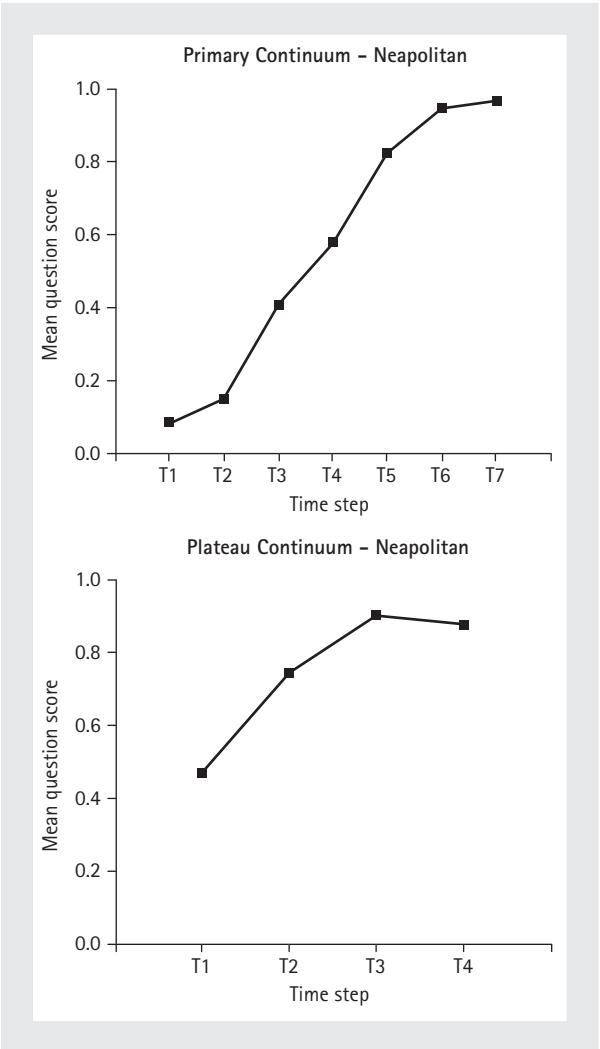

Figure 11.3.5. Results for mean question responses for the peak series (upper) and for the plateau series (lower).

Tonal alignment appears also to affect the identification of lexical contrast, such as that between geminate/singleton minimal pairs of Italian of the type *nonno/nono* 'grandfather/ninth' (D'Imperio, Petrone, and Nguyen 2007), since later alignment induced more 'nonno' responses in an identification task with durationally ambiguous stimuli. In fact, listeners seem to expect different alignment realizations in varying segmental structures. In a comparison of early and late falls in Dutch, which have clearly different alignments of both H and L targets, the voiceless obstruents in

the syllable coda were found to push the targets of the perceived pitch accents to the left, whereas the opposite was true for sonorant codas (Rietveld and Gussenhoven 1995). That is, the same physical target alignment can lead to the perception of different pitch accents if the coda consonants are changed.

Recent studies on the tonal marking of the French Accentual Phrase (AP) by Welby (2003) and Spinelli et al. (2007) show that the L target of the LH initial rise, which can be secondarily associated with the left edge of this constituent (or with a later syllable edge), can be employed as a cue to word segmentation in non-word pairs such as *mélamondine* and *mes lamondines* 'my lamondines.' Similar findings have been reported for English (Ladd and Schepman 2003) regarding the impact of L target location in L+H*. When this accent is located on the last name of pairs with an ambiguously affiliated consonant, such as *Norma Nelson/Norman Elson*, later alignment values are found for pairs where the consonant is in coda position, which appears to help listeners disambiguate syllabic affiliation.

11.3.4 Conclusion

The studies reviewed in this section show that aligning tonal targets and recovering this alignment on the part of humans involves complex mechanisms which may only be captured by thorough language-specific descriptions. The communicative function of speech might be the key within this complex picture, as de Jong (2007) rightly points out. It might well be, as he proposes, that there is a conflict between *hardware* requirements (physiological and higher-level cognitive and/or attentional constraints) as compared to *shareware* requirements, originating from the essential fact that language is principally meant to be understood. Hence, despite model predictions that are too heavily based on universal (auditory and/or articulatory) constraints, the ultimate answer to alignment variability might simply be that speech is learned, and is not merely determined by any physiological, attentional, or functional constraint in and of itself. More research is needed to find the universal thread binding the language-specific data, by shedding light on the interaction between universal and language-specific constraints on tonal alignment.

CHAPTER 12

PHONOLOGICAL REPRESENTATIONS IN LANGUAGE ACQUISITION: CLIMBING THE LADDER OF ABSTRACTION

BENJAMIN MUNSON,
JAN EDWARDS, AND
MARY E. BECKMAN

This chapter provides a rich discussion of the development of phonological representations. The authors draw together multiple lines of research addressing early category learning and the development of these categories into a linguistic system. They integrate work done from the perspective of production, perception, and sociolinguistic indexing.

12.1 Introduction

The physical world in which humans reside is limited to four dimensions, but the mental world in which our knowledge of language resides is not similarly limited. Individuals' knowledge of speech sounds comprises representations of information in multiple sensory domains, including representations of the auditory characteristics of the sounds that they have produced and have heard others producing, of the visual characteristics of the sounds they have seen others producing, and of the tactile, kinaesthetic, and somatosensory characteristics of sounds that they have produced.

This information is represented at multiple levels of abstraction in multiple domains of interpretation. Individuals who know English can interpret the duration of the interval of aperiodic energy between the release of a dorso-velar closure and the onset of voicing in a following front vowel as indexing a particular pattern of coordination among the gestures for the dorsal closure, the glottal opening, and the following vowel posture. This coordination pattern, in turn, can be interpreted as indicating the talker's intention to invoke a particular term in the system of paradigmatic phonation-type contrasts (a long voice-onset time indicates that the sound is part of the voiceless series of stop consonants) as well as a variety of syntagmatic facts about the utterance in which the syllable occurs, such as its meter (the /k/ is aspirated in *tom͡cat* but not in *bu͡cket*) and its prosodic phrasing (the voice-onset time in /k/ is shorter in *tom͡cat* than in *Tom's c͡at*). The word form [k^hæt] in an utterance of *Tom's c͡at* is indexed to the class of entities *felis catus*, just as the word form [k^hɪti] is in an utterance of the variant form *Tom's kitty*. Moreover, specific pronunciations of that word form are indexed to attributes about the talkers who produced them. Details of the pronunciation of [æ], for example, can be interpreted as indexing a talker's age and sexual orientation (Johnson 2006; Smith et al. 2008), in perhaps a similar way to the way that the choice of *cat* over *kitty* can be interpreted as indexing the hearer's age, social relationship to the talker, and attitudes towards the class of entities *felis catus*. Details of the pronunciation of the word form as a whole, similarly, may be interpretable as indexing whether the word form refers to the class of entities *felis catus* or a different class of entities: female *homo sapiens* including Tom's wife *Kat*, short for *Katherine*, as differentiated from Dick's sister *Kat*, short for *Katrina* (e.g. Jurafsky et al. 2002; Gahl 2008).

Thus, the categories that are indexed by an utterance of the word *cat* are of at least five types, including (1) the categories of intentions to posture the lips, tongue, glottis, etc., and to coordinate gestures for making different postures at the articulatory-motor level, (2) the spectral patterns and auditory events that a listener parses to perceive the talker's articulatory intentions, (3) the terms in paradigmatic and syntagmatic contrast in the grammar of word forms and phrases of the talker's and listener's shared language, (4) the classes of entities, properties, and events that are indexed by particular word forms and phrases, and (5) the types of social

identities and relationships that are the larger cultural context for utterances that are produced by a talker and perceived by a listener who share a language.

This chapter discusses representations of the sound structure of language across the first part of the lifespan, from the time *in utero* at which the auditory system begins functioning, to early adulthood. Specifically, we focus on how phonological development involves building progressively more abstract structures, starting with raw sensory encodings of the acoustic input that are first encountered in the womb, to the articulatory representations that begin in the first year of life, to the abstract representations that continue to develop throughout the lifespan. A more detailed theoretical justification for our approach can be found in Beckman et al. (2007). Our proposal is similar to an independently developed one presented in Pierrehumbert (2003). Our theoretical stance is that representations are latent variables. That is, representations are objects that can never be observed, but can only be inferred from individuals' overt behaviors. For example, consider aspects of a representation for the word *kitty*, which describes it as comprising four phonemes grouped into two CV syllables, grouped into a trochaic foot. Positing such a representation could relate specific observable aspects of the pronunciation of this word, such as the aspiration of the word-initial /k/, or the realization of the medial /t/ as a flap, and the fact that hearing an utterance of this word can prime the production of phonologically related words like *city* or *kidder* or prosodically related words like *sofa*, to the observation that the word can be segmented into two syllables and four phonemes by many literate speakers of English. While the representation links together a wide variety of observable physical properties and behaviors, however, the representation itself is never observed or even observable. This view is motivated in large part by observations about children's development, namely, that development involves children building phonological representations progressively as the consequence of producing and comprehending speech.

From this standpoint there are two mistakes to try to avoid in the discussion of representations of children's knowledge of speech sounds. One is to rely solely on observational methods, such as phonetic transcription, that inherently invoke models of fully formed adult phonologies. The other is to overinterpret data from other types of observational methods in terms of representations that are motivated by accounts of adult behaviors.

12.2 Developmental changes in speech-sound knowledge

Building a model of phonological acquisition begins with the detailed study of children's knowledge of different aspects of sound structure of language, as well as

of developmental changes in this knowledge. This section will review some of the basic facts about these developmental changes, calling specific attention to (a) the laboratory phonology methods that these studies have used, and (b) the implications of the findings for phonological representations, particularly as they relate to the five sources of categories described in our introduction. In reviewing these facts, we will be using the term *language-specific* to identify evidence that knowledge is generalized from experience with hearing and producing a specific language. When we use this term, we will mean specific to a given ambient language, and not just specific to the capacity for human language in general. For example, evidence that infants growing up in different speech communities behave differently is incontrovertible evidence of learning something about the language's sound structure, and is not just a development of the general capacity of primates (or of all animals) to produce and perceive vocal gestures for social purposes. There is evidence of language-specific categories at the lowest rungs of the "ladder of abstraction" well before there is evidence of knowledge of categories at the higher levels that are more typically associated with the term "phonological" in literature on the phonetics-phonology interface from the last century (see, e.g. Pierrehumbert 1990; Keating 1996). In light of the early evidence of language specificity, an axiomatic approach to this interface is particularly counterproductive for understanding the development of phonological knowledge. That is, with Pierrehumbert et al. (2000/this volume), we think that the sources of "categoriality" are a proper object of study in their own right (see also Beckman and Edwards 2000a). However, that debate is beyond the scope of this chapter, and we will use both "phonetic" and "phonological," as appropriate for the context, in referring to language-specific knowledge at different levels of the ladder of abstraction.

12.2.1 Perception

The development of knowledge of the sound structure of the ambient language begins very early. Infants begin to perceive speech in a language-specific way even before they are born, as evidenced by the finding that newborns prefer listening to the native language of their mothers as compared to another language with different prosodic characteristics (Nazzi et al. 1998). In the first few months of life, infants are also able to perceive many consonant contrasts (cf. Eimas et al. 1987, for a summary of this work). However, the contrasts they can perceive are not language-specific and, in fact, non-human animals can also perceive these same contrasts (e.g. Kuhl and Miller 1978; Kuhl and Padden 1983; Lotto et al. 1997). The first strong evidence of language specificity in perception of speech segments (as opposed to prosody) is observed at around 6 months when infants demonstrate a preference to listen to vowels that are more similar to native-language vowels (Kuhl et al. 1992) than ones that are not. By about 9 months of age, infants stop

attending to consonant contrasts that are not in their native language (e.g. Werker and Tees 1984a). This ability to home in on language-specific contrasts appears to be one of the first linguistic measures of how well an infant is learning the ambient language. Kuhl et al. (2005) found that English-speaking infants who were better able to discriminate a native consonant contrast (between /ta/ and /pa/) at 7.5 months had larger vocabularies at 2 years, relative to infants who were better able to discriminate a non-native contrast (between a Mandarin affricate and fricative). Houston and Jusczyk (2003) showed that infants as young as 7.5 months encode talker-specific information when listening to speech, such that infants were more accurate in recognizing words when produced by the same talker they had been familiarized with on a prior day.

As discussed in section 12.2.3, children's early productions may be subject to phonetic preferences. Two recent studies showed that individual differences in production preferences are mirrored in preferences in perception. Vihman and Nakai (2003) report that the strength with which infants prefer to listen to one contrast over the other is correlated with differences in the production of the same contrasts: children preferred to listen to contrasts that they did not produce. DePaolis (2006) replicated this finding, but demonstrated that it was only operational in children with relatively more advanced vocal development. Children with less advanced development preferred to listen to contrasts that they preferred to produce. Together, Vihman and Nakai (2003) and DePaolis (2006) suggest a reciprocal relationship between early production preferences and attention in perception.

Toward the end of the first year of life, children begin to develop a receptive vocabulary. An influential study by Werker and Stager (2000) found that less phonetic detail is available to young children when a string of sounds can be interpreted as a label of an object than when simply listening to the same string. Fourteen-month-old infants readily discriminate syllables such as [bɪ] versus [dɪ] when they are paired with complex visual displays such as checkerboard patterns, but do not discriminate them when they are paired with pictures of novel objects that the syllables could be labeling. This asymmetry is not shown by younger infants, suggesting that it does not emerge until children have learned a critical mass of vocabulary items. Subsequent work by Werker et al. (2002) supports the suggestion that this asymmetry in discrimination is linked to vocabulary development, in that only those infants whose parents reported them producing at least 25 words (or comprehending at least 200 words) demonstrated the tendency. Those with smaller expressive or receptive vocabularies did not.

Beyond the first year of life, children's speech perception abilities continue to develop. Much of this work has been informed by the hypothesis that children's phonological representations gradually become more fine-grained as they learn more and more words and must differentiate among them (e.g. Metsala and Walley 1998). Hazan and Barrett (2000) tested 6- to 12-year-old children on acoustic continua associated with a variety of consonant contrasts and found

that it was not until age 12 that children's identification functions were as steep as adults. Clayards et al. (2008) argued that the slope of identification functions reflects the shape of the experienced distributions of sounds being categorized. In this model, children's increasingly steep identification functions reflect their accrual of more sharply peaked distributions of sounds along different sensory parameters.

A number of studies have investigated differences in the weighting of acoustic cues by children and adults. In a series of studies, Nittrouer and colleagues (e.g. Nittrouer and Studdert-Kennedy 1987; Nittrouer 1992, 1996; Nittrouer and Miller 1997) found that children's identification functions differed more strongly for /s/-vowel and /ʃ/-vowel sequences stimuli that differed in formant transitions than did adults'. Hence, children's fricative identification appears to rely more on formant transitions, while adults' judgments relate more to the fricative noise. Mayo et al. (2003) found that developmental differences in the influence of formant-transition and fricative-spectra cues for this same contrast were related to developmental changes in phonological awareness. Their seven-month longitudinal study of 5-year-old children showed that improvements in phoneme segmentation and blending (typically the most demanding of the tasks used to measure phonological awareness) occurred before shifts in cue-weighting. Furthermore, phonological awareness, as measured at the earliest time periods of the study, predicted cue-weighting strategies measured at the latest time period of the study.

Burnham (2003) reports a related finding for a measure of language-specific speech perception: the difference in degree of match to the sharp S-shaped curves of "categorical" identification functions between a native-language and very similar other-language continuum. This measure has a U-shaped curve in development, which peaks at around 6 years of age for English-speaking children, and is highly correlated with the same measures of phonological awareness that Mayo and colleagues used, as well as with measures of reading comprehension. This result provides further evidence of the protracted development of language-specific perception skills in the development of phonological categories at higher levels of abstraction that can be tapped in the acquisition of literacy in languages with alphabetic writing systems.

In sum, studies in this section underscore how perception both shapes and is shaped by the acquisition of language-specific phonological categories. Infants' prodigious early perception abilities allow them to encode the acoustic-phonetic detail needed to uncover the consonant and vowel categories relevant for the language being acquired. These in turn shape infants' subsequent perception of the same sounds in a manner that facilitates other aspects of linguistic processing, such as spoken-word recognition and word learning. At the same time, the extremely early onset of language specificity in perception is an important reminder of the importance of looking at the input to the child.

12.2.2 Input

The field of language acquisition has come a long way from the days when the only thing that was said about the input was that children needed a little of it. Input to very young children (variously called "baby talk," "motherese," and "infant-directed speech" or IDS) has now been studied in detail for more than thirty years. Many characteristics of IDS have been observed consistently across a range of studies in a number of languages using different methodologies. IDS relative to adult-directed speech (ADS) has a higher pitch, greater pitch range, shorter utterances, slower rate, and simpler syntax (e.g. Garnica 1977; Fernald and Simon 1984; Lieven 1994; Fernald et al. 1998). One aspect of IDS that is relevant for the establishment of phonological representation is the observation that speech sounds in IDS are hyperarticulated (e.g. Fernald 2000) relative to the forms in ADS. However, this aspect of IDS is not found consistently across studies. While many studies (e.g. Kuhl et al. 1997; Liu et al. 2003) have found that IDS relative to ADS has more extreme formant frequencies for the point vowels /i, a, u/ in many languages, at least one language, Norwegian, does not show this pattern (Englund and Behne 2006). Moreover, the evidence for hyperarticulation of specific consonant contrasts is inconsistent. For example, some studies have found a greater contrast in VOT between voiced and voiceless stop consonants in IDS relative to ADS (Malsheen 1980), some studies have found a smaller contrast in IDS (Sundberg and Lacerda 1999), and still other studies have found no difference (Baran et al. 1977). It has been suggested that hyperarticulation of consonant contrasts might be found in IDS to older infants, but not to younger infants, since older infants would be better able to make use of this information. In support of this claim, Sundberg and Lacerda (1999) found a smaller contrast in VOT for voiced and voiceless stops for IDS to 3-month-olds and a larger contrast in IDS to 11- to 14-month-olds. Cristià (2009) found a similar result in her study of IDS to 4- to 6-month-olds and 12- to 14-month-olds. She examined the contrast between /s/ and /ʃ/ as measured by differences in the first spectral peak of the fricative spectrum and found a smaller contrast between the two fricatives in IDS than in ADS to the younger group and a greater contrast in IDS than in ADS to the older group.

Cristià (2009) suggests that IDS may facilitate speech and language development either through its "affective" or its "informational" component. The affective component of IDS is signaled by its higher pitch and its greater pitch range, both of which are characteristic of positive emotion in speech (Scherer 2003). Infants attend to IDS, at least in part, because they prefer to listen to speech with positive emotion. For example, they prefer "happy" ADS to neutral IDS (Singh et al. 2002). The informational component of IDS relates to factors such as the shorter utterance durations, simpler syntax, and hyperarticulated vowel and consonantal contrasts, which may help infants learn linguistic contrasts. In support of this claim, Thiessen et al. (2005) found that 6- to 8-month-old infants were able to segment speech

into words that they had heard in IDS, but not in ADS. A small number of studies demonstrate that individual differences in IDS predict individual differences in infant speech perception. Liu et al. (2003) found a positive correlation between the size of the vowel space in IDS to 6- to 8- and 10- to 12-month-old Mandarin-acquiring infants and these infants' ability to perceive the contrast between a Mandarin sibilant affricate and fricative. Cristià (2009) found that 12- to 14-month-old infants who were better able to discriminate between /s/ and /ʃ/ had mothers who produced more distinct /s/ and /ʃ/ categories in their IDS.

In sum, the input that children receive during language acquisition provides them with at least some support for early vowel and consonant category formation, in that the signal may exaggerate the parameters that allow the child to uncover the categories in the signal itself. Much research continues to be needed in this area. Though Cristià (2009) demonstrates a relationship between individual differences in category distinctiveness in adults' production and in the perception of children receiving this input, we know of no other study that reports this, nor any study reporting similar effects on children's subsequent productions.

12.2.3 Production

Children's vocal production changes dramatically in early development, perhaps most so during the first two years of life. In the first six months of life, children's productions progress from reflexive vocal behaviors (like crying and fussing) to sustained vocalizations suggesting independent control of respiration and phonation (Oller 1980; Stark 1980). Transitions in this early stage are likely driven by the growing autonomy of the different anatomical and physiological systems used in speech production, as well as neural control of speech production that is separate from control of the same structures in non-speech tasks. Toward the middle of the first year of life, infants' vocalizations become both more varied and more speech-like. This phase culminates in children beginning the rhythmic articulatory movements of early canonical babbling. Phonetic transcription reveals strong consonant-vowel co-occurrence restrictions in this early babbling. These restrictions support an understanding of early babbling as a simple rhythmic wagging of the jaw, with different gross static tongue postures superimposed on the jaw cycle as a whole (MacNeilage and Davis 1990). These co-occurrence constraints gradually relax. The gradual decoupling of consonant place and adjacent vowel quality reflects the infants' growing ability to control tongue movement separately from jaw movement within the jaw cycle.

Toward the end of the first year of life, infants begin to tailor their vocalizations to the characteristics of the language being acquired, as shown in the distribution of consonant places and manners of articulation in the babbled speech

(de Boysson-Bardies and Vihman 1991), the formant patterns in vocoid portions (de Boysson-Bardies et al. 1989), and the tempo and melody of the babble (Levitt and Utman 1992; Whalen et al. 1991). For example, Whalen et al. found that the babble of French-acquiring infants was more likely to contain the rising glissandi characteristic of adult French than was the babble of English-acquiring ones.

The phase in which children babble overlaps with the phase in which they produce their first words. As shown by Vihman et al. (1985), the phonological characteristics of babble and first words are qualitatively very similar: children's production preferences in babbling correspond closely to their preferences in producing real words. Schwartz and Leonard (1982) found that children's early word learning is sensitive to production capacities, with children learning words that contain sounds over which they have productive control more readily than ones over which they do not. More recently, Storkel (2006) reported a similar relationship for children with larger-sized lexicons, demonstrating that children with less rich phonological knowledge continue to restrict their word learning to forms that contain sounds over which they have productive control beyond the fifty-word stage.

During the early stages of multisyllabic word and multisyllable utterance production, languages' prosodic structure also influences production accuracy substantially. It is well documented that English metrical feet have a predominantly strong-weak structure (Cutler and Carter 1987). Children's early productions often delete weak syllables so that the resulting productions fit this foot structure. This is true both in multisyllabic words and multiword utterances (Gerken 1994b).

Children's early words are coarse approximations of the adult forms. Transcriptions of toddlers' word productions are characterized by systematic errors such as deletions and substitutions relative to the target pronunciation. Transcription analyses of children's productions over the pre-school years typically show a rapid change in production patterns until they reach adult-like levels of accuracy for even the most challenging sounds. In large-scale normative studies of speech-sound acquisition in American English, this occurs by approximately age 6 (e.g. Smit et al. 1990).

Acoustic analyses of children's productions give a somewhat different picture of development. Detailed acoustic studies reveal that children's productions of sounds that are transcribed as substitutions for a target sound are often acoustically intermediate between and distinct from both the target sound and correct productions of the substitute. For example, Baum and McNutt (1990) showed that children's frontally misarticulated /s/ were distinct from their productions of /θ/, as well as other children's correct productions of /s/. Scobbie et al. (2000) found that children's productions of target /st/ clusters, transcribed as /t/ or /d/, were acoustically distinct from correct productions of /t/ and /d/. Two recent studies found durational differences between productions that had an apparently

deleted syllable or segment and ones that did not. Carter and Gerken (2004) found that children's productions of verbs followed by a trisyllabic weak-strong-weak word with a deleted initial syllable were longer than those followed by a correctly produced strong-weak word. Song and Demuth (2008) showed that vowels were lengthened prior to deleted final consonants relative to productions without deletions. These findings and others suggest that phonological acquisition is continuous. Consistent with this, Li et al. (2009) demonstrated that Japanese- and English-acquiring children's production of anterior sibilant fricatives involves gradually greater acoustic differentiation across early phonological development.

A parallel body of research has examined developmental changes in speech-sound duration and trial-to-trial variability in children's productions as a way of understanding developmental changes in speech-motor control. The logic underlying these studies is that longer duration and greater trial-to-trial variability indicates less mature motor control. A well-established finding is that children's speech segment durations are longer and more variable than adults' (Eguchi and Hirsch 1969; Smith 1978, 1992; Kent and Forner 1980; Smith and Kenney 1994; Smith et al. 1996; Lee et al. 1999; Munson 2004). This is true even after children have acquired perceptibly accurate speech production. Kent and Forner (1980) showed that speech of children up to 12 years of age is measurably slower than that of adults. Lee et al. (1999) showed that changes in variation in formant frequencies decrease with age throughout the teenage years. Munson (2004) showed that trial-to-trial variability in the spectral characteristics of /s/ decrease throughout the 3- to 8-year-old age range. This finding also holds for kinematic parameters. Smith and Goffman (1998) found that children produce lip movements with greater trial-to-trial variability than adults. Goffman (1999) showed that stability in lip movements differentiated between children with different overall levels of linguistic development. Children with primary language impairments (i.e. language impairments that occur in the absence of a clear predisposing condition) produced non-words with more variability than children with typical language development.

In sum, studies of children's speech production suggest that phonological development takes place over an extensive time period, not simply the first few years of life. It involves the acquisition of productions that are sufficiently adult-like as to be perceived and transcribed as accurate, and it also involves the development of adult-like speech-motor control. It is notable that the time course of development is shown to be considerably more protracted when production accuracy and motor control are assessed by acoustic analysis and direct kinematic measures, than when it is assessed by transcription alone. Hence, studies of the development of production abilities that focus solely on transcribed speech are likely to underestimate the duration and complexity of this facet of phonological development. Developing tools to measure continuous change in children's speech is crucial to gain a fuller

picture of phonological development. These tools include acoustic and kinematic analyses, as well as novel auditory-perceptual rating scales, discussed in further detail in section 12.3.2.

12.2.4 Higher-level phonological knowledge

The studies reviewed in sections 12.2.2 and 12.2.3 document developmental changes in parametric phonetic knowledge of the distribution of sounds in two of the primary phonetic domains, articulation and acoustics. There is a parallel set of more abstract representations that further categorize the speech signal. This section discusses the development of these representations.

The emergence of these more abstract representations appears to be tightly yoked to the developmental expansion of the lexicon. Beckman and Edwards (2000b) and Edwards et al. (2004) studied the relationship between the development of abstract phonological representations and vocabulary growth. These studies examined children's repetition of sequences of sounds embedded in non-words. We reasoned that children could repeat a non-word-embedded sequence of sounds that occurs in many words, such as the sequence /ft/, by referencing the articulatory and acoustic representations for this sequence in known words in which it occurs, like *after* and *fifty*. In contrast, children's ability to repeat sequences of sounds that occur in few or no words, such as the sequence /fk/, cannot be made in reference to lexical knowledge. Repetition of these sequences would be supported by the existence of representations of objects like /f/ and /k/ that had been abstracted away from the sequences in which they occur. Consistent with this reasoning, Edwards et al. (2004) found that the discrepancy between the accuracy of repetition of high- and low-frequency sequences (the *frequency effect*) decreased monotonically in children aged 3:0 (years:months) to 7:11. Further analyses found that the frequency effect was predicted statistically by measures of vocabulary size. Munson, Edwards, and Beckman (2005) showed that the predictive relationship between vocabulary size and the frequency effect held when developmental changes in real-word speech production accuracy and speech perception were controlled statistically. That is, changes across development in the magnitude of the frequency effect appear to be distinct from developmental changes in parametric phonetic knowledge. Recent work by Zamuner (2009) partly replicated this finding and showed that the association is most robust for sounds in word-initial position.

These findings suggest that the emergence of abstract phonological representations in childhood is tied to developmental changes in vocabulary size. One interpretation of the mechanism that underlies this association is that increases in vocabulary size lead to a reorganization of the lexicon along dimensions of phonological similarity. These dimensions become *de facto* representations of the sublexical units like phonemes and syllables. Beckman and Edwards (2000a) and Beckman et al.

(2007) hypothesized that there is a reciprocal relationship between the emergence of these representations and word learning. Their emergence allows children to interpret novel word forms as combinations of known categories. This ability then allows children to form representations for novel strings more efficiently than if these strings were interpreted solely relative to existing articulatory and auditory representations. The fact that the emergence of abstract representations is yoked to developmental changes in vocabulary size gives a clue to *why* they emerge: one of their functions might be to allow the word learner to parse unfamiliar words as sequences of stored sublexical units during word learning. This in turn would facilitate the learning of new words. This conjecture would predict not only the relationship between lexical size and the integrity of sublexical units, but a relationship in turn between the integrity of these units and the trajectory of vocabulary development. A similar argument is made by Pierrehumbert (2003) and by Metsala and Walley (1998), the latter of whom showed that vocabulary growth is related to the development of the metaphonological abilities that support early reading ability. This finding that has not yet been documented in longitudinal research, though cross-sectional studies of word learning have documented that children are more likely to learn high phonotactic probability non-words than low probability ones (Storkel 2001).

Three themes can be extracted from this brief discussion. First, the development of higher-level phonological knowledge is a protracted process. Second, multiple sources of evidence suggest that the development of these representations is yoked to developmental changes in vocabulary size. Third, higher-level phonological knowledge is multilayered, and each of these layers is abstracted progressively further away from the parametric phonetic encodings.

12.3 Emerging areas of research

The areas of research that we see as promising are those that expand on our knowledge of the encodings that children make and the generalizations that they impose on them. This includes cross-linguistic research and research on populations with atypical speech and language abilities. These expand on our knowledge base by examining a fuller range of variation in speech sounds and speech-sound knowledge than is possible in the study of typical speakers of only one language. We also see as promising studies that look at the different types of generalizations that children make over the parametric phonetic signals they have encoded, including those about the attributes of speakers who produce the speech they have heard. Finally, we regard as promising new techniques in measurement and analysis that allow us

to better understand the speech children produce, and the cognitive architecture that allows them to make generalizations about it.

12.3.1 Rethinking universals

The studies reviewed in sections 12.2.2, 12.2.3, and 12.2.4 demonstrate that development involves the accrual of knowledge in the primary sensory domains of audition, vision, proprioception, and somatosensory perception, as well as the development of progressively more abstract representations of this information. This section considers which aspects of these processes are universal, and which are language-specific. As discussed earlier, some language specificity is evident very early in development, even before the transition from babbling to first words. Productions become increasingly language-specific throughout early language development. Rvachew et al. (2006) showed language-specific expansion of control of the vowel space in the second year of life in infants acquiring Canadian French or Canadian English. Buder and Stoel-Gammon (1994) found language-specific productions of /t/ in 2.5-year-old children acquiring Swedish or English. Swedish children produced a more diffuse spectrum for /t/ than English children, mirroring differences between the adult languages. Li et al. (2009) demonstrated language specificity in 2- and 3-year-old children's productions of /s/, mirroring the cross-language differences seen in adults' productions of that sound.

Other recent studies have examined the claim, first articulated in Jakobson (1941), that the order in which children acquire speech sounds is relatively stable across languages and reflects universal constraints on the development and change of phonological systems. More recently, optimality theorists have made a similar claim, namely that all markedness constraints should outrank all faithfulness constraints in early child speech (e.g. Demuth 1995b). However, this claim was not supported by the results of Vihman and Velleman (2002) who examined spontaneous word productions of twenty children (five each for English, French, Japanese, and Welsh). They found that markedness constraints did not dominate faithfulness constraints, and furthermore, they observed language-specific differences in the ranking of markedness and faithfulness constraints. In another study, in which the same word-initial consonant-vowel sequences were elicited across four languages, Edwards and Beckman (2008a) examined 2- and 3-year-old children's productions of lingual obstruents in Cantonese, English, Greek, and Japanese, and found substantial differences in the acquisition of what are ostensibly the same sounds across these languages. Multivariate statistical analyses showed that both language-specific constraints (specifically, the frequency of occurrence in the language) and language-universal constraints (presumably relating to universal ease of production and perception) were needed to account for the patterns across languages.

One example of a cross-linguistic asymmetry is in the acquisition of obstruent place of articulation in English and Japanese. As reviewed by Beckman et al. (2003), large-scale studies of normal phonological development have found different orders of acquisition of /s/ and a corresponding post-alveolar sound: /ʃ/ is acquired later than /s/ in English, /ɕ/ is acquired earlier than /s/ in Japanese. Similar cross-linguistic asymmetries are found in the acquisition of /t/ and /k/. These findings run contrary to the assertion that coronals like /t/ and /s/ have a privileged status in phonological acquisition (e.g. Stemberger and Stoel-Gammon 1991).

Clearly, there are substantial differences in phonological acquisition as a function of the language being acquired. What is less clear, however, is the extent to which the mechanisms that promote phonological category formation and abstraction across languages are similar. Some of the mechanisms described in Section 12.2.2 seem likely candidates for universal applicability in at least spoken languages. Vocal-tract acoustics do not differ from language to language, and thus the non-linear mapping between articulation and acoustics presumably enforces the same discretization of the phonetic space regardless of the (spoken) language being acquired. Moreover, the finding that infants tend to impute categories when given structured input is presumably a reflection of statistical learning abilities that are common to all humans. Maye et al. (2002) and Maye et al. (2008) demonstrated that hearing infants can learn phonetic categories from non-random distributions of voice onset time (VOT) in stops. This learning would presumably occur regardless of the language being acquired. The finding also implies that if a contrast among phonological categories is to be learnable from the input, the shapes of the acoustic-phonetic distributions must reflect the category structure. Of course, the parameters used by languages to convey categories are language-specific. For example, VOT has different utility in characterizing voicing contrasts across languages, as demonstrated by Kong (2009), among many others. However, finding that VOT distributions do not mirror the category structure of the language does not necessarily mean that voicing cannot be learned from the distributions of acoustic parameters such as VOT; it might indicate instead that VOT is not the (sole) parameter whose distribution is useful for inducing voicing categories in that language.

Other mechanisms of abstraction might be candidates for other kinds of language specificity. Consider Edwards et al.'s (2004) finding that phonotactic probability effects in non-word repetition are linked to developmental changes in vocabulary size. To our knowledge, this link has been explored in only one language other than English. Bréa-Spahn (2009) examined relationships between phonotactic probability, vocabulary size, and non-word repetition accuracy in a dialectally diverse group of children acquiring Latin American and Caribbean varieties of Spanish. Bréa-Spahn found that vocabulary didn't mediate the size of the phonotactic probability effect on non-words in children acquiring Spanish in the way

that it had been found to do previously in English. One possibility, suggested by Bréa-Spahn, is that this negative result reflects the nature of the vocabulary test used in her study, which measures knowledge of words that are highly frequent in *written* Spanish across different spoken dialects. Another possibility is that the Spanish lexicon is less conducive to the kinds of generalization that are measured by phonotactic probability effects in the non-word repetition tasks of Edwards et al. and Bréa-Spahn. Spanish has longer words than English, which are composed of simpler syllable shapes than those of English. These words are consequently less confusable with one another than English words. A lexicon with longer, less confusable words might not exert the same pressures on the language learner to form representations with the same level of detail about the specific consonants and vowels that compose each different syllable as does the English lexicon, which comprises shorter words with more complex syllable shapes that allow more types of minimal contrast, such as *cat* versus *scat* and *cats* as well as *cat* versus *pat* and *cap* or *cat* versus *coat* and *curt*. One can imagine even more striking differences between language pairs that differ even more than Spanish and English do (see also Vitevitch and Rodríguez 2005).

The hypothesis that different lexicons lead to qualitatively different types of generalizations is partly supported by recent work by Beckman and Edwards (2010). Beckman and Edwards showed that the frequency of occurrence of different consonants in the ambient language lexicon predicts appreciably different proportions of variance in children's production accuracy of those consonants across a typologically diverse set of four languages. In two of the languages, English and Cantonese, consonant frequency predicts a substantial proportion of variance in children's word-initial consonant production accuracy. In the other two languages, Japanese and Greek, consonant frequency predicts much less variance in consonant production accuracy. English and Cantonese both have large vowel inventories and many monosyllabic words whereas Japanese and Greek both have only five vowels and very few monosyllabic words. These differences lead us to ask whether consonant segments extracted away from the following vowel context are the relevant type of representation for evaluating the relationships between category frequency and accurate production. An answer to this complex question can only come from large-scale cross-linguistic studies of speech-sound development that are informed by rigorous analyses of how languages carve the parametric phonetic space into categories, how these categories function in composing the words of the language, and how the earliest words are distributed along dimensions of similarity and contrast in the overarching structural principles that organize the lexicon for rapid parsing and incorporation of new word forms. A number of promising ongoing research projects are examining this, including the παιδολογος project (Edwards and Beckman 2008) and cross-linguistic work by Vihman and colleagues (e.g. Vihman et al. 2007).

12.3.2 Measurement within and across languages

A second emerging area in research regards the measurement of children's production accuracy and conformity to adult norms. As reviewed in Section 12.2.2, transcription underpredicts patterns of variation in children's speech. Children's productions of what listeners perceive as identical productions are often acoustically distinct (Macken and Barton 1980; Scobbie et al. 2000). However, there are many consonant contrasts for which no standard acoustic measure is yet available. Moreover, even when there is a well-developed acoustic measure (such as VOT for voicing contrasts), acoustic analysis also underpredicts cross-language differences in whether and when children's productions will be perceived as accurate by native speakers of the target language (as demonstrated, e.g. by Kong 2009). Thus it is important to develop measures that can assess children's perceived speech production abilities in more robust detail than most transcription protocols allow.

Recent work with this aim has demonstrated that very subtle differences among types of productions can be assessed perceptually if the right measurement tools are used. For example, Schellinger et al. (2008) examined adults' perception of children's productions of /s/ and /θ/ which had been elicited via real-word and non-word repetition tasks. Children's productions were carefully transcribed by a phonetically trained native speaker of English, and tokens were chosen from six different transcription categories: correct [s] for /s/ productions, [s] substitutions for target /θ/, productions coded as intermediate between [s] and [θ] but closer to [s], intermediate productions judged to be closer to [θ], [θ] substitutions for target /s/, and correct [θ] for /θ/ productions. A group of naïve native-English-speaking listeners were presented with these fricatives and were asked to rate how close they were to prototypical /s/ and /θ/ endpoints using a visual analog scaling (VAS) method. VAS was implemented by presenting listeners with a double-headed arrow bounded by the text "the 's' sound" and "the 'th' sound" and asking them to click at the location on the line that represented where they thought these sounds fell relative to the two endpoints. Quite strikingly, listeners' click locations discriminated among the same six different types of fricatives differentiated by the transcriber.

Other studies have shown that individual listeners' VAS judgments are well correlated with the acoustic parameters that differentiate between endpoints; that they are robust even to different levels of task difficulty (Munson, Kaiser, and Urberg-Carlson 2008; Kaiser et al. 2009); and that they are psychometrically superior to other continuous measures, such as direct magnitude estimates of phoneme goodness (Urberg-Carlson et al. 2008). These measures hold great promise as ways to capture variation in children's productions that reflect a level of phonetic detail not captured by typical transcription protocols, and which better reflect the norms of a community of listeners, rather than the perceptions of a single transcriber.

Such work can also shed light on some of the apparent cross-linguistic differences discussed in Section 12.3.1 (Edwards and Beckman 2008a). For example, Munson, Li, Yoneyama, Hall, Beckman, Edwards, and Sunawatari (2008) examined whether the cross-language asymmetry in fricative acquisition between English and Japanese that Beckman et al. (2003) found is attributable, in part, to differences in how adults in the ambient language environment interpret children's productions. Li et al. found that Japanese-speaking adults accept a narrower range of children's productions (in a two-dimensional acoustic phonetic space) as acceptable tokens of /s/ than do English-speaking ones. This asymmetry suggests that language specificity in how speakers parse the acoustic phonetic space is an additional source of cross-language differences in acquisition. It also emphasizes the importance of not relying solely on native-speaker transcribers' judgments when measuring phonological development.

12.3.3 Socioindexical learning

A third emerging area of inquiry is reflected in studies of children's acquisition of socially relevant variation in language. This work builds on concepts and methods in the growing subfield of sociophonetics, as summarized by Docherty and Foulkes (2000), Hay and Drager (2007), Foulkes (2010), and Docherty and Mendoza-Denton (this volume), among others. This work has shown that individuals can mark their membership in different social groups through distinctive patterns of phonetic variation, and that listeners often use this variation to identify attributes about speakers. Much of this work has examined variation relative to macrosociological categories such as race, ethnicity, socio-economic status, and gender; however, more recent work has examined variation as it relates to local variation, such as social groups within schools (i.e. Drager 2008). The influence of social variation on perception extends beyond tasks in which listeners make inferences about speaker attributes. Studies have shown that listeners' phonetic identification and spoken-word recognition can be biased by the social attributes they impute to the speakers they are listening to (Strand and Johnson 1996; Johnson et al. 1999; Niedzielski 1999; Strand 2000; Drager 2006; Babel 2009; Munson 2009; Staum Casasanto 2008).

Understanding the acquisition of sociophonetic variation is important for at least two related reasons. First, learning the form-meaning mappings for social variants provides an additional level of complexity to the task of phonological acquisition. Given that lexical learning is one source of the richly articulated hierarchy of multiple levels of category formation in regular phonological development, we would predict that sociophonetic learning might drive the development of an additional level of categorization of sounds. That is, when the child hears radically varying productions of ostensibly the same word form in the same prosodic context, and

hears the same pattern of variation in another word form that shares a common subpart, the child is forced to abstract away some category structure for the common subpart that equates its variant forms. A variant of this argument is made by Foulkes et al. (2005). Second, just as abstract lexical phonological representations provide a scaffold for early word learners to parse systematic contextual variability in interpreting new strings of sounds, so might socioindexical stereotypes provide a scaffold for parsing phonetic variability that might otherwise appear to the learner to be random. That is, acquiring knowledge of the social categories allows the child to "parse out" the indexical function of the variation, to further promote unity of the different variants into one category at the cross-cutting level of abstraction from the lexicon. Put simply, socioindexical learning might impact the nature of phonological representations very directly and very significantly.

Children become aware of social variation in speech early in development, as illustrated by Patterson and Werker's (2002) finding that infants become sensitive to relationships between face gender and voice gender late in the first year of life. One large-scale study on the acquisition of socially relevant variation in production is provided by Docherty et al. (2006). Docherty et al. examined variation in medial /t/ variants in the variety of English spoken in and around Newcastle-upon-Tyne, England. Variants of /t/ in that dialect are stratified by sex, with women producing a pre-aspirated variant in medial position more often than men. Children demonstrate this sex asymmetry in production of this variant by about 3.5 years. Smith et al. (2007) examined the acquisition of standard and non-standard pronunciations of the MOUTH vowel (i.e. the vowel in the word 'mouth', using Wells's [1982] lexical-set notation) in a town in Northern Scotland. They found that children acquired the standard [ʌʉ] variant before the non-standard [uː] variant, and that the rate of children's production of [uː] was correlated with the rate that the caregiver used it.

Recently, Li, Kong, Beckman, and Edwards 2008 compared the acquisition of two gender-marked variants in a cross-sectional study of 2- to 5-year-old children. Li et al. showed that the three-way contrast among /s/, /ɕ/, /ʂ/ in Mandarin emerges in 2- and 3-year-old children, and is robustly present already in most of the older 3-year-old children. However, a special gender-marked variant of /ɕ/ only was found only in the 5-year-old girls. That is, the socioindexical marker is acquired only after the regular phonological categories (i.e. those that are used to convey lexical contrast) are acquired. Li and Kong also examined children's production of a gender-marked voicing category in Japanese obstruents. Li and Kong found that not even 5-year-old Japanese-acquiring boys had yet acquired the ability to produce stops with a true voicing lead, which is a gender-marked category in that language. However, true voiced stops in languages such as European French and Thai tend to be acquired late (see e.g. Allen 1985; Gandour et al. 1986). That is, the socioindexical marker had not yet been acquired, perhaps because it involves a regular phonological category that tends not to be adult-like until late in the preschool years.

One hypothesis that emerges from Li and Kong's work is that certain socioindexical markers might be acquired only after children have mastered production of the phonological categories themselves. This result in production is consistent with findings on the development of vowel perception in the first year of life. By about 6 months of age (Kuhl et al. 1991), infants prefer to look at images of faces whose postures match those required to produce the vowels that they are hearing (i.e. infants prefer to view an image of an adult with spread lips matched with the vowel /i/ than with the vowel /u/ or /ɑ/). However, at 6 months, they have no preference for viewing an image of a male with spread lips and a male voice producing /i/, relative to that of an image of a female with spread lips and a male voice producing /i/. It is not until about 10 months of age that infants show a preference for viewing an image of an adult talker matched with a voice of a talker of the same gender (Patterson and Werker 2002). Crucially, the preference for matching the face and voice gender emerges *after* the preference for matching a facial posture and vowel quality emerges, at least at the group level. The time course of regular phonological development and indexical development is clearly an important topic that is ripe for future studies.

12.3.4 Unified models of representations in children with disorders and typically developing children

Models of typical language are made more powerful if they can also account for the abilities of individuals with atypical speech and language abilities, i.e. individuals with speech and language impairments. This section reviews selected studies on the phonological abilities of children with speech and language impairments in light of the models and studies discussed earlier. First, consider children with speech sound disorder (SSD, sometimes referred to as phonological disorder [PD] or phonological impairment [PhI]). SSD is defined as significantly below age-level speech production in the absence of a clear medical or psychosocial etiology, such as hearing loss, intellectual impairment, structural anomaly, or a disorder of neuromotor control.

The error patterns that children with SSD make are often very systematic, and mirror those made by younger children acquiring the same language. SSD provides an opportunity to understand the factors that contribute to variation in pronunciation while holding other factors known to affect pronunciation, like dialect and age, consistent. One consistent finding is that children with SSD have poorer speech perception ability than their age peers. Rvachew and Jamieson (1989) reported that children with SSD had less categorical perception of synthetic /s/-/ʃ/ and /s/-/θ/ continua than their typically developing peers. Edwards et al. (2002) showed that children with SSD require more acoustic information than their peers to discriminate minimal pairs than typically developing age peers, and that their speech

perception is proportionate to the severity of phonological impairment. Children with SSD also have poorer production motor control than their peers, as shown by Edwards's (1992) finding that they compensate more poorly for biomechanical perturbation than their peers with typical development. Hence, the perception difficulties of children with SSD appear to reduce their access to one important source of categories, distributions in the parametric phonetics. They also have reduced knowledge of articulatory-acoustic correspondences.

Three recent studies examined the relationship among measures of the speech-perception ability, speech production ability, and other measures of the phonological abilities of children with SSD. Munson, Edwards, and Beckman (2005) found that children with SSD do not differ from age-matched peers without SSD in the magnitude of the phonotactic probability effect in the non-word repetition task of Edwards et al. (2004), but did differ in the perception measure of Edwards et al. (2002). Again, measures of vocabulary size were better predictors of the magnitude of the phonotactic probability effect in non-word repetition than any other measure, including chronological age. Munson et al. (2010) found that speech perception deficits better discriminated between children with and without SSD than did performance on two naming tasks that had been shown previously to index the robustness of lexical and phonological relationships. Rvachew and Grawburg (2006) analyzed relationships among speech perception, vocabulary size, and phonological awareness in a large cohort of children with and without speech-sound disorder. Rvachew and Grawburg measured children's receptive vocabulary size, their ability to discriminate between accurately and inaccurately produced tokens of sounds in real words, their speech production accuracy, and their ability to make explicit judgments about the sound-structure of words (i.e. phonological awareness). Consistent with Munson et al.'s studies, Rvachew and Grawburg found that children's speech production accuracy was predicted by their speech perception ability. They further found that speech perception predicted both vocabulary size and phonological awareness, and that vocabulary size predicted phonological awareness even after the relationship of those two variables with speech perception was accounted for statistically. Hence, though there is ample evidence that children with SSD do not readily learn acoustic-phonetic distributions, they appear not to have specific difficulty in learning abstract phonological representations from their (presumably poorly phonetically specified) lexicon.

Second, consider children with primary language impairment (LI, sometimes referred to as Specific Language Impairment [SLI]). LI is defined as severe difficulty acquiring morphosyntactic, semantic, and lexical aspects of language in the absence of a clear predisposing condition. Though the primary presenting characteristics of children with LI are in aspects of language other than phonology, numerous research studies have shown that these children have difficulties in aspects of sound-structure. Indeed, Bishop and Hayiou-Thomas (2008) suggested that these seemingly subtle difficulties may be the most clearly heritable aspects of LI. Children

with LI have problems perceiving speech (Ziegler et al. 2005), though the magnitude of this deficit is likely exaggerated by the use of synthetic speech stimuli in these experiments and the memory demands inherent in the tasks conventionally used to measure it (Coady et al. 2007; Coady et al. 2005). Moreover, children with LI produce speech with greater kinematic variability than typically developing children (Goffman 2004). Munson, Kurtz, and Windsor (2005) reported that children with LI have larger phonotactic probability effects in non-word repetition than children with typical development. These were similar in magnitude to the effects in a younger group of children matched for vocabulary size, suggesting that the phonological representations of children with LI were no worse than would be predicted from the number of words they know. That is, children with LI have deficits not only in mapping the parametric phonetic space, but also in acquiring words. These lexical-acquisition difficulties lead to fewer opportunities to form abstract representations of the sound-structure of language.

Together, studies of SSD and LI mirror many of the findings from many of the studies of typically developing children, in that they show the primacy of speech perception in speech production, and the relationship between the size of the lexicon and the type of phonological generalizations that children make.

12.3.5 Computational modeling

Models of normal functioning are made more powerful if they can be implemented computationally. Many existing models focus on learning of relationships between different aspects of the framework we have presented in this paper. For example, Plaut and Kello (1999) present a neural network model of the emergence of phonetic representations for sounds, modeling these as patterns of stable activation in a layer of hidden nodes mediating between acoustic representation and articulatory ones, that are sharpened by the link to semantic contrasts in the lexicon. Oudeyer (2005a) and Westermann and Miranda (2004) present two very different models of how articulatory-acoustic relationships might be learned. One model of the development of higher-level phonological knowledge (here, the learning of gradient and potentially violable constraints on the sound structure) is presented by Boersma and Hayes (2001).

In contrast, relatively few models have attempted to explain the development of different types of phonological knowledge simultaneously. There are two notable recent exceptions. The first of these is Redford and Miikkulainen (2007), a model of the emergence of different syllable structures in development based on a combination of articulatory knowledge, perceptual knowledge, and the rate of lexical access. Oudeyer (2005b) models the development of one aspect of higher-level knowledge, phonotactic constraints, from a model of articulatory-acoustic associations. Again, this topic is ripe for future research, which will benefit from our

increasingly sophisticated understanding of parametric phonetics, acoustics, the lexicon, and the mechanisms of learning, as well as advances in the computational methods needed to develop viable models.

12.4 Conclusion

This chapter presented a framework for characterizing phonological knowledge along multiple sensory domains and at multiple levels of abstraction. Central to this framework is understanding how the information in the primary sensory domains are encoded, and the factors that promote the emergence of higher-order representations that parse this variation into categories. We believe that this framework is useful in understanding how phonological representations change in development. Indeed, the studies that we reviewed suggest that phonological development can be understood as the gradual development of progressively more abstract structures in individuals' representation of language. Perhaps most importantly, though, we hope that the framework and findings reviewed in this chapter convince the reader that the status of something like /s/ as a category is enforced by multiple factors. Facts about the sound's articulation and acoustics, as well as how it functions in the lexicon and in socially situated communication all contribute to its cognitive representation. Moreover, there is ample evidence that adults' knowledge of these categories reflects their lifetime of producing and hearing sounds in words and in social communication. Researchers across disciplines should be cognizant of this when invoking categories like /s/ in other types of inquiry. The behavior of these categories—how they evolve during language change, how they are perceived, and how they are accessed—cannot be studied outside of considering how and why they arise, both in phylogeny and ontogeny.

CHAPTER 13

CHANGES IN REPRESENTATIONS

This chapter addresses the nature of historical change, focusing on how the motivations and mechanisms for language change are informed by experimental work. Chitoran frames these issues by reviewing the question of the source(s) of "naturalness" in phonology. Harrington discusses the role of synchronic variation in setting the stage for diachronic changes. Kirchner introduces an exemplar theory-based model of change, explaining processes of "phonologization."

13.1 The nature of historical change

Ioana Chitoran

13.1.1 Introduction

The nature of historical change is one of the earliest concerns of experimental approaches to phonology. "Why" sound change happens and "how" are key questions that, by the nature of the object of study, invited an experimental perspective, and inspired researchers to incorporate in their approach increasingly sophisticated methods of experiment design, technological advances, and computational models. Since linguistic change can rarely be observed and studied directly, an alternative is to replicate it in the laboratory or to model it, and these endeavors have proven fruitful. This body of work reminds us that the establishment of the Laboratory Phonology conference series in 1987 was the confirmation of a research program, rather than its inception.

This section deals with the representation of historical change. This section addresses three issues. In Section 13.1.2 I review proposed sources of sound change in phonetic variation stemming from speech production and perception, and I discuss their implications for representations. In Section 13.1.3 I discuss the relationship between synchronic and diachronic systems, focusing on the development of phonologization models and on the related controversy surrounding the issue of naturalness. In Section 13.1.4 I outline promising new directions in experimental work on sound change.

13.1.2 Listener-perceived vs. speaker-produced variation: Implications for representations

Approaches to sound change are primarily structured along the major distinction between the *initiation* and the *propagation* of change. The discussion will be limited

to the first aspect, which is directly relevant to the issue of representations.[1] All models agree that sound change happens because variation in speech exists. Where they differ is in determining the relative importance of a particular type of variation: the phonetic variation inherent in the signal produced by the speaker (production-oriented change), or the variation perceived by the listener (perception-oriented change).

Two models of phonetic variation have been influential in explaining sound change: Lindblom's H&H theory (Lindblom 1990; see Harrington, this chapter) and Ohala's phonetic listener-based model. Both are built on extensive experimental work and have defined the main issues and research questions regarding sound change.

Ohala (1974, 1981, 1990, 1992, 1993, 1995) asserts the crucial role of the listener in initiating certain sound changes, citing similar intuitions of early phoneticians in this respect (e.g. Sweet 1874; Durand 1955). For Ohala, the main source of sound change is the misapprehension of the signal by the (possibly inexperienced) listener. Under this view sound change takes place in the acoustic-auditory domain rather than in articulation. Ohala agrees that synchronic variation in speech is found in articulation, but for a sound change to actually take place, to cause a shift in the sound system, the presence of variation alone is not sufficient. The auditory system of the listener is needed to process the signal in such a way that it extracts crucial information from it. In arguing that variation in speech production is not (yet) sound change, Ohala points out that much of this variation is phonetically predictable, and consequently factored out by the experienced listener. This type of variation cannot trigger sound change unless a listener fails to compensate for it, takes the signal at face value, and thus produces a new form, different from the one intended by the speaker. This is the scenario that Ohala labels *hypocorrection*. The listener fails to correct the phonetically predictable variation. A large body of experimental work has confirmed that listeners regularly compensate for phonetically predictable variation, and fail to do so when the variation is not predictable (Mann and Repp 1980; Beddor et al. 1986; more recently Harrington et al. 2008). This last study shows that the ongoing fronting of /u/ in Southern British English can be traced to the effects of a preceding anterior consonant, but that younger speakers fail to compensate for this coarticulation, thus triggering a shift in the boundary of the /u/ category. This is observed in both their production (their /u/ is more fronted than that of older speakers) and in perception (their /i-u/ category boundary is shifted towards /i/). The result is a generally fronted /u/.

[1] Attempts to explain the propagation of change are of course equally important, but are beyond the scope of this section. See, most recently, Crawford (2009) for discussion of innovative modeling approaches to sound change propagation.

A smaller category of sound change falls under the scenario that Ohala calls *hypercorrection*, whereby the listener performs an unnecessary, inappropriate correction of the signal, and ends up producing a new form. Hypercorrection often results in dissimilation. An example is vowel backing after a palatal glide in Slavic (Ohala 1990): /stoj+ā/ 'stand' becomes *stoj+ū*, because listeners mistakenly attribute the frontness of the final vowel to the preceding front glide, and correct it by backing the vowel. The general scenario underlying both of these sound changes can be characterized as "misunderstanding in sound change," in the words of Labov (1994), implying a mismatch between production and perception.

Ohala's model of sound change assumes rich phonetic representations, incorporating details of acoustic-articulatory relations, aerodynamic principles, and principles of how our auditory system extracts information from the acoustic signal. This phonetic model necessarily assumes that language learners have direct access to phonetic detail. At the same time, however, work by other researchers suggests the need to further enrich representations by accommodating symbolic aspects such as hierarchical structure (cf. Pierrehumbert 1990), based for example on ample evidence that the phonetic realization of segments is differentially affected by different levels of prosodic structure (Keating et al. 2003, among others; see Fougeron 1999 for a review). The question of representations more generally is currently one of the most exciting ongoing debates in the field, fueled by several interesting and strikingly different models. Some of these are discussed in the course of this section.

A major difference between Lindblom's and Ohala's models is their (non-)teleological aspect. For Lindblom sound change results from the interaction of two goals: articulatory economy and enhancement of perceptual contrast. For Ohala the initiation of sound change is instead non-optimizing and non-teleological, although he agrees that its spread may well be. Both models have been subsequently integrated by Blevins (2004) in a model which classifies sources of sound change into three categories: *change*, *chance*, and *choice*. In Blevins's CCC model the categories *change* and *chance* derive from Ohala's model, and refer to sound change via misperception and misapplication of a phonetics-phonology mapping. The category of *choice* follows Lindblom's model and refers to change stemming from synchronic variation along a continuum of careful to casual speech (hyper- to hypoarticulation). Blevins maintains Ohala's non-teleological view, except for instances of *choice*. Her argument is simply that the articulatory and perceptual goals are redundant as an explanation, and therefore cannot be definitively proven. It has been proposed, for example, that the common change from palatalized velars to palatals is motivated by the goal of maximizing a phonological contrast phonetically. But Guion (1998) demonstrated, in a series of production and perception studies, that velars before front vowels are easily confusable with palato-alveolar affricates, more so than with velars before back vowels. This invites then the

simple explanation that this sound change is due to purely perceptual conditioning, eliminating the need to refer to a principle of maximizing contrast, which in turn would require further explanation.

So far the emphasis here has been on the listener as a source of sound change. But other phonetic accounts of sound change have placed more emphasis on production. Goldstein (1983) examines common patterns of vowel shift, explaining how patterns of variability consistent with this type of sound change may emerge from the resonance properties of the vocal tract under essentially random articulatory variability. This hypothesis is tested in a simulation using the Haskins Laboratories articulatory synthesizer (Rubin et al. 1981). The proposal is consistent with the difference between stable and unstable regions identified in Stevens's (1972, 1989) quantal theory.

Building on such earlier studies, the model of Articulatory Phonology subsequently developed by Browman and Goldstein (1986, 1990a, onward) includes detailed accounts of sound change. Even though the theory is not developed specifically as a theory of sound change, it has contributed insightful explanations and makes clear predictions about the way in which patterns of speech production can change a phonological system. Specifically, Browman and Goldstein (1991) propose that many cases of sound change can be analyzed consistently with Ohala's model, as well as with their dynamic definition of articulatory gestures as shared primitives ("common currency") in characterizing both phonological patterns and phonetic actions. For example, in Browman and Goldstein's model, reductions in gestural magnitude can account for lenition phenomena, and variable gestural overlap accounts for assimilation and deletion patterns.

But the more interesting cases are still the perceptually based sound changes. In a gestural model these involve reassignment of gestural attributes among temporally overlapping gestures, and misparsing of articulatory movements. Gestural reassignment captures the listener's failure to correctly identify the source of a particular property of the signal, as in Ohala's model. One of Browman and Goldstein's examples of such a failure is the historical change of /x/ to /f/ in words like *cough* and *tough*, pronounced [ko▫x]/[to▫x] at the stage when the consonantal change took place (Browman and Goldstein 1991). Given increased overlap between the second element of the diphthong [▫] and the velar fricative [x], the lip-rounding gesture of the former co-occurs with the frication of the velar gesture. If the frication is attributed by the listener to the labial gesture rather than to the tongue body (velar) gesture, then a labial fricative is more likely to be perceived. The shortening of the diphthong into a monophthong also follows from this analysis, as the narrow labial gesture for rounding is no longer attributed to the offglide.

Gestural misparsing can also explain cases that involve the apparent insertion or deletion of a gesture. An example of the former is the "spontaneous" vowel nasalization in Hindi, in the absence of a nasal consonant (Ohala and Amador

1981): Sanskrit *švāsa* 'breath' > Hindi [sās]. The acoustic and perceptual account attributes the change to the high air flow volume through the open glottis for the fricative, reinterpreted by the listener as nasalization. The alternative gestural account relies on the finding that velum height for oral constriction gestures varies directly with the constriction degree (Bell-Berti 1980), so that in a sequence [sa] the velum lowers rapidly from consonant to vowel. This rapid velum lowering may be misinterpreted as an intended velum-lowering gesture, and misattributed to nasalization. Note that the two accounts are not entirely equivalent, because the gestural one assumes an articulation that is already present (rapid velum lowering) rather than being a perceptual mirage.

The opposite case, misparsing with a deletion effect, is also attested, for example in Shona /kumwa/ > /kumɣa/ (Ohala 1981a). The labiality of [w] can be entirely attributed to the preceding [m], factoring it out and keeping only the velar component. This analysis is entirely equivalent to Browman and Goldstein's gestural one. Two successive labial gestures for [mw] result in a very similar overall lip movement as one single labial gesture. The listener can thus attribute this pattern to a single gesture instead of a pair.

By and large, these models are similar in that they agree on the necessity of rich phonetic representations to capture the phonetic variation identified as the source of the change. Where they can differ is on the issue of speaker-produced vs. hearer-perceived variation. It appears to be harder, in a gestural account, to accommodate those less common cases where an exclusively perception-based explanation has been proposed. This invites concrete experiments targeting the specific nature of the phonetic representations proposed—acoustic or articulatory—that are best suited for a model of variation and sound change. This particular issue, in turn, is especially relevant for understanding the process of phonologization, discussed next, whereby phonetic variation becomes part of the grammar. The question of representation is non-trivial in this respect, since phonologized patterns require symbolic representation.

13.1.3 Relationship between synchronic and diachronic systems: Phonologization and phonetic naturalness

All models agree that sound change has its source in synchronic variation, and its effects can cause not just differences in pronunciation, but changes in the phonological patterns learned by the listener. For this reason, many models of sound change have focused crucially on the process of phonologization. I start from a schematic definition of phonologization: the shift from high phonetic variability to low variability, followed by the development of a new contrast. Most models agree on these basic elements (e.g. Hyman 1976; Kiparsky 1995; Hajek 1997; Blevins 2004; Kirchner et al. 2010; Kirchner, this chapter).

The most widely cited and explicitly formulated model of phonologization is Hyman (1976). The process involves two steps: phonetic variation leading to phonological variation (phonologization), and phonological variation leading to distinctive variation (phonemicization). Hyman's model clarifies and reconciles the production vs. perception perspective: "Phonological change is perception-oriented, even though the seeds for a change may be articulatory" (Hyman 1976: 416).

Hyman's model has the clearest predictive value. Blevins (2004) mistakenly criticizes it for predicting that the evolution of a new contrast always implies the loss of former contrasts, but Hyman does not actually make this claim. What is crucial to this model and the reason why it works, are its three steps, schematized below:

(1)

1	2	3
intrinsic phonetic variation	*extrinsic variation*	*distinctive phonemic stage*

If we follow the three steps in the classic example of the bifurcation of Southeast Asian tone, they correspond to the following changes:

-step 1: voiced and voiceless consonants determine f0 perturbations on following vowels;
-step 2: f0 perturbations are exaggerated, and no longer attributable to universal phonetics. A rule develops: á → ǎ/C [+voice] _____

The transition from 1 to 2 marks the phonologization stage, and at this point the system can go either way. It may keep [pá] and [bǎ] with both the voicing distinction and the two tones in complementary distribution. This is the predicted outcome if the f0 and voicing distinctions are perceived to an equal degree. If, on the contrary, the primary distinction perceived is the f0 rise instead of the consonant-voicing distinction, then the split tones are predicted to survive at the expense of the voicing distinction. This is the attested outcome, and it marks the phonemicization stage (see also Janda 2003), which is the transition from 2 to 3:

-step 3: distinctive split tone develops, consonant-voicing distinction is lost.

Hyman explicitly states the optionality of the contrast loss: "... accompanying every phonologization is a potential dephonologization" (Hyman 1976: 410).

Hyman's model is the most clear to date, despite being developed over three decades ago. There are, however, several alternatives in representing the details of the phonemicization stage as we rethink the nature of phonological representations and categoriality in trying to capture the shift from higher to reduced variability. Kirchner (this chapter) proposes an exemplar-based phonologization model. Another promising way of modeling reduction in patterns of variation, without recourse to a major change in representation, is the gestural model of articulatory phonology. The model can handle precisely the general scenario discussed here by

virtue of its units, articulatory gestures, conceived of as dynamic targets, and thus is well placed to capture changes in variability.

One frequently debated issue related to phonologization is naturalness. All sound change is phonetically plausible because it stems from phonetic variation. In that sense it is *natural*. At the same time, much of synchronic variation reflects diachronic change, so there is also a sense in which one might expect the same naturalness to play a role in synchronic phonology. The issue of naturalness has received much attention from several generations of linguists, starting with early generativists (Stampe 1979; Donegan and Stampe 1979), and has occasionally steered theoretical approaches toward a more functionalist or reductionist angle. The conclusion reached at this point is that naturalness is relevant in the diachronic dimension, but once a pattern is phonologized it becomes independent of its original articulatory and/or acoustic sources. Janda (2003) proposes in fact the term *dephoneticized*, which best captures this particular stage. At this point the pattern may be subject to different principles for which naturalness is irrelevant. This simply means that a debate about the phonetic naturalness or unnaturalness of a phonologized pattern becomes a moot point. Many phonological patterns cannot be directly attributed to phonetic principles such as perceptual salience, ease of articulation, perceptual recoverability (for the most explicit arguments see Hyman 1975, 2001; Anderson 1981).

A series of recent studies have helped clarify this question by examining the loss of the original conditioning environment in phonologization. They show explicitly how phonetically natural as well as unnatural factors interact in phonologization. It is generally understood that exaggeration of production variation severs it from its coarticulatory source. Subsequent loss of the source is taken as an indication of phonologization. Recent studies by Beddor (2007) and Beddor et al. (2007) have shown that the loss of coarticulatory sources is affected by the larger segmental and phonotactic context, which is, of course, language-specific.

The studies by Beddor and her colleagues focus on the loss of the nasal in coarticulatory vowel nasalization in VN sequences. They reveal a systematic covariation between the duration of the nasal consonant and the extent of nasalization in the preceding vowel: in a VN sequence the shorter the nasal consonant, the longer the extent of nasalization on the vowel. This is explained as temporal sliding of the velum-lowering gesture toward an earlier onset in the preceding vowel. This covariation is in turn conditioned by other phonetic elements in the extended context: (i) the voicing of a consonant following the VN sequence; (ii) the duration of the vowel in the sequence. These patterns are summarized in (2).

(2) Trade-off relations between extent of V nasalization and N duration

(i) American English:	Ṽ	**N**	C	*example*
	short	long	voiced	[spɛ̃nd]
	long	**short**	voiceless	[spɛ̃nt]

(ii) American English:	$\tilde{V}$	**N**	*example*
	tense	**short**	[s$\tilde{i}$n]
	lax	long	[sĩn]
Thai:	$\tilde{V}$	**N**	*example*
	V:	**short**	[bẽ:n]
	V	long	[bẽn]

A similar pattern is reported for Italian (Ohala and Busà 1995; Busà 2003), involving the presence of a fricative vs. a stop following VN. There are, therefore, clearly identifiable contexts where N shortens systematically, due to articulatory overlap between a vowel and an adjacent nasal.

Based on the trade-off in production, the authors predicted equivalence in perception. Since the overall amount of nasalization tends to remain constant, listeners should not be sensitive to its location (on the vowel or on the consonant). Perception studies were carried out with speakers of two languages with different timing relations: American English (a language with trade-off in production) and Ikalanga, where no trade-off relations were found, and where NC is traditionally analyzed as pre-nasalized ^{n}C. Natural English and Ikalanga stimuli were edited to co-vary N duration with extent of V nasalization. Listeners of both languages showed sensitivity to the overall amount of nasality, regardless of where it is located in the signal (V or N), supporting the authors' perceptual equivalence prediction. English listeners, however, performed significantly worse on stimuli where vowel nasalization had been kept constant, as they were expecting a trade-off relation between V nasalization and N. They showed the same sensitivity to co-variation when the voicing of C in VNC was varied. Many listeners predominantly perceived a nasalized vowel, especially in the $VNC_{voiceless}$ context, with a shorter nasal.

These results have important implications for diachrony. The perceptual equivalence of nasality on either V or N predicts that coda nasals are resistant to loss, since N can still be perceived even with relatively little nasality. Cross-linguistically this prediction is borne out in languages where the loss of coda N is known to have been slower than loss of other coda consonants. At the same time, however, co-variation in production between V nasalization and N along with perceptual sensitivity to it, predicts that listeners can predominantly perceive a nasalized vowel and no N in the relevant context. It can thus counteract coda stability. This is the scenario that would facilitate listeners' (learners') reconstruction of $\tilde{V}$ for /VN/, leading to loss of N as coarticulatory source.

These studies show convincingly that the loss of the conditioning environment in sound change is not a direct consequence of phonologization, but rather the result of repeated interaction between synchrony and diachrony. Certain phonetic contexts favor shorter nasals and heavier V nasalization, leading to systematic synchronic alternations, which in turn condition a sound change, and a phonological change of introducing contrastive vowel nasalization in the system. At the

same time, language-specific timing relations and language-specific phonotactics ("unnatural" factors) can affect the course of the sound change by determining to what extent a given context occurs in a language.

13.1.4 Experimental work on sound change: New directions

The studies presented above illustrate the combined research directions advocated by Ohala (1974: 353): ". . . there are (. . .) two types of experiment: that of the man-made controls and that of the nature-made controls." Ohala's own model is built primarily on human-made controls, the nature-made ones being much harder to come by. But whenever these are available they offer a valuable testing ground for hypotheses, allowing the study of historical change in a living laboratory of synchronic variation. These are the types of studies I would like to highlight next. All are experimental phonetic studies of sound patterns in one language that can inform historical development in that same language or a different one.

The potential of this crossover approach was noted early on by Pierre Delattre (1946) in a paper whose very title heralds Ohala's direction: "Stages of Old French phonetic changes observed in Modern Spanish." The paper is essentially a list of thirty-one well-documented sound changes of Old French for which an equivalent phonetic stage can be found in Modern Spanish, a language that is not a direct descendant. Delattre advocates the study of diachronic change by observing patterns of synchronic variation. At the same time, such studies can answer more general questions about speech production. A relevant example is glide strengthening in initial position: /j/ > /ɟ/ or /dʒ/ (later reduced to a fricative in Modern French) and /w/ > /gw/, as in Gmc [wadja] 'wage' > Gallo-Roman [**gw**aɟja] > Mod. French *gage* [gaʒ]. Synchronic variation in Spanish reflects this pattern: e.g. [**gw**esos] is frequently heard for *huesos* 'bones.' Experimental investigations of such patterns can indeed clarify many historical questions. In the case of the Old French palatal glide it is hypothesized that it went through a stage where the affricate was palatalized. If a similar co-occurring palatalization is observed in Spanish, it would strengthen the hypothesis regarding the diachronic change in French. The labio-velar glide raises a question with even broader implications for typology and for speech production. If strengthening involves increased constriction at one end, and there are two ends available (the lips or the tongue dorsum), what makes languages choose one over the other?

A number of phoneticians have responded unwittingly to Delattre's call, through sophisticated experimental studies that have accomplished even harder tasks: contributing to the reconstruction of sound systems whose historical development is much more sparsely documented than that of French. An excellent example is the thorough acoustic analysis of Athabaskan stops by McDonough (2003) and McDonough and Wood (2008), which informs the historical evolution of that

system. In Athabaskan languages the stop series has a three-way contrast: aspirated, unaspirated, ejective—traditionally written *d, t, t'*, *g, k, k'*. The authors argue that the aspirated stops are not phonetically aspirated, but are instead affricates with a velar release [tx, kx]. They consider this to be the native, inherited pattern. While this analysis had already been suggested by earlier linguists (Haas 1968; Young and Morgan 1980, 1987), McDonough's (2003) acoustic study of Navajo stops and McDonough and Wood's (2008) investigation of stops in five different Athabaskan languages confirm the analysis experimentally. In these languages aspirated stops are shown to have long, heavily fricated releases. A quantitative analysis of the *t* [tx] release spectrum finds it to be no different from that of the velar fricative [x]. The sound change *t* > *k* in some Athabaskan languages is therefore best interpreted as [tx] > [kx], as in Navajo [tx^{w}o] vs. Jicarilla [kxo] 'water.' The detailed acoustic analysis of synchronic Athabaskan data thus reveals valuable information about the historical development of the sound system of this language family.

Another successful instance of this approach is Moreton and Thomas's (2007) instrumental analysis of diphthong raising in American English, which they compare to the better known case of Canadian Raising. Their study challenges the view that the sound change arises from the Great Vowel Shift (GVS). They propose instead that it begins as voicing-conditioned variation in the offglide (rather than variation in the nucleus, predicted by GVS). They hypothesize that voiceless codas favor assimilation of the /ai/ nucleus to the offglide, resulting in raising, while voiced codas favor assimilation of the offglide to the nucleus, resulting in lowering. Testing requires a longitudinal study, showing that at some point assimilation to the offglide overtakes assimilation to the nucleus.

While such a study is no longer possible on the Canadian variety of English, the authors used recordings of speakers from Cleveland, OH, where similar alternations are observed. Comparative acoustic analyses of archival and new recordings showed that F1 of nuclear [a] lowers by 2 percent before voiceless codas for speakers born before 1910, and by 14 percent for speakers born after 1965. At the same time, the offglide itself shows a similar rise over time. These results favor the new hypothesis (voicing-conditioned assimilation to the offglide increases over time) over the GVS hypothesis, whereby the nucleus would be primarily affected. The authors have compared phonetic variation in one variety of English to the similar pattern of a completed sound change in another variety. This methodology has led to radical rethinking of a well-known diachronic explanation in light of new synchronic data.

A similar approach is used by Chitoran and Hualde (2007) in comparing phonetic variability across five modern varieties of Romance to determine the historical development of vowel sequences and diphthongs—e.g. [pi.a]-[pja]. By interpreting the experimental results in the diachronic context of each language, the authors show that each synchronic variation pattern reflects a different stage within the evolution of a hiatus-diphthong contrast, so that all varieties are staggered at different points along a hiatus-diphthong continuum. Systematic duration measurements

revealed phonetic factors common to all varieties, that are responsible for the variation: initiality effects (vocalic sequences are longest in word-initial position), proximity to stress effects (stressed vocalic sequences are longest, followed by pre-tonic sequences, and by pre-pre-tonic ones—*di**á**cono* > *di**a**gráma* > *di**a**gonál*). The different patterns currently observed in the five Romance languages are shown to follow from the interaction of three independent factors: (i) a general articulatory tendency for hiatus to resolve to diphthongs, due to the relative stability of diphthongal articulations; (ii) phonetic (prosodic lengthening) effects which inhibit the shift from hiatus to diphthong; and (iii) system-internal (lexical attractor) effects of pre-existing diphthongs in a language from different historical sources.

13.1.5 Conclusions

Even though laboratory methods cannot be applied directly to sound change, a significant component of experimental work has directly informed our understanding of the interplay between diachrony and synchrony. Replicating sound change in the laboratory, testing hypotheses about sound change in one language against synchronic variation in another language, and computational modeling of sound change, all allow for the investigation of subtle details of diachronic change. These methods have significantly deepened our understanding of change. At the same time, new findings in this area invite a rethinking of gradience and categoriality as the turning point in the question of representations. Returning to the definition of phonologization as the shift from high to reduced variability, the emerging view is that of categoriality as illusion, while phonological systems evolve by organizing gradience.

13.2 The relationship between synchronic variation and diachronic change

Jonathan Harrington

13.2.1 Introduction

The dramatic effect of sound change on phonology can suddenly be brought into sharp focus by place names in which the relationship between spelling and

pronunciation can be especially opaque. Consider, then, the possibly amused reaction of a local inhabitant to an unwary visitor who asks for directions to the English villages of Cholmondeston, Happisburgh, or Wrotham, that is to /ˈtʃʌmstn̩/, /ˈheɪzbrə/, or /ˈruːtəm/. The extent of synchronic variation can be no less dramatic, as exemplified by the numerous phonetic forms associated with different meanings of *I do not know* that can reduce to little more than a sequence of three nasalized schwas differing minimally in phonetic height (Hawkins 2003, this volume).

Diachronic and synchronic variation are evidently linked, especially at points in the utterance at which synchronic variation is high: moreover, as will be discussed later in this contribution, diachronic change may be propagated by imitation. For example, the considerable variation synchronically in syllable-final compared with syllable-initial stops (e.g. Byrd 1996b) has a reflex in the greater tendency for diachronic place assimilation (Latin: *scriptu* > Italian *scritto*) and vowel-consonant blending (e.g. the development of nasal vowels in French) to occur in domain-final position (Ohala and Kawasaki 1984; Ohala 1990b; Hock 1992). Similarly, Beckman et al. (1992) show how many prosodically induced diachronic changes such as intervocalic stop lenition and the deletion of weak vowels and syllables can be explained using the same mechanisms of gestural overlap, hiding, and truncation that form part of modeling speech production synchronically in the task dynamic model (see also Browman and Goldstein 1992).

One of the aims in developing a phonetic model of sound change is not only to find evidence for such diachronic-synchronic relationships but also to explain the way in which fine-grained, continuous variability in speech communication can give rise over a much longer timescale to the change from one abstract symbolic category into another. A question closely related to this is the following: if sound change is not planned, i.e. is not teleological at the level of conversational interaction (Ohala 1993b; Lindblom et al. 1995), then how does it come to be that, far from being random, similar sets of patterns of phonological change such as vowel chain shifting can be observed in languages and their varieties? As a first step towards answering these difficult issues, we present a brief overview of the types of synchronic variation that are likely most relevant for understanding sound change.

13.2.2 The nature of synchronic variation

Synchronic variation is ubiquitous and occurs for a number of different reasons. First, the fact that we can speak intelligibly when performing different activities such as giving a lecture, talking while riding a bicycle, running, or taking part in immediate compensation experiments (Riordan 1977) in which one or more articulators are artificially constrained, shows the plasticity of the speech production system in adapting to different environments (Lindblom 1990). Secondly, variability is predicted from the non-linear relationship between articulation and acoustics:

for example, a back vowel like [u] can be produced with a variable constriction location without the variability having very much effect on the resonances which are critical for the vowel's perceptual identification (Stevens 1989; see Iskarous, this volume). Thirdly, the evidence for cue trading in perception experiments (Repp 1981) suggests a certain degree of variability is tolerated in the production of speech: thus the intervocalic stop voicing distinction can often be achieved by co-varying the extent of voicing in the closure, the duration of aspiration/frication, and the direction of the first formant trajectory (Lisker 1986).

A fourth type of synchronic variability, which has been central to Ohala's (1993b) model of sound change, comes about because of the biological and physical constraints on the speech production and perception mechanisms and the interaction between the two. Thus, laryngeal tension at the onset of voiceless stops tends to carry over synchronically into the following vowel, causing the fundamental frequency to be raised: such variation has been shown diachronically to be related to the phonological development of tone in many Asian languages (Hombert et al. 1979, also Chitoran, this chapter). It is often possible to relate this type of synchronic variability to cross-linguistic patterns in the distribution of sounds. Consider that both the infrequent occurrence of high, compared with low, nasalized vowels and the tendency for high vowels to lower if they are nasalized diachronically (e.g. the development of Latin *una*/*unus* into feminine /yn/ but masculine /œ̃/, rather than /ỹ/, in French) can both be related to the same perceptual constraint introduced by nasalization: nasalized vowels produced with a high tongue position are nevertheless perceived to be phonetically lower because of the introduction of a nasal formant intermediate in frequency between the first two oral formants (Beddor et al. 1986; Wright 1986).

An important characteristic of the above fourth type of variability is that it arises involuntarily due to factors like biomechanical inertia and limitations on the perceptual system (and this is one of the main reasons why Ohala 1993b argues that so much of sound change is not cognitive and therefore not teleological). But there is a fifth type of synchronic variation that evidently does not fit into this non-cognitive category and that is more directly associated with a range of meanings that are conveyed by the speaker. Speakers evidently vary the clarity of their speech in relation to how predictable the speech signal is for the listener (Lindblom 1990). This is different from the fourth type of variability, both because a speaker has control over the extent of reduction of a phrase or word (as the earlier example of Hawkins 2003 *I don't know* demonstrates) and because it can provide listeners with cues about the "newness" of the information (Fowler and Housum 1987). Included within this group might also be phonetic variation due to different kinds of prosodic structure, such as the syllable-dependent "clarity" of /l/ in many English varieties, the degree of strengthening of consonants at different phrase boundaries (Keating et al. 2003), as well as the numerous cues arising from conversational interaction (Local 2003), such as the phonetic markers to indicate whether a speaker

has reached the end of a speaking turn. It is this fifth type of variation that is primarily implicated in Lindblom et al.'s (1995) model of sound change discussed in further detail below.

The preceding type of variation is primarily linguistic. But there is also a sixth type of non-linguistic variation that provides information about the speaker, including the emotional state and attitude of the speaker, as well as regional and social information which have been the primary data in many sociolinguistic investigations of sound change (Labov 1994, 2001). But this type of variation has not until recently found its way into phonetic models of sound change primarily for the reasons amplified in Docherty and Mendoza-Denton (this volume) that sociophonetic variability has been marginalized in developing cognitive models of speech production and perception.

13.2.3 Phonetic models of sound change

One of the most influential accounts of the relationship between synchronic variability and diachronic change is due to Ohala (1981a, 1993b) in which, as discussed in further detail by Chitoran (this chapter), many sound changes are attributed to the listener's misparsing of coarticulation. This idea is also central to the analysis of sound change in Browman and Goldstein (1991), based on temporally overlapping gestures in the framework of articulatory phonology (Browman and Goldstein 1992), also discussed by Chitoran.

Like Browman and Goldstein (1991), Lindblom et al. (1995) give greater emphasis to the role of the speaker in sound change as well as to the idea of volition. A central aspect of their model is that sound change arises along the continuum from hypo- to hyperarticulation. In normal conversational interaction, listeners typically attend to *what* is being said (the "what" mode) whereas *how* something is said is not usually the focus of attention. It is when the how mode is especially active that a listener may sample a new pronunciation variant and add it to the lexicon. A prediction of their model is that a listener might add a new pronunciation variant at points of hypoarticulation when the variability in the speech signal is high: this is because, given that these also tend to be points of low information content, the "what" mode is to a large extent disengaged, as a result of which the "how" mode is active. Thus, a very interesting aspect of Lindblom et al.'s (1995) model is that it makes quite explicit how information redundancy, high production variability, perception, and sound change might be interconnected. In Lindblom et al. (1995), the lexicon is assumed to include multiple variant pronunciations sampled from those that are perceived from language use in everyday conversation, and it is this aspect of their model that also foreshadows the similar idea in exemplar theory (Pierrehumbert 2003, 2006a) that the lexicon stores considerable amounts of non-redundant information and fine phonetic detail.

Since words and phrases of high frequency are more likely to be produced in a semantically redundant and therefore hypoarticulation context, and since this is also one of the main contexts in which new pronunciation forms are presumed to be absorbed into the lexicon, then, according to Lindblom et al. (1995), high-frequency words should undergo sound change earlier than low-frequency words (see Hooper 1976 and Philips 1984, 1994 for compatible evidence). The idea that the progress of sound change is linked to lexical frequency is also central to Bybee's (2008) usage-based model which is founded on the idea that linguistic structure is created as language is used. In this model, sound change takes place in words and phrases as a result of the reduction of gestures through repetition. The change to lexical items is modeled in an exemplar framework in which a cluster of new phonetic variants for a word are updated. Bybee's (2008) model and indeed any conceivable model of sound change in exemplar theory is founded on the "*fact* [emphasis added] that articulatorily motivated sound change takes place earlier in high-frequency words than in low-frequency words" (Bybee 2008: 115). However, the empirical evidence showing that sound change applies earlier in high-frequency words is still somewhat sparse. For example, the analysis by Dinkin (2008) of an extensive amount of formant data from the Telsur survey of American English (Labov et al. 2006) found little evidence to suggest that the Northern cities vowel shift is affected by lexical frequency and it is also questionable whether these effects, if they exist at all, really are lexical (Cohn 2005).

In both Ohala (1993b) and Lindblom et al. (1995), sound change at the level of the individual is phonetically *abrupt* because the change from the old to the new pronunciation takes place in one discrete jump, rather than in gradual increments between the two. On the other hand, regular sound change for Labov (1981, 1994) is the result of the *gradual* transformation of a single phonetic property in a phonetic space. Mowrey and Pagliuca (1995) also favor an interpretation of sound change as gradual and consider that claims for abruptness are an artifact of orthographically based, categorical approaches to sound change. In Blevins (2004), sound change that originates from "choice" or "chance" in her model can occur "without noticeable changes in pronunciation or with gradual changes in pronunciation." Mowrey and Pagliuca (1995) present arguments based on neuromuscular activation to show how even metathesis may be gradual. Recent experimental evidence from performance errors is relevant in this regard in showing how many perceived categorical errors are gradient, such as /k/ for /t/ substitutions in which the tongue-dorsum gesture for /k/ intrudes incrementally upon the tongue-tip raising gesture for /t/ (Goldstein, Pouplier, Chen, Saltzmann, and Byrd 2007; Pouplier 2008).

The acoustic analysis of the Christmas broadcasts of Queen Elizabeth II in Harrington et al. (2000) and Harrington (2006, 2007) showed not an abrupt, but incremental phonetic change within the same individual over a fifty-year period. Moreover, these changes in the Queen's vowels were gradual and over a long time period (in some cases of over thirty years) and quite possibly at such a slow rate

(e.g. an estimated 60 Hz per decade for /u/) that they may well be imperceptible within a short time period. Mielke (2007) quite rightly comments that these averages per decade might obscure different changes within words or within individual years, and proposes instead an analysis by year or by word. It is of course very difficult to do this, because the Queen did not always produce the same content words in abundance from one year to the next. However, some formant data for the most frequently occurring word with an /æ/ nucleus per annum in the Christmas broadcasts, *family*, is shown in Figure 13.2.1 over a twenty-year period between 1952 and 1972. These data are suggestive of a gradual change of around 250 Hz between 1952 and 1964 in /æ/ within the same individual producing the same word. The change is not likely to be due to the physiological effects of aging (in which F1 lowers rather than raises—see Harrington et al. 2007) and it is in the same direction as the community change in which the /æ/ was reported to be becoming more open in this period (e.g. Gimson 1966).

This type of diachronic phonetic change found within the same individual seems to be quite reminiscent of the Neogrammarians' analysis of sound change as incremental and quite possibly not perceptible, at least not from year to year.

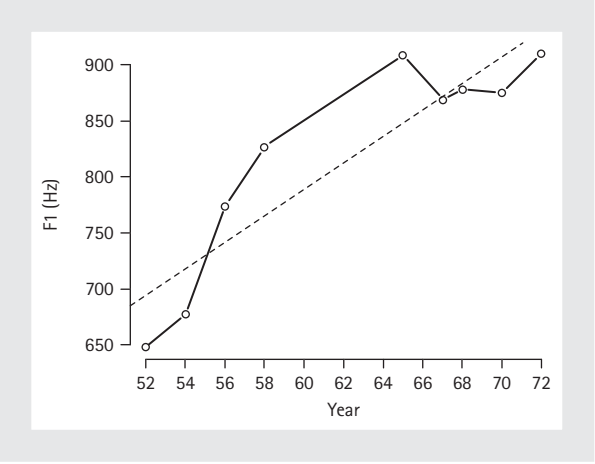

Figure 13.2.1. F1 at the temporal midpoint of /æ/ averaged across all productions of *family* that occurred in any one year. The data are from Queen Elizabeth II producing the annual Christmas broadcasts (Harrington et al. 2000). All productions of *family* were prosodically accented. Data points are only shown when there were at least three tokens of *family* in any one year.

13.2.4 The relevance of imitation for modeling sound change

Both the above studies of the Christmas broadcasts as well as other longitudinal investigations (e.g. Sankoff and Blondeau 2007) show that adults beyond the critical age of language acquisition are influenced by diachronic change taking place in the community. Labov (2006) comments that the extent of adaptation is greater in children than in adults and may even diminish in adulthood with increasing age. This is, as Docherty and Mendoza-Denton (this volume) note, an issue that needs further investigation. Another is how these kinds of phonetic adaptations to community changes in adults come about at all. Some results and conclusions from recent studies on speech imitation may begin to provide a solution to this issue. For example Silverman (2006) argued on the basis of acoustic and perceptual data that the sound change by which rounding spreads diachronically across velars, but not alveolars, in Trique could be accounted for by imperfect copying or imitation, but not by an intentional modification of phonetic variants.

In the imitation paradigm, subjects' speech is compared before and after they have performed a task such as shadowing or listening to another speaker whose speech may have been artificially manipulated in some way. There are now various experiments showing that the subjects' speech production is shifted subtly towards the speech that they have listened to, after they have engaged in the task. One of the first to show this was Goldinger (1998), who demonstrated a shift, as judged from whole-word perception experiments, towards the shadowed speech. Shockley et al. (2004) replicated this experiment and additionally showed that the imitation in the shadowing experiment was phonetic: subjects shifted their VOT in the direction of shadowed material in which VOT had been artificially lengthened. More recently Nielson (2007) extended these experiments by demonstrating just such imitation in a listening task in which subjects were recorded before and after they listened to a speaker whose VOT in /p/-initial stops had been lengthened. Nielson (2007) showed not only that subjects' VOTs were lengthened after listening to these stimuli, but also that the imitations generalized to /k/, even though /k/-initial words had not been part of the stimuli they had listened to. Interestingly, Nielsen (2007) did not find an imitation effect when subjects listened to stimuli with shortened VOTs, possibly because any such imitations would encroach too much upon the acoustic-phonetic space of the corresponding voiced stop phoneme. In quite a different kind of experiment, Pardo (2006) demonstrated a phonetic convergence between interlocutors who took part in a conversation in the Map Task paradigm (Anderson et al. 1991). Finally, subjects in Delvaux and Soquet (2007) had to name an ideogram, X, in sentences like *il y a une X dans le pot* 'there is an X in the pot' produced in a different variety of Belgian French. Their attention was therefore on X, but what was measured was the /o/ in *pot* 'pot' which was repeated from trial to trial. Thus, imitation took place in a repetitive and semantically predictable context. Recall from the discussion earlier that this is exactly the kind of context in which the pool of variants is likely to be sampled, leading to a potential sound change in

the model of Lindblom et al. (1995). This is a context in which the "how" mode is strongly activated, allowing novel pronunciations to be suggested to the listener, just as they would have been in the experiment by Delvaux and Soquet (2007), given the strong phonetic differences between the two varieties in the production of /o/.

The production-perception mechanisms that might be responsible both for these kinds of subtle unwitting shifts in imitation and regular sound change are interpreted by Goldinger (1998) in terms of an episodic/exemplar model of speech perception in which lexical items are built out of auditory traces of words accumulated in long-term memory. However, Pardo (2006) rejects an exemplar-based analysis of the greater convergence because her speakers' imitations were not tied to any specific lexical item. Moreover, since features and phonemes in exemplar theory are supposed to be emergent statistical generalizations across word-based stored exemplars, it is not clear how the results in Nielsen (2007) or Shockley et al. (2004), in which phonemes or even features are imitated over a short timescale, could be accounted for by a shift of stable feature- or phoneme-based generalizations that have been built up over the speaker-hearer's lifetime, at least not without invoking an adaptable phonological processor in speech perception and production that is independent of the lexicon (e.g. McQueen et al. 2006).

In a direct-realist model by contrast, imitation can be modeled as a natural consequence of the presumed "common currency" of layered gestures that are invoked in both speech production and speech perception. As discussed in Fowler (2000), perceiving gestures might serve as a prime or goad for their imitation in production, analogous to the spontaneous imitation of facial expression (see also Fowler et al. 2008 and Shockley et al. 2009 for further discussion). Moreover, based on analyses of the remarkably slight discrepancy between choice and simple reaction times in speech-shadowing tasks, Fowler et al. (2003) propose that speech perception can have an immediate influence on speech production without recourse to cognitive processing. For these reasons, a certain degree of imitation is predicted to occur in the direct realist framework as an automatic consequence of perceiving the same abstract speech gestures that control speech production. This direct link between perception, action, and imitation is also central to Sancier and Fowler's (1997) explanation of the slight shift in the VOT of Portuguese and American English stops produced by a bilingual Portuguese-English speaker after the speaker had left the USA to spend several months in Brazil. They model this change as the result of a realignment of the speaker's laryngeal-supralaryngeal phase relationships, induced by the perception of gestures in the ambient Portuguese language environment.

13.2.5 Imitation and sociophonetic constraints

At the same time, imitation cannot be entirely automatic. For example, Mitterer and Ernestus (2008) found that subjects imitated primarily phonologically relevant

detail, suggesting a somewhat looser coupling between production and perception than implied by Fowler et al. (2003). In addition, Babel (2009) has recently shown that the extent of subjects' imitation is conditioned by social factors: for example, they tend to imitate speakers more if they have a positive attitude towards them. As far as sound change within the individual is concerned, we also have to explain, as Labov (2006) comments, not only why younger speakers seem to adapt their speech to a greater extent in moving to a new community than older speakers, but also that some adults adapt very little or perhaps not all. Moreover there is so far no explanation for the incompleteness of sound change in the Christmas broadcasts: in Harrington, Fletcher, and Beckman (2000), we found that there was shift towards, but not an attainment of, less aristocratic, mainstream RP vowels and certainly no evidence that the Queen adopted what were, for much of the twentieth century, stigmatized phonetic variants such as the London Cockney glottal stop in place of syllable-final /t/.

There might therefore be a sociophonetic regulatory system that prevents imitation and sound change from applying blindly. The model of Lindblom et al. (1995) incorporates various forms of feedback to evaluate the potential sound change for its articulatory, auditory, and sociophonetic cost. The first two of these could be conceived of in terms of the regulatory feedback or feedforward systems that have been proposed at the level of the speaker and hearer (e.g. Guenther and Perkell 2004) but Lindblom et al. (1995) also invoke a more abstract community-level feedback which measures and regulates the sociophonetic consequences of the potential imitation and sound change (see also Pierrehumbert 2003 for a similar idea). The sociophonetic regulator would presumably resemble a filter allowing imitation to pass from hearer to speaker, except for a "blacklist" of allophones whose copying was prohibited. But this seems to accord the speaker a great deal of volition in deciding what to imitate, which may be incompatible with the results from some of the studies discussed above showing that imitation takes place largely without the speaker's awareness.

Perhaps the way forward is to abandon the idea of a sociophonetic regulator and instead to recast the mechanism of imitation and the transmission of incremental sound change as a by-product of the way that speakers use language in conversation to interact with each other in solving cooperative tasks (Giles et al. 1991b; Clark 1996; Garrod and Doherty 1994). For example, Garrod and Pickering (2009: 293) discuss how, when subjects interact with each other in some form of coordinated action or task, then imitation and entrainment are likely "at many different levels, from basic motor programs to high-level aspects of meaning." They also emphasize that the influences are largely automatic, so that interactants are typically unaware of the alignment processes. In this model, the linguistic and motoric alignment of speakers facilitates cooperative action: that is, the success of solving a task collaboratively requires the development of a macroscopic structure in which the individual speakers' action plans are fused in a common goal, and it is this shared plan between

interactants that either brings about, or is facilitated by, an alignment between them at various linguistic and motoric levels.

As Krauss and Pardo (2006) comment, while these types of models are informed principally by linguistic imitation, they also make a number of interesting predictions concerning phonetic convergence that have for the most part not been tested. One of them is that the more speakers are able to engage successfully in cooperative tasks, then the more likely it should be that they influence each other resulting in greater phonetic adaptation. Conversely (and compatibly with the results from Babel 2009 discussed above), adaptation should be less likely when cooperation fails, or in tasks involving speakers who are unsympathetic to each other. Also, the influence on adaptation and sound change of a more passive medium such as television might be expected to be comparatively negligible (but see Stuart-Smith 2006) precisely because there is no macroscopic action plan between the television and the recipient. These issues might help explain why speakers differ in their extent of adaptation when exposed to a different variety for a longer period. Finally, the reason why the Queen has not embraced Cockney-style glottal stops would not be because of an internalized sociophonetic monitor banning the uptake of these allophones, but instead because the opportunities for Her Majesty to engage in conversational, cooperative task-solving with members of the Cockney-speaking community have probably been quite scarce in the last fifty years.

13.2.6 Modeling sound change in an interactional, self-organizational system

The previous section suggests, then, that imitation is a consequence of cooperative interaction between two individuals and that sound change may be a derivative of such imitation. But how exactly can we explain how gradient synchronic variability ultimately produces phonologization, i.e. a categorical shift? From the above considerations, it is immediately apparent that if we view imitation and sound change as a consequence of joint cooperative action between speakers, then any model based on versions of the speech chain, in which the speaker is compartmentalized from the listener, is not likely to yield many interesting solutions to this problem.

An alternative approach—and one that is more compatible with models of cooperative interaction discussed in the preceding section—is to situate both sound change and the relationship between phonology and phonetics in terms of a model of self-organization (see Lindblom et al. 1984 for an early application of self-organization to speech). As discussed in Oudeyer (2006), a common theme in systems of self-organization, which have also been used to predict many phenomena in nature such as the formation of ice crystals and the cathedral-like formation of termite nests, is that a macroscopic structure emerges as a consequence of the interaction between subcomponents of the system (Shockley et al. 2009). Thus for

speech, computational models consisting of agents with simplified vocal tracts and hearing systems are sometimes used to show how phonological structure emerges from the cumulative effect of many similar, imperfect imitations over time between speakers and hearers (e.g. de Boer 2001). In this kind of model, a language's phonology is not determined by innate principles of universal grammar but is instead just one of the many possible solutions to the way in which convergence arises from speaker-hearer interactions (see Bybee 2008 for a related interpretation that linguistic universals and specifically the principal of structure preservation in Kiparsky's 1985 theory of lexical phonology are not innate but arise through the interaction and change). As a specific example, the notion in universal grammar of phonologically unmarked vs. marked is recast in Kochetov's (2008) self-organizational model as stability vs. instability respectively in the relationship between the production and perception systems. Notice also that in this kind of model there is no sense in which phonology precedes, or is translated into, phonetics. Thus in Gafos (2006) and Gafos and Benus (2006) both the discrete and continuous aspects of complex systems are related using the same formal language of the mathematics of nonlinear dynamics whose properties of differential equations are used to express the relationship between category stability and change (see also Gafos and Goldstein, this volume).

Blevins and Wedel (2009) present a model of sound change based on self-organization, albeit within the exemplar framework (see also Wedel 2007). In their model, sound change is the outcome of the opposition between noise in the production-perception feedback loop which can cause category instability by introducing new phonetic variants and a so-called reversion to the mean which, through processes such as motor entrenchment (Zanone and Kelso 1997) and the perceptual magnet effect (Kuhl et al. 1991), maintains category stability. Blevins and Wedel (2009) argue that, in accord with one of the main principles of self-organizational systems, sound change occurs when the phonological system is in an unstable state. Similarly, Kochetov (2008) used computer simulations to show that the combination of complex vowel inventories and secondary consonant articulations is unstable as far as production-perception relationships are concerned: as a result, the language self-organizes to a more stable system with either rounding contrasts in the vowels or secondary articulations in the consonants (or neither of these).

In a self-organizational model, change can, but need not, be driven by social forces. This is because, even within a socially stable system, there is nevertheless a randomness in the way speakers and hearers interact with each other, and this kind of noise can be shown mathematically to push a phonological system from one stable state to another (e.g. Gafos and Benus 2006). Alternatively, a stable phonological system can be made unstable because of the changing speaker-hearer interactions that might result either from a realignment of the social structure in the community or from dialect contact. Notice that in neither case is there

any sense in which sound change is teleological or planned, precisely because the randomness of speaker-hearer interactions implies that there is an unpredictability in the phonological reorganization that they give rise to.

13.2.7 Concluding remarks

Sound change seems to be imperceptible and non-teleological at the level of speaker-hearer interactions but organized and apparently purposeful at a macroscopic level. The following components of phonetic models of speech production and perception have been especially useful for modeling this dichotomy. (1) Listeners make unintentional errors in processing the acoustic signal in which coarticulation is misparsed. (2) Phonetic variants may be especially salient at points of high information redundancy in the speech signal. (3) Speakers and listeners imitate each other unwittingly and this may be one of the mechanisms that incrementally transmits sound change. (4) Imitation may be a consequence of shared macroscopic action plans between a speaker and listener in conversational interaction, and may be affected by social forces. (5) A self-organizational model is most likely to be compatible with (4), in which phonological category stability and change are emergent properties of speaker-hearer interactions.

Finally, the gradient modeling of the shifts that result both from sound change (Browman and Goldstein 1991) and the perception of the ambient linguistic environment (Sancier and Fowler 1997) in a gestural model provides a way of thinking about sound change that goes beyond the categorical shift from one IPA allophone to another, and that may turn out to be as fruitful as the recent recasting of categorical performance errors as gradient shifts of gestural intrusion and reduction (e.g. Goldstein, Pouplier, Chen, Saltzman, and Byrd 2007; Pouplier 2008).

13.3 Modeling exemplar-based phonologization

Robert Kirchner

Chitoran (this chapter) presents several laboratory phonology studies of sound change and phonologization which compel a re-examination of standard assumptions about gradience and categoricality. Harrington (this chapter) presents evidence of the incremental nature of sound change. In this section, we confront the

implications of these results for standard phonological theory, and suggest a way forward.

13.3.1 W(h)ither phonology?

At the nexus of the phonetics-phonology interface and synchrony-diachrony issues lies the problem of phonologization, standardly conceived as a diachronic development whereby gradient phonetic patterns come to be reanalyzed as patterns over symbolic representations (Hyman 1975, 1976). Renewed attention to phonologization, particularly in the evolutionary phonology framework of Blevins and Garrett (1998, 2004), has cast doubt on the very centerpiece of modern phonological theory, the markedness constraints of Optimality Theory. Simpler grammatical models are possible, the argument goes, if the phonological formalism need not concern itself with questions of typological markedness or phonetic naturalness, leaving that job to diachronic interaction with the phonetic component, which is needed in any case, as is argued in Chitoran (this chapter). Consider a phonetically sensible rule such as k → k^j/__{i, e}. Phonological systems tend to include rules like this, rather than, say, k → m/__{i, e}, and the former is widely attested as a sound change, simply because it arises from phonologization of gradient coarticulation, whereas there is unlikely ever to be a pattern of phonetic variation between /k/ and an [m]-like allophone to serve as grist for reanalysis. A phonological markedness constraint favoring assimilatory dorsal fronting is therefore superfluous. The research programs of phonetically based Optimality Theory (e.g. Hayes et al. 2004) and functional phonology (Boersma 1998), though coming at this issue from the opposite direction—incorporating phonetics more tightly into phonological theory and analysis—seem, ironically, to confirm the evolutionary phonology verdict. Striking resemblances have been found, in every domain of phonological typology,[2] between the substance of well-attested phonological patterns and lower-level phonetic variation, which relate straightforwardly to phonetic considerations such as articulatory undershoot, gestural overlap, aerodynamics, auditory salience, etc., such that there appears to be no domain of pure phonological markedness, autonomous from phonetics.

At this point, it is useful to remind ourselves exactly what work the phonology module (now divested of responsibility for markedness generalizations) does in this division of labor. The reason we speak of some patterns as being phonologized is that they display *categoricalization* and *stabilization*, which are difficult to account

[2] With the probable exception of metrical phonology, which seems to reflect a rhythmic cognitive faculty (cf. Tilsen 2009) distinct from articulatory and perceptual phonetic considerations. This rhythmic faculty, however, does not serve as an example of autonomous phonological markedness, insofar as rhythm is found in many extralinguistic domains of human (and animal) behavior, such as limb movement.

for in terms of purely phonetic factors. The notion of articulatory undershoot (Lindblom 1963), for example, can explain gradient vowel reduction, where the degree of centralization varies continuously with speech rate (and any other factors affecting articulatory displacement/velocity). But it cannot, by itself, account for categorical reductions of the sort discussed by Crosswhite (2004), e.g. a distribution of vowels with a cluster of points around [ə], and other clusters around full vowel values, but with few points in between (see generally Pierrehumbert 1994b on the instrumental interpretation of discrete vs. gradient variation). Nor can phonetic factors such as undershoot explain why phonologized processes are conditioned by coarse phonetic context, particular relatively stable cues such as stress placement, rather than fine phonetic detail which may vary from token to token, such as precise vowel duration. In a typical categorical vowel reduction, for example, the [ə] fails to revert to a full vowel even in slow, careful speech, when articulatory velocity considerations are less pressing. Both categoricalization of the variation and stabilization of the context can be accounted for by assuming that the phonologized reduction pattern is stated over a different level of representation from the gradient pattern.

To answer the question posed in the previous paragraph: this is in fact the *only* work that phonology appears to be doing—if by "phonology" we mean a symbolic level of representation for sound patterns and its attendant theory—and it does it by brute force. The observation of categorical and stable behavior is obtained simply by stipulating that the structural descriptions of phonologized patterns are limited in reference to a small set of discrete, symbolic units. Moreover, this assumption does not come with any intrinsic account of how phonologization occurs. At some point, under this story, speakers reanalyze patterns of variation, from numeric to symbolic terms; but what mechanism induces this shift, and what factors in the original phonetic pattern is it sensitive to? And if phonologization is merely an arbitrary reassignment of a pattern from one level of representation to the other, why don't we observe this development in reverse: "phoneticizations" of originally stable categorical patterns?[3]

Indeed, this standard view, on closer examination, encounters a number of immediate difficulties. How do we reconcile this abrupt shift from numeric to symbolic patterns, which the standard view presupposes, with the incremental nature of sound change, discussed in Harrington (this chapter)? Moreover, is the distinction between phonetic and phonologized patterns really as clear-cut as the foregoing discussion implies? Phonologization might instead be a matter of degree, ranging

[3] The editors suggest that near mergers might represent such a case of phoneticization. Near mergers, however, involve blurring of a lexically conditioned distinction (in some or all contexts), not a contextually conditioned pattern of variation going from categorical to gradient application. Phoneticization, in my intended sense, would correspond to, e.g. a final devoicing alternation pattern which is categorical, perhaps neutralizing, at one stage of a language, and a variable, partial devoicing pattern, sensitive to fine phonetic detail, at the next stage. In all the controversy about incomplete neutralization in final devoicing (see e.g. Warner et al. 2004), no one has suggested a historical development from categorical to gradient application as the explanation.

from low-level, slightly speaker-controlled variation at one end of the spectrum, to categorical, stable, perhaps somewhat morphologically conditioned alternations at the other. The two-level assumption forces a choice between phonetic and phonological analyses of any given pattern, thereby precluding elegant treatments of partially phonologized patterns (cf. Pierrehumbert et al. 2000/this volume; Cohn 2006). As an example of the latter, consider consonant lenition in Florentine Italian (Giannelli and Savoia 1979; Kirchner 2004):

- Voiceless stops, /g/, and affricates /tʃ/ and /dʒ/ obligatorily lenite to continuants in "weak position" (i.e. roughly intervocalic within an intonational phrase);
- but the outcome of this lenition varies gradiently from close fricative to Ø, depending on place of articulation, speech rate, and register;
- additional consonants undergo various forms of lenition in weak position in faster/more casual speech;
- and lenition expands beyond weak position in faster/more casual speech.

The categorical aspects of this pattern, spirantization of voiceless stops, /g/, and the affricates, are just the tip (indeed, three separate tips) of an iceberg of quantitative phonetic variation. I suspect that partial phonologizations will prove, upon sufficiently close examination of patterns in a broad range of languages, to be the rule rather than the exception.

To state the problem in another way: Pierrehumbert (1994) observes that virtually every case of gradient allophonic variation which phoneticians have investigated has proven to be, in some respects, language-specific (cf. discussion of Beddor 2007 and Beddor et al. 2007 in Chitoran, this chapter). How do these gradient patterns arise? There must be some mechanism whereby purely physiologically determined (and therefore language-independent) patterns of variation come to be incrementally enhanced, in language-specific ways. Here the phonologization problem resurfaces in a slightly different guise; but in this case we cannot attribute the development of the pattern to a shift in level of representation, for this is all quantitative variation within the phonetic component. On the other hand, a model of this quantitative enhancement of low-level variation presumably could handle phonologization as well. Categoricality can be regarded merely as an advanced stage of enhancement, "the discrete limit of [a] continuous process," as Pierrehumbert et al. (2000/this volume) put it, without resort to a symbolic level of representation.

In sum, the assumption of a symbolic phonological level of representation is neither necessary nor sufficient for an account of phonologization. The time is ripe, therefore, to consider the outlines of a theory of phonology in which sound patterns (categorical, gradient, and intermediate) are stated not over symbols, but directly over numeric auditory and articulatory signals. This move promises to resolve the debate between evolutionary phonology (which objects to the massive redundancy of phonologized markedness constraints that replicate phonetic factors), and phonetically based Optimality Theory (which objects to a phonological theory that

makes no markedness claims, and that pushes markedness issues into the realm of unformalized meta-theory). The new phonological theory directly includes quantitative phonetic factors in the scope of its formalism; if the research program is successful, phonological markedness generalizations would emerge from the interaction between direct phonetic factors and the pattern-generalizing properties of the speech-processing system.

13.3.2 Exemplar theory

Contemporaneous with this debate about phonological markedness constraints vs. phonetic factors, an alternative conception of phonology (indeed, of grammar generally), namely usage-based, or exemplar theory, has been put forward, most comprehensively in its application to phonology by Bybee (2001), and most explicitly by Pierrehumbert (2001a). The essence of exemplar theory in phonology is massive storage of exemplars: memories of individual experiences of speech, including fine phonetic detail. Linguistic categories are not represented as symbols, but as "clouds" of exemplars associated with category labels. Speech recognition involves a calculation of distance in phonetic space between an auditory stimulus and the stored exemplars, and the application of a classification rule to these distances. Pierrehumbert, for example, adopts the *k*-Nearest-Neighbors rule (see Mitchell 1997: ch. 8, for a machine-learning perspective on kNN and other "instance-based" classification rules). If $k = 10$, and the ten exemplars closest to the stimulus have the word category labels {'pit,' 'pet,' 'pit,' 'pit,' 'pet,' 'pit,' 'peat,' 'pit,' 'pet,' 'pit,' 'pit'}, then the modal category, 'pit,' is chosen as the output of recognition.

Exemplar-based speech production, in turn, involves generation of an output based on mean phonetic properties of the exemplars of the target category. Taking the notion of speech production seriously, this output should include a motor plan, i.e. a matrix of vocal tract muscle group activation levels over time. The output would also include an auditory target signal; comparison of the auditory target signal to actual auditory self-perception provides an error signal for feedback purposes (Moore 2007); cf. Flemming (1995), arguing for parallel auditory and articulatory representations in phonology. At this stage of exemplar theory's development, however, modelers have either used toy numeric representations (e.g. Pierrehumbert 2001; Wedel 2004), or acoustic signals (e.g. Johnson 1997b; Kirchner et al. 2010) as proxies for the auditory/motor signals which are in principle required.

The production output may also be influenced by phonetic pressures, such as bias towards articulatory reduction. Moreover, the exemplars may be weighted by recency and semantic appropriateness, or tagged with indexical or pragmatic variables, for purposes of production and recognition.

13.3.2.1 *What are the categories?*

Word token frequency effects discussed in Harrington (this chapter), such as Bybee's (2001) *every-memory-mammary* reduction pattern, and Goldinger's (1996, 2001) low-frequency imitation result, motivate the identification of the exemplar category labels with *words*. On the other hand, an adequate exemplar-based theory of phonology must be able to capture sound patterns that pertain to smaller domains than the whole word, most obviously *segments*. Pierrehumbert (2002) therefore assumes that exemplars are also parsed into sub-word-level phonological unit categories, such as segments. A whole-word exemplar of 'pit,' for example, might simultaneously be classified, in its initial portion, as an exemplar of the segment [p^h], etc. I argue below that this resort to *a priori* phonological unit categories is in fact unnecessary, given a production model in which portions (of any size) of exemplars can be compared to portions of other exemplars, for purposes of pattern generalization (see Section 13.3.3).

13.3.2.2 *Motivation for exemplar theory*

For an overview of the growing body of (principally experimental) literature motivating exemplar-based phonology, see Gahl and Yu (2006) and articles contained therein, as well as Port (2007) and Johnson (2007). See also Moore and Maier (2006), and De Wachter (2007), for exemplar-based approaches to automatic speech recognition. A few illustrative effects are considered below.

13.3.2.2.1 *Incremental sound change*

Assume a variable phonetic bias which causes the output for a particular word to deviate from previous exemplars, to a greater or lesser extent, in direction X. This output is immediately classified as a new exemplar of the target word, shifting the category mean subtly towards a more X-like pronunciation. This is incremental sound change at the word level.

13.3.2.2.2 *Frequency sensitivity*

Assume that the phonetic bias above is one of articulatory reduction. The more often a word is produced, i.e. the higher its token frequency, the more often its outputs are subjected to the reduction bias, the more the word category mean shifts towards a reduced pronunciation. This is Bybee's *every-memory-mammary* reduction pattern.

13.3.2.2.3 *Imitation: recency*

Assume exemplars are weighted by a factor which decays over time. Recent exemplars therefore have a stronger influence on production outputs than older exemplars, particularly if the target word is of low token frequency, because there are fewer countervailing exemplars within the word category to resist the recent exemplars' influence. This is the Goldinger effect.

13.3.2.2.4 *Imitation: individual and sociophonetic variation*

In pragmatic contexts where imitation of a particular individual or identification with a particular group is important to the speaker, exemplars tagged as productions of that individual, or other group members, can be up-weighted, resulting in a temporary shift in productions towards speech characteristics of that individual or group (see Harrington, this chapter; Stuart-Smith 2007a). Similarly, hearers can tune their perception to speaker and sociophonetic variation by up-weighting stored exemplars of the speaker, or of other speakers with similar social characteristics.

13.3.2.2.5 *Generalization to other words*

As noted above, any phonologically adequate production mode must be able to access and compare portions of word exemplars, not just whole words, to one another. One basic thing that phonological knowledge allows us to do, as humans, is to produce words that we have never uttered before, e.g. repeating a word just learned from another speaker. At the point of hearing this new word (and recognizing it as such), the individual acquires an exemplar encoding an auditory experience of the word, but no corresponding articulatory experience. Without articulatory information for this word, no motor plan can be output to the speaker's vocal tract. This deficiency can only be overcome by generalizing: in exemplar terms, by assembling a novel motor plan for the target word based on portions of exemplars of other words with similar auditory targets. Let us assume that this technique allows us to compare, *inter alia*, segment-sized portions of exemplars to one another, and to pool similar exemplar portions in order to form generalizations. Thus, for all the word-level effects described above, it should be possible to model corresponding segment-level effects.

13.3.2.2.6 *Phonologization as pattern entrenchment*

A basic property of the production model is that it generates outputs based on mean properties of the relevant clouds of exemplars. Whatever the initial variance, generating exemplars in the neighborhood of the cloud mean results in the cloud's distribution sharpening about the mean, i.e. progressively less variance. This result is called "pattern entrenchment" in the exemplar literature.

Putting the entrenchment idea together with the segment-level generalization story above, we must ask how conflict is resolved if a temporal portion of the exemplars of the target word exhibits one phonetic pattern, but the preponderance of similar portions of exemplars of other words exhibit a contrary pattern. Consider a pattern of affrication of /t/ before a high front vowel. Assume the existence of a word whose exemplars happen not to conform to this pattern (e.g. [ati]), whereas most exemplars of many other words conform to the pattern ([bot^{s}i], [t^{s}igama], [arat^{s}iketo], etc.). In a usage-based approach, pattern strength is assumed to depend on a trade-off between *similarity* (inversely proportional to

distance) and *frequency*. In the scenario sketched above, the exemplars within the target word cloud are all relatively similar to each other in their entirety, favoring an output which conforms to the word cloud pattern, i.e. [ati]. By comparison, the other exemplars containing a [t^{s}i] sequence are not very similar to one another globally, and consequently they exert no unified pull on the output, except for the pattern itself. Considerations of global similarity therefore favor maintenance of the word-level pattern. On the other hand, there are many more exemplars containing a [t$^{(s)}$i] sequence than exemplars of the target word. Considerations of frequency therefore favor extension of the affrication pattern to the output. The winning pattern would depend on the actual numbers in the exemplar corpus. If the target word has a high token frequency, it may resist the general pattern (cf. Bybee's 2001 observation that high token frequency licenses phonological and morphological exceptionality in words).[4] Otherwise, the output will succumb to the affrication pattern, resulting in a new exemplar, [at^{s}i] for this word. The mean for this word category accordingly shifts slightly in the direction of affrication. The affrication outcome in future productions of this word now has the combined pull of the external exemplars, and this internal exemplar as well, making further affrication outcomes increasingly likely for this word (assuming there are no relevant competing patterns), and eventually obligatory. This is lexical diffusion of the affrication pattern. Once the pattern reaches a critical mass among the words of the lexicon, the foregoing dynamics make it inevitable that the pattern will become obligatory for all words that meet its structural conditions, hence Neogrammarian regularity as the end stage of the sound change. Moreover, this generalization of the pattern is independent of its phonetic origin: rather than occurring to varying degrees depending on tongue blade trajectory and intraoral pressure, we arrive at an affricate allophone which is stably conditioned by a following high front vowel. This is phonologization.

13.3.2.2.7 *Word vs. segment recency effects*

Putting the phonological generalization story above together with recency weighting, a recency effect should be stronger if the recent exemplars and the target word are of the same type, and weaker if they only share similar portions, such as a segment. In the latter case, the recency effect will be diluted by the greater number

[4] Bybee further claims that the propensity of a pattern to generalize depends on type frequency (how many word types instantiate the pattern). Type frequency is not directly computable in a pure exemplar-based model, as there are no types per se, only sets of tokens. Cliff and Kirchner (in progress), however, show that a type frequency effect emerges from an interaction between aggregate token frequency (the total number of tokens instantiating the pattern) and similarity. In brief, if the exemplars instantiating a pattern are phonetically diffuse, as would be the case if they are scattered over a large number of types, a target word need have little in common with each of the pattern-bearing exemplars in order to be subject to the pattern. If however the pattern-bearing exemplars are tightly clustered into a few types, the target word will either have to conform to one of the types in all respects, effectively becoming homophonous with it, or it will escape the pull of the pattern entirely.

of relevant non-recent exemplars. But with sufficiently high recency weighting, the recent exemplars could have an observable effect even on the production of different words containing the same segment, the experimental effect reported by Delvaux and Soquet (2007) discussed in Harrington (this chapter). But how do we reconcile Goldinger's recency effect, which was limited to low-frequency words of the same type as the stimuli, with Delvaux and Soquet's recency effect, which extended beyond the stimulus word types? It may be significant that Delvaux and Soquet's effect reportedly persisted on the order of several minutes, whereas Goldinger's persisted for two weeks. In exemplar-based terms, the recency factor might decay to the point where the stimuli's effect becomes negligible at the segmental level, but still is strong enough to exert an effect at the whole-word level.

13.3.2.2.8 *Caveat*

It must be borne in mind that the foregoing claims about the behavior of an exemplar-based speech-processing model are merely conjectures on what seem to be likely outcomes, given input patterns with certain characteristics. Most assessments of exemplar theory's capacities, particularly in the experimental literature, unfortunately remain at this impressionistic level. Clearly, the phonological results of an approach involving computation of numeric similarities over various differentially weighted clouds of exemplars cannot be demonstrated with any rigor in the absence of an explicit computational model, implemented and tested on real speech data sets.

13.3.2.3 *Treatment of the time dimension*

Exemplar theory's development, however, has suffered from the lack of an explicit model capable of applying to real speech. The problem is that speech is variable-length time-series data. A slow-speech exemplar of a word may be considerably longer than a fast-speech exemplar. The discussion of exemplar-based comparison above assumed that it was possible to calculate the distances among a set of exemplars. But how does one calculate a distance between items of unequal length?

The recognition side of the model is not the central problem. A number of exemplar-based recognition models have been put forward, from Johnson's (1997) original X-Mod to the large-vocabulary continuous automatic speech recognition system of De Wachter (2007). A recognition model merely needs to assign a category label (or a sequence thereof) to an input signal based on its similarity to the variously labeled speech exemplars in memory. That is, matching whole-word exemplars to whole-word exemplars. This distance calculation can be done optimally using classic dynamic time warping (DTW), a well-understood computational technique for aligning two variable-length signals, locally stretching or shrinking subsequences within one to best fit the other (see generally Sankoff and Kruskal 1983).

Most of the phonologically interesting attributes of exemplar theory, however, pertain to the production side of the model. Production involves a harder problem: generation of a concrete output signal from a target word category (or a sequence thereof). We have already observed that an adequate exemplar-based processing model must be able to identify patterns obtaining over portions of exemplars. Thus, we have the further problem of deciding how to identify and compare similar portions of different exemplars, which may likewise be of variable length, and may begin and end at different points in relation to the start or end of the exemplar. Pierrehumbert's (2001) exemplar model deals only with static, fixed-dimensional data, and so does not address the variable-length problem, in either recognition or production. It is not clear how Pierrehumbert's model might be extended to real speech. To recap, the production system needs to be able to generalize, but how can it generalize over a collection of unique, variable-length speech signals?

One response to this problem, adopted (but not computationally fleshed out) in Pierrehumbert (2002) and Wedel (2004), is to appeal to less time-variable units, such as segments. Segments can be characterized, albeit crudely, in terms of relatively static phonetic targets. Thus, if our exemplar system parses signals into segment as well as word categories, we can pool together all exemplars of, e.g. /s/, reduce these to fixed-dimensional vectors representing the phone "target" (perhaps with contextual target measurements as well), abstracting away from temporal variation within the exemplars. We can now generate an output based on an average of these fixed-dimensional vector values. However, this move raises the non-trivial problem of how these categories come to be established. Moreover, this segmentation into *a priori* phonological units seems contrary to the spirit of exemplar theory. Phonological units such as segments are simply local patterns obtaining over speech signals, involving relatively stable correlations between auditory cues and articulatory gestures. Such units, like all phonological patterns, should emerge bottom-up from comparison over the exemplars, rather than being treated as primitives. Moreover, this approach fails to do justice to the rich dynamic structure of speech.

Rather than segmenting the dynamic signal into quasi-static chunks, one might adopt a dynamic computational technique *ab initio*. DTW, useful for calculating distances between whole exemplars for recognition purposes, can also be used, with certain enrichments, to solve the problem of identifying similar portions within exemplars, without resorting to *a priori* phonological unit categories. The remainder of this section describes the **P**honological **E**xemplar-**B**ased **L**earning **S**ystem (PEBLS) model of Kirchner et al. (2010).

13.3.3 PEBLS

To generate an output for a given word, PEBLS begins, as in Pierrehumbert's model, by randomly selecting an exemplar from this word class for use as the input. The

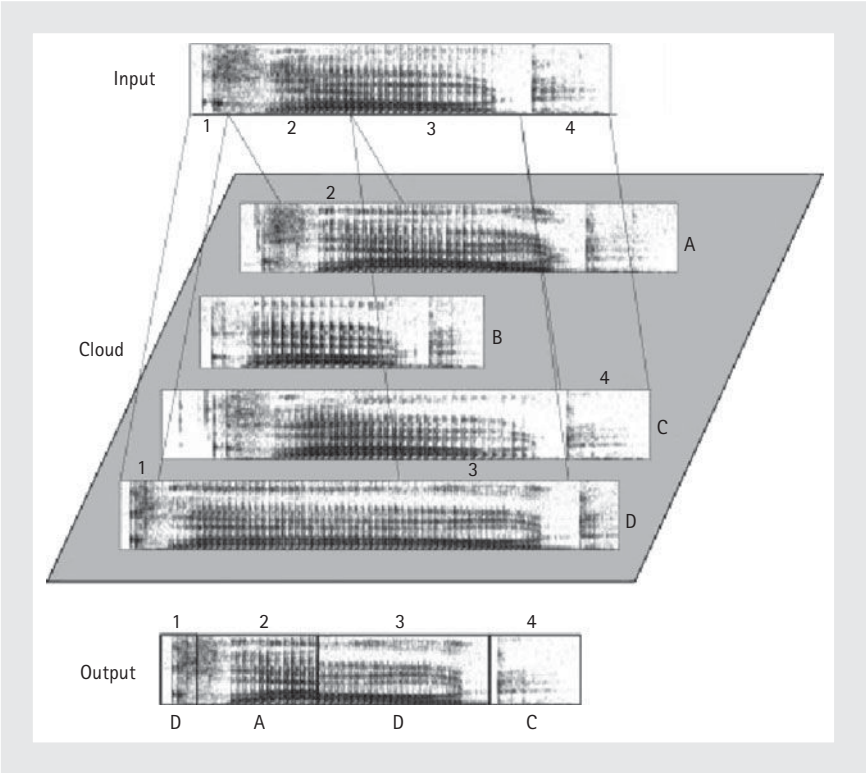

Figure 13.3.1. Output as alignment of input with cloud. Numbers indicate corresponding subsequences within the input and cloud, and the concatenation of these subsequences which form the output. Letters show the particular exemplar from which each output subsequence was taken.

production problem can now be cast as finding an optimal *alignment* between the input and the word cloud. That is, the output is constructed from subsequences of the cloud exemplars which more-or-less correspond to subsequences of the input, and which more-or-less reflect typical subsequences (i.e. generalizations) within the cloud, as schematically represented in Figure 13.3.1.

The challenge lies in specifying an alignment criterion that can find these subsequences. PEBLS builds upon the DTW technique, with two particular innovations. Firstly, whereas DTW aligns a whole signal to another whole signal, PEBLS allows alignment of any frame[5] of the input with any frame of any exemplar within the

[5] That is, the signal is pre-processed into a spectrographic or quasi-spectrographic representation (in Kirchner et al. 2010, actually a mel-frequency cepstral representation), where the representation consists of a sequence of frames, each frame representing an analysis of the acoustic signal during a fixed-width time window. The similarity calculation used throughout PEBLS is based on Euclidean distance between frames.

cloud—transitioning forward or backward in time within any given exemplar, or from part of one exemplar to another. This move permits PEBLS in principle to find alignments of subsequences of one exemplar with subsequences of another exemplar, as suggested in Figure 13.3.1—that is, to pool data on a less-than-whole-exemplar basis. Intuition suggests, though, that some transitions are more permissible than others, namely transitions similar to those instantiated within the cloud. To compute this permissibility, an *intra-cloud transition network* is constructed: a similarity matrix of the entire cloud to itself, offset by one frame. Cell (i, j) of this matrix thus encodes not the similarity of frame i to j, but the similarity of i to the frame that immediately precedes j. By means of this transition network, PEBLS takes into account not only how the input aligns with each exemplar in the cloud, but how the cloud aligns with itself—getting emergent structure from self-similarity within the data.

Secondly, whereas DTW simply finds the minimum-distance alignment, PEBLS requires an alignment that *generalizes*, reflecting frame sequences which are in some sense prototypical of the cloud. To capture the generalization effect, the alignment criterion must incorporate some measure of the frequency of similar subsequences within the cloud. This problem is analogous to the statistical notion of confidence, that a particular sample reflects the distribution of an underlying population. This confidence sensitivity can be obtained by hierarchically clustering the vector of alignment scores from the previous frame at each dynamic programming step, and selecting the cluster that maximizes a function of the cluster's mean similarity, size, and variance. The criterion thus involves a potential trade-off between similarity and density (i.e. size over variance): a high-similarity but atypical alignment may lose to a somewhat lower-similarity alignment if drawn from a much higher-density cluster. This is PEBLS's implementation of the similarity-frequency trade-off for assessing the relative strength of competing patterns, discussed in 13.3.2.2.6 above.

Kirchner et al. (2010) report that PEBLS, trained on a small corpus of recordings of short nonce words, generated outputs that reflect pattern entrenchment. That is, outputs for words tended to conform to the prevailing pattern within the word cloud, even when a pattern-violating input was selected. Moreover, when the model was applied iteratively, pattern-violating outputs became increasingly rare, eventually ceasing altogether. The model thus showed a diachronic progression from a variable to an obligatory pattern.

13.3.4 Conclusions

PEBLS provides a solution (though perhaps better solutions remain to be discovered) to the modeling problem which has hindered the development of exemplar theory, namely how to generate a composite output from a collection of unique, variable-length signals. PEBLS further provides the first explicit model of exemplar-

based pattern entrenchment using real speech signals. Many of the conjectured capacities of exemplar-based phonology (see 13.3.2.2) remain to be established, such as modeling of recency effects, generalization of patterns outside the word class, modeling of sociophonetic variation, and modeling of top-down semantic and pragmatic effects.

Inasmuch as PEBLS computes a global optimization for the output, there exist close parallels to Optimality Theory (or more directly, to harmonic grammar). The alignment process described above is analogous to OT enforcement of correspondence constraints. A more elaborated version of PEBLS would include soft constraints reflecting phonetic pressures as part of the optimization criterion, e.g. an energy minimization imperative. In PEBLS then, as in OT, phonological patterns would arise from conflict between constraints favoring current patterns (including patterns within the word class, as with IO faithfulness), and constraints favoring phonetic naturalness. PEBLS, however, computes over numeric signals rather than symbolic representations.

This approach is presented as a possible way forward for phonology, addressing the legitimate critique of evolutionary phonology by abandoning symbolically stated (therefore pseudo-phonetic) markedness constraints. At the same time, this exemplar-theoretic approach does not relegate markedness concerns to unformalized meta-theory, but rather seeks to model markedness effects explicitly, through interaction between direct phonetic constraints and the pattern-entrenching dynamics of an exemplar-based speech-processing system.

PART IV

INTEGRATING DIFFERENT PERSPECTIVES: INSIGHTS FROM PRODUCTION, PERCEPTION, AND ACQUISITION

The goal of this part is to demonstrate how integrating different domains and perspectives provides insights into the understanding of phonological questions, including lexical and sublexical representation, variation, and phonological patterning. The contributions advance our understanding of the relationship between production and perception, as well as the relationship between acquisition of language and the human capacity for language.

This page intentionally left blank

CHAPTER 14

INSIGHTS FROM PERCEPTION AND COMPREHENSION

HOW PERCEPTUAL AND COGNITIVE CONSTRAINTS AFFECT LEARNING OF SPEECH CATEGORIES

LORI L. HOLT

REPRESENTATIONS OF SPEECH SOUND PATTERNS IN THE SPEAKER'S BRAIN: INSIGHTS FROM PERCEPTION STUDIES

NOËL NGUYEN

This chapter provides rich reviews of experimental research on questions of category learning and the role of speech perception in formation of phonological representations. Holt highlights the importance of looking for convergent evidence through parallel methods using both controlled and naturalistic data to advance our understanding of category learning. Nguyen emphasizes how "the role of both detailed phonetic characteristics and abstract phonological categories in speech perception" (pp. 367–8) are critical to modeling perception.

14.1 How perceptual and cognitive constraints affect learning of speech categories

Lori L. Holt

14.1.1 Introduction

Categorization is an important facet of speech communication. However, we do not yet have a complete understanding of how speech categories are learned in infancy or adulthood. In part, this is because it is not feasible to entirely control and manipulate speech to observe consequences of different patterns of experience. Converging methods of cross-language observation, laboratory-based training of speech and non-speech categories, and animal learning models provide a means of balancing the competing demands of ecological validity and experimental control to reveal how auditory and cognitive constraints affect speech category learning. The present section describes these approaches and explains how they inform us about how general perceptual and cognitive constraints affect learning speech categories.

14.1.2 Speech categorization

To extract a message from speech, listeners must accomplish two complementary perceptual feats. They must discriminate linguistically relevant acoustic variability and generalize across irrelevant variability. To put it another way, listeners must *categorize* speech in a manner specific to their language (Lotto 2000). Since the mapping of acoustic variability is language-specific, these categories must be learned from experience with speech.

Young infants from different language communities respond to speech in a way that is more similar than different, discriminating sounds without respect to whether they are phonemically distinctive in the ambient language (Jusczyk 1997 for review; Demuth and Song, this volume; Munson et al., this volume; Maye, this volume). In stark contrast, adults have difficulty discriminating even highly acoustically distinct differences between some non-native sounds (Iverson, this volume; Smiljanic, this volume). Japanese-speaking adults, for example, are poor at discriminating English /r/ versus /l/ (Miyawaki et al. 1975), although 6- to 8-month-old Japanese-learning infants discriminate the sounds as well as English-learning infants (Kuhl et al. 2006). Older infants discriminate acoustic differences between native sounds even more effectively than early in development (Kuhl et al. 2006), but no longer very accurately distinguish many non-native sounds (see Werker and Tees 1999; also McMurray and Farris-Trimble, this volume).

Experience with the native language shapes speech perception already in infants' first year (Werker and Tees 1999). The experience-dependent change in speech perception is thought to reflect the influence of native-language speech category learning and has been described as "warping" of perceptual space (Kuhl et al. 2008). Imagining perceptual space as a multidimensional topography, the landscape appears to be relatively flat in early infancy with any discontinuities arising from general auditory processing. The mapping from acoustics to perceptual space is closely related to the raw acoustic differences among speech sounds; and infants' speech discrimination is mostly independent of the native language environment. Speech category learning warps perceptual space to reflect regularities of native speech and infants begin to perceive speech relative to native-language categories rather than solely according to psychoacoustic differences, with regions of increased within-category similarity reducing discrimination and regions of reduced between-category similarity enhancing discrimination (Liberman et al. 1957; Iverson et al. 2003).[1] In the multidimensional perceptual topography, categories can be envisioned as basins in which there is limited perceptual discriminability flanked by between-basin peaks representing regions of exaggerated perceptual discriminability (Spivey 2007).

[1] Although within-category differences are diminished, it is now understood that speech categorization is not entirely "categorical" (Liberman et al. 1957; Harnad 1990). Infants (McMurray and Aslin 2005) and adults (Lotto et al. 1998; McMurray et al. 2008) remain sensitive to within-category acoustic variation. Speech categories exhibit graded internal structure such that speech sounds are treated as relatively better or worse category exemplars (Miller and Volaitis 1989; Johnson et al. 1993; Iverson and Kuhl 1995; Iverson et al. 2003).

This perceptual warping aligns with native language regularities, promoting efficiency in native speech categorization. However, once the perceptual system has committed to a native-language parse of perceptual space, it can be quite difficult for adults to learn non-native categories that do not align with native-language regularities (Best 1995; Flege 1995). For example, Italian has a single category /i/ in a region of perceptual space that accommodates both /i/ and /ɪ/ in English. The warping of perceptual space to accommodate a single Italian /i/ category comes at some detriment to later learning English (which requires discrimination between /i/ and /ɪ/ in the very same region of perceptual space; Flege et al. 1999a). Of note, sounds that fall in regions of perceptual space not inhabited by native sounds (like Zulu clicks for English listeners) avoid interactions with native categories and continue to be well-distinguished in adulthood, presumably as a result of their acoustic differences (see Smiljanic, this volume; Escudero, this volume). In contrast, experience-dependent decreases in non-native discrimination are most evident among speech contrasts similar to those of the native language (like English /r/-/l/ for Japanese listeners; Best 1995; Flege 1995).

It is remarkable that infants begin to form speech categories without an indication of how many categories exist in the native language and without significant exposure to these sounds in isolation (see Vallabha et al. 2007). Nonetheless, speech category learning appears to be well underway well before infants speak the first word or develop a significant lexicon (Jusczyk 1997, for review). Although the groundwork for speech category learning begins in infancy, there is a lengthy developmental course whereby even 12-year-olds have not reached adult levels of speech categorization for some native sounds (Hazan and Barrett 2000). Lexical development, learning to read, and continued development of perceptual expertise with speech are all likely contributors along this protracted developmental course (see Munson et al., this volume).

Despite an appreciation for the profound influence of categorization on speech processing, we do not yet have a complete understanding of how speech categories are learned in infancy or adulthood. At least part of the reason for this is that it is not feasible to entirely control and manipulate speech experience. Natural cross-language comparisons for adults and infants, like those described above, are the standard upon which our understanding is based and they have provided an appreciation of the range of behaviors to be accounted for by any theory. But, without controlled manipulation of experience, models of speech category learning are more descriptive than predictive.

14.1.3 Converging methods

Given the difficulty in manipulating and controlling speech experience, it is useful to take a converging methods approach, investigating auditory and cognitive

constraints on speech category learning from multiple, coordinated perspectives that vary in the extent to which they emphasize ecological validity, or naturalness, of experience versus experimental control of experience. Ideally, the coordinated approaches converge and it is possible to develop predictions from tightly controlled laboratory experiments that may be tested in natural speech communication. The sections below describe representative research findings that exemplify multiple converging methods for investigating speech category learning.

14.1.3.1 *Naturalistic cross-language comparisons*

Cross-language comparisons provide a natural experiment in differing histories of speech experience (see Smiljanic, this volume; Iverson, this volume; Escudero, this volume). The review above highlighted some of what is known about infant speech category learning from this approach. Among adult learners, some degree of plasticity for learning non-native speech categories is maintained, although its expression is critically dependent on the amount and quality of the second-language input and its interaction with the speech categories learned for the first language (Flege 1995). Flege and MacKay (2004) report that native speakers' ability to discriminate non-native vowels is best predicted by self-estimated amount of first-language usage, with lower usage predicting better second-language performance. In fact, non-native perception among adults arriving earlier in the second-language environment and using their first language less often was statistically indistinguishable from that of native listeners. Flege suggests that the learning mechanisms that guide first-language speech category learning remain intact through adulthood but that first- and second-language processing share common resources and mutually influence one another. Non-native speech categories are perceived through the lens of the perceptual space warped by learning native categories. On the whole, naturalistic cross-language studies indicate plasticity in adult non-native speech category learning but, as is the case for infant learning, the mechanistic details of this learning remain unclear.

14.1.3.2 *Laboratory-based speech category training studies*

Wrestling with the issue of control over experience, some studies have taken the approach of manipulating short-term speech experience in the laboratory. Artificial "languages" comprised of speech tokens manipulated to have special characteristics have been used widely as a tool in understanding infant language acquisition (e.g. Saffran et al. 1996; Thiessen 2007), including speech category learning (Maye et al. 2002). In these studies, infants hear speech possessing well-controlled regularities and their responses thereafter reveal the influence of this short-term experience. Using this approach, Maye et al. (2002) reported that infants who heard a bimodal

distribution of speech sounds exhibited exaggerated speech discrimination compared to infants who heard the same speech sounds sampled unimodally. This suggests a categorization-like warping of perceptual space as a result of short-term exposure to different distributions of speech.

A limitation of these approaches, to date, is that they leave open important issues about how category learning proceeds with more natural speech. Whereas infants experience a rather continuous stream of speech, laboratory-based experiments typically solve the segmentation problem for the listeners by presenting isolated instances (e.g. Maye et al. 2002). Although input regularities can guide segmentation (Saffran et al. 1996), the extent to which distributional regularities support speech category learning in unsegmented speech remains unknown (Pierrehumbert 2003a). In addition, artificial languages tend to be rather simple with one or several acoustic dimensions defining categories. In natural speech, infants must contend with highly multidimensional input. Future research must determine the extent to which distributional learning scales to more natural speech category learning challenges.

Kuhl and colleagues (2003) have taken a step in this direction by exposing 9-month-old English-learning infants to Mandarin Chinese across twelve play sessions with a Mandarin-speaking adult. This exposure was sufficient to reverse the decline in Mandarin speech discrimination observed among infants exposed instead to English-language play sessions. Perhaps telling of the mechanisms involved in infant speech category learning, the preservation of Mandarin speech discrimination was observed only with live play sessions and not from exposure to the same speech via audiovisual or audio recordings. Mere exposure to distributional regularities may not be enough to direct learning in more natural circumstances. It seems likely that a combination of factors, including distributional regularity in speech input (Holt et al. 1998) and the potential for socially driven feedback (see Goldstein and Schwade 2008, 2009) influence early speech category learning, but details of these mechanisms remain to be discovered.

Laboratory-based speech training among adults learning a second language also informs our understanding of speech category learning (e.g. Jamieson and Morosan 1989; Logan et al. 1991; Pisoni et al. 1994; Bradlow, Akahane-Yamada, Pisoni, and Tohkura 1999; McCandliss et al. 2002; Iverson et al. 2005; Goudbeek et al. 2008). Some early attempts to train adults on non-native categories included discrimination training with little acoustic variance in the training set. Although listeners learned to discriminate training stimuli, they typically could not transfer learning to natural speech or to different contexts (Strange and Dittmann 1984). Recent research has underscored the importance of acoustic variability. Including multiple speakers and phonetic contexts in training seems to aid generalization (Jamieson and Morosan 1989; Lively et al. 1993; Bradlow, Akahane-Yamada, Pisoni, and Tohkura 1999; McCandliss et al. 2002; Iverson et al. 2005). In such studies, participants tend to improve in reliably categorizing non-native speech over the course

of training with learning persisting across months and generalizing to speech production in some studies (Bradlow, Akahane-Yamada, Pisoni, and Tohkura 1999). However, extensive training is necessary to evidence learning and the final level of achievement typically has not been equal to that of native listeners (Logan et al. 1991; Bradlow et al. 1999; Lively et al. 1993). Thus, training studies provide evidence of plasticity in the adult system to support category learning, although the system is clearly not as flexible as in infancy.

Such studies also have begun to make mechanistic predictions about adult learning. For example, McCandliss et al. (2002) hypothesized that the perceptual warping apparent in native-speech category learning produces neural circuits committed to processing native-language speech categories. Hearing similar non-native sounds activates native circuits, thereby further reinforcing them. Counter-intuitively, training listeners with non-native sounds may reinforce existing *native* categories because perceptually similar non-native sounds activate neural circuits supporting native categories. By this logic, McCandliss et al. (2002) predicted that training with highly exaggerated instances of non-native speech falling outside native perceptual space and then incrementally adjusting training stimuli to be more representative of the non-native categories may facilitate learning. Their results support this prediction, but also indicate a role for explicit feedback in learning (Tricomi et al. 2006) suggesting a more complex set of learning mechanisms (see also Goudbeek et al. 2005, 2008).

Many studies, including most cited above, have investigated Japanese adults learning English /r/ vs. /l/, an adult speech category learning problem that is notoriously challenging. Other speech categories appear to be more easily learned by non-native listeners (Pisoni et al. 1982; Polka 1992) and this may be predicted by the relationship between first- and second-language categories and their interaction (Best 1995; Flege 1995).

Even among more easily learned categories, there are enormous individual differences in adult speech category learning (Golestani and Zatorre 2009), making it difficult to draw sweeping conclusions about the degree of adult plasticity. Although it is not yet the norm for studies to investigate individual differences in detail (although see Perrachione et al. 2011), it seems likely that research can capitalize on individual differences to understand more about auditory and cognitive constraints on speech category learning (Slevc and Miyake 2006).

14.1.3.3 *Laboratory-based non-speech category learning studies*

One way to gain experimental control over listeners' histories of experience is to create novel sound stimuli with which listeners have no experience and for which listeners possess no *a priori* categories. Training listeners to categorize such artificial non-speech sounds makes it possible to exert control over and have knowledge of listeners' entire history of experience with the sounds, thus providing the

opportunity to investigate explicitly the general perceptual and cognitive constraints on auditory processing that may influence speech categorization.

The literature in this area has produced insights about auditory category learning relevant to speech categorization. It has long been observed that vowels and consonants exhibit different patterns of categorization and discrimination, with vowels tending to be perceived more continuously, with less abrupt categorization boundaries and less sharp discrimination peaks than consonants (Repp 1984; Schouten and Van Hessen 1992). Mirman et al. (2004) examined whether general auditory constraints on processing the differing spectrotemporal acoustics of vowels and consonants might play a role by training listeners to categorize non-speech sounds modeling rapidly changing acoustic dimensions of consonants or steady-state acoustic dimensions of simplified vowels. Patterns of non-speech discrimination and categorization mirrored those of the speech stimuli they modeled. General characteristics of auditory sensory memory may play a role: more quickly decaying perceptual memory traces for rapidly changing sounds relative to steady-state sounds could account for this pattern for both speech and non-speech.

Many accounts have suggested that infants' initial parse of the perceptual space relies upon natural "boundaries" in auditory processing that arise from discontinuities in the mapping from acoustics to audition. The most compelling case is a proposed discontinuity in auditory temporal processing that may influence voicing perception (Pisoni 1977). Examining the question of how discontinuities would interact with experience, Holt et al. (2004) trained listeners to categorize non-speech sounds varying along this perceptually discontinuous acoustic dimension. Learning was facilitated when the sound input distribution boundary aligned with the perceptual discontinuity relative to when listeners were forced to categorize across the perceptual discontinuity. However, listeners did eventually learn in the latter situation. Thus, basic auditory constraints on perceptual processing may provide an initial parse facilitating categorization, but learning is flexible enough to overcome perceptual biases.

Non-speech category learning studies also highlight how task influences category learning. Discrimination training (explicit comparison of stimuli) and categorization training (responding to acoustically variable instances as category members) warp listeners' perception of non-speech stimuli in different ways. Discrimination training increases listeners' sensitivity to small distinctions among stimuli thereby working against categorization (Guenther et al. 1999). This insight from non-speech learning is important because it is common for studies of speech perception to use discrimination tasks as indices of categorization, taking heightened discrimination between pairs as an indication of a category boundary. Guenther et al.'s non-speech auditory training study indicates that discrimination training is not equivalent to category learning and it has implications for interpreting the fact that Japanese listeners trained to discriminate English /r/-/l/ do not generalize well to natural speech categories (Strange and Dittmann 1984).

Another characteristic of speech categories is their multidimensionality; typically, numerous acoustic dimensions co-vary with speech categories. The dimensions defining perceptual space are not equivalent and some acoustic dimensions play a greater role in determining the category than do others. For example, both formant frequency and vowel duration co-vary with English /i/-/ɪ/, but native listeners rely much more on formant frequency than vowel duration (Hillenbrand et al. 2000).

Relative perceptual cue weight develops across childhood (Nittrouer 2004) and is native-language specific. Whereas native English listeners rely primarily on formant frequency for /i/-/ɪ/, non-native listeners often rely more on duration (Flege et al. 1997). These findings and others make clear the importance of learning to perceptual cue weighting, but exactly how perceptual weighting relates to details of speech experience remains unclear.

The control over experience afforded by non-speech categories allowed Holt and Lotto (2006) to test the kinds of input distributions that affect perceptual cue weighting. They trained adults with explicit feedback to categorize two novel non-speech input distributions drawn from a two-dimensional acoustic space defined by the rate at which sine wave tones repeatedly increased and decreased in frequency (Modulation Frequency, MF) around a particular base frequency (Center Frequency, CF). Although the dimensions were psychoacoustically equated and the stimuli defining the categories were sampled such that the dimensions were equally informative to the categorization task, listeners relied much more upon CF than MF for categorization. This bias allowed Holt and Lotto to investigate how different stimulus-training sets influence perceptual cue weighting. Moving the distributions closer along the preferred CF dimension, thereby making CF a less reliable categorization cue, had no effect. However, making CF more variable within each category distribution caused listeners to rely on MF instead of CF in categorization responses. It appears that the variance experienced across an acoustic dimension is significant to perceptual cue weighting. An implication of this finding is that use of an inefficient acoustic dimension in non-native speech categorization may be lessened by experience with substantial variability along this dimension.

An issue in the above studies is their use of feedback. Laboratory-based non-speech category training tends to rely on explicit feedback atypical of natural speech learning, which does not seem to involve explicit category labels, or explicit feedback (e.g. Jusczyk 1997). Goudbeek and colleagues (2005, 2008) have used non-speech categories to investigate the role of feedback in auditory category learning, reporting that without explicit feedback listeners find it very difficult to learn categories defined by multiple acoustic dimensions. This is curious, considering that highly multidimensional speech categories appear to be learned by infants without explicit feedback.

It seems likely that speech category learning during first-language acquisition involves complex relationships among acoustic speech and various simultaneous

events in the environment. Of course, these naturally complex interactions are difficult to control, making it challenging to infer mechanism. Video games are an immersive environment in which researchers can maintain control over auditory experience while manipulating complex multimodal relationships that the sounds have with other perceptual events, thus involving participants in the functional use of sound without explicitly training them in auditory categorization or giving them explicit feedback for category learning.

Wade and Holt (2005) developed a space-invaders-style videogame in which visual creatures were each associated with a category of sounds designed to model some of the multidimensional complexity of speech categories, without sounding like speech. To succeed in the game, participants had to learn the relationship between each creature and the corresponding sound category, although this was never made explicit. Similar to the process of learning to treat acoustically distinct speech signals as members of the same speech category, listeners gradually learned that perceptually discriminable creatures' sounds were functionally equivalent in the game. After thirty minutes of game play, listeners' responses indicated significant category learning and generalization to novel sounds. Though there was no explicit feedback, participants were able to learn the complex auditory categories incidentally, suggesting that functional use of sound and multimodal relationships among sounds and other perceptual dimensions may be significant in complex, multidimensional category learning. Of interest to understanding how sound categories are represented by the brain (Nguyen, this chapter), neuroimaging methods reveal that learning to categorize non-speech sounds in this way recruits brain regions typically associated with speech processing (Leech et al. 2009) and warps the perceptual space in a manner similar to that observed among infants learning native-language speech categories (Liu and Holt 2011).

To date, there are relatively few non-speech auditory category learning studies that address the challenges most relevant to speech category learning. It is not yet well understood to what distribution statistics listeners are sensitive, how feedback of various forms may influence speech category learning, how acoustic dimensions are perceptually weighted or how task affects the warping of perceptual space. The ability to carefully manipulate experience with non-speech categories provides an opportunity to investigate these issues in greater depth to discover constraints on auditory learning relevant to speech categorization.

14.1.3.4 *Non-human animal speech category training studies*

Speech category training studies with non-human animals offer some of the same benefits of experimental control over experience present for non-speech learning studies with humans. Animals as diverse as birds, macaques, and chinchillas can discriminate speech (Dewson 1964; Burdick and Miller 1975; Kuhl and Miller 1975;

Morse and Snowdon 1975; Dooling and Brown 1990) and there is a rich history of using animals to probe speech perception absent speech experience. These studies have defined general auditory constraints on speech perception (Kuhl and Miller 1975, 1978; Kluender and Lotto 1994; Dooling et al. 1995; Lotto et al. 1997; Sinnott et al. 1998). For example, Lotto and colleagues (1997) found that Japanese quail trained to peck in response to /ga/ peck more heavily to a perceptually ambiguous sound between /ga/ and /da/ when it is preceded by /al/. This context-dependent response pattern exactly mirrors the pattern of context-dependent speech categorization among human listeners that is taken as evidence of perceptual compensation for coarticulation (Mann 1980). The existence of parallel perceptual responses to speech in humans and non-humans suggests general auditory constraints on speech processing may contribute to speech perception challenges like compensation for coarticulation (Lotto et al. 1997), trading relations (Kluender and Lotto 1994), categorical perception (Kuhl and Miller 1975), and discrimination of prosodic qualities of speech (Ramus et al. 2000).

Animal models also allow controlled investigation of the effects of experience on speech processing (Kluender et al. 2005). For example, Kuhl et al. (1991) reported that monkeys do not show the patterns of graded internal vowel category responses indicative of perceptual warping that are observed for human adults and infants (Grieser and Kuhl 1989), perhaps indicating species specificity in this aspect of speech categorization. However, the monkeys had no experience with speech. When Kluender and colleagues (1998) provided birds experience with vowel input distributions, birds' subsequent responses were graded and highly correlated with human listeners' graded categorization responses to the same sounds. Experience with the distributional characteristics of speech categories is essential in producing graded responses to speech indicative of perceptual warping, a hypothesis only testable with the control over experience afforded by animal models.

Control over animals' speech experience allowed Holt et al. (2001) to determine that the relationship of fundamental frequency (f0) and voicing (with higher f0 associated with voiceless categories in English and other languages; see Kingston and Diehl 1994) is not an obligatory influence of f0 on voicing arising from perceptual constraints, but rather is more likely due to the learnability of covariation between these acoustic dimensions. It arises only when animals experience correlation between the acoustic dimensions during training. Kluender et al. (1987) found that Japanese quail learn the complex mapping among multiple acoustic dimensions defining English alveolar stop consonants and generalize to speech never heard in training. This category learning was impressive because there were no invariant acoustic cues among the stimuli that could define category membership. Thus, the high multidimensionality of speech categories can be accommodated by rather simple learning processes such as those available to quail.

The issue of how feedback influences speech category learning is important in interpreting evidence from animal studies because most methods rely on explicit feedback in training animals to respond to speech. However, even with animal training paradigms that require explicit feedback, it is possible to learn about characteristics of unsupervised learning. In the Kluender et al. (1998) study mentioned above, birds responded to tokens from one of two vowel categories. All vowels were equivalent in training in that response to each vowel elicited the same feedback. Nonetheless, birds' responses mirrored distributional characteristics of the vowel input distributions such that the birds responded to some vowel exemplars more robustly than others. This aspect of animal learning cannot arise from the feedback and appears to reflect something general about distributional learning.

In sum, studying animal learning can serve as a means of understanding how general auditory capacities and general learning mechanisms may solve some of the challenges of speech category learning. Prototype effects (Kuhl et al. 1991; Kluender et al. 1998), lack of acoustic invariance and multidimensional learning (Kluender et al. 1987), perceptual warping by categorization (Kluender et al. 1998), perceptual segmentation (Hauser et al. 2001), and the effects of correlation among acoustic dimensions (Holt et al. 2001) are characteristics of speech category learning that have been illuminated by animal learning models.

14.1.4 Conclusion

To understand category learning, we must understand both the perceptual mechanisms involved and how they are affected by experience. There are important unresolved questions in speech category learning, ripe for research. Ultimately we must explain how experience alters the perceptual space among infants learning speech categories and, in doing so, shapes the learning challenges encountered by adults learning non-native speech categories. There is a need to better define distributional learning and to delineate its mechanisms, including the role of feedback. We must understand exactly what it means to "warp" a perceptual space and discover the representations that inhabit the space. Moreover, we must interpret individual differences, where they exist, and attend to the role higher-level cognitive constraints like attention, working memory, and decisional processes play in guiding first- and second-language speech category learning.

Although there is much work ahead to understand speech categorization, much has been learned in recent years. The representative research reviewed here shares the aim of understanding the mechanisms of speech category learning by investigating the cognitive and perceptual constraints listeners bring to the task and indicates the promise of a converging methods approach to move our models from descriptive to predictive.

14.2 Representations of speech sound patterns in the speaker's brain: Insights from perception studies*

Noël Nguyen

14.2.1 Introduction

Over the last few years, considerable advances have been made in our understanding of the processes that allow listeners to perceive and extract meaning from speech. To a significant extent, these advances have been facilitated by instrumental techniques used in conjunction with more standard experimental procedures. For example, eye-tracking (e.g. Allopenna et al. 1998; Creel et al. 2008; Dahan et al. 2008; Speer, this volume) and the tracking of hand movements (Spivey et al. 2005) now make it possible to continuously follow the dynamics of speech processing as the speech signal unfolds over time. Likewise, event-related brain potentials studies (Dehaene-Lambertz 1997; Phillips 2001; Molfese et al. 2005) and brain-imaging studies (Scott 2003; Scott and Johnsrude 2003) provide crucial new insight into the cerebral underpinnings of speech perception and comprehension (see Idsardi and Poeppel, this volume for a review). Another major development concerns the fact that speech perception studies now go well beyond the processing of speech sounds produced by an individual speaker in a laboratory setting, and extend to spoken language in the context of its primary site of occurrence (Local 2003), that is, social interaction. Fragments of conversational speech are used as material in perceptual tests for example, and more generally, rigorously controlled experimental designs have been combined in a variety of innovative ways with large-scale investigations of spontaneous speech data.[2] Yet another important development relates to the increasingly large number of studies of the contribution to speech perception of visual articulatory information associated with movements of the speaker's face (Massaro 1998). These advances in speech perception research have contributed to establishing new links with speech technology, among other disciplines, and to triggering the development of automatic speech recognition systems whose design partly mirrors the way in which speech is processed by human listeners (Moore 2007; Scharenborg 2007).

* Thanks are due to Cheryl Frenck-Mestre, Pauline Welby, one anonymous reviewer, and the editors, for helpful comments on an earlier version of this section.

[2] In a different but complementary perspective, Holt (this chapter) shows how the growing focus on ecological validity leads studies using laboratory speech to be increasingly combined with work on more natural speech communication.

In extending beyond the limits of its traditional domain, speech perception research has shed new light on what has consistently constituted one of the most important issues for laboratory phonology: the way in which speech sound patterns are represented in the speaker/listener's brain. Indeed, conjectures have been made for quite a long time—well before the inception of laboratory phonology itself—about what such representations might be. In the following, a short overview will be presented of what recent speech perception studies tell us about the form and function of these mental representations for speech sound patterns.

14.2.2 The representational and mapping problems

Speech sounds are highly variable; yet listeners extract meaning from speech effortlessly and successfully in most circumstances. To account for this seemingly paradoxical fact of our everyday experience, theories and models of speech perception and comprehension have to deal with two central problems. The first problem relates to how speech sound patterns are represented in the listener's memory. The second problem concerns the way in which access to such representations is achieved by the listener from the input speech signal. I will refer to these as the *representational* and the *mapping* problem, respectively. The solutions offered by speech perception models to these two problems are inevitably intertwined (Hawkins, this volume): for example, the mapping mechanism is likely to take a different form if the sound shape associated with each word in memory is specified as a function of the context of occurrence for that word, as opposed to being context-independent.

According to an approach that long prevailed (see Klatt 1989, for a historical overview), speech perception involves retrieving invariant properties relative to distinctive features and phonemes, independent of the variability shown by the corresponding speech sounds both within and between speakers, and it is in terms of these invariant properties that words are represented in the listener's mental lexicon. In this approach, a clear demarcation is posited between the surface phonetic form of a word and the underlying phonemic representation for that word. Variations in the production of speech sounds attributable to inter-individual anatomical differences are assumed to be factored out at an early stage of perceptual processing by means of a speaker normalization procedure.

It is now generally considered that this approach does not offer a satisfactory characterization of a number of major properties that have been found to co-occur in the speech perception system. One of these properties is the remarkable robustness to alterations in words' surface shapes caused by various phonological processes. To explain how words can be successfully recognized in spite of these alterations, researchers have moved away from the traditional approach, to propose either more sophisticated mapping mechanisms, alternative lexical representations,

or both. Another no less remarkable property is the sensitivity shown by listeners to indexical information about the speaker's individual and social identity. This phenomenon appears to be largely inconsistent with phoneme-based models of speech perception, and a major challenge today is understanding how sensitivity to speaker indexical information may combine with perceptual robustness despite phonological variation. Other recent studies have raised yet more questions for the traditional view by suggesting that so-called fine phonetic detail can be perceptually relevant for listeners. These properties now attributed to the speech perception system are successively discussed in the next sections. We start with the perceptual processing of phonological variation.

14.2.3 Spoken word recognition in the face of phonological variation

Words can show substantial variations in their surface form under the influence of a variety of phonological phenomena such as assimilation or deletion (Anttila, this volume). A well-known example is provided by word-final coronals which take the place of articulation of a following labial or velar consonant in English[3] among other languages (Nolan 1992), as in *green boat* [gɹiːm bəʊt]. It has been a matter of much discussion how listeners can correctly identify words in which the final segment has undergone regressive place assimilation. Indeed, this phenomenon has come to form a key test for speech perception models whose domain of application extends much beyond the processing of assimilation itself (see, among others, Darcy 2003; Ranbom and Connine 2007; Gaskell and Snoeren 2008; Pitt 2009; Lahiri, this volume).

In a series of studies (Lahiri and Marslen-Wilson 1991, 1992; Lahiri and Reetz 2002; Eulitz and Lahiri 2004; Friedrich et al. 2008; see also Fitzpatrick and Wheeldon 2000; Wheeldon and Waksler 2004; as well as Lahiri, this volume), Lahiri and colleagues have gathered both behavioral and brain-imaging data suggesting that assimilation does not have a disruptive effect on word recognition, and that listeners in fact have little or no sensitivity to it. In the Featurally Underspecified Lexicon (FUL) model of word recognition proposed by Lahiri and colleagues, this is attributable to the fact that each word is associated in the mental lexicon with a highly abstract phonological representation, which is underspecified for certain features such as [coronal]. As a result, variations that a surface form may show with respect to these features do not prevent it from remaining consistent with the underlying phonological representation.[4] According to the featural underspecification

[3] Recent work by Dilley and Pitt (2007), however, showed that regressive place assimilation in word-final coronals in conversational speech in English is less frequent than has been previously assumed.

[4] For example, a so-called no-mismatch relationship will be said to exist between the surface form [gɹiːm] and the phonological representation for *green*, because the coronal place of articulation of the

hypothesis, it is because of the way in which lexical representations are tailored that listeners are able to successfully recognize phonological variants of a given word, and this applies irrespective of the context in which these variants are encountered.

By contrast, Gaskell and his colleagues (e.g. Gaskell and Marslen-Wilson 1996; Gaskell 2003; Gaskell and Snoeren 2008; Snoeren et al. 2009) have emphasized the role that context may play in the perceptual treatment of assimilation. Their view is that listeners retrieve the underlying form of a word by means of an inference process that aims to reverse the effect that assimilation may have had on how this word was produced. Studies conducted by the authors have revealed that this compensation-for-assimilation process is triggered in the context of a viable environment only, as opposed to an unviable one (in the case of word-final coronals in English, environmental viability refers to a subsequent word-initial labial or velar consonant). In the context-independent and representational FUL approach, assimilatory variation is dealt with as the input word form is mapped onto the lexicon, while the phonological inference account assumes that compensation for assimilation occurs at a pre-lexical level and may extend to novel words or non-words.

Although assimilation is traditionally characterized as causing a categorical change in the value of a segmental feature, studies on the phonetic realization of segments that occur in a context appropriate for assimilation have found a variety of patterns from non-assimilated forms through cases of partial assimilation to fully assimilated forms (e.g. Browman and Goldstein 1990a; Ellis and Hardcastle 2002). When assimilation is incomplete, cues to the underlying identity of the target segment are still available to listeners in the speech signal (e.g. Surprenant and Goldstein 1998; Snoeren et al. 2008). In yet a different model of the perception of assimilation, namely the feature-cue-parsing model developed by Gow (2001, 2002b, 2003), listeners are assumed to be tuned to these fine-grained acoustic cues, which provide them with information about both the (partially) assimilated segment and the assimilating segment. There is evidence showing, for example, that when a word-final alveolar is assimilated to the following velar, as may be the case in *lead covered*, differences in F1 and F2 frequency extending throughout the preceding vowel and possibly even further towards the beginning of the word can be found, relative to a /g/-final word such as *leg* in *leg covered* (Nolan 1992; Local 2003). Such differences may contribute to explaining why assimilated alveolars are perceptually recoverable from speech (Wright and Kerswill 1989). This is consistent with recent work showing more generally that perceptually relevant cues to the identity of a given segment are spread over an interval that can extend well outside the segment's most prominent boundaries in the speech signal (e.g. West 1999; Hawkins and Nguyen 2004). The feature-cue-parsing model contends that listeners are in most circumstances able to directly identify segments from

final consonant in *green* is assumed to be underspecified in the lexicon. Thus, listeners are expected to interpret both [gɹiːm] and [gɹiːn] as referring to the word *green*.

speech, and that neither underspecified lexical representations nor phonological inference mechanisms are necessary for word recognition.

Work on the role of phonological variation in word recognition has recently turned to another potentially influential factor, namely the listener's degree of exposure to this variation. It is well known that lexical frequency has an important function in both production and comprehension, e.g. high-frequency words are recognized more rapidly than low-frequency words (see Jurafsky 2003, for a review). There is now evidence that frequency effects apply to the phonological variants for a given word. For example, Connine (2004) showed that American English listeners are biased towards perceiving the word *pretty* (as opposed to the non-word *bretty*) to a greater extent when presented with a speech sound sequence that contains the more frequently produced voiced alveolar flap ([pɪɪɾi]), compared with the less frequent voiceless alveolar stop ([pɪɪti]) in intervocalic position. According to Connine and her colleagues (Connine 2004; Connine and Pinnow 2006; Ranbom and Connine 2007; Connine et al. 2008), this is inconsistent with a model of word recognition in which 'pretty' has a single underlying phonological representation with a voiceless alveolar stop, which is recovered by the listener from the flapped variant. Rather, Connine and colleagues have argued that the main phonological variants of a word are jointly stored in the listener's mental lexicon and that each of these variants has a strength in memory that reflects the listener's frequency of exposure to that particular form. In Connine and colleagues' proposal, phonological variation is therefore directly encoded in the lexicon, in the form of a repertoire of alternative phonological representations for each word, contrary to the featural underspecification theory (in which the representation for each word is unique) and both the inference process and the feature-cue-parsing theory (which assume that phonological variation is factored out at a pre-lexical stage of processing).

As we have seen, the models of speech comprehension reviewed above diverge from one another in quite a large measure with respect to the characterization of how phonological variation is dealt with by the listener. One major bone of contention relates to how much of this variation is incorporated into the lexicon, and how much is abstracted away during lexical access. In that respect, an opposition has arisen between the abstractionist viewpoint, as embodied by the FUL model for example, and the exemplar-based viewpoint, according to which each word is associated in the lexicon with a list of exemplars that each reflect a particular context-dependent realization for this word. Studies on indexical effects in speech perception, to which we now turn, have allowed further advances in this debate.

14.2.4 Indexical effects in speech perception

Speech contains a large variety of properties which relate to the speaker's physical, psychological, and social characteristics, and which have been referred to as

indexical properties (Abercrombie 1967). As listeners, we are highly sensitive to these properties. For example, we are able to recognize the correspondence that may exist between a specific phonetic pattern and the speaker's social status (Labov 1966; Foulkes and Docherty 2006; Stuart-Smith 2007b). Little or no attention was paid to indexical properties in traditional models of speech perception which assumed that listeners concentrate on the most prominent acoustic cues relative to phonemic contrasts as a first step towards extracting meaning from speech. The role of indexical properties in the recognition of the speaker's individual and social identity was undisputed, but these properties were assumed to be processed independently of phonemic cues. However, studies over the last twenty years have revealed that speech sounds are processed differently by listeners depending on the speaker's perceived gender (Johnson et al. 1999; Strand 2000), age and social class (Hay, Warren, and Drager 2006), idiolect (e.g. Norris et al. 2003), and dialect (Niedzielski 1999; Evans and Iverson 2004; Hay, Nolan, and Drager 2006; Dahan et al. 2008), and that indexical properties have a significant impact on spoken-word recognition itself. These findings have contributed to reshaping our conceptions of the listener's mental representations for speech patterns.

An early demonstration of the effects of indexical properties on word recognition was provided by Mullenix et al. (1989). These authors showed that response speed and accuracy in a word recognition task both decrease when listeners are presented with words produced by multiple speakers rather than a single speaker. The fact that it is more difficult for listeners to identify words originating from more than one speaker may be accounted for in a way consistent with traditional speech perception models, if one supposes that decreased performance in the multiple-speaker condition is due to the greater amount of computing resources consumed by a speaker normalization mechanism that comes into play prior to lexical access. However, an alternative explanation has been proposed that relies on the assumption that speaker-specific phonetic characteristics are encoded by listeners in long-term memory as spoken words are being processed (e.g. Pisoni 1993; Bradlow, Nygaard, and Pisoni 1999). Empirical evidence from experimental studies has accumulated in support of this proposal. For example, Palmeri et al. (1993) found that it is easier for listeners to recognize that a word has already been presented to them (i.e. an "old" word as opposed to a "new" one) when both tokens of that word were produced by the same speaker rather than by different speakers. Likewise, Goldinger (1996) showed that prior exposure to a word facilitates later recognition of that word to a greater extent when the speaker is the same as opposed to different across the two repetitions. On this account, the Mullenix et al. (1989) effect is attributable to the fact that memory encoding of speaker-specific phonetic characteristics takes more time and resources when listeners are exposed to voices from a larger variety of speakers.

The finding that indexical properties come into play in word recognition has lent strong support to so-called exemplar-based models of speech perception and

understanding (Goldinger 1996, 1998; Johnson 1997c; Coleman 2002). While these models draw on a long-established line of research in cognitive psychology, their introduction into the field of speech perception is relatively new and they stand in stark contrast to the traditional phoneme-based approach. Exemplar models take the view that, for each encountered token of the word, an exemplar forms in the listener's memory that includes all the perceptual and contextual details specific to that token. These include sensory-motor, semantic, and pragmatic characteristics, but also indexical information about the speaker's identity and the situation of occurrence, to mention but a few properties. Exemplars are therefore deeply anchored within their context of occurrence in the largest possible sense and this has drastic implications for how spoken language may be represented in the brain (Bybee and McClelland 2005). In non-analytic models such as Johnson's (1997c, 2005) XMOD, exemplars have no internal structure, and are conceived as unanalyzed auditory representations associated with whole words. More abstract representations associated with words' sound shapes are assumed to exist, but only as the result of a pattern formation process in the online processing of speech. In this view, "abstract phonological structure is a fleeting phenomenon—emerging and disappearing as words are recognized" (Johnson 1997c: 111), as opposed to being more permanently stored in memory, as is assumed in abstractionist models.

Central to the exemplar approach is the assumption that indexical properties are integral to how words are represented in the mental lexicon, along with lexically contrastive phonetic properties. More generally, talker-specific information and linguistic information are viewed as being processed in an integrated fashion by the listener (Nygaard 2005). However, experimental evidence suggests that listeners are not sensitive to at least some aspects of the talker's voice in spoken word-recognition. For example, artificially produced variations in words' overall acoustic amplitude do not affect response accuracy in a word-recognition (Sommers et al. 1994) or word-monitoring (Magnuson and Nusbaum 2007) task, and overall amplitude does not seem to be retained in long-term memory as a perceptually relevant aspect of the words' surface forms (Church and Schacter 1994; Bradlow, Nygaard, and Pisoni 1999). Thus, listeners seem to employ a mechanism that filters out overall amplitude and potentially other acoustic characteristics prior to the long-term storage of surface forms. More generally, representations for words in memory must be, to a certain extent at least, abstract, since it is estimated that the auditory trace of speech fades away after about 400 ms (Pardo and Remez 2006). Recent repetition-priming studies (McLennan et al. 2003; Luce and McLennan 2005; McLennan and Luce 2005) indicate that both abstract phonological representations and talker-specific exemplars may in fact coexist in memory and come into play at different stages in spoken word recognition. Specifically, early processing would be dominated by abstract representations, whereas talker-dependent information would exert an influence at a later stage of processing.

14.2.5 Beyond the lexicon: Perceptual relevance of fine phonetic detail

In recent years, research has increasingly focused on the listener's sensitivity to properties of the speech signal that are generically referred to as "fine phonetic detail" (FPD; see Hawkins, this volume). This research suggests that FPD has a significant impact on speech perception and understanding, at least in some circumstances. FPD includes allophonic variation, sometimes specific to certain words or classes of words (Pierrehumbert 2002). Fine phonetic detail is designated as such in the sense that it is to be distinguished from the local and most perceptually prominent cues associated with phonemic contrasts in the speech signal. Crucially, FPD "rarely, if ever, is a major contributor to distinguishing the citation form of lexical items from one another. In other words, FPD is subphonemic phonetic variability that contributes to phonological or other contrasts that distinguish meanings, but not necessarily lexical items" (Hawkins 2010).

Recent studies on the role of FPD in spoken-word recognition have provided evidence that perceptually relevant allophonic variation includes vowel-consonant acoustic transitions (e.g. Marslen-Wilson and Warren 1994), within-category variations in voice onset time (Allen and Miller 2004; Andruski et al. 1994; Ju and Luce 2006; McMurray, Tanenhaus, and Aslin 2009), long-domain resonance effects associated with liquids (West 1999), and graded assimilation of place of articulation in word-final coronals (e.g. Gaskell 2003).[5] To a certain extent, however, the fact that listeners are sensitive to allophonic variation was established much earlier. For example, studies conducted in the 1970s and 1980s consistently showed that coarticulation between neighboring segments provides listeners with perceptually relevant cues to segment identity (and by extension to word recognition). A well-known example is regressive vowel-to-vowel coarticulation in English, which allows the identity of the second vowel to be partly predictable from the acoustic cues associated with it in the first vowel (Martin and Bunnell 1982). Between these early studies and more recent research on fine phonetic detail in speech perception, however, a change in the focus of interest has occurred. Whereas the former centered on the role of coarticulation in phoneme identification, the latter extends the potential influence of fine phonetic detail to higher levels of processing, ranging from lexical access and syntactic parsing to the interpretation of conversational interaction. Central to this line of research is the assumption that information contained in fine phonetic detail can percolate up to the lexical level and above, contrary to an approach to speech perception in which access to meaning from the speech signal is accomplished through the mediation of a sequence of abstract intra-lexical units from which fine phonetic detail is left out.

[5] The studies cited here were conducted on either American or British English.

Some of FPD-oriented research takes the view that the central place typically attributed to the lexicon in theories of speech understanding has led to an overemphasis on short-domain phonetic properties related to phonemic contrasts. It is argued that shifting the focus away from the lexicon facilitates the identification of fine phonetic properties associated with phonological contrasts that are spread over long temporal windows, and/or which perform functions other than lexical differentiation. Thus, the phonetic exponents of phonological contrasts (including fine phonetic detail) have been shown to differ in function words as opposed to content words, a phenomenon attributed to the fact that function words and content words form two different systems of contrastivity, with a restricted inventory and less variation in phonological structure for function words (Local 2003). Likewise, it is now established that fine phonetic detail is related in systematic ways to the time course of conversational interaction, and in particular to patterns of turn-taking and cooperative exchange of information in a conversation (e.g. Local 2003; Plug 2005; Ogden 2006; Szczepek-Reed 2006).

The phonetics of conversational interaction (e.g. Couper-Kuhlen and Ford 2004) is an area in which evidence for the role of FPD in speech perception is growing. In the course of such an interaction, the behavior of each talker can evolve with respect to that of the other talker in two opposite directions: it may become more similar to the other talker's behavior (a phenomenon referred to as convergence) or more dissimilar. Convergence effects have been shown to be systematic and recurrent, and manifest themselves under many different forms, which include posture (e.g. Shockley et al. 2003), head movements and facial expressions (e.g. Estow et al. 2007; Sato and Yoshikawa 2007), and, as regards speech, vocal intensity (Natale 1975), pitch curve (Gregory et al. 1993; Bosshardt et al. 1997), and rate of speech (Giles et al. 1991). These phenomena may facilitate conversational exchange by contributing to setting a common ground between speakers (Giles et al. 1991). Recent studies (e.g. Pardo 2006) have indeed shown that perceived similarity in pronunciation between talkers increases over the course of the interaction and persists beyond its conclusion. Conversational interaction therefore seems to have long-lasting effects on the pronounced form of words, and this may be taken as indicating that words are stored in memory in a form that is highly dependent on their context of occurrence. More specifically, it now appears that the representations associated with words in the mental lexicon for each talker may dynamically evolve during conversation under the influence of the other talker's speech patterns, and retain the traces of that influence once the conversation has ended.

14.2.6 Conclusion: Towards hybrid models of speech perception and understanding

Experimental evidence is now available that provides support for the role of both detailed phonetic characteristics and abstract phonological categories in speech

perception. This has opened the way towards the development of new models of speech perception and understanding that aim to bridge the gap between the exemplar-based and abstractionist approaches. For example, Tuller and her colleagues (Tuller et al. 1994; Tuller 2004; see also Nguyen et al. 2009) have proposed a model that uses concepts from the theory of non-linear dynamical systems to account for the mechanisms involved in the categorization of speech sounds, and according to which there are two complementary aspects to speech perception. On the one hand, speech perception is assumed to be a highly context-dependent process sensitive to the detailed acoustic structure of the speech input. On the other hand, it is viewed as a non-linear dynamical system characterized by a limited number of stable states, or attractors, which allow the system to perform a discretization of perceptual space and which are associated with abstract perceptual categories. The recent development of so-called hybrid models (Hawkins 2003; Luce and McLennan 2005; McLennan and Luce 2005; Pierrehumbert 2006a; Hawkins 2010b) is also governed by the assumption that detailed phonetic properties and abstract phonological categories combine in the representations for words in memory.

CHAPTER 15

EMERGENT INFORMATION-LEVEL COUPLING BETWEEN PERCEPTION AND PRODUCTION*

BOB MCMURRAY AND
ASHLEY FARRIS-TRIMBLE

In this chapter, the authors argue for a new view of the production-perception coupling. The mechanisms involved are investigated through three case studies: one addressing online processing mechanisms, the second considering statistical learning mechanisms, and the third arguing for the role of parsing as a unifying mechanism. They conclude that much of this coupling is emergent, with perception playing a major role.

* The authors would like to thank Keith Apfelbaum, Jennifer Cole, Matt Goldrick, Kristine Kovack-Lesh, Joe Toscano, and Allard Jongman for discussions during the development of these ideas. This research was supported by NIH grant DC008089 to Bob McMurray.

15.1 Introduction

Speech perception is difficult. Speech arrives at a high rate, and acoustic cues are variable and context-dependent (e.g. Delattre et al. 1955; Öhman 1966; Nguyen, this volume). As a result there are few unambiguous points in the signal. These cues are mapped onto a large set of similar words (Marslen-Wilson 1987; Luce and Pisoni 1998), creating opportunities for competition and confusion.

If we assumed that the acoustic signal consistently reflected talkers' intentions, speech perception might be easier. Yet this is only half the problem. Speech production requires articulator sequencing (Fowler 2007; Port 2007b; Gafos and Goldstein, this volume) and involves competition between representations (e.g. Dell 1986; Goldrick and Blumstein 2006), potentially leading to misproductions. There is nonlinearity in the articulatory/acoustic interface: for some categories, large ranges of articulation yield similar acoustics (Johnson, Ladefoged, and Lindau 1993; Perkell et al. 1993), while for others, small changes create qualitatively different sounds (e.g. Stevens 1999; Stevens and Keyser 2010). Thus, understanding the content created by production is crucial for understanding speech perception.

Successful communication requires parity between talkers and listeners (Liberman and Whalen 2000): the message transmitted must correspond to the message received. Achieving parity is generally assumed to require a common set of units. The talker must encode the message in the same elements that the listener uses to decode it. This raises two questions: how is parity achieved, and where does it come from?

This chapter addresses these questions from the standpoint of perception. Prior approaches emphasize parity in the *units* of perception-production, which is an implicit consequence of a longstanding assumption that perception computes a single discrete unit (e.g. a phoneme or feature) at any given point in time.[1] In contrast, we will show that if perception computes a probability distribution over many possible units, *information-level* parity emerges as a by-product of processes like online competition and statistical learning. We demonstrate this in three case studies examining online processing, development, and their interaction. They illustrate that even if we assume no explicit perception-production coupling, but embrace a distributed and probabilistic model of perception (and the mechanisms that undergird it), information-level coupling emerges.

[1] Approaches to speech that compute multiple units at different times (e.g. unpacking a series of overlapping gestures, Fowler 1984) still implicitly make this assumption. The critical distinction is whether for any cue or set of cues in the signal (e.g. a VOT of 35 ms), does the system output a single item (e.g. a voiceless sound) or a distribution (it is 80% likely to be voiceless and 20% likely to be voiced).

15.1.1 How is parity achieved?

A hotly debated issue in speech perception is whether there are common units for perception and production. Is speech perception sensitive to the properties of production, and/or is production sensitive to the demands of perception? Such coupling could provide a set of common units in the service of communicative parity.

Some argue that production is fundamental: perception is encoded in terms of articulatory gestures (Fowler 1996) or the motor commands that generate them (Liberman and Mattingly 1985). Such arguments were initially made in response to the lack of invariant cues in the signal (Liberman and Mattingly 1985) and/or theoretical claims about the nature of perception (Fowler 1996). However, gestural encoding also has been argued to serve as a source of parity (Liberman and Whalen 2000).

Others place the burden on acoustics: articulation is planned in terms of acoustic or perceptual targets (Perkell et al. 1997), and languages emphasize perceptually salient contrasts (Ohala 1996). This leaves the perceptual system unconstrained, while the input is shaped to capitalize on its properties. This approach was initially motivated by the demands of articulation and the properties of the aerodynamic-acoustic interface, but it also offers parity, with acoustic features as the shared units.

Others sidestep the issue of coupling, treating perception as independent of production. Such approaches emphasize general auditory principles (e.g. Holt and Lotto 2008) or principles of memory and statistical learning (e.g. Goldinger 1998; Pierrehumbert 2003a). They argue that parity does not require strong perception-production coupling: parity in meaning is sufficient (Diehl et al. 1991). As long as listeners recover the intended words, the units are irrelevant.

This debate implicitly emphasizes one type of coupling, what we term *unit-level* coupling. Are the units of the message, the alphabet so to speak, articulatory or acoustic? In one sense, this still frames speech as a sequence of units, even if these units are dynamic, non-linear, and overlapping.

There is a more general approach, however, which focuses on information more than structural units. On this view, production is understood to create a non-uniform distribution of possible speech signals—not all sounds are equally likely. Thus, whatever the units are, perception "expects" the information content it receives to reflect the informational (or distributional) properties of production. We term this *information-level* coupling, reflecting the information-theoretic emphasis on the statistical distribution of units rather than their identities (e.g. Shannon 1948).

Work by MacNeilage and Davis (2000; MacNeilage et al. 2000) offers an example. They show that consonant-vowel (CV) syllable structure is basic to articulatory organization due to biomechanical constraints (the jaw functions as an oscillator). These constraints lead particular CVs to be favored over others in both infants' early vocalizations and in the words of many languages (coronals/front vowels and

velars/back vowels). Information-level coupling, then, could arise if the perceptual system expects CVs over VCs and if it processes particular CVs differently, regardless of whether these are encoded as acoustic features (e.g. Stevens 2002) or articulator movements (e.g. Browman and Goldstein 1986).

At any point in processing, information-level coupling emphasizes the statistical distribution of possibilities, rather than a single unit. Speech production is shaped by processes like lexical competition, articulation dynamics, the structure of the language, and the aerodynamic-acoustic interface. As a result, for any phonetic category there is a distribution of observed forms across utterances. For perception to be sensitive to this distribution, then, it must encode more than a single gesture or feature. Rather, it must recover the *likelihood* that various events caused the signal. For example, rather than determining if a given segment is a labial, the perceptual system must determine the likelihood that it was generated by a labial, a coronal, or a velar. As a result, perception may be better described as encoding a distributed representation of possible interpretations, rather than a single unit. Such representations are common in connectionism (e.g. McClelland and Elman 1986), and the emphasis on the likelihoods of causal events is in the spirit of Bayesian approaches (e.g. Clayards et al. 2008; Feldman et al. 2009), although we do not take a strong stance on Bayesian approaches (cf. Nearey and Hogan 1986; Toscano and McMurray 2010).

The probabilistic and distributed nature of perception downgrades the issue of individual units,[2] emphasizing statistical structure instead. It also expands the scope of inquiry beyond speech perception or production. The informational properties of the signal reflect not only articulatory factors, but cognitive factors like competition between words (Dell 1986; Rapp and Goldrick 2000; Goldrick 2007), or the fact that phonetic cues are conditioned on lexical (Connine 2004; Connine et al. 2008) or prosodic (e.g. Salverda et al. 2003; Salverda et al. 2007) factors. In perception, the structure of the lexicon can shape perception (Magnuson et al. 2003; McClelland et al. 2006; Newman et al. 1997), e.g. in English, after /fulɪ/, /s/ is less likely than /ʃ/ as *foolish* is a word and *fooliss* is not. Finally, parallelism is central to theories of word recognition (McClelland and Elman 1986; Marslen-Wilson 1987), so such theories may offer insight to information-level coupling.

Indeed, the objective of spoken communication is to translate meaning to articulation, and sound to meaning. Thus, to understand perception-production coupling in this broader communicative context, it is important to consider perception and production at their broadest levels, the complete mapping between meaning and articulation, and between acoustics and words. For our purposes, we simplify

[2] In fact, two distributed representations using different component units (e.g. gestures vs. acoustic features) can often be easily transformed into each other, as long as they are sufficiently dense (multiple units involved to represent any given input), and individual units have high resolution. Thus, when considering the information content of speech as a distributed representation, gestural and acoustic representations may be somewhat isomorphic and the particular choice of units may not have dramatic consequences for downstream processing.

this by assuming words as a small tractable unit of meaning (though see Hawkins, this volume).

15.1.2 Where does perception-production coupling come from?

Articulatory approaches to perception largely assume perception-production coupling as a design principle (acoustic goal approaches do the same for production). Liberman and Mattingly (1985) argue that an articulatory encoding of speech is the only way to solve the invariance problem. Similarly, Fowler and Rosenblum (1991) as well as recent work by Rizzolatti and Arbib (1998) on mirror neurons, argue that coupling arises out of the general principle that perception is organized to detect the cause of sensory stimulation (direct realism and/or embodiment). Under these views, perception-production coupling is the product of evolution (Rizzolatti and Arbib 1998), hardwired or "innately specified" (Liberman and Mattingly 1985: 3).

What if we start from a different design principle: the idea that the perceptual system developed to recognize meaning from acoustic input (e.g. Gow and McMurray 2004)? Starting here, could coupling be observed?

It is possible that coupling emerges from complex processes built around independent demands. To illustrate, consider Strogatz and Stewart's (1993) example of pendulum clocks. They describe how two pendulum clocks on the same wall eventually phase-lock such that both pendula reach their peaks and troughs simultaneously, even when the oscillation of one is perturbed to push it out of phase. Neither clock has an explicit mechanism coupling it to the other clock—both were designed simply to keep time. However, pendula are sensitive to subtle vibrations on the wall, and the result is emergent coupling, a by-product of complex systems sensitive to many forces.

In speech perception, similar coupling may be observed. Perceptual learning processes that undergird phonological development (e.g. Jusczyk 1993; Werker and Curtin 2005; Holt, this volume) evolved in the service of grouping sounds and recognizing words. Yet principles like statistical learning (Maye et al. 2002; McMurray, Aslin, and Toscano 2009), which underlie perceptual learning, could create information-level coupling because those statistical distributions were created by speech production. Similarly, word recognition systems built solely to recognize words (e.g. McClelland and Elman 1986) often entail real-time dynamics like competition, which are also posited in production (e.g. Dell 1986). The resulting complementarity may lead to information-level coupling.

Like the clocks, perception-production coupling may be *emergent*, a by-product of two complex, yet independent systems (production and perception) that are sensitive to subtle signals. Perception-production coupling is no less useful (for perception) or interesting (for phonology) as an emergent property than as a

design principle. In fact, emergence may offer a better explanation of the source of coupling by eliminating the need for oversimplified accounts of development and evolution in which such coupling is "built in" (cf. Lickliter and Honeycutt 2003 and Spencer et al. 2009).

15.1.3 Overview

This chapter pursues these ideas by asking two questions. First, we ask whether there is evidence of information-level coupling in speech perception. Second, we ask whether such coupling is emergent from processes like real-time word recognition and development. We start from the minimal assumption that mapping sounds to meaning (words) is the primary goal of speech perception, and we must only assume parity at the word-level for successful communication. We do not suggest this as a theory. Rather, starting here offers a better position from which to identify emergence when we find information-level coupling in theories that were developed without reference to it. Finally, we note that emergence is crucially supported by counterexamples—places where the same processes that give rise to information-level coupling also yield phenomena that look uncoupled. Such examples highlight the fact that coupling is not an organizing principle, but a by-product.

Three case studies address these issues. The first argues that perception and production systems based on similar online processing mechanisms (McClelland and Elman 1986; Dell 1986) result in what look like systems tuned to each others' demands. The second demonstrates that statistical learning mechanisms (e.g. Maye et al. 2002) underlying development cause the perceptual system to internalize the regularities created by production, again showing informational coupling. The third shows how simple processes that compensate for talker and coarticulation by evaluating incoming speech relative to expectations (Cole et al. 2010; McMurray and Jongman 2011; McMurray et al. 2011) build on interactive activation and statistical learning to partition out both indexical and articulatory variance in the signal.

15.2 Case study 1: Online processing

Information-level coupling may arise as a consequence of a common processing principle: interactive activation. This commonality may lead perception to reflect the distributional properties of production, even if coupling is not an organizing principle for either system. Analogous work on word recognition and speech errors offers a clear illustration.

A fundamental issue in spoken-word recognition is time. The acoustic material comprising a word unfolds over time, and at early points there is ambiguity. For example, after hearing only [bi], listeners expect 'beach,' 'beak,' or 'beat,' and it is not until the final consonant that these can be disambiguated. Work on word recognition employs these facts to study the mapping and disambiguation processes that identify words in real time (e.g. Marslen-Wilson 1987; Allopenna et al. 1998; Dahan and Gaskell 2007). Such work has yielded a remarkable consensus: from the earliest moments of perception multiple words are activated in parallel, activation is updated by continually arriving material, and words compete for recognition.

This process has been described by interactive activation models (McClelland and Elman 1986; Luce et al. 2000; Figure 15.1 below), in which a small set of units are activated corresponding to the perceptual input. Activation then spreads to phonemes and words, resulting in the parallel activation of multiple interpretations of the signal at each level. These compete and feed back to affect processing at earlier levels as the system arrives over time at a single decision. In this dynamic process, bottom-up activation flow, competition, and feedback occur in small increments at each time-point.

In such models there are no distinct processing stages—processes like phonetic categorization occur simultaneously with lexical access. Activation is graded: it depends on bottom-up factors (e.g. match to the input) and top-down factors (e.g. lexical structure). Finally, representation at any level is distributed. Activating /b/, /i/, and /k/ at the phoneme level partially activates 'beat' and 'bead' (since they are connected to two of these phonemes), and activates 'beak' more (it is connected to all three). Thus, the activation *across* the lexical layer offers a probabilistic encoding of the array of possible inputs and something approximating their relative likelihoods given the signal.

Interactive models largely emphasize the mapping between features or phonemes and words (though see Elman and McClelland 1986). However, broadly construed, this framework proposes that intermediate states of perceptual processing immediately cascade to higher levels. That is, lexical candidates start to build activation before categorization of continuous acoustic cues is completed.

Supporting evidence comes from recent eye-tracking studies asking whether fine-grained acoustic differences are reflected in lexical activation (McMurray et al. 2002; McMurray et al. 2008; see also Andruski et al. 1994; Speer, this volume). In one such study, participants heard a token from one of six nine-step voice onset time (VOT) continua spanning two words (e.g. 'beach'/'peach,' 'bump'/'pump') and clicked on a picture of the word they heard, selecting it from a screen containing the target ('beach'), its competitor ('peach') and two unrelated objects. Their eye movements were measured as they did this, since the likelihood of fixating each object exhibits a fairly close correspondence to lexical activation dynamics (Tanenhaus et al. 1995; Allopenna et al. 1998).

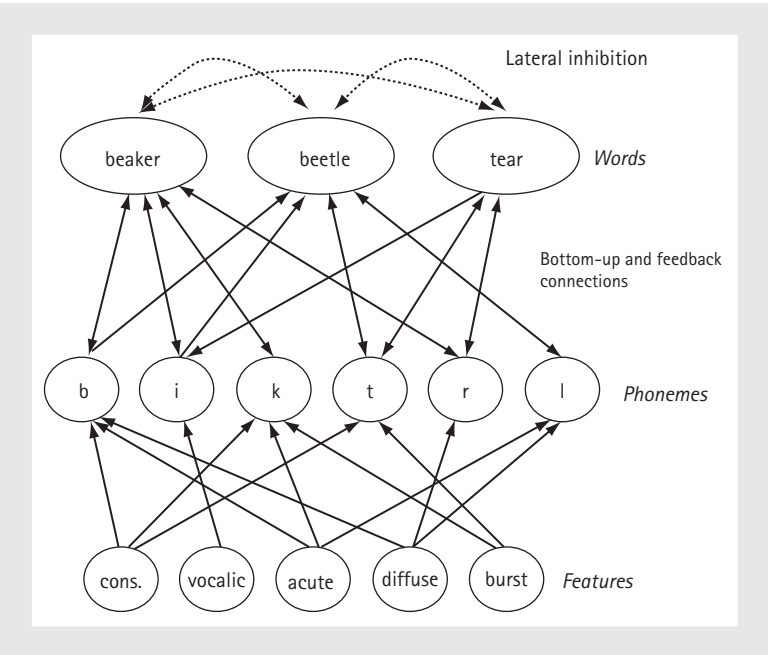

Figure 15.1. A schematic diagram of TRACE (McClelland and Elman 1986). The model receives input when particular feature nodes are activated indicating the detection of the corresponding cue. Activation spreads to the phonemes that are connected to the active feature nodes, and phonemes inhibit each other. Activation then spreads from phonemes to the appropriate words, which also compete. Finally activation feeds down from lexical representations to the phonemes that are connected to them and the cycle continues.

Participants made increasingly more fixations to the competitor as VOT approached the category boundary, even when trials in which the subject clicked the competitor were excluded (Figure 15.2). That is, the activation for the competitor object gradiently reflected acoustic differences *within a category*. This supports the claim that lower-level processes (e.g. speech categorization) are not completed before higher ones (e.g. lexical access) begin; rather their intermediate states cascade in real-time. Thus, the pattern of activation for lexical candidates potentially reflects a distributed probabilistic encoding of the *acoustic* signal, particularly given the vast number of similar words in the lexicon. After hearing 'beak,' for example, the relative activation for 'peak' reflects VOT; 'geek' reflects variation in place of articulation cues; and 'back' reflects vowel height.

How does this parallelism between production (via acoustics) and perception lead to information-level coupling? This is vividly seen in complementary studies

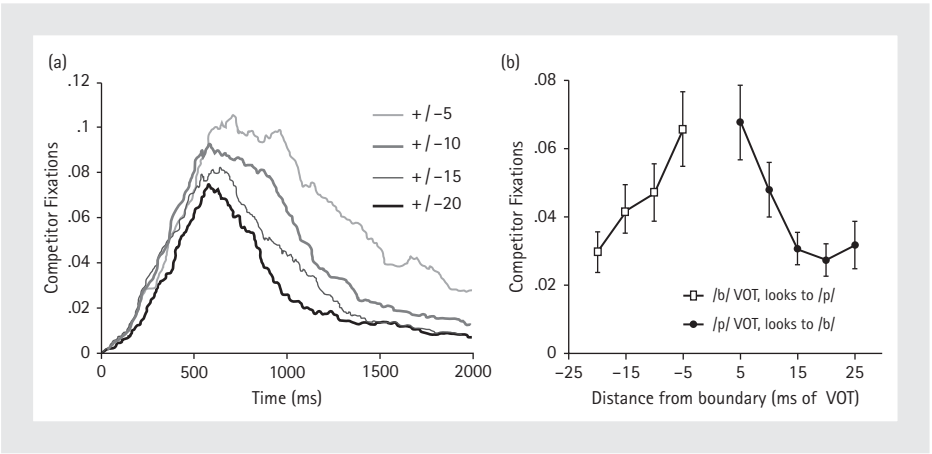

Figure 15.2. Results from McMurray et al. (2002). (a) Fixations to competitor (e.g. *peach*, when the target had a low VOT) as a function of time and distance from each subject's category boundary. (b) Average fixations to the competitor (area under the curve) as a function of distance from the category boundary (adapted from McMurray et al. 2002).

by Goldrick and Blumstein (2006) and McMurray, Tanenhaus, and Aslin (2009). Goldrick and Blumstein (2006) point out that interactive activation also underlies *speech production* (Dell 1986; Rapp and Goldrick 2000). Thus, they predicted that the production system should also exhibit a continuous cascade to the level of the signal (see also Gafos and Goldstein, this volume). To test this, they elicited misproductions of stop consonants using a tongue-twister paradigm and showed that when voiced consonants were mispronounced as voiceless, their VOTs were 8 ms *shorter* than a true voiceless consonant. Conversely the VOTs of voiceless consonants mispronounced as voiced were 5 ms longer than true voiced consonants.[3] This implies continuity between phonological and articulatory processes. Competition between the voiced and voiceless units is not completed prior to articulatory planning, resulting in an intermediate form. As a result, these cascading processes have consequences for the statistical distribution of speech cues.

If the perceptual system could harness these regularities, information-level coupling would emerge. If the first segment of a word were mispronounced, the perceptual system might make an early commitment to an incorrect word (Figure 15.3a).

[3] They also found that the vowel length (a secondary cue to voicing) was unaffected by mispronunciations: voiced sounds mispronounced as voiceless had the same longer vowel lengths as an underlyingly voiced sound. That is, despite the incorrect VOT, the underlying form was still completely observable in the vowel length. Thus, the degree to which a cue is sensitive to competition from the intended articulation is a function of its importance to signaling the contrast, its position (vowel length comes later in the word), or its interactions with other cues (e.g. prosody).

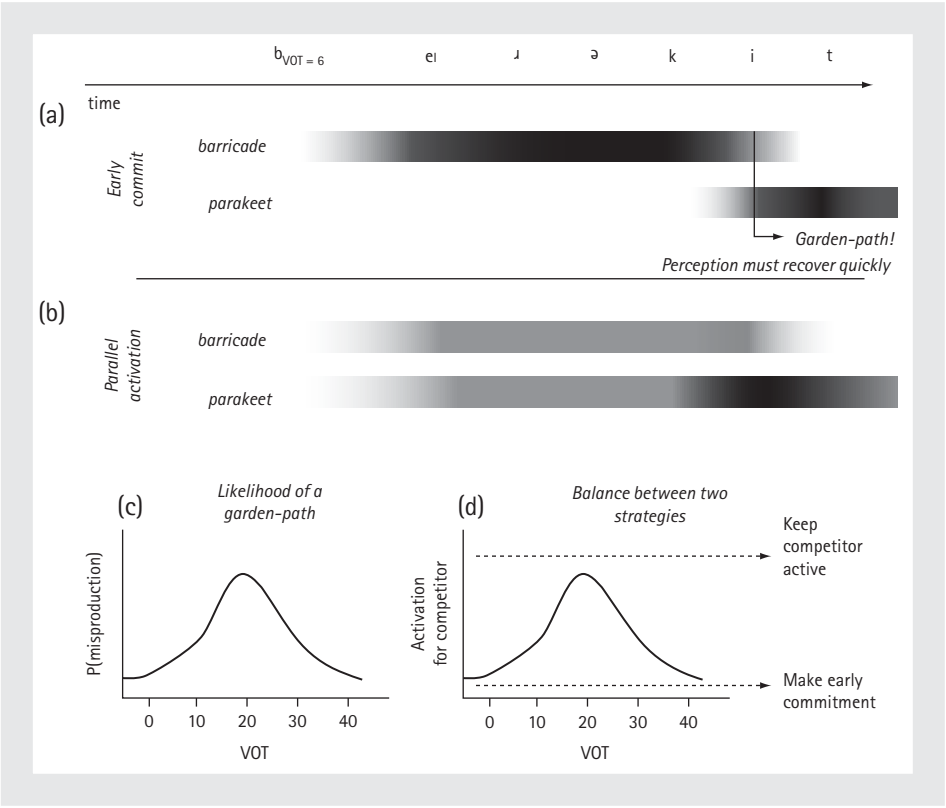

Figure 15.3. Strategies for coping with misproduction (and other sources of uncertainty). Activation for 'parakeet' and 'barricade' as a function of time if (a) listeners make an immediate commitment to a single candidate, or (b) listeners partially activate and wait for the disambiguating information. (c) Goldrick and Blumstein (2006) suggest that the probability that a given phoneme was mispronounced is a function of its VOT—unambiguous VOT values are more likely to be correctly pronounced. (d) Thus, a useful strategy would be to modulate the commitment (activation) to the competitor (e.g. the balance between the two strategies) as a function of VOT: listeners should commit to a candidate when they are sure the input is not a mispronunciation and maintain both options when it may be.

When this commitment was revealed to be inconsistent with later material, the system would have to revise this interpretation, a costly "garden-path." For example, if 'barricade' were mispronounced with a /p/, the system might commit to 'parakeet' and be unable to revise several phonemes later at '-ade.' However, if the system was sensitive to fine-grained differences in VOT, it could maintain partial activation for both options, allowing it to quickly activate the correct one when the disambiguating information was heard (Figure 15.3b). Goldrick and Blumstein (2006)

show that the likelihood that an utterance was mispronounced is a function of its VOT: mispronounced forms have VOTs closer to the boundary than correct ones (Figure 15.3c). Thus, if the system modulated competitor activation as a function of distance from the category boundary, it would "hedge its bets," keeping both competitors available when a misproduction is likely, but committing when it is not (Figure 15.3d).

The gradiency found by McMurray et al. (2002), a consequence of interactive activation, could contribute to coupling. A VOT of 8 ms (likely a mispronounced /p/) will lead to more activation for the voiceless competitor than 0 ms. However, for listeners to take advantage of this gradiency, it must persist long enough to participate in recovery when the disambiguating material arrives.

McMurray, Tanenhaus, and Aslin (2009) assessed this by presenting subjects with ten-step VOT continua ranging from 'barricade' to 'parricade' and 'parakeet' to 'barakeet,' words that have on average 250 ms between the VOT and the disambiguating material. The subjects' task was to click on the referent from a screen containing the two competitors ('barricade,' 'parakeet'), two unrelated words and an X (to indicate a non-word, e.g. 'parricade'). Eye movements generated *before* the point of disambiguation (e.g. '-ade' in 'barricade') reflected a gradient commitment on the basis of VOT—the likelihood of fixating was linearly related to VOT (Figure 15.4b). More importantly, the eye movements generated *after* the point of disambiguation (POD) showed that even if subjects committed (looked) to the (incorrect) competitor prior to the POD, they were faster to recover if the VOT was closer to the target (Figure 15.4a), suggesting that the early gradiency was retained for several hundred milliseconds,[4] affecting recovery.

These complementary studies imply that listeners take advantage of statistical regularities of misproduction—they make a gradient commitment based on continuous cues, and then reap the benefits if they need to recover. If the distribution of such errors constitutes a component of the information content of speech production, then perception appears geared to it. But speech errors are just one domain in which such hedging would be useful; ambiguity in VOT, for example, can also be created by talker differences (Allen et al. 2003), prosody (Fougeron and Keating 1997), and speaking rate (Kessinger and Blumstein 1998). Keeping competitors active in proportion to their likelihood is an efficient strategy for coping with variability created by speech production more generally.

However, this coupling is not an organizing principle of either system. Under this view, perception and production are independent, though they operate by similar principles. Coupling, then, is a by-product of the fact that in both cases (1) graded

[4] Moreover, it did not appear that subjects were ever committing to a sublexical interpretation. The point of disambiguation averaged 250 ms, so subjects must have been maintaining a gradient interpretation at least that long. Simulations with TRACE extended this, showing that TRACE fails to account for these data until phoneme inhibition (the process by which TRACE makes a discrete decision at sublexical levels) is eliminated.

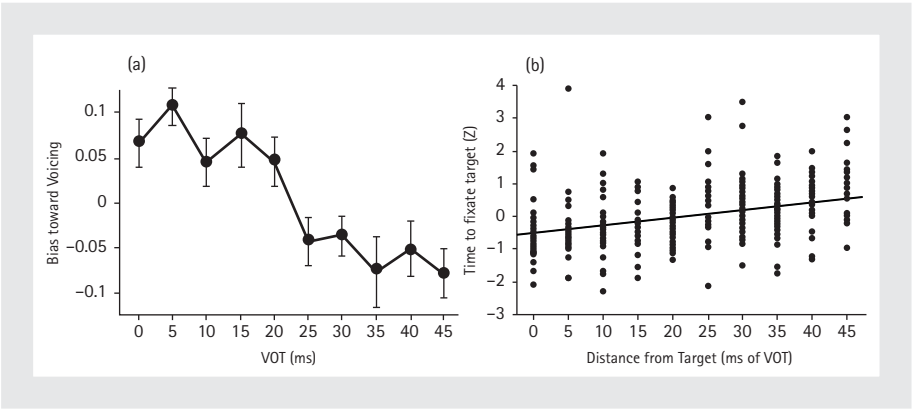

Figure 15.4. Results from McMurray, Tanenhaus, and Aslin (2009). (a) The probability of making an initial commitment to the voiced sound as a function of VOT. (Figure adapted from McMurray, Tanenhaus, and Aslin 2009.) (b) The latency between the point of disambiguation and the subject's correct fixation. This was fastest when the VOTs were at the target (e.g. a 0 ms 'barricade' or a 45 ms 'parakeet') and increased with distance.

activation continuously cascades between lexical and perceptual or articulatory processes; and (2) multiple items are maintained in parallel. In fact, interactive activation has been posited in domains like music perception (Bharucha 1987), object segregation (Vecera and O'Reilly 1998), and visually comparing objects (Goldstone and Medin 1994), where no coupling is needed. It is a general class of processes, not specialized to speech. Nonetheless, the fact that both production and perceptual systems operate by these principles allows coupling to emerge. Continuously cascading processing means that competition in production is reflected in the acoustic signal; similarly, consideration of lexical candidates continuously reflects the input. Serendipitously, this is advantageous for perception.

The independence of these systems (and hence emergent coupling) is illustrated by two examples. First, in production, words in high-density neighborhoods are produced faster than those in low-density ones (Vitevitch 1997, 2002); however in perception, neighborhood density exerts a slowing effect (Luce and Pisoni 1998). Second, due to lexical feedback, perception is good at coping with non-words that differ from the target by a single feature (Ganong 1980; McMurray, Tanenhaus, and Aslin 2009; McClelland et al. 2006 for a review). If coupling were a design feature, misproductions should favor non-words to take advantage of these perceptual processes and avoid neutralizing contrast. Yet, misproductions are more likely to form a word (e.g. Dell and Reich 1981; Dell 1986). In both cases, differing effects of production and perception arise from analogous interactions between lexical structure and online competition, suggesting that interactive activation may be

a more useful way to characterize each system than coupling. Nonetheless, the Goldrick and Blumstein (2006) and McMurray, Tanenhaus, and Aslin (2009) work shows how common processing principles lead to perceptual processes that appear tuned to the informational properties of production.

Yet recent results imply coupling as a goal of the system. Clayards et al. (2008) employed the eye-tracking paradigm of McMurray et al. (2002) to measure sensitivity to differences in VOT. However, the distribution of VOTs across trials was not uniform. One group of subjects heard VOTs centered consistently around two prototypes, with little within-category variation; for a second group, prototypes were the same, but there was more variation around them. While evidence of a gradient representation was observed in the latter group, the group hearing more consistent data showed little sensitivity to fine detail. Listeners appeared to explicitly tune their expectations about VOT to the distribution in the input, with the consequence that when VOT is variable, the lexicon is ready with a gradient, parallel representation. In the next case study, we examine mechanisms that might underlie such tuning and show that this may be an emergent result of basic learning processes.

15.3 Case study 2: Developmental mechanisms

Our second case study illustrates that informational coupling need not be inherent in the perceptual system, but can instead arise over the course of development via domain-general statistical learning mechanisms.

A fundamental problem in the development of speech perception is how infants learn the phonological categories of their language. This has received enormous attention (e.g. Eimas et al. 1971; Kuhl 1983; Werker and Tees 1984; Werker and Polka 1993; Werker and Curtin 2005), and a standard story has emerged. Soon after birth infants can discriminate most of the sounds of the world's languages (though not all: Eilers and Minifie 1975; Eilers et al. 1977), but by twelve months, the set of contrasts has been pruned to only those of their native language.

A compelling account for this change is a type of statistical learning mechanism, distributional learning (see Holt, this volume and Maye, this volume for a discussion of this and other factors). Acoustic analyses consistently show that across talkers and contexts, acoustic cues cluster around prototypical values. The distribution of VOTs in English in word-initial position, for example, shows two clusters, one centered at 0 ms (voiced sounds) and one centered at 60 ms (voiceless sounds) (Lisker and Abramson 1964; Allen and Miller 1999; Figure 15.5).

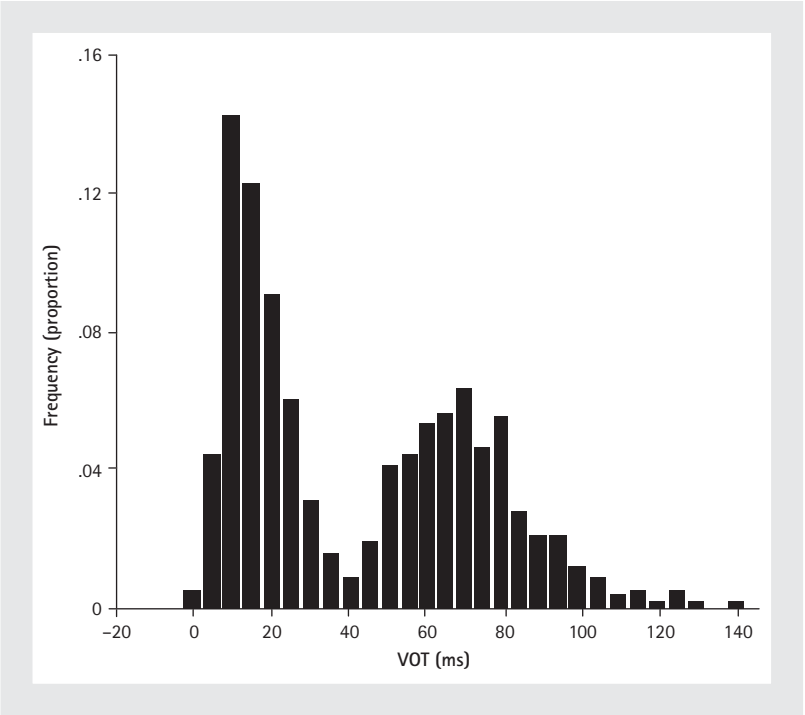

Figure 15.5. The frequency of individual VOT values in English (data adapted from Allen and Miller 1999; figure reprinted from McMurray, Aslin, and Toscano 2009, with permission).

Such distributions can also be seen in analyses of the vowel space (Peterson and Barney 1952; Hillenbrand et al. 1995), formants of approximants (Espy-Wilson 1992), and multiple cues for fricatives (Jongman et al. 2000). Simply eyeballing a plot of the statistical distribution of a cue (a histogram, e.g. Figure 15.5) is often sufficient to identify the number of categories, their shape, and location. Distributional learning suggests that infants learn speech categories in the same way, counting the likelihood of individual cue values, and using simple clustering techniques to identify the categories.

Work by Maye et al. (2002) suggests infants can employ such learning. They exposed infants to a series of syllables in which VOT varied according to a unimodal or bimodal distribution. Infants exposed to the unimodal distribution failed to discriminate pairs of VOTs from this continuum, but those receiving bimodal input succeeded, suggesting that the statistical structure of the input can diminish or create phonetic contrast (see also Maye et al. 2008).

Distributional learning can account for many of the developmental patterns outlined above. It describes how infants acquire categories, and with some simple assumptions, it explains discrimination performance. McMurray, Aslin, and Toscano

(2009) instantiated distributional learning in a simple computational model and demonstrated that it can account for the discrimination and loss of non-native contrasts, the enhancement of difficult contrasts, and the discrimination of prototypical from non-prototypical exemplars of the same category.

Distributional learning also gives rise to informational coupling. Studies by Miller and colleagues have used goodness ratings to map large regions of the voiceless category (e.g. VOTs from 0 to 240 ms) and have shown graded prototype effects that correlate well with the statistical distribution of the cues (Miller and Volaitis 1989; Miller 1997 for a review; Miller and Eimas 1996 for work with infants). Thus, gradient patterns of lexical competition observed in the eye-movement studies of McMurray et al. (2002; McMurray and Aslin 2005 for infants) may derive directly from statistical distributions, a by-product of *developmental* mechanisms. This persistent developmental/learning process can also account for Clayards et al.'s (2008) results in which manipulating cue distributions alters adults' sensitivity to fine-grained detail. Even if informational coupling is not an explicit goal, ongoing statistical learning (a remnant of development) will give rise to it.

Such coupling leads to precise expectations based on the statistics of production. For example, across languages, short-lag VOTs (typically interpreted as voiceless sounds in languages with pre-voicing, and voiced sounds in languages with aspiration) exhibit narrow categories, but pre-voiced and aspirated stops show wider categories. This is due to both articulatory constraints and contextual factors (speaking rate exerts greater influence on pre-voiced and long-lag VOTs: Kessinger and Blumstein 1998). Using statistical learning one could extract such regularities and "learn" to tolerate less variance in the short-lag category. This is a clear example of information-level, rather than unit-level, coupling—whether the dimension is encoded in terms of gestural timing or an acoustic cue, the statistical properties are the same, and learning gives rise to similar results.

One could argue that such learning evolved to achieve coupling, that coupling is not emergent, but an explicit goal of statistical learning. Two pieces of evidence suggest this is not the case: work on infant-directed speech (IDS), and work on dimensional selection.

First, if statistical learning serves the explicit goal of achieving coupling, caregivers might modulate their speech to help it (in effect putting some of the onus on production). Such modulation in IDS could suggest that coupling is a design principle, while its absence is consistent with an emergent property of independent perception and production systems. Kuhl et al. (1997; see also Werker et al. 2006) offer evidence on this. They measured three point vowels (/i/, /ɑ/, and /u/) in English, Russian, and Swedish mothers' speech in free conversation with their infant and an adult. Across languages, there was greater separation among the point vowels in IDS than adult-directed speech (ADS). By distributional learning accounts, this

should facilitate the acquisition of these contrasts, suggesting that mothers tune their production to match their children's needs (e.g. Vallabha et al. 2007).

However, other cues suggest a more complex story. Englund (2005) examined VOTs produced by six Norwegian mothers and found that for *both* voiced and voiceless sounds, VOTs lengthened in IDS. IDS did not differentially enhance the voicing contrast, and may have made it more difficult (in IDS voiced velars and alveolars had VOTs of 25 ms—almost voiceless for ADS). This is more consistent with the effect of slower or hyperarticulated speech than with intentional enhancement of phonetic contrasts.

We have replicated this result with twenty caregivers and a laboratory task rather than free conversation, and found similar results (McMurray et al. in preparation): VOTs increase overall in IDS (Figure 15.6a). However, when speaking rate is factored in (e.g. the ratio of VOT to vowel length), the effect of IDS disappears (Figure 15.6b). Thus, IDS-induced changes in VOT may derive more from general rate changes rather than caregivers manipulating the acoustic contrast. Additionally, in a preliminary analysis of the vowel space, we found that while the point vowels do expand in IDS, interior vowels move as well (Figure 15.6c), and not clearly in a direction consistent with enhancement (c.f. /ɑɪ/ and /ɑɹ/). More importantly, the variance increases dramatically (see Figure 15.6d, 15.6e) such that it may be harder to distinguish sounds, even though the means are further apart. Table 15.1 illustrates this with a series of logistic regressions that discriminated pairs of vowels on the basis of F1 and F2. Performance was similar for the IDS and ADS classifiers, and a benefit for IDS was only observed in half of the models. Thus, the increase in variance may outweigh separation of the means in some cases.

Table 15.1. Performance of a series of logistic models trained to classify pairs of vowels in either ADS or IDS. Shown is percentage correct for each model. For each contrast, the better performing model is shown in bold

	Contrast	ADS	IDS	IDS Cost/Benefit
Close Contrasts	/iɹ/ vs. /eiɹ/	**71.4**	69.1	−2.3
	/æ/ vs. /ɑɪ/	71.0	**72.5**	+1.5
	/ɑɪ/ vs. /ɑɹ/	**81.7**	76.3	−5.4
	/ɑɹ/ vs. /ɝ/	79.3	**80.6**	+1.3
	/ɝ/ vs. /ʌ/	66.2	**75.6**	+9.4
Middle	/ʌ/ vs. /oʊ/	88.0	**93.1**	+5.1
	/eiɹ/ vs. /ɑɹ/	**92.5**	90.4	−2.1
Far	/iɹ/ vs. /oʊ/	99.0	**100**	+1
	/iɹ/ vs. /æ/	**94.7**	92.6	−2.1
	/æ/ vs. /oʊ/	**99.3**	99.0	−0.3

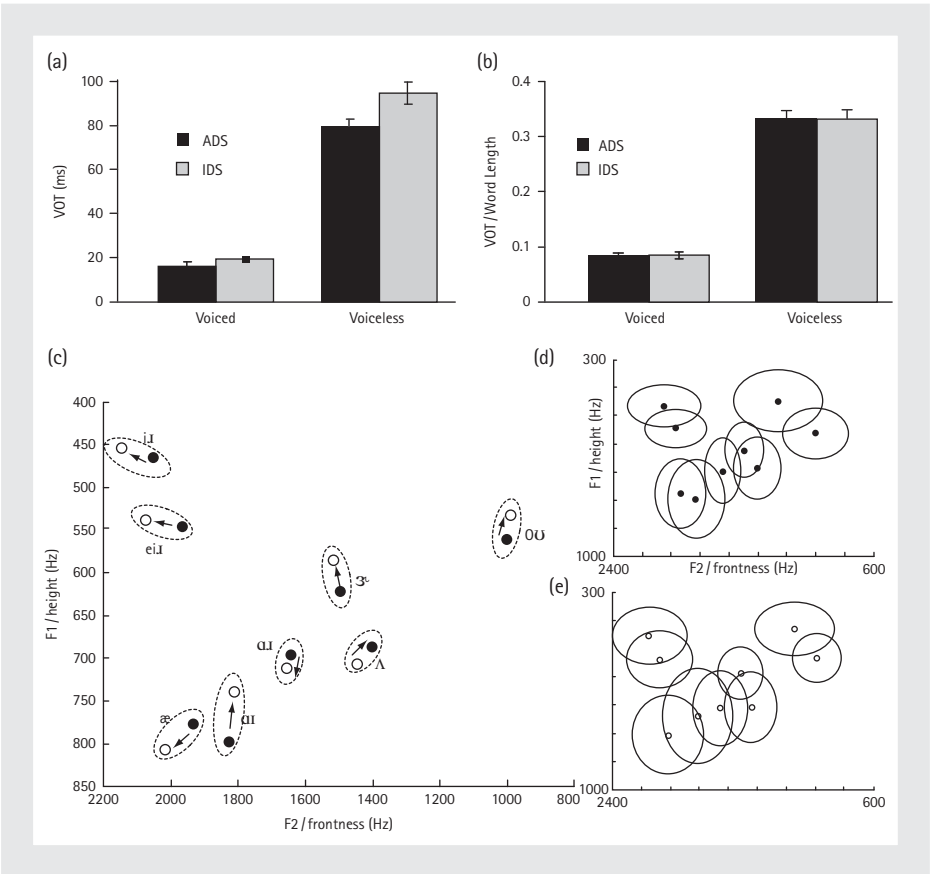

Figure 15.6. Phonetic analyses of IDS (McMurray et al. in prep). (a) VOTs for voiced and voiceless sounds in IDS and ADS. (b) Voicing when computed as the ratio of VOT to word length. (c) Preliminary analysis of the vowel space (the direction of the X and Y axes have been reversed to reflect height and backness). (d) and (e) Standard deviations (plotted as ellipses) for each category in ADS (d) and IDS (e).

Thus, in production, differences in IDS do affect the statistics, and infants are likely to be sensitive to this. Sometimes this benefits learning (e.g. the expansion of the point vowels), other times it may not (the increase in VOT and the variance in the vowels; see Kirchhoff and Schimmel 2005 for an engineering analysis). This suggests that enhancement is not a core principle of the developmental system, but rather an almost accidental property of articulatory factors and distributional learning.

A second argument for emergence comes from cue weighting. If distributional learning evolved to support coupling, one might expect the system to be biased to

do this for specific, relevant cues and to ignore irrelevant cues like talker or pitch, for the purposes of establishing phonetic categories. To the extent that this dimensional selection emerges on its own, the coupling arising from distributional learning may be emergent.

A recent series of studies in our lab (Rost and McMurray 2009, 2010) suggests that (1) the ability to ignore irrelevant cues is not a given—it develops late, and perhaps *after* infants can discriminate the categories of their language; and (2) a different form of statistical learning accounts for this ability. A puzzling finding in early word learning is that 14-month-olds who have mastered the ability to discriminate minimal pairs (e.g. [buk] / [puk]) cannot map them onto two different referents (Stager and Werker 1997; Werker, Cohen, Lloyd, Casasola, and Stager 1998). This was initially described as an effect of resource limitations or task demands (Fennell and Werker 2003), or unbalanced lexical competition (Swingley and Aslin 2007).

Rost and McMurray (2009) hypothesized that if speech categories were not fully developed, then hearing only a single token (or a few highly similar ones) may not offer enough support for phonological learning (Figure 15.7a). Thus, consistent with approaches to L2 acquisition (Lively et al. 1993) and visual categorization (Quinn et al. 1993; Oakes et al. 1997), they exposed infants to multiple exemplars of 'buk' and 'puk,' produced by different talkers (Figure 15.7b). This increases task difficulty, but could augment learning by providing more bottom-up information to separate the developing categories. Results favored this view, with infants exposed to multiple exemplars learning the words.

Follow-up experiments (Rost and McMurray 2010) asked what component of the variability was responsible. The contrastive cue, VOT, varied between talkers and exemplars, and this variation could support the sort of distributional learning described above. This was tested using a single talker, but manipulating VOT to create a bimodal distribution (Figure 15.7c). Infants failed in two replications. Only when VOT was held constant (within a word) and the *non-contrastive* cues (e.g. pitch and timbre) were varied did infants succeed (Figure 15.7d). Variability along irrelevant dimensions was necessary for infants to determine the invariant dimension (VOT). Thus, even at this relatively late age, infants don't know what cues to attend to—this must be acquired.

This makes it challenging to argue that the ability to perceive gestures or distinctive features is innate, but easy to see where it might come from. As children learn to downweight irrelevant information, they zero in on the correct cues, and use distributional learning to acquire their within-dimension organization. Thus, emergent information-level coupling may apply to the dimensions on which information is to be found (cues) as well as to organization within them (categories). Distributional learning is not constrained to informative dimensions—these too must be discovered.

But is dimensional weighting explicitly encoded in the system, or does it emerge from other processes? A number of approaches hypothesize that cues receive

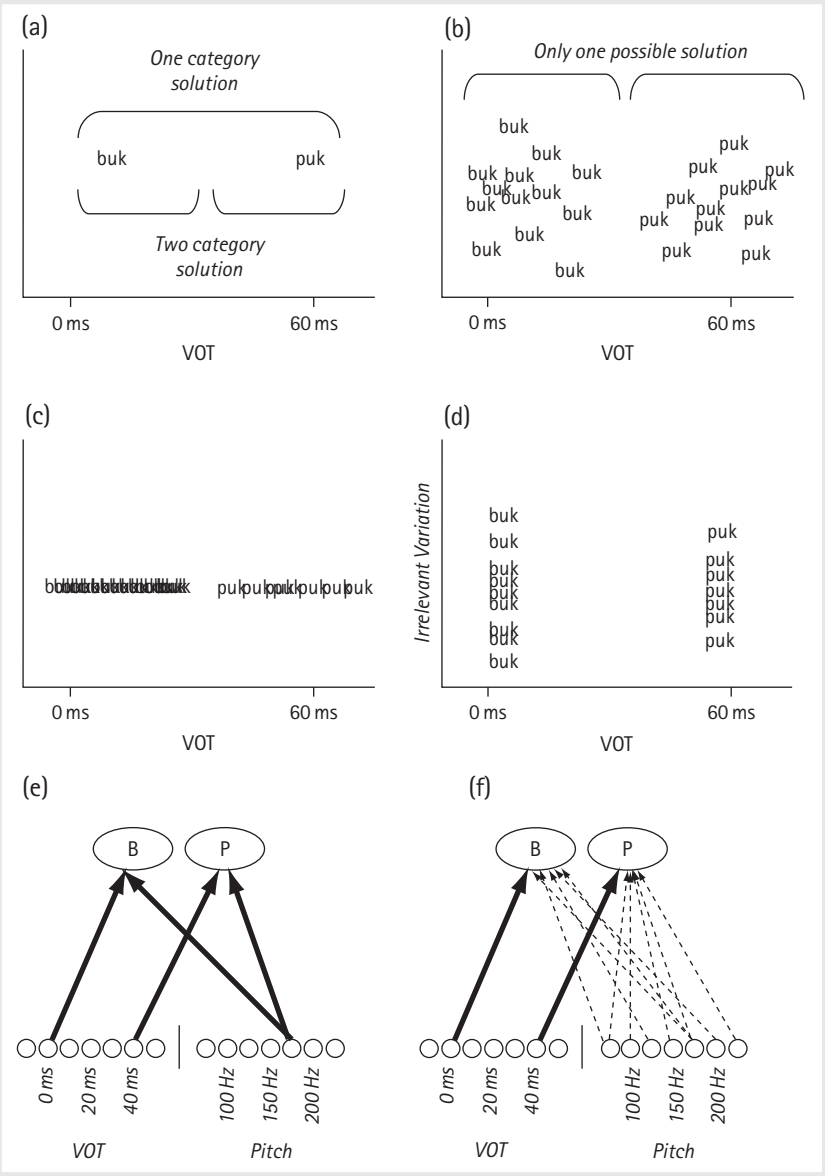

Figure 15.7. Conditions tested by Rost and McMurray (2009, 2010). (a) Single exemplars yield ambiguity in categorization. (b) Multiple exemplars varying in VOT (X-axis) and other factors (Y-axis) yields clear categories. (c) Variation in VOT alone did not yield evidence of successful learning. (d) Variation in speaker alone resulted in two categories. (e) An associative account of dimensional weighting. When pitch is constant during training (i.e. a single speaker) this may result in strong (erroneous) associations between pitch and category. (f) Variable pitch during training weakens the association between pitch and category, allowing the true VOT categories to emerge.

explicit weights that apply equally to the whole dimension (Ernst and Banks 2002; Jacobs 2002; Toscano and McMurray 2010) and are explicitly estimated from the statistics of the input. However, a compelling alternative view holds that weighting emerges *implicitly* in the links between individual values along a dimension (e.g. individual pitch or VOT values) and words or phonemes (Apfelbaum and McMurray 2011). Under this account, if only two VOT values are heard (voiced and voiceless) these values will be strongly associated with their respective categories. However, for more variable cues (e.g. pitch), associations will be spread across the whole dimension, and hence be weaker (Figure 15.7e, 15.7f). The result is that any individual pitch contributes less to categorization than any individual VOT value. In this case, the lower weighting for pitch emerges from the lack of consistent associations between individual pitch values and words.

Statistical learning drives the perceptual system to become attuned to the regularities of speech production. Is this sufficient to account for listeners' abilities? Remez (2005) argues that statistical learning resembles a roomful of actuaries: counting every possible combination of events, but never extracting higher structure. We agree—the system must do more than estimate event frequencies. But computational implementations have always assumed this. McMurray, Aslin, and Toscano (2009; see also Vallabha et al. 2007), for example, explicitly situate statistical learning in the context of learning categories (estimating the parameters of a Gaussian distribution). This model does not count individual VOTs—it uses them to estimate categories. Even with this, the model still fails without competition, a feature of many connectionist implementations (Rumelhart and Zipser 1986; Guenther and Gjaja 1996; McMurray and Spivey 2000; McMurray, Horst, Toscano, and Samuelson 2009). Competition forces these models to make a *decision* about competing abstractions of the signal in order to learn. Thus, distributional learning alone is insufficient, but combined with other processes (like competition found in interactive activation) it tunes the perceptual system to the regularities of production.

Moreover, distributional learning does not appear pre-built to attend to particular cues, nor does caregivers' behavior reflect any explicit enhancement of the input. Thus, distributional learning may be a general process, and information-level coupling a serendipitous consequence. This idea is supported by the domain generality of such learning. Rosenthal et al. (2001), for example, demonstrate that visual categories, for which no communicational parity is required, can be learned via distributional statistics. Moreover, the use of variability to weight dimensions is seen in many domains where coupling is not pivotal, like rule-learning (Gómez 2002), depth perception (Atkins et al. 2003), and classic learning theory (Bush and Mosteller 1951; Restle 1955; Bourne and Restle 1959). Finally, the associative account of dimensional weighting (Apfelbaum and McMurray 2011) suggests that what appears to be a complex dimensional selection process may be the product of many simple associations between cue values and words, associations one would need for word recognition anyway.

Finally, there are sources of information that may not be clearly articulatory, but could also leave a statistical trace, and hence usefully shape the expectations of the system. For example, individual talkers have characteristic VOTs (Allen et al. 2003). As we examine in our final case study, taking advantage of such regularities may require both online processing and statistical learning.

15.4 Case study 3: Compensation, expectations, and parsing

The final case study asks how speech perception can identify categories when confronted with multiple influences on a cue. We propose that a more general version of parsing (Fowler 1984), based on computing cues relative to expectation, may give rise to sophisticated coupling.

A recurring challenge in speech perception is variability. Variability is essential for emergent informational coupling, particularly over developmental time, yet it is also a challenge. However, variability is not simply noise. Variability typically reflects the overlap of multiple influences; for example, F2 is affected by place, voicing, and talker. How does the listener cope with and harness this variability in real time?

Consider vowels. Figure 15.8a shows the first and second formant frequencies from a corpus of /ɛ/ and /ʌ/ reported by Cole et al. (2010). The two clearest clusters don't correspond to these vowels: the top right encompasses male talkers, and the bottom left, females. Where F1 usually reflects tongue height and F2 backness, to a naïve observer, these cues more readily distinguish classes of talkers.

If the system were constrained to articulatory dimensions, this would not be a problem—this variance would be ignored. However, a less constrained statistical learning system (such as the one we have argued for) would extract these non-articulatory categories (gender) first. Moreover, even if the listener had appropriate categories, in perception the overlapping variance due to gender would mean a substantial number of sounds would be miscategorized.

Parsing is an approach introduced by Ohala (1981a) and developed by Fowler and colleagues (Fowler and Smith 1986; Pardo and Fowler 1997) to address the issues raised by overlapping sources of variance. In Fowler's account, at any given time, the articulator position reflects both preceding and upcoming segments (coarticulation), resulting in ambiguity at that point in time. Parsing determines possible causes of this signal (considering a range or distribution of possible causes). This allows the system to attribute sources of variance in the target segment to its

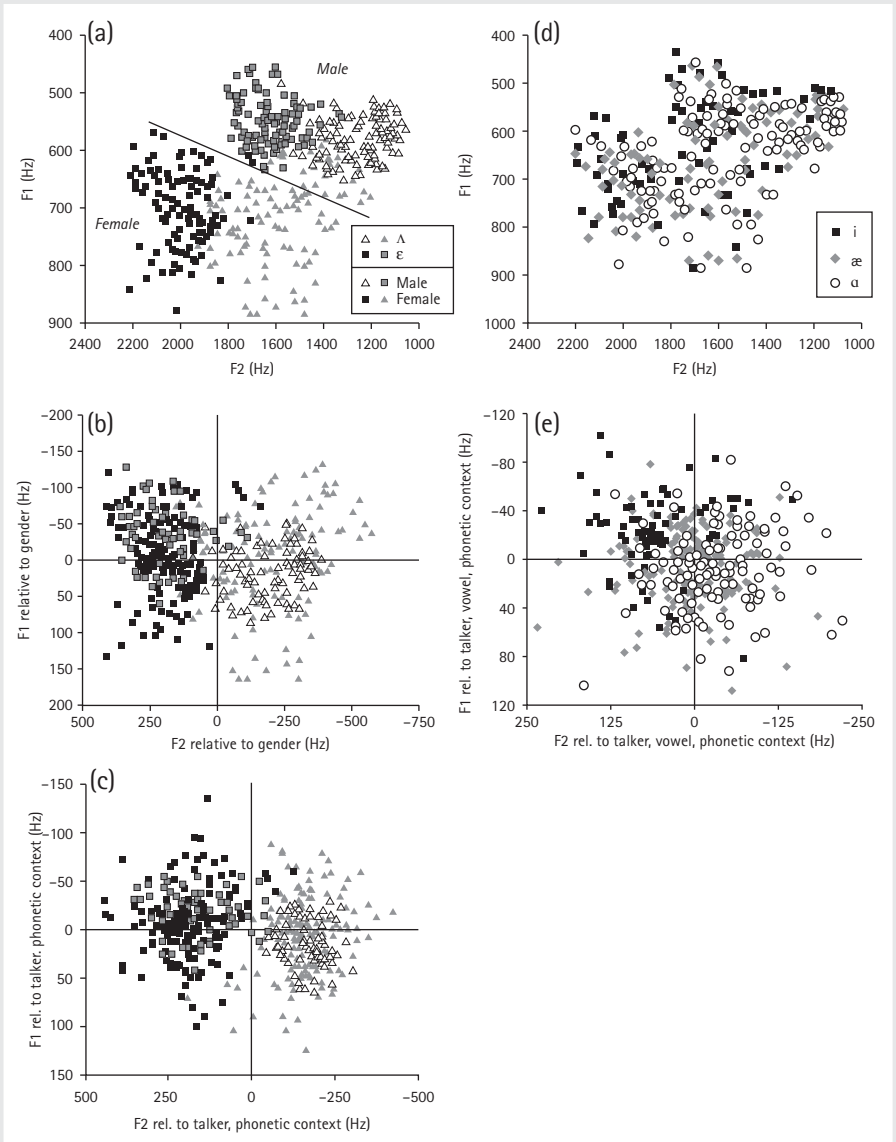

Figure 15.8. F1 and F2 for /Λ/ and /ɛ/, from Cole et al. (2010). (a) Raw values show little clustering by vowel, but large clusters by talker gender. (b) F1 and F2 relative to expectations-derived from gender. 0 Hz represents the expected F1 or F2 for that talker, with positive values meaning higher frequencies than expected. Now, two categories emerge. (c) F1 and F2 once additional variance due to individual talkers, the neighboring consonant and vowel, have been accounted for. (d) Raw F1 and F2 coded by the anticipatory vowel context. (e) F1 and F2 after parsing out talker, the target vowel, the neighboring consonant, and formant frequencies.

neighbors, resulting in a less ambiguous target, while simultaneously using these attributions to predict the upcoming gesture, or to clarify the prior gesture.

For example, in English, talkers typically lower their velum during the vowel preceding a nasal consonant, resulting in a partially nasalized vowel. Fowler and Brown (2000) showed that listeners parse the nasalization as a property of the upcoming segment (since English does not have phonemically nasalized vowels), allowing them to recognize the underlying oral vowel, and use the residual nasality to anticipate the nasal consonant (see also Krakow et al. 1988). Gow (2003; see also Ohala 1981a) argues that parsing does not have to be gestural—grouping acoustic features on the basis of similarity is sufficient, assuming that similar acoustic cues are likely to be caused by the same phonological event.

Thus, processes like parsing can contribute to coupling, allowing variable statistical distributions to be appropriately assigned meaningful causes and simultaneously harnessing this variance to do perceptual work. A series of recent studies built on the idea of parsing to develop a more complete account of compensation for contextual variability: Computing Cues Relative to Expectations (C-CuRE; Cole et al. 2010; McMurray and Jongman 2011; McMurray et al. 2011). In this model, potential causes such as the talker or neighboring phoneme are identified and used to form expectations about likely cue values (e.g. what pitch would be expected for this talker). New cue values are coded relative to these expectations to achieve a generalized version of parsing that is sensitive to a range of articulatory and non-articulatory factors.

Thus far, however, work on parsing has typically only looked at two overlapping causes; and most importantly, it has not assessed the quantity of information in the signal that could be gleaned from parsing. That is, it has not been examined with respect to the distribution of cues seen in the input. This requires analyzing a large corpus of speech in which multiple sources of variation can be parsed, and a way to determine the relative value of parsing over using the raw signal.

Cole et al. (2010; see also McMurray et al. 2011) offer a first step in this direction. They examined a corpus of measurements of /ʌ/ and /ɛ/ produced by ten talkers in various consonantal contexts, flanked by three point vowels. Parsing in the C-CuRE framework was modeled with linear regression. Informally, when working with discrete independent variables (e.g. talker gender), the regression formula computes the mean value of each group (male/female). This is the expected cue value for that group. If we recode individual data points as their difference from the group mean (the residual), we have a new set of relative cue values in which the variance due to that factor has been removed. As a result, a given F1 measurement is recoded as "high for a male" or "low for a female," rather than just low or high.

Figure 15.8b shows the same F1/F2 measurements after parsing the effect of gender; that is, after F1 and F2 are recoded relative to their expected values for that gender. Here, the variance due to gender is no longer seen and two vowel categories emerge—most of the /ʌ/s have higher than average F2s, and most of the /ɛ/s are

lower. Compensating for variance due to the consonant and upcoming vowel makes these clusters more and more apparent, allowing the categories to be distinguished for 95 percent of the tokens (Figure 15.8c).

A similar benefit is observed if we examine the same cue values, but with reference to the upcoming context vowel. Initially, the raw values don't vary systematically as a function of context vowel (Figure 15.8d). Vowels produced before an /i/ are distributed uniformly across the space relative to /ɑ/ and /æ/, and a logistic regression could only predict the correct vowel with 28.6 percent accuracy (chance = 25 percent). However, after parsing out talker, place, voicing, and the target vowel, a significant effect of vowel context can be seen (Figure 15.8e): /i/s shift the target vowel toward lower F1s and higher F2s, /ɑ/s toward higher F1s and lower F2s, and /æ/s toward higher F1s and F2s. After parsing in the C-CuRE framework, the same classifier predicted the upcoming vowel at 39.4 percent.

C-CuRE thus provides a partial answer to how one identifies individual tokens as well as clusters in the input—by progressively accounting for sources of variance in the signal, the underlying structure is made statistically more discriminable (see McMurray and Jongman 2011; McMurray et al. 2011). Most importantly however, this simple approach to compensation can deal with variance due to talker-specific (e.g. indexical) effects as well as variance due to articulatory causes (e.g. V-to-V and C-to-V coarticulation). In fact, talker was perhaps the most important source of variance in this corpus, accounting for 82.4 percent of the variance in F1 and 40.8 percent in F2. Thus, parsing, even when not extracting gestures, can substantially contribute to speech perception. This reinforces information- rather than unit-level coupling.

Thus, as we observed with both statistical learning and interactive activation, apparent coupling between perception and production can emerge also from these compensation mechanisms, although mechanisms like relative cue encoding (parsing in the C-CuRE framework) are capable of achieving more than just coupling. Parsing does not care what you call the units. They could be gestures, phonological features, acoustic groups, or even talkers—the relevant facts are the patterns of covariance between these units and the cues. Lexical or phonological regularities could also provide a source of information. If the vowel can be unambiguously identified as /ʌ/ due to a lack of minimal pairs, the system may be able to parse more quickly and more accurately (see Gow and McMurray 2007).

Similar operations are also seen in domains beyond speech. Listeners can simultaneously cancel out interference from reflected sound sources (echoes) and use them to make inferences about room acoustics (Clifton et al. 2002); and infants can use their knowledge of one object to segregate it from novel objects in a visual scene (Needham and Baillargeon 1998). Neither example makes sense in a coupling framework. Rather, in C-CuRE, parsing is a general perceptual process for coping with variance, which is underscored by our ability to model it with something as simple as linear regression. Auditory contrast effects (e.g. Kluender et al. 2003; Holt

2006; Alexander and Kluender 2008) can also be seen as interpreting a local cue with respect to its average value over some window, though at a lower level of auditory processing. Thus, C-CuRE (or something akin to it) may serve as a general mechanism in multiple domains and levels of processing.

Distributional learning, interactive activation, and relative cue encoding may, in the end, be related. In order to recode cue values relative to expectations for parsing, listeners must have access to the means and variances of the possible categories (e.g. the mean F1 for the talker and his range of variation). These could be attained via distributional learning. Further, broadly described, C-CuRE also requires activating categories of some kind, in order to generate expectations which in turn influence perceptual processing at lower levels. Such interactions are hallmarks of interactive activation (see Fowler and Smith 1986). Interactions between C-CuRE or parsing, statistical learning, and interactive activation have not been considered either computationally or empirically. Yet, given the possibility of such interactions and their computational power to achieve information-level coupling, this is the next step.

15.5 General discussion

Each of the three case studies we have presented draws on work that is largely uncommitted with respect to unit-level coupling. Interactive activation, distributional learning, and C-CuRE or parsing can operate on gestural, acoustic or other classes of inputs, and do not require parity or coupling as a goal. However, by understanding these *processes* of perception and development, we see how information-level coupling emerges. All three processes ultimately show sensitivity to the statistical distributions created by speech production. Yet, like the coupled clocks, even if we assume that speech perception only cares about extracting meaning, and speech production only cares about communicating it, coupling at the information level still emerges.

Interactive activation describes both perception and production as continuously cascading processes, parallel activation and competition. While production and perception systems are independent, our discussion of the consequences of gradiency in speech perception for handling misperception shows how apparent coupling can emerge. Similarly, *distributional learning* assumes that perceptual development is also independent of production. Yet speech production creates regularities in the input that can be internalized by distributional learning. Finally, *parsing in the C-CuRE framework* may pick up where statistics leave off. By attributing portions of the phonetic variability to different causes, overlapping statistical

distributions can be separated and the leftovers used to make inferences about other things, both articulatory and non-articulatory.

And yet all three mechanisms are more general. All three processes characterize many domains of cognition and have consequences that are not directly motivated by a perception-production linkage. While statistical learning may largely learn the distributions created by articulatory factors, it is not limited to a specific type of cue. C-CuRE, though it can be used to identify articulatory causes, can be equally useful in coping with talker-related variation or harnessing lexical structure. It could be a powerful mechanistic bridge between gestural and episodic approaches. In all three cases, when these general processes are applied to speech they yield effects that look like a fundamental coupling between perception and production.

These three mechanisms may be deeply related. Statistical learning is insufficient without the goal of building a category (not simply counting occurrence), and more importantly, without competition—key elements of interactive activation. Interactive activation requires connections mapping the range of input values onto these categories, connections that could be acquired by distributional learning. Parsing requires the means and variances of each cue (the domain of statistical learning) and has a similar information flow between perceptual input and multiple competing categories. Examining such interrelationships, both empirically and computationally, is an important next step (see McMurray, Horst, Toscano, and Samuelson 2009; Mirman et al. 2006 for potential models).

The approach outlined here has focused on how production impacts perception, not the converse. Even within our discussion of speech perception, we've emphasized mechanisms that are based in the mapping between continuous acoustic cues and words. These are clearly not sufficient to describe perception as we've broadly construed it. Complex processes at lower levels of perception are important to speech, processes like streaming, auditory contrast, and short-term adaptation. However, some of these may operate by the same principles: Auditory retuning mechanisms (e.g. Guenther et al. 2004) may operate by statistical learning. Auditory contrast effects in which preceding tones can influence the interpretation of phonetic cues (e.g. Holt 2006) could be handled by relative cue encoding that treats frequency as a function of the distance from the expected frequency. Similarly, there are complex processes at higher levels, such as the haptic McGurk effect (Fowler and Dekle 1991), or effects on perceptual learning (Kraljic, Samuel, and Brennan 2008). These too may ultimately be implemented by statistically tuned interactive mechanisms, perhaps with additional sources of information to account for such coupling, though this may push the limits of our account. Finally, coupling may need to be more explicit in production abilities, where production must anticipate the acoustic results of a motor command, particularly during development. Here, connectionist models like those of Plaut and Kello (1999) and Guenther (1995) offer an excellent platform to examine the emergence of phonology from similar principles like statistical distributions and competition dynamics.

Nonetheless, these case studies illustrate the power of a different way of approaching the problem of linking perception and production. Ultimately, our case studies treat perception as resulting in a distributed, probabilistic representation, with the result that the nature of the units may be less important than their statistical properties. By focusing at this information level, and on the *processes* of perception, production, and development, we can see how production and perception become coupled without making strong theoretical assumptions about the nature and source of the units. We do not explicitly argue against unit-level coupling, but it's unclear whether this framing of the question buys us anything except recurring theoretical debate. Moving toward distributed representations may make these polarizing distinctions less important, emphasizing where broad agreement can be reached: production leaves a complex signature on the acoustics, and perception would do well to capitalize on it. While basic processes like statistical learning, interactive activation, and parsing are clearly domain-general processes, they support the kind of specialization necessary to achieve information-level coupling between perception and production.

CHAPTER 16

INSIGHTS FROM ACQUISITION AND LEARNING

The contributions in this chapter discuss the role of language development in early acquisition, multilingualism, and second-language acquisition, and consider how these inform our understanding of core phonological questions. Together they paint a picture of the critical role of both production and perception in the learning of phonological systems and show how such acquisition studies provide insight into the nature of adult phonological structure.

16.1 How phonological representations develop during first-language acquisition*

Katherine Demuth and Jae Yung Song

16.1.1 Introduction

Little is known about the phonological representations that underlie children's early productions, and why variability in production persists even as the child progresses toward the mastery of language. Much of the early research on phonological acquisition focuses on children's production of segments, showing variability both between and within individuals (e.g. Smith 1973; Ferguson et al. 1992; Vihman 1993, 1996). Some of this research identified children's early motor-control limitations as a means for understanding their variable productions (e.g. MacNeilage 1980; Lindblom 1992). Others have shown that within-speaker variability is influenced by the frequency of lexical and syllable patterns in the ambient language (e.g. Beckman and Edwards 2000b; Levelt et al. 2000; Roark and Demuth 2000; Edwards et al. 2004; Storkel 2004; Zamuner et al. 2004; Munson et al., this volume). Still others have shown that the phonological contexts in which words and morphemes appear can have an enormous effect on whether a morpheme is apparently produced or not (e.g. Panagos et al. 1979; Bennett and Ingle 1984; Echols and Newport 1992; Gerken and McIntosh 1993; Rvachew and Andrews 2002). Variable processes of coda deletion and coda cluster reduction are also subject to contextual variation within certain dialects. This has been investigated for adults speaking American and British English (e.g. Roberts 1997; Foulkes et al. 2005; Docherty et al. 2006), and African-American English (e.g. Stockman and Vaughn-Cooke 1989; Wolfram 1991; Moran 1993; Bailey and Thomas 1998; Rickford 1999; Stockman 2006). Such adult

* We thank our collaborators Stefanie Shattuck-Hufnagel and Lucie Ménard for their contributions. This work was funded in part by NICHD grant #R01 HD057606.

variation is an important issue to keep in mind when examining child phonological and morphological development across dialects.

In this section, we argue that children's phonological representations as evidenced by their productions may be more intact than often assumed. We also suggest that conducting fine-grained acoustic analysis of child (and child-directed) speech holds the potential for better understanding children's developing phonological representations, and the factors that influence variability in production over time. We review below some of the traditional methods used, discussing some of their limitations, and then discuss recent laboratory phonology research examining the development of phonological representations as evidenced through production.

16.1.2 Traditional methods and some limitations

16.1.2.1 *Observational/longitudinal studies*

Many investigations of children's phonological development have been observational case studies, where longitudinal data is collected and developmental trends assessed. Some consist of diary studies (e.g. Deville 1891), whereas others consist of tape-recorded and transcribed child speech, using either orthographic (Brown 1973) or phonetic (IPA) form (Smith 1973). Such studies provide useful albeit impressionistic information about a child's language development, upon which many theoretical claims have been made. In fact, acoustic analysis is critical to fully analyze these data. For example, if the transcription indicates that the child produced no coda consonant on the word *dog*, it is impossible to know if the child's representation was really CV, or if there might have been vowel lengthening, indicating that the child has some knowledge of the "missing" coda consonant. Nonetheless, these types of spontaneous, longitudinal corpora can be extremely useful as pilot data for forming hypotheses about aspects of phonological development, which could be investigated under more controlled, experimental conditions. They are also useful in documenting individual differences in phonological development. However, for any specific research question there may not be enough tokens of the right type from spontaneous speech corpora to fully assess the extent of children's phonological knowledge.

16.1.2.2 *Experimental production studies*

Some of the concerns about sparse data can be addressed in cross-sectional experimental studies using elicited imitation or elicited production methods. This provides the opportunity for exploring children's phonological and morphological knowledge under controlled contexts at a given point in time. For example, Kirk and Demuth (2005) compared 2-year-olds' acquisition of segmentally similar

consonant clusters at the beginnings and ends of words (*ski* vs. *ask*, *ax*). They found that children were better at producing consonant clusters word-finally, especially when these decreased in sonority (i.e. *ask* [æsk] was produced more accurately than *ax* [æks]). However, it is also possible that some of children's early cluster errors could be due to articulatory difficulty. For example, Kirk (2006) examined 2-year-olds' coda productions in monosyllabic and disyllabic non-words. They found better coda production in monosyllabic words, and in the final and stressed syllables of disyllabic words. Similarly, Song et al. (2009) found better production of third person singular *-s* in utterance-final compared with utterance-medial position—for both 1;10-year-olds and 2;3-year-olds (with the older children doing better overall). Taken together, these results suggest that children are more accurate in producing coda consonants in stressed and final syllables, which are longer in duration, thus allowing more time to complete the full articulation. On the other hand, many experimental production studies have not necessarily examined the data from a more fine-grained acoustic perspective.

16.1.2.3 *Experimental perception studies*

There is a wealth of literature examining the development of infants' perceptual abilities (see Holt, this volume; Munson et al., this volume; Maye, this volume). Some of this literature has focused on the development of native vs. non-native speech contrasts in infants under 1 year of age, showing that this can be influenced by segmental frequency effects (e.g. Anderson et al. 2003). It has been found that 19-month-olds have detailed subphonemic phonological representations that encode cues for place, manner, and voicing (White and Morgan 2008). However, mapping novel words onto objects appears to be challenging for 14-month-olds, indicating a heightened processing load that is only overcome around 20 months (Stager and Werker 1997; Swingley and Aslin 2000). There have also been several studies showing cross-linguistic differences in listening preferences for different types of lexical stress (see Nazzi et al. 2006 for review). However, there has been little investigation of infants' preference for other types of phonological units (though see Jusczyk et al. 2002).

In summary, much has been learned over the past twenty years about the course of phonological development. However, the methods used all exhibit certain limitations. The longitudinal studies have typically lacked an accompanying acoustic record and tend to involve small case studies. Likewise, many cross-sectional production studies have typically not exploited information in the acoustic signal when assessing children's phonological competence. They therefore miss potential covert contrasts the child may be making, presenting an incomplete and potentially misleading picture of what children know about phonological structure. They also tend to focus on one age, with little attention to development. Finally, although a few phonological issues have been examined in infant perception studies, these

typically use non-linguistic measures, such as listening times or listening preference. Many of the experimental studies also report only group data, making it difficult to assess individual differences in phonological development. Nonetheless, our understanding of how and when children begin to develop grammatical competence at different levels of phonological structure is quickly evolving through the use of more widely available laboratory phonology methods, promising new and exciting results in the coming years.

16.1.3 Contributions from laboratory phonology

As outlined above, one of the challenges to understanding the development of phonological knowledge is that children sometimes make acoustically measurable distinctions corresponding to constrasts in adult speech but that are not perceived by the adult. This includes making subtle VOT distinctions for target voiced vs. voiceless onset stops, both of which tend to sound voiced to the adult ear (e.g. Macken and Barton 1980; Scobbie et al. 2000), and extrinsic vowel-duration distinctions before apparently missing voiced vs. voiceless codas (Weismer et al. 1981; Stoel-Gammon and Buder 1999). Young 1–2-year-olds have also been found to use spectral and durational cues to distinguish /gr/ from /gl/ in onset clusters (both heard as /gw/ by adults; Kornfeld 1971). Thus, children may acquire adult-like phonological contrasts earlier than often assumed, despite the fact that their early words often deviate from the adult form. Below we review further such evidence and discuss several possible factors that affect young children's production.

16.1.3.1 *The development of syllable and prosodic word structures*

Researchers have noted that children's early word shapes follow a systematic course of development. Drawing on data from English and Dutch (Fikkert 1994), Demuth (1995b) identified four stages in the development of words, suggesting that similar stages of development might be found in the acquisition of all languages. For example, Dutch-speaking children's early words expand from core syllables (CV) (e.g. [fa] for *olifant* 'elephant') to minimal words (bimoraic feet, i.e. CVC, CVV(C), or CVCV in Dutch) (e.g. [faut] 'elephant'), and eventually to larger, more complex phonological words (e.g. [olifant] 'elephant') as they progress in language acquisition.

While exploring four English-speaking children's attempts to produce coda consonants in monosyllabic CVC words such as *dog* [dɔg], Demuth et al. (2006) found that two of the children from 1–1;6 often lengthened the vowel when the coda is omitted, or added heavy aspiration or an epenthetic vowel to codas that were produced (e.g. CV ~ CVC ~ CVC^{h} ~ CVCV). Similar findings have been reported from corpus studies of other 1–2;6-year-olds (Vihman and Velleman 1989;

Goad and Brannen 2003). This raises the question of the nature of children's early syllabic representations, and whether these include coda consonants at all. Goad and Brannen (2003) proposed that heavy aspiration (typically appearing syllable-initially at this stage) occurring on the final consonant provides support that young children have only CV structure, and that apparent codas are actually onsets to an empty-headed syllable.

In contrast, Demuth et al. (2006) proposed that these children have a highly ranked NoCoda constraint, forcing output forms of CV, CVV, or CVCV. Given the high instance of vowel lengthening in the absence of a coda, they also suggested that English-speaking children may have an early awareness of word-minimality effects, where well-formed English lexical items must take the form of a bimoraic foot (Demuth 1995b; see Fikkert 1994 for similar explanations of early epenthesis in Dutch). Thus, children's early use of vowel lengthening and the addition of an epenthetic vowel (e.g. *dog* /dɔg/→ [dɔː] ~ [dɔgɔ]) could both be understood in terms of children's attempts to meet word minimality. Under all these approaches the assumption was that children had early limits on syllabic (and prosodic word) representations, and that this began to change around the age of 2–2;6, as more target-like coda consonants were produced. However, these studies did not conduct acoustic analysis to further investigate these issues.

Some of the first studies to explore these issues acoustically came from an investigation of word productions from 1–2-year-old Japanese-speaking children (Ota 1999). Since Japanese is a mora-timed language, the issue of when these children become sensitive to moraic structure was of interest. Using durational measurements, Ota (1999) found that Japanese learners showed moraic compensation when they omitted the coda consonant, lengthening the vowel, in effect to constitute two moras of structure. Specifically, he showed that the short vowel that preceded a missing coda nasal (CVØ) was significantly longer than a short vowel in an open syllable (CV) for all three children under investigation. Interestingly, such an asymmetry in vowel duration was not found when onset consonants were deleted, suggesting that the deletion of non-moraic segments does not lead to the compensatory lengthening of vowels. The findings suggest that Japanese children have an awareness of moraic representations or moraic weight of codas even when they cannot reliably produce the word-final consonants.

Similarly, Song and Demuth (2008) examined three English-speaking children's compensatory lengthening of vowels in the context of missing codas (e.g. *dog* /dɔg/ → [dɔː]). Languages like English require well-formed content words to contain a bimoraic foot with either a coda consonant (e.g. *tin* [tɪn]), or a tense (long or bimoraic) vowel (e.g. *tea* [ti]) (Hammond 1999). Thus, if lengthening selectively occurs with lax (monomoraic) vowels but not with tense (bimoraic) vowels, this would support the hypothesis that compensatory lengthening serves to preserve bimoraic or minimal word structure. However, if lengthening occurs across the board, this would indicate that increased vowel duration compensates for the omitted

segment. The results showed that 1–2-year-olds lengthened both long and short vowels, suggesting that lengthening was compensating for the missing segment rather than the timing unit, i.e. mora (Stemberger 1992). This suggests that learning some of the language-specific constraints on prosodic word structure may take longer than previously assumed (Demuth 1995). However, it also provides support for the notion that these children have some representation for the missing coda.

16.1.3.2 *Limitations on the articulatory control of onset and coda consonants*

So far, we have provided evidence that children can exhibit adult-like representations of words even when their word production is not yet adult-like. This raises the question as to the nature of the factors that affect children's early productions. It is possible that there is a speech-planning explanation for these findings. That is, children might have a coda in their phonological representation, but not yet having the articulatory gestures needed to execute CVC especially within a multi-word utterance. The findings reported above, where morphemes were more accurately produced utterance-finally compared to utterance-medially, provide some support for this position. Furthermore, vowel epenthesis appears most often following voiced codas, and aspiration noise appears most after voiceless codas (Demuth et al. 2006), suggesting that processes of speakers using acoustic cue enhancement might be involved (Keyser and Stevens 2006). That is, the child may be trying to ensure that cues to the voicing of the final consonant are clearly perceived although the cues might not be quite adult-like.

Weismer et al. (1981) found that children who apparently "omit" word-final stops nevertheless produce a stop allophone in word-medial position (e.g. *do(g)* vs. *doggy*), indicating that /g/ must be part of the lexical representation of *dog*. This raises the possibility that some children's early attempted codas may include coda closure, but lack the acoustic cues expected by an adult. We are currently conducting acoustic and ultrasound analyses to see if there is evidence for an incomplete closure gesture at early points in development. If such covert contrasts were found, it would suggest that the acquisition of coda contrasts is a gradient rather than a discrete process, with "quasi codas" produced en route to full coda articulation (cf. Hewlett and Waters 2004). In addition, although most typically developing English-speaking children reliably produce coda releases by the age of 2, there is still some variability in the acoustic realization of coda stops. We are currently pursuing investigation of these issues in the acoustic record of both children and adults to better determine the development of acoustic cues to phonological contrasts, and the extent of individual variation (Demuth et al. 2009).

Further acoustic evidence of articulatory challenges faced by learners comes from Imbrie (2005), who compared ten children's variable productions of the onset stops /b, d, g, p, t, k/ at 2;6–3;6 using durational, amplitude, spectral, formant,

and harmonic measurements. When these acoustic measures were interpreted in terms of the supraglottal, laryngeal, and respiratory actions that give rise to them, comparison with adult productions of the same words showed that children have acquired appropriate positioning of their primary articulators for producing a stop consonant. However, the children's gestures were still far from achieving the adult pattern even by the age of 3;6. For example, at this age children are still learning to adjust the tongue body during stop production, and the higher compliance of the articulators, smaller articulator size, and high subglottal pressure results in more tokens that have multiple release bursts and bursts that are shorter than those of the adult speakers. Longer VOT times and highly variable f0 suggest that children are still learning to adjust vocal fold stiffness and glottal spreading, as well as intraoral pressure. High variability in amplitude across an utterance suggests they are also still learning to control subglottal pressure. Thus, these children were less consistent than adults in controlling and coordinating certain aspects of their articulatory gestures, articulator stiffness, and respiration, though some aspects of the children's speech did become more adult-like over the course of the year of the study (see McGowan and Nittrouer 1988 and Nittrouer et al. 1989 for similar findings for fricatives).

Using the same methods, Shattuck-Hufnagel et al. (2011) examined children's coda consonant productions, focusing on cues to voicing distinctions. The findings indicate that 2;6–3;6-year-olds exhibit systematic acoustic cues to coda-voicing contrasts (e.g. *dog* vs. *duck*): an observable voice bar was more likely to precede voiced codas, whereas vowel glottalization was more likely to precede voiceless codas. Results from both 1;6–2;6-year-olds and their mothers' child-directed speech show similarities; the voice bar appeared more frequently before voiced compared to voiceless codas (Demuth et al. 2009). For mothers, the duration of the voice bar was also longer for voiced codas, and children showed a trend in this direction. However, only mothers showed a significantly higher use of vowel glottalization before voiceless codas. Thus, although younger children produce some acoustic cues to coda-voicing distinctions, other cues take more time to become adult-like.

These findings raise questions regarding the relationship between early articulatory gestures and phonological representations. Regarding tongue gestures, Gick (2007) examined ultrasound recordings of an 11-month-old child imitating productions of /r, l, w/. In accord with results discussed so far, he found that the child's production employed distinct articulatory traces and acoustic cues for each phoneme, despite the fact that the percept was not completely adult-like. On the other hand, Ménard et al. (2006) found that French-speaking 4-year-olds' CVC syllables were produced using different types of lip gestures than those of adults, and that children's stressed and unstressed syllables were less differentiated than for adults.

Preliminary study of two Canadian French children (aged 1;11 and 2;3) explored these issues in children's monosyllabic (CV, CVC) and disyllabic (CVCCVC)

familiar words (Ménard and Demuth in preparation). The older child produced many word-medial codas, and had a distinct movement of the tongue for final VC as compared with final V. However, the younger child did not produce any codas in the disyllabic words he attempted. Furthermore, his vowels in the resultant CVCV productions were almost twice the duration of other vowels, showing compensatory lengthening. In addition, his tongue moved slightly toward the end of the vowel. This appears to be an articulatory gesture related to the attempted medial consonant, as confirmed by spectral analysis. These studies lay the groundwork for a more comprehensive investigation of young children's articulatory gestures using ultrasound. They also suggest the importance of conducting close acoustic and gestural analysis of apparently coda-less CVC utterances, for evidence of non-adult-like cue patterns and how they change as children master adult-like pronunciations.

16.1.3.3 *Context effects on the production and comprehension of grammatical morphemes*

Some of the variable production and comprehension of both phonological units and grammatical morphemes may be influenced by the prosodic context and/or utterance position in which they occur. Children have long been known to exhibit within-speaker variability in the production of English inflectional morphemes (e.g. Brown 1973). Many researchers suggest that this is due to incomplete semantic or syntactic representations. However, our recent study of third-person singular *-s* found that children were much less likely to produce this morpheme when it is a part of phonologically more complex codas (*hits* vs. *sees*), and in utterance-medial position as compared to utterance-final position (Song et al. 2009). This demonstrates that some of the within-speaker variability in the production of inflectional morphemes may be due to phonotactic complexity and positional effects. Hsieh et al. (1999) raise the possibility that this particular morpheme may be shorter in utterance-medial position. This could mean there is less time to produce it in utterance-medial position, resulting in more omission. Acoustic analysis of our stimuli used for both elicited production and comprehension experiments with 2-year-olds indicates that medial *-s* is indeed shorter than final *-s* (Song et al. 2009). This appears to have a negative effect on both production and comprehension of utterance-medial *-s* (Sundara et al. 2011). Interestingly, the effects of position are evidenced at the gestural level as well, in both older children (5–7-years-olds) and in adults (e.g. Nittrouer et al. 2005).

Gerken (1996) provided elicited production evidence for 2-year-olds showing earlier production of articles that are prosodically licensed as part of a disyllabic trochaic foot ([*hits the*]$_{\text{Ft}}$ [*piggy*]$_{\text{Ft}}$ vs. [*catches*]$_{\text{Ft}}$ *the* [*piggy*]$_{\text{Ft}}$). We examined longitudinal data to determine if 1–2-year-olds' use of articles would exhibit the same prosodic contextual effects in spontaneous speech. The results were confirmed for four of five children (Demuth and McCullough 2009). Interestingly, acoustic analysis of the productions from the fifth child showed a strong connection between

prosodic organization and article production; her articles were produced as separate prosodic words at age 1;10, then became prosodified as part of a bimoraic foot (like the other children) at the age of 2. Little is known about young children's prosodic organization of grammatical morphemes, and how this develops over time. It is also unclear if children might go through a stage of development where they leave a "prosodic trace" for unrealized grammatical function items, such as that found in the omission of unfooted syllables for words like (*Lu)cinda* (Carter and Gerken 2004). Such a finding would provide additional evidence that children have some representation for the syllables and morphemes they omit. This is obviously a rich area for further research, using both longitudinal and cross-sectional methods.

16.1.4 Recent developments and future directions

16.1.4.1 *New methods*

New technological developments such as more accessible audio/video recording equipment and speech analysis software are beginning to address some of these limitations of previous longitudinal corpora. For example, the CHILDES database (MacWhinney 2000) now allows for both Unicode IPA transcription and the linking of audio/video files to the transcription record. This means that new databases, such as the Providence Corpus (English; Demuth et al. 2006) and the Lyon Corpus (French; Demuth and Tremblay 2008), are being donated with the audio files attached, allowing for a close examination of phonological and morphological development over time. This will permit much more extensive study of the acoustics of child and child-directed speech, and how this develops for the mothers and their children aged 1–3 years. The inclusion of the mother's speech in these corpora is particularly important, serving as a baseline for understanding the nature of the input.

Ultrasound methods are only now starting to be used to explore the nature of children's early phonological representations. With a small ultrasound probe placed under the chin, it is possible to collect both acoustic recordings of child speech and video recordings of tongue movements in a non-invasive manner. This can provide some idea of the types of articulatory gestures being made, and the extent to which these may be incomplete. For example, some children exhibit protracted problems with the production of glides, producing only some of the required articulatory gestures (Bernhardt et al. 2005). This method therefore holds the potential for providing a better understanding about the articulatory underpinnings of phonological development, and possibilities for remediation.

16.1.4.2 *Future directions*

To adequately address the nature of language acquisition it is critical to know more about the input children hear. Some suggest that child-directed speech is a

form of "clear speech", with larger vowel space (Kuhl et al. 1997) and less reduction of segments (e.g. *want (h)im*) than those typically found in adult-directed speech (e.g. Bernstein-Ratner 1982, 1987). However, studies of Dutch child-directed speech suggest more reduction of vowels in grammatical morphemes (van de Weijer 1998). It would be therefore extremely helpful to know more about the acoustic/phonological properties of child-directed speech and the possible connections with individuals' phonological development, as well as how both change over time. Computational techniques are currently being developed that could eventually approximate an automatic alignment of phonemes with the acoustic signal, making it possible to examine a large amount of child-directed speech (e.g. Sjölander 2003). This in turn could shed light on the nature of the acoustic input language learners actually hear, providing a better understanding of how and when children develop adult-like phonological representations, both perceptually, and in production.

In summary, we have examined evidence from laboratory phonology research showing that children under the age of 3 have more robust phonological representations of syllable structures and words than evidenced from impressionistic studies of production. Examining the shape of children's early syllable and word productions is crucial to improving our understanding of the emergence of phonological representations. Several new data sources and methods are now making it possible to conduct laboratory phonology studies of phonological development in ways that were not possible before. This has brought with it an increasing number of studies from different languages, enriching our understanding of the acoustics of phonological development in a cross-linguistic context. The next decade promises to be an exciting one, with much more research on phonological development using laboratory phonology techniques. The results should provide a clearer picture of the course of phonological development at various levels of structure, and the implications this holds for later language development more generally.

16.2 Speech processing in bilingual and multilingual listeners

Paola Escudero

16.2.1 Introduction

In this section, I discuss the acquisition of speech processing skills in bilingual and multilingual populations. The focus on processing is first motivated by the fact that

it has received relatively little attention within the domain of phonology. In addition, it seems reasonable to assume that multilinguals' abilities to understand their languages should precede their abilities to produce them, and therefore the study of speech comprehension places us closer to multilinguals' underlying language skills. Smiljanic (this chapter) suggests that the evidence so far shows that improvement in perceptual abilities does not seem to transfer uniformly to the development of production abilities. However, Escudero (2005) reinterprets the same studies and shows that it is likely that perception develops first and needs to be in place before production development can occur.

It is also important to investigate speech processing in speakers of more than one language, since it constitutes a highly complex process even within monolingual populations. This complexity has been emphasized in recent phonological (Escudero 2005; Boersma and Hamann 2009) and psycholinguistic (Cutler 2008) studies which agree on the fact that speech processing involves at least two separate processes, i.e. speech perception and word recognition, and two different representations, i.e. pre-lexical and lexical. Phoneticians commonly study speech perception, which involves a *pre-lexical* mapping of the raw acoustic signal onto the speech sounds (or phonemes) of a language, while psycholinguists commonly study the *lexical* mapping of sounds onto the words of a language, i.e. word recognition. Cutler (2008) gives ample evidence confirming the fact that listeners make use of pre-lexical and lexical representations and that these representations are accessed via separate processes. Although these two processes have mostly been studied separately (e.g. Storkel and Morrisette 2002), Cutler states that phoneticians and psycholinguists have increasingly been interested in investigating the interrelations between the pre-lexical and lexical components of speech processing. Laboratory phonologists should also take such a comprehensive approach when empirically and theoretically accounting for the workings of speech processing.

What are the specific problems facing a second-language learner, a bilingual, or a multilingual, speaker when learning to perceive the sounds of a language and to recognize words containing such sounds? In the remainder of this contribution I will address three topics that shed light on this matter. The emphasis is placed on *sequential* and *simultaneous* bilinguals,[1] who are more commonly referred to as second-language learners and bilinguals respectively, mainly because, to date, there have been very few studies in the domain of speech processing which targeted speakers of more than two languages or multilinguals. In Section 16.2.2, I review the factors affecting speech perception performance in bilinguals and multilinguals; specifically, bilinguals' linguistic background and the influence of the experimental tasks with which they are presented. In Section 16.2.3, I describe the evidence showing that sound perception and word recognition do not seem to go hand in hand

[1] The term "sequential bilingual" refers to a speaker of two languages who acquired his or her second language after the first, either as a child or adult. This term contrasts with "simultaneous bilingual" which refers to a speaker of two languages who acquires them simultaneously from birth.

in first-language acquisition and that the interrelation between these processes is also problematic for L2, bilingual, and multilingual speakers. The final section deals with the influence of orthography on the perception of sounds and the recognition of words in second and third languages. Despite L2 researchers' acknowledgment of the strong influence of written systems on speech processing, laboratory phonologists have only recently considered the systematic study of orthographic influences on bilingual and multilingual speech processing. The contribution ends with a summary and brief discussion of future directions for furthering understanding of this important area.

16.2.2 Factors affecting L2, bilingual, and multilingual performance

Here we consider how the age of acquisition, language proficiency, the language in which testing is conducted, and cross-linguistic influences between the bilingual's and multilingual's languages affect speech perception performance.

16.2.2.1 *Bilinguals' age of acquisition and language proficiency*

An important question related to bilingual and multilingual linguistic knowledge is whether children or adults who speak two languages have one or two linguistic systems, a question also addressed in Smiljanic (this chapter). A noteworthy variable when considering this question is the age at which the bilinguals' languages are acquired. Specifically, it seems that a difference needs to be made between *sequential* bilinguals who acquired their second language after their first, either as children or adults, and *simultaneous* bilinguals who acquired two languages at the same time from birth. Behavioral studies conducted within the domains of speech perception and production show differential performance for these two types of bilinguals. On the one hand, speech perception studies with sequential bilinguals support one of the main hypotheses in Flege's Speech Learning Model (Flege 1995, 2003) which states that bilinguals possess a common phonological space for their two languages. For instance, Caramazza et al. (1973), Williams (1979), and Flege and Eefting (1987) found that advanced adult L2 learners have perceptual category boundaries for VOT (Voice Onset Time) with a value that is intermediate between the values of the monolingual VOT boundary in the two languages. Similarly, Pallier et al. (1997) found that Spanish-Catalan bilinguals, who acquire Catalan in their childhood but are dominant in Spanish, did not perform like Catalan-dominant bilinguals because they could not accurately perceive the contrast /e/-/ɛ/ which is found in Catalan but not in Spanish.

On the other hand, Sundara and Polka (2008) found that simultaneous bilinguals, but not early L2 learners, seemed to accurately separate the production of

their two languages. Specifically, Canadian English (CE)-Canadian French (CF) bilinguals could discriminate the voicing differences between /dV/ tokens produced by CE speakers and those produced by CF speakers. In contrast, monolingual CE, monolingual CF, and early sequential bilinguals of CF with CE as their first language could not discriminate between CE and CF productions of the same syllable above chance. Similarly, Burns et al. (2003) and Burns et al. (2007) found that the discrimination of the English VOT distinction between /b/ and /p/ in simultaneous bilingual English-French 10–12-month-old infants was similar to that of monolingual English infants of the same age. In addition, Sundara and Polka (2008) found that bilingual 10–12-month-old infants performed similarly to age-matched English monolingual infants in the discrimination of English /d/–/ð/.

Thus, speech perception studies show that simultaneous bilinguals' speech processing cannot be distinguished from that of monolinguals in their two languages, at least when they are very young (10-12 months of age) or as adults.[2] This, in turn, suggests that they may have separate systems for handling the processing of their two languages.[3]

As pointed out by Idsardi and Poeppel (this volume), neurophysiological research can provide great insight into bilingual and L2 speech processing, including the question of whether bilinguals possess one or two systems for their languages and whether they perform like monolinguals. Many neurophysiological studies in the domain of bilingual speech perception have been conducted using the mismatch negativity (MMN) component of the event-related brain potential, which can be used to examine how the brain organizes phonological categories (Näätänen et al. 1978). This component is measured pre-attentively, as opposed to the attentive measure of all the behavioral studies reviewed above, through auditory exposure to the target sound distinctions while listeners read a book or watch a silent movie. The results so far suggest that sequential bilinguals, specifically Hungarian immigrants in Finland (Winkler et al. 2003) and simultaneous Swedish-Finnish bilinguals (Peltola et al. 2007) process vowels by means of an intertwined phonological system that handles both of their languages. It seems that the pre-attentive measure of

[2] Sundara, Polka, and Genesee (2006) found that, although adult simultaneous bilinguals and monolinguals performed similarly in the discrimination of the English /d/–/ð/ contrast, 4-year-old bilinguals were poorer at discriminating the same contrast than age-matched monolinguals. In speech production, Sundara, Polka, and Baum (2006) demonstrated that adult bilinguals and monolinguals differed in their production of the stop contrast /d/–/t/. However, the authors also report that monolingual French and English listeners did not distinguish between bilingual and monolingual productions of /d/ and /t/ tokens.

[3] Speech production studies show that bilingual children with different linguistic backgrounds produce language-specific differences in VOT for the consonant /t/ (Watson 1990; Khattab 2000; Johnson and Wilson 2002), but not for /d/. Sundara, Polka, and Baum (2006) showed that this production problem is developmental because they found that adult bilinguals do produce a VOT difference between English and French /d/. McLeod and Stoel-Gammon (2005) also show that adult bilinguals produce VOT values within monolingual ranges.

bilingual speech processing reveals that even simultaneous bilinguals do not process their two languages independently. In the next section, it will be shown that these somehow contradictory results between behavioral and neurophysiological studies can be explained by the specific task presented to the bilingual.

Another important factor in bilingual speech processing is the level of proficiency that bilinguals have in each of their languages. Elman et al. (1977) found that, unlike the majority of their bilinguals who exhibit intermediate perception similar to that found in Caramazza et al. (1973), two "strong" or more proficient bilinguals who were sequential adult L2 learners had VOT perceptual boundaries that matched the monolingual perception of each of their languages. Similarly, Escudero and Boersma (2002) and Escudero (2005, 2009) showed that advanced, but not beginning or intermediate, learners of Spanish and Canadian French had vowel perception similar to monolingual listeners for Spanish /i/–/e/ or Canadian French /æ/–/ɛ/ respectively. These examples suggest that sequential bilinguals can perform like monolinguals in the perception of second-language consonants and vowels. In the next section, we will see that these seemingly contradictory results depend on the type of language setting in which the bilinguals perform the speech perception task.

16.2.2.2 *The task presented to the bilingual*

In bilingual infant perception, monolingual-like performance seems to depend on the sensitivity of the task to infants' perceptual abilities. Using a task similar to the head-turn preference procedure (see Maye, this volume), Bosch and Sebastián-Gallés (2003) showed that language-specific phonetic discrimination of the Catalan /e/–/ɛ/ contrast is delayed in Spanish-Catalan bilingual infants because at 8 months of age only monolingual Catalan infants were able to discriminate this contrast, while both groups of infants could discriminate the same contrast at 10 months. However, Albareda et al. (2011) show using a more sensitive paradigm such as the Anticipatory Eye Movement paradigm (McMurray and Aslin 2004), that 8-month-old Spanish-Catalan bilinguals are able to discriminate the Catalan contrast.

In line with the differences described in the previous section, psycholinguistic studies demonstrate that the amount of L1 or L2 activation during bilingual speech processing depends on factors such as language proficiency and dominance, and, especially, the language used during the task (Marian and Spivey 2003a, b). Grosjean (2001) suggests that the bilingual's languages can be activated selectively or in parallel as a function of the amount of use of the two languages during task instructions or in the stimuli presented. For instance, Kroll and Sunderman (2003) and Marian and Spivey and Spivey (2003b) showed that bilinguals have differential lexical activation depending on which of their two languages is used during the testing session.

Escudero (2005, 2009) found that early and late sequential French-English bilinguals perceived the Canadian French /æ/–/ɛ/ contrast differently depending on

whether they listened to it in a testing session solely conducted in their first language (Canadian English), or in another testing session conducted solely in their second language (Canadian French). Similarly, Escudero and Boersma (2002) and Boersma and Escudero (2008) showed that advanced learners of Spanish with Dutch as their first language performed similarly to Spanish monolinguals when classifying Spanish /i/ and /e/ within a solely Spanish setting.

In the neurophysiological studies reviewed above (Winkler et al. 2003; Peltola et al. 2007), the role of the language setting or context was controlled for in order to investigate the extent to which the bilinguals or L2 learners' phonological systems relate to one another. However, these pre-attentive results seem to contradict behavioral results which have been gathered controlling for the language used in a testing session, because in pre-attentive studies sequential and simultaneous bilinguals do not seem to have differential performance when perceiving sounds in their two languages. The answer to this controversy is given by Lehtola et al. (2007) who found that a group of simultaneous bilinguals similar to those tested by Peltola et al. (2007) appeared to be able to process the sounds of their two languages by means of two phonological systems when an attentive behavioral task was also included within the pre-attentive MMN testing session. It remains to be seen whether attentive and pre-attentive methods should be combined using the MMN technique, as has been done by Schafer et al. (2005) for monolingual children, in order to get closer to knowing whether bilinguals process their languages using the same or different phonological systems and whether their performance is comparable to that of monolinguals of either language.

16.2.2.3 *Cross-linguistic influence and proficiency in multilingual performance*

Very few studies have examined the acquisition of third or fourth (L3, L4, etc.) languages and the majority of the studies conducted so far have concentrated on speech production rather than speech processing. Within L2 learning, it is well known that the learner's L1 prominently influences L2 performance, to the extent that native-like performance can be achieved depending on how the sound systems of the two languages relate to one another. However, much less is known about whether L3 learning is influenced by the L1, L2, or both. Cenoz et al. (2001) review a number of studies on L3 acquisition, mainly in the domains of syntax and semantics, and suggest that the typological or linguistic distance between the learner's three languages determines which of the previous two will influence L3 learning. Many studies have shown that there is a tendency to activate an L2 when learning to produce the sounds of an L3, because an L2 has been learned more recently than the L1, which may lead to its prominent use in L3 learning (Williams and Hammarberg 1998; Dewaele 1998; Wrembel 2007).

Another factor which has been suggested to play a role in L3 acquisition is the level of proficiency in the learners' languages. Gonzalez Ardeo (2001) found an L3 speech production advantage for bilinguals with high proficiency or high exposure to their L2. However, in the domain of speech perception, Gallardo del Puerto (2007) found no effect of language proficiency on the English consonant and vowel perception of Spanish-Basque bilinguals. The author suggests that this finding is due to the fact that both Spanish and Basque have similar vowel and consonant systems and that therefore either of the two languages can be used to aid L3 acquisition of English. Simon et al. (2010) and Escudero et al. (forthcoming) show that Spanish (L1) learners of Dutch (L3) who are highly proficient in English (L2) are more accurate in perceiving Dutch vowels than Spanish learners who have only basic knowledge of English. Again it seems that the degree of similarity between English (L2) and Dutch (L3) does affect the potential benefit to L3 learning.

16.2.3 The relation between speech perception and word recognition

Here, we consider whether the difficulties with speech processing in monolingual, bilingual, and multilingual language acquisition are found at the pre-lexical or at the lexical levels. First, the complexity of the relationship between speech perception and word recognition in monolingual first-language acquisition is addressed and it is demonstrated that children do not fully master the two processes involved in speech processing until later in life. Then, it is shown that simultaneous and sequential bilinguals may have problems with one or both processes but that their performance can be close to that of monolinguals. Finally, the few available studies on L3 sound perception and word recognition are discussed.

16.2.3.1 *Learning of minimal pairs in monolingual and bilingual children*

Werker and colleagues have shown that infants younger than 17 months are unable to learn to associate two pictures to two different words if the words differ in a single consonant, i.e. if they constitute a minimal pair as in, e.g. /bin/ and /din/ (Werker, Cohen, Lloyd, Casasola, and Stager 1998; Werker et al. 2002). Importantly, the same studies show that these infants have no trouble distinguishing the minimal pair in a purely discrimination task where no word learning and recognition is involved. According to the speech-processing approaches within phonology and psycholinguistics reviewed in the introduction, these results may indicate that young infants have trouble accurately mapping minimally different pre-lexical representations to their lexical counterparts.

Unlike the consonant studies described above, infants younger than 17 months seem to be able to learn and recognize words that differ in some vowel contrasts

(Curtin et al. 2009). Specifically, they could learn to associate the minimal pair /dit/ and /dɪt/ to two different novel objects but not /dit/ and /dut/ or /dut/ and /dɪt/. Additionally, it seems that the difficulty in learning minimally different words or *lexical neighbors* continues later in life, as shown by Swingley and Aslin's (2007) lexical competition study with 1;6-year-olds, Storkel's (2009) word-learning study with 1;4–2;6-year olds, and Giezen et al.'s (under review) word-learning study with 5–6-year-olds. Importantly, Escudero and Benders (2010) review a number of recent studies showing that the type of testing paradigm used to examine infant and children's early word recognition seems to heavily influence their performance.

Most of the studies reviewed in Section 16.2.2 suggest that bilingual children have similar speech perception performance to age-matched monolinguals. The question is whether they have similar difficulties as monolinguals when learning words that constitute a minimal pair. Fennell et al. (2007) taught the novel words /bɪ/ and /dɪ/ to monolingual English, English-French, and English-Cantonese infants and found that the monolingual but not the bilingual infants could learn the minimally different words at 17 months. In contrast, Mattock et al. (2010) report that English-French bilinguals and not French or English monolinguals succeeded at learning the words /bos/ and /gos/ at 17 months. To explain these contradictory findings, Mattock et al. suggest that the type of contrast and its phonetic realization across languages may be an important factor in bilingual infants' performance. The consonant contrast is phonemic in the bilinguals' two languages, while the consonants of the words /bɪ/ and /dɪ/ in Fennell et al.'s study show considerable phonetic variation across the languages and the vowel /ɪ/ is only phonemic in English and not in French or Cantonese.

16.2.3.2 *Recognition of minimal pairs in sequential bilinguals and multilinguals*

It has been suggested that sequential bilingual's difficulty in perceiving L2 sounds resides in the fact that they do not have distinct lexical representations for those sounds. Pallier et al. (2001) found that Spanish-Catalan bilinguals activated both lexical entries /dona/ (s/he gives) and /dɔna/ (woman) when presented with either of the Catalan words, which may be due to the fact that Spanish has the vowel /o/ but not /ɔ/. Cutler and Otake (2004) found similar results when Dutch listeners were presented with words containing /æ/ and /ɛ/, probably because Dutch only has /ɛ/. The authors of both studies interpret their findings as evidence of a lexical problem, i.e. the bilinguals have a single lexical representation for both words containing two different sounds. However, the problem may reside in their pre-lexical processing rather than in their lexical representations and processing.

In an attempt to separate the role of word recognition and perception, Curtin et al. (1998) found that English listeners could discriminate the two Thai distinctions voiced vs. voiceless-unaspirated and voiceless-unaspirated vs.

voiceless-aspirated depending on the nature of the task: if the task was perceptual they could discriminate the latter contrast better than the former, but if the task was lexical the opposite was true. The authors claim that this is because English has a voicing distinction, represented lexically by the feature [±voice], which distinguishes voiced and voiceless-unaspirated stops, but not an aspiration distinction. However, as was shown in Section 16.2.2, the task presented to the bilinguals matters: in a follow-up to Curtin et al., Pater (2003) found that when the word recognition and perception tasks were more equal, English listeners' performance was comparable for both Thai contrasts.

Perhaps a more sensitive methodology to investigate the interrelation between speech perception and word recognition in bilinguals is that of Weber and Cutler (2004), who used eye-tracking technology to measure how listeners evaluate incoming auditory input over time. The authors used words whose first syllables minimally differ in the English contrast /æ/–/ɛ/, which was previously shown to be difficult for Dutch-English sequential bilinguals (Cutler and Otake 2004; Cutler et al. 2004). The results showed that Dutch listeners looked longer and more frequently at a picture of, for instance, a *pencil* when the target word was *panda* than at a less confusable distractor (e.g. *beetle* when the target word was *bottle*), which may suggest that they have the same representation for words containing the two English vowels. However, when the Dutch listeners heard, for instance, *pencil* they did not look at the picture of the *panda*. Thus, the authors infer that these bilingual listeners have encoded the /æ/–/ɛ/ contrast lexically because they show an asymmetry in their inaccurate patterns of recognition of words containing these vowels. This means that these bilinguals may have no problem with differentiating the first syllables of words containing this contrast at a lexical level but that they perceive the two vowels as equal, i.e. they have different lexical representations but a single perceptual representation for the two vowels. The question that emerges from these results is how learners can encode lexical contrasts that they cannot auditorily perceive. An answer to this question will be given in the next section.

As for the learning of minimally different words by multilinguals, Simon et al. (2010) and Escudero et al. (forthcoming) show that Spanish learners of Dutch, who have English as their second language, learn words containing Dutch vowel contrasts that do not exist in Spanish less accurately than words containing Dutch vowel contrasts that have a similar counterpart in Spanish. In addition, other studies have shown that these listeners have problems perceiving novel Dutch vowel contrasts (Escudero and Wanrooij 2010). As mentioned in the previous section, a high proficiency in L2 English facilitated the learning of Dutch words containing sound contrasts that exist in Dutch but not in Spanish, and that have similar counterparts in English. That is, the Dutch vowel pairs used by Simon et al. and Escudero et al. had acoustic properties which matched the multilinguals' L2 vowel system rather than their L1 vowel system. In this case, it seems that sequential multilinguals are able to beneficially transfer vowel categories from a previously learned language.

16.2.4 The role of orthography in bilingual and multilingual speech processing

It is well known that sequential adult bilinguals are influenced by the orthography of their first language (L1). Much of foreign speech accent seems to have its source in L1 spelling conventions. Apparently, L1 grapheme-phoneme correspondences are quite entrenched in our linguistic knowledge. Recently, Escudero and Wanrooij (2010) found that orthography influenced sound perception by Spanish adult learners of Dutch with low and high proficiency in Dutch. That is, when orthography was available in the response categories, listeners were more accurate in the perception of the novel Dutch contrast /a/–/ɑ/ than when they heard tokens of these vowels in a task where the target stimuli and the response categories were presented only auditorily. The authors show that this positive difference is due to the fact that the Dutch vowels differ not only in vowel quality but also in length, and that their spelling represents such a duration difference; i.e. *aa* and *a*. Spanish learners of Dutch seem to be able to exploit this vowel duration difference (cf. Escudero et al. 2009). In addition, the authors show that for other contrasts the influence of orthography may be negative, i.e. it leads to lower accuracy.

In a first attempt to empirically examine the role of orthography in bilingual word recognition, Escudero et al. (2008) used the same eye-tracking paradigm to test Weber and Cutler (2004) and Cutler et al.'s (2006) hypothesis that Dutch-English bilinguals acquired differential lexical representations through orthography. Escudero et al. taught two groups of native Dutch speakers, who had a high proficiency in English, twenty English non-words, which followed the same pattern as the words used in Weber and Cutler, i.e. the first syllables contained the English /æ/-/ɛ/ contrast. One of the groups learned the words only by listening to their auditory forms and looking at their pictures, while the other was also presented with their orthography. Only the group who learned the words with their orthography looked at the picture of a word containing /ɛ/ and not at the picture of the word containing /æ/, which suggests that the availability of spelled forms results in the establishment of lexical contrasts that can be used in auditory word recognition. In addition, the results suggest that learners may not be able to encode a lexical contrast for auditorily confusable L2 words if they are learned only on the basis of their auditory forms.

The studies mentioned above show that adult sequential bilinguals seem to be able to transfer their L1 orthographic representations when learning a novel contrast with orthographic representations that match those of their first language. This suggests a tight link between pre-lexical and lexical auditory representations and orthographic representations. This type of auditory and visual connection has been previously demonstrated in the domain of visual word recognition (e.g. Van Orden 1987; Ota et al. 2009).

As for multilingual listeners, it would be interesting to investigate how the spelling forms of the learners' L1 and L2 languages influence the learning of words in a third language. Escudero and Simon (in preparation) set out to examine whether the orthographic effects in speech perception found by Escudero and Wanrooij extend to word learning by varying the availability of orthographic information in the same word-learning task as that used by Escudero et al. (forthcoming), reviewed in the previous section. Preliminary results show that Spanish learners of Dutch who have English as their second language are both positively and negatively influenced by orthographic information in similar ways as was found for speech perception, i.e. orthographic information leads to higher accuracy for some contrasts while it leads to lower accuracy for others.

16.2.5 Summary and final remarks: What do we know so far and what is still to come?

This contribution has illustrated the existing empirical evidence for a number of aspects of bilingual and multilingual pre-lexical and lexical speech processing. It was shown that simultaneous bilinguals, but not early or late sequential bilinguals, can perform like monolingual listeners when perceiving vowels and consonants. This was the case when they were tested within behavioral speech perception studies but not when using neurophysiological pre-attentive methods, such as electroencephalography. Further, early and late sequential bilinguals can perform like monolinguals if the study is conducted solely in the language in which they are being tested, which avoids the activation of their other language and promotes monolingual-like performance. Additionally, neurophysiological studies which include an attentive task yield monolingual performance in simultaneous bilinguals. Studies conducted with speakers of more than two languages show that the typological closeness of the multilinguals' languages determines cross-linguistic interactions between them during L3 acquisition.

The learning of minimally different words is a complex matter because it involves the mastery of both pre-lexical and lexical processing. In that respect, it seems that children master the learning and processing of minimal lexical pairs after their second year of life. Importantly, monolingual and bilingual early recognition of minimal pairs is influenced by the type of sound contrast involved, i.e. either vowels or consonants, and the sensitivity of the testing paradigm to reveal children's perceptual abilities. As for adult sequential bilinguals and multilinguals, it seems that they could have problems with perceiving sound contrasts that only exist in their L2 or they could have problems encoding lexical differences between words containing those contrasts. More studies comparing different contrasts and using different methodologies for examining the performance of the same type of bilinguals or multilinguals would shed more light on this issue. In addition, individual

differences seem to be the norm rather than the exception in L2, bilingual, and multilingual populations; these ought to be brought to light and explained.

Since bilingual and multilingual speech processing can be positively or negatively influenced by orthography, future research should further examine orthographic influence, as it applies to both pre-lexical and lexical processing, and whether it needs to be modeled as part of the knowledge underlying speech processing (cf. Simon and Herreweghe 2010). Another important question for additional research is whether sources of non-orthographic visual information can result in the same speech-processing effects, positive or negative, as shown with orthography.

16.3 Second-language speech learning

Rajka Smiljanic

16.3.1 Introduction

During the process of mastering their first language, infants become uniquely attuned to the distributional patterns of the sounds in their ambient language and less attentive and less sensitive to the phonetic dimensions of sound contrasts not found in the language input (Werker and Tees 1984a). As a consequence, adult second-language learners who already have a system of phonological contrasts in place as part of their linguistic knowledge encounter difficulties in acquiring and processing a non-native language. Second-language acquisition thus represents a quantitatively and qualitatively different process from first-language acquisition, involving often effortful retuning and realigning of the existing linguistic system to the sound structure of the non-native language. A broad goal of research on second-language learning is to understand the processes by which a language learner comes to perceive and produce speech sounds in a non-native language. Furthermore, this research seeks to understand how these processes change over time and what role exposure to a second language plays in the learning process. Second-language learning research, in general, draws from and informs cognitive, biological, developmental, social, linguistic, and educational perspectives.

The section begins by considering significant findings over the past few decades in the second-language speech perception and production domains. We focus on some important insights in the segmental domain and some newer results from investigations of connected speech and suprasegmental phenomena. Theoretical frameworks that have been proposed to account for a wide range of empirical findings are also discussed. As will become evident, substantial progress has been

made in advancing our knowledge in all of these domains. However, these advances also raise a number of new questions. Throughout the review, we highlight some of these open questions. We end by noting some additional active and future areas of second-language speech-learning research.

16.3.2 Second-language speech perception

Early cross-language research amply demonstrated the profound effect that linguistic experience has on second-language (L2) learning. In the perceptual domain, which had been a focus of a lot of early cross-language speech studies, it has been repeatedly shown that adult listeners have difficulty discriminating phonetic contrasts that are not used distinctively or are phonetically realized differently in their first language (L1) (see Strange 1995 for review). Such perceptual advantage for native over non-native speech sound contrasts was taken to demonstrate that the adult speech perception mechanism is geared to process the native language in the most efficient way while at the same time contributing to adult learners' difficulty in acquiring L2.

Although the basic premise of an L2 processing disadvantage remains valid, other work has since uncovered considerable variation in perceptual difficulties with non-native sound contrasts, showing that discrimination accuracy can vary from chance to native-like levels (Polka 1991, 1992; Best 1993, 1994; Best et al. 1988, 2001, 2003). For instance, Farsi velar and uvular stops (Polka 1992) and Zulu voice and click place contrasts (Best et al. 1988) are easily distinguished by native English speakers, although these contrasts are not present in their L1 and these listeners had no prior experience with them. In contrast, English listeners' discrimination of Hindi retroflex and dental stops and Nthlakampx velar and uvular ejectives is near chance (Werker et al. 1981; Werker and Tees 1984a). Some non-native vowel contrasts are difficult for L2 learners to distinguish even when they are similar to L1 contrasts in terms of their phonological features (e.g. Gottfried 1984). In contrast, Polka (1995) demonstrated native-like ease in discrimination of the German tense /u/–/y/ vowel contrast by English listeners without any previous experience with that contrast. Even though the listeners did not have two distinct L1 categories to map the non-native contrasts onto, they were sensitive to some phonetic aspects of the distinction between the members of the contrastive pair (they rated them as "good" and "poor" exemplars of a single native category). Findings such as those reported in Polka (1995) clearly demonstrated that non-native listeners' discrimination abilities are not constrained exclusively by phonological distinctiveness in their L1. Non-native speech perception is affected by fine-grained phonetic similarities (that do not reflect phonological contrasts in L1) and dissimilarities between the two sound systems in contact. Similar sensitivity to fine-grained phonetic similarities and differences between L1 and L2 categories has been demonstrated for non-native

consonant contrasts (Best et al. 1988), phonotactics (Flege and Wang 1989; Dupoux et al. 1999; Halle et al. 2003), coarticulatory patterns (Beddor et al. 2002; Bohn and Steinlen 2003; Levy and Strange 2008), and segmental context (Sheldon and Strange 1982; Lively et al. 1993; Schmidt 1996; Strange et al. 2001).

The notion that non-native listeners are sensitive not only to phonetic details that signal phonological contrastiveness, but also to fine-grained phonetic variation within categories, led to the exploration of underlying mechanisms in L2 phonetic learning. Research with native listeners has shown their sensitivity to contextual variation and within-category structure (e.g. Volaitis and Miller 1992; Allen and Miller 2001). Furthermore, learning L1 affects the weighting of perceptual cues such that perceptual sensitivity along an acoustic dimension is reduced near the distributional peaks of L1 category prototypes (Iverson et al. 2003; Kuhl et al. 1991). Acquiring one's L1 can also lead to greater perceptual sensitivity for some dimensions than for others (e.g. Francis and Nusbaum 2002; see also Holt, this volume). Such learned L1-appropriate weighting of acoustic cues is one source of perceptual difficulties in L2 learning. For instance, Japanese listeners' problems in differentiating between the English /r/–/l/ contrast relate to their lack of sensitivity to changes in the third formant frequency (F3), a primary cue for the native English listener, and their focusing instead on duration and changes in F2 (Miyawaki et al. 1975; Iverson et al. 2003). This leads to miscategorization of L2 sounds and a difficulty in production and perception of the novel contrast.

Other possible underlying sources of difficulties for L2 learners have been identified through examining the discrimination of English tense vs. lax vowels. While native English listeners seem to rely predominantly on spectral differences and only partially on the durational differences between members of the tense/lax pairs (Hillenbrand et al. 1995, 2000), non-native listeners of various language backgrounds, ranging from Spanish, Portuguese, Catalan, German, and Russian to Mandarin and Japanese, seem to weight duration more heavily (Flege et al. 1997; Rauber et al. 2005; Cebrian 2006; Kondarova and Francis 2008, forthcoming; Escudero et al. 2009). Importantly, some of these languages do not use phonemic vowel length in their L1 to differentiate vowel contrasts, suggesting that exposure to English, allophonic use of duration in L1, and/or some universal preference for duration (cf. Bohn's desensitization hypothesis, 1995) guide this perceptual bias. Combined, these results show that L2 listeners employ different cue-weighting patterns in L2 perception, compared with native listeners and sometimes with the patterns observed in their own L1. An important issue is how the experience-based and universal preferences interact in shaping different L1 and L2 cue-weighting strategies and accuracy in the processing of L2 contrasts.

The effect of linguistic experience and variability in levels of discriminability of non-native contrasts extend to phonotactics and prosody as well (Dupoux et al. 1997, 1999, 2001; Hallé et al. 2004; Burnham and Mattock 2007; Aoyama and Guion 2007; Francis et al. 2008). Looking at sound sequences, Dupoux and colleagues

(1999, 2001) showed that in both identification and discrimination tasks, Japanese listeners hear, for instance [ebzo], with an illegal CC cluster, as [ebuzo], a phonotactically permissible CVC sequence in their L1. The frequency of such a "repair" epenthesis may differ for other L1s and for different sound sequences (Davidson 2007b). Illustrating a difficulty of non-native speech perception at the level of non-native stress, Dupoux et al. (1997, 2001) showed that French listeners performed more poorly when perceiving contrastive stress compared with Spanish listeners, presumably reflecting the property of French fixed stress rather than the variable stress used contrastively in Spanish. The authors also found that the adult learners' difficulties varied depending on the task (ABX vs. AX discrimination task) and level of speech processing. Finally, with regard to the perception of tonal contrasts by speakers of a tonal vs. non-tonal languages, Halle et al. (2004) showed that French listeners' judgments were largely psychophysically based. In contrast, Taiwanese Mandarin listeners' judgments were based on a set of contrastive categories and were more categorical. The observed variation across types of prosodic contrasts and listener background languages further demonstrates that the perceived similarity between phonetic properties of the native and non-native sound structure plays a crucial role in adult learners' L2 perception patterns. More work is needed focusing on perception of phonotactic and prosodic phenomena and how they interact with other levels of linguistic processing.

16.3.3 Second-language speech production

Besides learning and tuning their responses to the relevant acoustic cues for L2 phoneme discrimination and identification, L2 learners need to learn how to produce new contrastive L2 sounds, their specific phonetic targets, coarticulatory and co-occurrence patterns in syllables and in words, and novel prosodic patterns, including stress, intonation, and rhythm. All these aspects of L2 pronunciation present L2 learners with significant challenges, resulting in pervasive accented speech patterns. In an early exploration of consonant production by Spanish learners of English, Flege (1991) found that those bilingual speakers who learned English as adults produced voice onset time (VOT) of English /t/ with intermediate values, i.e. between those found in monolingual Spanish and monolingual English speakers. The results were taken to indicate that these late learners did not succeed in establishing a new L2 category but rather used a different phonetic implementation rule for a single (i.e. merged English and Spanish /t/) phonetic category. Similar production difficulties were found for a variety of non-native consonant and vowel contrasts and L1–L2 pairings (e.g. Flege and Skelton 1992, 1995; Bohn and Flege 1992, 1997; Ingram and Park 1997; Flege et al. 1999a; Aoyama et al. 2004; Tsukada et al. 2005).

Analogous to the perceptual difficulty with novel sound sequences, production of novel sound sequences presents another level of difficulty for adult L2 learners. For example, Hansen (2004) showed that Vietnamese speakers had trouble producing /s/, /f/, /v/, /l/, and /ʃ/ phonemes in coda positions of English words, presumably due to their distributional restrictions in Vietnamese (see also Davidson et al. 2004; Davidson 2006b for English speakers' difficulty in producing novel consonant sequences in onsets not found in English). Interestingly, these studies revealed asymmetries in the degree of difficulty for these sequences, i.e. not all consonants in coda positions or consonant clusters were found equally difficult, despite the fact that all of them were absent from L1. These results suggest that language-specific generalizations derived over the classes of phonemes and phoneme sequences are crucial in accounting for the observed production patterns (see Davidson 2006b for discussion). Finally, difficulty in acquiring L2 rhythm and intonation, very salient aspects of foreign accent, has also been investigated, although this remains a largely understudied area of second-language learning (e.g. White and Mattys 2007; Jilka 2007). These studies demonstrated the need for better understanding of speech production beyond the level of individual segments, as potential sources of difficulty and foreign accent for L2 learners.

An important question that all second-language learning production and perception studies address concerns the nature of the phonological system(s) in L2 learners. Results from behavioral perception and production studies indicate that the two sound systems, L1 and L2, are interrelated. Most of the studies reviewed so far clearly show the effect of L1 on production and perception patterns in L2. An interesting and less explored aspect of this interaction is the effect that L2 exhibits on L1 (MacKay et al. 2001; Guion 2003; Cebrian 2006). For instance, Flege et al. (1987), Flege (1987), and Mack (1990) found that French learners of English produced their native language stops with longer VOT than is characteristic of French, although not with as long a VOT as that produced by native speakers of English. The effect of L2 on L1 appears to manifest itself in two ways: the L1 segment is "modified" in such a way as to make it more dissimilar from the new L2 category (Flege and Eefting 1987), or the L1 segment becomes more similar to the new L2 sound (Flege 1987). Some evidence suggests that the degree of interrelatedness of the two systems depends on the age of acquisition with simultaneous bilinguals more likely to develop two independent monolingual-like sound systems compared with early and late bilinguals (Guion 2003; Kang and Guion 2006; Escudero, this chapter). Note, though, that not all studies found that L1 and L2 systems are independent in early or simultaneous bilinguals (Sundara, Polka, and Baum 2006).

Another longstanding issue in L2 speech learning as well as in speech science and experimental phonetics concerns the nature and the relationship between the perception and production systems (see McMurray and Farris-Trimble, this volume).

A common assumption in both first- and second-language acquisition is that the development of adult (or native)-like phonetic perception precedes production abilities. An interesting counterexample to this assumption, as discussed by Goto (1971), Sheldon and Strange (1982), and Yamada et al. (1994), is that some Japanese learners were successful in producing distinct English /r/ and /l/ categories despite the fact that they could not reliably identify native tokens, i.e. their production abilities exceeded their perception abilities. Bradlow et al. (1997) tested the production-perception link by exploring whether success in perceptual training with a wide range of naturally produced stimuli varying across talkers and syllable positions, i.e. high variability training, led to an improvement in speech production by adult L2 learners. As expected, they found that training in /r/–/l/ identification resulted in perceptual learning of the novel contrast by Japanese learners. Importantly, they found that the knowledge of the new contrasts gained through perceptual training improved the learners' productions of the same contrasts as judged by the native English listeners. Finally, the results showed a high degree of individual variation in the level of learning in the two domains; i.e. there was not a uniform amount of improvement that transferred from learning through perceptual training to the production patterns by all learners. The finding that improvement in production occurred through perceptual exposure only, rather than through explicit instruction, was interpreted to support a unified mental representation for production and perception mechanisms consistent with the motor theory of speech perception (Liberman et al. 1967; Liberman and Mattingly 1985, 1989) and direct-realist approach (Fowler 1986; Best 1995). While these theories, as well as second-language learning models (Best et al. 1988; Flege 1987, 1992, 1995; Best 1994, 1995), provide frameworks for considering the transfer of perceptual learning to production, they still need to be refined to account for the lack of correlation between the degrees of learning in the two domains.

Finally, exploring the role of the environment and learner-related variables in determining how successful second-language learners are in achieving native-like levels in L2 processing has shed important light on non-native production and perception patterns. As described above, non-native speakers can have extreme difficulties with producing and perceiving certain non-native segmental and suprasegmental contrasts. Foreign accent can persist even for proficient speakers and even for early L2 learners (e.g. Flege and Hillenbrand 1987; Flege et al. 2006). Some adult learners, however, manage to achieve native-like levels in proficiency and pronunciation (Bongaerts 1999; Bongaerts et al. 2000; Birdsong 1992, 2007). These seemingly contradictory results underscore the importance of better understanding the role of variables, such as the age of acquisition, length of residence in L2-speaking country, relative amount of L1 and L2 use, quantity and quality of input from native L2-speakers, gender, motivation, social stigma associated with speaking with an accent, musical training etc. (e.g. Flege et al. 1995, 2006, 1999b; Bialystock and Hakuta 1999; Flege 1999; Piske et al. 2001; Cebrian 2006; MacKay et al. 2006;

Gottfried 2007; Mayr and Escudero 2010; see also Escudero, this chapter; and also numerous studies cited above).

16.3.4 Theoretical models

Several theoretical frameworks have been proposed to account for the underlying mechanisms that shape learners' difficulties with novel sound contrasts. One common assumption shared by these models is that an acquired native sound system and the native speech experience serve as organizing principles that systematically relate to the adult learners' non-native perception and production processes. One such model, the Native Language Magnet model (NLM), focuses on characterizing developmental changes in auditory perception from a universal to a language-specific perception during the first year of life (Kuhl et al. 1991, 1992; Iverson and Kuhl 1996; Kuhl et al. 1992; Kuhl and Iverson 1995). These changes reflect a reorganization or "attunement" of infants' phonetic perception to the contrasts that are linguistically functional in the ambient language. The ambient language input "warps" the underlying auditory-phonetic "space" in which phonological categories reflect the distributional properties of the native language system. One of the defining tenets of the model is that a prototype category acts as a magnet in that it "pulls" acoustically similar sounds towards it, simultaneously decreasing discriminability of tokens close to the prototype and increasing sensitivity to across-category differences. The listeners' ability to differentiate the phonetic variation near the prototype is in that way diminished compared to discrimination around non-prototypes. Applied to second-language learning, the "pull" of the native prototypes on similar non-native sounds results in a diminished discrimination of non-native categories. The inability to "carve up" the acoustic space along the dimensions relevant for L2 sound contrasts, along with the emphasis on acoustic cues transferred from L1, can lead to the formation of "wrong" category representations and longer processing times in second-language processing (Iverson et al. 2003).

The Speech Learning Model (SLM), which was developed specifically to account for second-language acquisition phenomena, proposes that perceptual similarity between native and non-native sound categories affects the degree to which L2 learners will be successful in producing and perceiving L2 sounds (Flege 1987, 1995). According to the model, "equivalence classification" determines the degree of similarity between L1 and L2 sounds, with similar L2 sounds being approximated more quickly at the beginning of the learning process due to their assimilation to L1 categories. However, more successful formation of new L2 categories and more accurate production and perception will arise with L2 sound categories less similar to the existing L1 categories, presumably due to less "interference" from the L1 categories. An important assumption of SLM is that the language acquisition processes remain intact over the lifespan, allowing L2 learners to apply the same

processes to L2 acquisition rather than losing these abilities at some critical point during development (Lenneberg 1967).

Another model developed to account for non-native speech perception, the Perceptual Assimilation Model (PAM), provides an assessment of L2 learners' perceptual difficulties in discriminating non-native sound contrasts within the existing native phonological system of contrasts (Best 1994, 1995; Best et al. 2001; Best and Tyler 2007). Patterns of assimilation of L2 contrasts to L1 categories present various scenarios by which non-native sound discrimination difficulties can be predicted. The direction and the degree of assimilation are determined by phonetic similarities between L1 and L2 sounds. For instance, two non-native sounds can be assimilated to two different L1 categories (two-category assimilation) or to a single L1 category (single-category assimilation), depending on the degree of the perceived similarity with L1 categories. Discrimination of non-native contrasts is expected to be very good in the former and poor in the latter case. Varied discrimination is expected for assimilation of two non-native sounds to a single L1 category in cases where one L2 sound is a "better" exemplar of the native category compared with the other member of the contrastive pair (category goodness difference). Finally, non-native sounds can be heard as speech but not be assimilable to any L1 categories (uncategorizable) or they can even be heard as non-speech sounds which fall outside the native phonetic space (non-assimilable). Discrimination in the last two cases can also vary widely depending on the saliency of the acoustic differences between the target L2 sound categories.

Despite many shared assumptions, these models differ in how they conceive of some aspects of the nature and mechanisms that underlie L2 speech-learning difficulties. For instance, SLM assesses adult learners' difficulties in acquiring single novel L2 sounds with an emphasis on production and on relatively experienced learners. On the other hand, PAM addresses perceptual difficulties of non-native contrasts, rather than single sound categories, through their assimilation to native categories. Furthermore, PAM was originally conceived to provide a framework for cross-language speech perception, i.e. perception of naïve non-native listeners, and only recently some details have been provided that account for L2 perception phenomena as well (Best and Tyler 2007).[4] Unlike the other two models, PAM makes specific claims about the nature of the underlying representations based on articulatory phonology (Browman and Goldstein 1986, 1989). On this view non-native listeners assimilate non-native sounds to native sounds based on detection of similarities in the articulatory gestures. All of the models, regardless of their different foci, have contributed greatly to our understanding of the processes shaping adult second-language learners' production and perception. Importantly, they have

[4] Recently, Second-Language Linguistic Perception Model (L2LP) was developed to address the entire developmental L2 perception process (Escudero 2005). This model also allows for the assessment of individual variation in L2 learning tasks.

generated specific questions and predictions that resulted in substantial research as discussed above. A variety of questions remain for the models and the research to address: Can these models be extended to account for perception/production of sound sequences and prosodic phenomena? How do the models of L2 sound processing link up to other levels of linguistic processing (e.g. lexical recognition)? How does acquiring an L2 lexicon exert influence over sound perception and category formation? How can individual variation in L2 attainment be incorporated into the models?

16.3.5 Final remarks

This review brings together some of the most important and most recent second-language-learning empirical findings and theoretical developments. Additional promising research areas include studies exploring the plasticity of speech perception and production mechanisms through various training paradigms (e.g. Strange and Dittmann 1984; Lively et al. 1993; Pisoni and Lively 1995; Yamada 1995; Bradlow et al. 1997, 1999; McCandliss et al. 2002; McClelland et al. 2002; Iverson et al. 2005; Francis et al. 2008; Kondaurova and Francis forthcoming), the notion of the cross-language phonetic similarity of sounds evoked by the theoretical models (e.g. Strange 2007; Park and de Jong 2008; Bradlow et al. 2007, 2010), effect of noise on non-native speech perception (e.g. Mayo et al. 1997; Van Wijngaarden et al. 2002; Van Engen and Bradlow 2007), the interaction of low-level acoustic information with information at higher-level structural and contextual information (e.g. Cutler et al. 2004; Bradlow and Alexander 2007), the effect of foreign-accented speech on intelligibility and speech processing (e.g. Munro and Derwing 1995; Rogers et al. 2004, 2006; Smiljanic and Bradlow 2007; Bradlow and Bent 2008), and of multilingualism on speech processing (see Escudero, this chapter), to name just a few. Another area contributing significantly to our understanding of L2 processing and facilitating raising of new questions concerns physiological and brain-imaging studies (see Idsardi and Poeppel, this volume). Future research should be extended to second-language processing in more naturalistic communication settings; i.e. exploring the role of social interactions and audiovisual information in second-language acquisition, looking at spontaneously produced speech, exploring how perception and processing are affected by more realistic goals and demands of everyday communication situations. This will allow us to find out whether insights discussed here extend to situations outside of common laboratory conditions. Finally, it is important to think about how our research findings can be used to inform language pedagogy, second-language teaching, and language policy, which have practical implications for everyday functioning of the growing population of non-native speakers.

This page intentionally left blank

PART V

METHODOLOGIES AND RESOURCES

The goal of this part is to highlight the diversity of methods, experimental paradigms, and resources that are the essence of the laboratory phonology perspective. The contributions provide a discussion of particular methodologies and resources that have proven useful, with attention to the types of theoretical issues these approaches have been (and can be) applied to.

This page intentionally left blank

CHAPTER 17

CORPORA, DATABASES, AND INTERNET RESOURCES

CORPUS PHONOLOGY WITH SPEECH RESOURCES

JENNIFER COLE AND
MARK HASEGAWA-JOHNSON

USING THE INTERNET FOR COLLECTING PHONOLOGICAL DATA

DAN LOEHR AND
LINDA VAN GUILDER

SPEECH MANIPULATION, SYNTHESIS, AND AUTOMATIC RECOGNITION IN LABORATORY PHONOLOGY

HENNING REETZ

PHONOTACTIC PATTERNS IN LEXICAL CORPORA

STEFAN A. FRISCH

Contributions to this chapter introduce a wide range of approaches to using large bodies of data for linguistic research. Cole and Hasegawa-Johnson emphasize the benefits of using large corpora to investigate phonological questions and discuss corpus creation and selection and tools for analysis. Loehr and Van Guilder discuss using the Internet for probing or generating speech data. Reetz discusses synthetic speech as an alternative to naturalistic data and the role of speech recognition in speech modeling. Frisch presents detailed discussion of the application of corpus analysis to questions about the structure of lexical knowledge.

17.1 Corpus phonology with speech resources

Jennifer Cole and Mark Hasegawa-Johnson

17.1.1 Introduction

This section introduces the methods of corpus phonology using speech databases, for the investigation of phonological variation, for understanding the phonetic underpinnings of phonological phenomena, and for research on the category structures of spoken language. We address the practical challenges in identifying an existing corpus appropriate for phonological research, such as the corpus resources mentioned in Loehr and Van Guilder (this chapter); Post and Nolan (this volume). The challenges of annotation that arise with the creation of a new corpus are also highlighted, with further discussion of annotation in relation to community standards taken up in Loehr and Van Guilder (this chapter). Finally, corpus research involves processing large speech databases, and this section ends with a discussion of the computational and statistical tools that are widely used in corpus studies, with related applications in the development of speech technologies.

Corpus analysis for phonological research involves investigation of the phonetic, phonological, and lexical properties of speech for the purpose of understanding the patterns of variation in the phonetic expression of words, and the distributional patterns of sound elements in relation to the linguistic context. In some respects corpus methods complement laboratory-based experimental methods in phonology, and for some fields of inquiry corpus materials are essential. The central role of speech databases and lexical corpora for the study of frequency and similarity in phonology corpus data is clearly demonstrated in Frisch (this chapter).

17.1.2 Phonetic considerations in phonological research

Phonology is concerned with characterizing the sound patterns of language, typically presented in terms of a *system* of contrastive sound elements (e.g. syllables, segments, features) and the *distribution* of those sounds in the make-up of phonological words and phrases. This focus on the sound system and the characteristic sound patterns of words is what distinguishes the study of phonology from the study of phonetics, as these two fields are traditionally construed. Yet the phonologist's perspective on sound systems is typically rooted in knowledge about the phonetic properties of the sound elements that make up a language, and reflects direct observation of the phonetic form of spoken words and phrases.

Considering the phonetic substance of phonological forms presents a challenge and an opportunity. The challenge arises from the inherent variability in the phonetic realization of a word, which can make it difficult to identify a unique description of its core phonetic properties. For instance, the English words *rapid* and *rabid* in careful pronunciation are phonetically distinct in their medial consonants, [p] and [b], but in casual pronunciation this distinction can be reduced, with an absence of voicing during closure for [b] and a shortening of the voice onset time for [p], rendering the two medial consonants phonetically very similar. This reduction of a phonological contrast poses a question about the nature of phonological encoding in long-term memory (i.e. the lexical form, see Chapter 8 this volume for discussion). If the phonetic detail related to reduced forms is not encoded in lexical representations, then the question shifts to address processing (see also Chapter 14 this volume, and Ernestus, this volume on this topic): what is the process by which a speaker/hearer establishes a mapping between the phonetic forms that are experienced and their encoding in the mental lexicon?

The variability of phonetic form is also a source of insight for phonology. Very often we can observe patterns of fine-grained phonetic variation that mirror phonological alternations or distributional restrictions. For example, the graded coarticulation of a vowel under the influence of a vowel in the upcoming syllable in English mirrors the phonological pattern of assimilation found, e.g. in local processes of umlaut or vowel harmony in other languages (Beddor et al. 2002; Cole et al. 2010). Observing patterns of "low-level" (i.e. sub-phonemic), gradient phonetic variation sheds light on how the phonetic context of a sound element can shape phonological patterns that restrict the occurrence of that element, and there is growing interest in uncovering the bases of phonological sound patterns in properties of phonetics and speech processing (e.g. Archangeli and Pulleyblank 1994; Blevins 2004; Hayes et al. 2004).

17.1.3 Motivating corpus analysis for phonology

17.1.3.1 *Variation and phonetic form*

In order to explore the variable phonetic substance of phonological elements, and the influence of phonetics in shaping sound patterns, the phonologist must go beyond the analysis of citation forms and examine words in connected speech in corpora that represent variation due to different speech styles (Hirschberg 2000; Yuan et al. 2005; see also Ernestus, this volume) and speaker variation (Foulkes 2010; see also Docherty and Mendoza-Denton, this volume). All this variety is part of the everyday experience of language for the speaker/hearer, and comprises the phonetic basis over which phonological patterns are learned.

When speech is used for communicative goals, as in the everyday use of language, it is produced with prosodic patterns that convey the information structure and pragmatic context of an utterance, and prosodic context is also known to affect the phonetic realization of words (e.g. Wightman et al. 1992; van Bergem 1993; Kochanski et al. 2005; Calhoun 2006; Turk and Shattuck-Hufnagel 2007; Cole et al. 2007; Yoon 2007; see also Turk, this volume; Frota, this volume). In addition, though prosodic features are present in all forms of speech regardless of style, the expressive content of spontaneous speech gives rise to a particularly rich variety of prosodic patterns, different from read speech (e.g. Nakatani et al. 1995; Schafer et al. 2005).

17.1.3.2 *Phonology in relation to linguistic structure and usage*

Spontaneous speech produced in communicative contexts offers the best opportunity to observe a wide range of phonetic variability; yet although researchers have devised methods for eliciting spontaneous speech in a laboratory setting (see Warner, this volume; Post and Nolan, this volume), by engaging subjects in controlled communicative tasks (e.g. Anderson et al. 1991; Hirschberg and Nakatani 1996; Schafer et al. 2005; Brown-Schmidt and Tanenhaus 2008; Khan 2008), the resulting data is (by design) less varied than speech produced in casual conversation. For direct observation of spontaneous, conversational speech, researchers turn to speech, databases for corpus analysis.

A speech corpus not only provides a basis for investigating variability in phonetic form, but it also provides a rich resource for studying the relationship between phonological form and other levels of linguistic structure. For instance, it has long been known that the sound patterns of a language may be sensitive to syntactic context (Kisseberth and Abasheikh 1974; Chen 1987) and may reflect discourse organization (Grosz and Hirschberg 1992). Clearly, evidence for any interaction between phonology and "higher" levels of linguistic structure must come from observation of whole phrases, multi-phrase utterances, and entire discourses. Similarly, a variety of syntactic, pragmatic, and discourse contexts are required to understand the phonology of intonation and prosody, and corpus materials have been widely used in such work (Post and Nolan, this volume).

Usage frequency is another factor known to influence the phonetic form of words. Greenberg and Fosler-Lussier (2000), Bybee (2001), and Bell et al. (2003), among others, have shown that words that occur frequently in speech have a higher incidence of consonant lenition and vowel reduction compared to low-frequency words. Usage statistics are calculated based on large corpora, which also provide plenty of data that illustrate the effects of usage on phonetic form. Bybee (2001) has also shown that patterns of phonetic reduction that arise in high-frequency words can be phonologized, resulting in stable synchronic sound patterns. By examining phonetic variation in relation to usage frequency, it is possible to identify

patterns that may be precursors to future sound change; see also Chapter 13 this volume.

17.1.4 Choosing a corpus

There are several considerations in choosing a speech corpus for phonological research. The first concerns the goal of the research and the availability of an existing corpus. A researcher interested in the effect of the given/new distinction on phonetic form may want to see how repeated mention affects the phonetic properties of words. This requires a corpus where speakers talk on the same topic for an extended period, incorporating multiple utterances, or multiple conversation turns in the case of dialogue, because repeated mention of a word is more likely in an extended discourse. A suitable corpus might be one consisting of extended interviews such as the Buckeye Corpus (Pitt et al. 2007); dialogues that are focused on a topic that sustains interest over time, as in the Switchboard corpus (Godfrey and Holliman 1997) or CallHome corpus (Canavan et al. 1997); or dialogues over lengthy tasks that require repeated mention of objects, places, or other things that are present in the task domain, as in the HCRC Map Task Corpus (HCRC Map Task Corpus 1993; see also Anderson et al. 1991). On the other hand, if the research goal is to investigate how speakers accommodate to the phonological and phonetic patterns of another person's speech, it would be essential to choose a corpus in which speakers with different speech patterns are engaged in interactive dialogue, such as the Fisher corpus (Cieri et al. 2004, 2005), which consists of telephone recordings from over 11,000 conversations between English speakers, representing a wide range of age groups and regional dialects, including non-US and foreign-accented varieties of English.

The corpora cited above are examples of speech databases for English (and in the case of CallHome, for other languages as well) that are in the public domain; they are disseminated to the public by a distributor, often with a licensing fee. The alternative to using an existing corpus is for the researcher to build a corpus from scratch, by recording speech samples directly from speakers recruited for that purpose. The advantages to using an existing, published corpus are savings in time and money, and with some corpora, access to a much larger database than a single individual researcher could construct. A further advantage to working with a corpus in the public domain is the possibility of building on the work others have done using the same corpus, or using prior results as a benchmark for testing new research methods.

Disadvantages of using existing corpora usually arise when the goals of the research are not adequately served by the speech materials available in existing corpora. For example, at the time of this writing, there is no publicly available database of dysarthric speech that surpasses the one compiled by H. Kim and her

colleagues containing just about one hour of speech for each of eighteen talkers (Kim et al. 2008). Likewise, to investigate the phonological structures of a non-standard dialect, the researcher may need access to speech that is produced in a social setting and register that is conducive to the use of that dialect. Speech samples that are recorded in a formal laboratory setting, or through interaction with an unfamiliar investigator who is not part of the target speech community, may fail to fully exhibit the characteristics of the dialect. A related limitation is the simple fact that there are no existing corpora for most languages, and similarly few databases for non-standard or non-prestige varieties of any language.[1] Using a portable digital voice recorder, spontaneous conversational speech data in any speaking style or language may be recorded with no substantial technical effort; most of the effort in acquiring a corpus is spent contacting subjects, acquiring their legal consent, creating a task description that will keep subjects talking long enough to collect the desired speech sample in the desired speaking style, and finally, transcribing the data.

Although spontaneous speech databases are especially relevant to the study of phonetic variation, existing corpora of read speech are appropriate for some research needs. Thus, the Boston University Radio Speech corpus (Ostendorf et al. 1996) is useful for research on prosody because it comes with a detailed, manually produced prosodic transcription, and a phone-level transcription, both of which are aligned with the audio signal. This corpus has been used for research on the acoustic correlates of prosodic features in American English, as they are represented in this style of professionally read speech (e.g. Dainora 2001; Choi et al. 2005; Kim and Cole 2005; Cole et al. 2007; Yoon 2007).

17.1.5 Corpus transcription

17.1.5.1 *Metadata and orthographic transcription*

In order for a speech database to be useful for phonological analysis, it is necessary to have some additional information about the content of the speech. Linguistic metadata will provide information about the speakers, such as sex, age, ethnicity, and region of residence. Metadata may also provide information about speaker recruitment and recording procedures.

[1] The Linguistic Data Consortium currently distributes speech databases for these languages: Arabic, Croatian, Czech, Dschang, English, Farsi, German, Hindi, Japanese, Korean, Mandarin, Nbomba, Portuguese, Russian, Spanish, Tamil, Turkish, Urdu, and Vietnamese. See also Loehr and Van Guilder (this chapter) for other languages and resources.

The most ubiquitous and, often, most useful type of annotation available for any speech corpus is its orthographic word transcription. Using an orthographic transcription together with a pronunciation dictionary, it is possible for the researcher to use simple text search tools in order to find places in the database where specific phonological structures of interest may have occurred, and to focus manual post-hoc analysis exclusively on the selected segments. At its simplest, the transcription is a separate document that specifies the words of each utterance in the database in running text. Much more useful are transcriptions that are time-stamped, so the beginning and end of each word (or sentence, or talker-turn) is indicated, allowing the researcher to locate that word/sentence/turn in the corresponding audio file. A useful method for producing partially time-stamped orthographic transcriptions is to segment the speech data at every silence longer than some threshold (e.g. 500 ms), and then to give the pre-segmented waveforms to transcribers for annotation.

Some corpora do not come with transcriptions, and the researcher must create one, as of course must be done for any corpus that is created by the researcher; working efficiently, it is possible for most annotators to transcribe utterance units in about four times real time, i.e. four minutes of transcriber time for every minute of speech. Although word transcription may seem like a very simple task, in the case of conversational speech complications arise due to disfluencies, hesitations, and speech repairs, or from poor signal quality. For these reasons, transcriptions almost always include questionable entries, where reasonable people disagree about what they hear in the recording. There are also a surprising number of orthographic ambiguities in the transcription of spoken English, e.g. numerical expressions, word fragments, idioms, discourse markers, and proper names each typically have two or more common transliterations. To minimize the impact of errors and uncertainties on the reliability of the transcription, transcription projects will typically rely on a written protocol for the treatment of disfluencies, errors, and ambiguous entries, which is used to train the transcribers (e.g. Linguistic Data Consortium 2009).

17.1.5.2 *Transcription of subword units: Phones and features*

When the research plan is to investigate phonetic variation at a level smaller than the word, such as the phone or syllable level, an additional layer of transcription is needed to identify such units within each word. Phone-level transcription is the most common subword level that is labeled in existing corpora, but transcriptions of this sort for large databases (anything over about 1,000 words) are rare. Because it is a very time-intensive task that requires phonetic training, phone-level transcription is rarely done by hand. Rather, an initial pass at transcription is made with the use of automated methods. Working from an orthographic word-level

transcription, the phones for each word can be retrieved from a digital pronunciation dictionary and automatically inserted into the transcription, as a further specification for each word. This step is followed by a procedure of *forced alignment*, by which each phone in the dictionary form of a word is mapped onto some portion of the acoustic signal for that word.

Forced alignment is done using algorithms from Automatic Speech Recognition (ASR), and is most successful when each phone associated with the word in its dictionary form is actually fully pronounced. But this is not always the case, and indeed, full pronunciation is not even typically the case for words in spontaneous speech (Greenberg and Fosler-Lussier 2000). Forced alignment can be improved by systems that explicitly model the most common patterns of pronunciation variation, but much more research is needed in this area to improve the reliability of the time-aligned phone labeling using this method. Some of the corpora mentioned above use forced alignment followed by a process of manual correction which can correct many if not all of the resulting errors (e.g. the Buckeye corpus). Manual correction is still a slow and costly procedure, but this dual approach using automatic labeling with manual correction is often an excellent compromise to the much more costly alternative of a full manual transcription.

The need for a digital dictionary for the use of forced alignment means that automatic phone labeling can be applied only to those languages for which such resources exist. Fortunately, there are efforts underway to produce such resources for an increasing pool of languages (e.g. Hussain et al. 2005).

Looking below the level of the phone, transcription can also specify smaller units such as phonological distinctive features or articulatory gestures. For example, phones specified for a given word in the pronunciation dictionary can be mapped onto distinctive features, and then automatic methods can be used to locate the distinctive features in the speech stream using acoustic landmarks. This approach has been demonstrated for many of the distinctive features used to encode lexical contrast (Stevens 2002; Livescu et al. 2007).

17.1.5.3 *Prosody transcription*

Corpus-based analyses have proved beneficial for the study of speech prosody, but introduce the need for an additional level of prosodic transcription. Using transcription methods such as the Tones and Break Indices (ToBI) system (Beckman et al. 2005), the locations of phrasal prominence and phonological phrase boundaries are identified, along with a tonal specification marking the associated pitch movement (see Post and Nolan, this volume for other approaches to prosody transcription). Prosody transcription is a complex task that incorporates the transcriber's auditory impression of prominence and phrasal juncture with

visual inspection of the graphical speech display (including at least the pitch track, waveform, and spectrogram), and requires specialized training. In the case of ToBI transcription it is also a slow task, taking anywhere from 10 to 100 times the duration of the speech recording, and requires first having a reliable time-aligned word transcription. And, like other forms of transcription, prosody transcription is error-prone and different transcribers can perceive the prosodic features of an utterance differently. Reliability studies of several ToBI transcription projects show that agreement rates between transcribers are impressively high—Pitrelli et al. (1994) report agreement rates of up to 81 per cent for tone label, and 92 per cent for the break index coding the level of phrasal juncture—but the potential for errors and uncertainty remains.

Many researchers have looked at ways to automate prosody transcription, primarily by identifying a set of acoustic correlates of prosody and using these features to train a classifier that takes as its input the word sequence, the acoustic speech signal, and sometimes additional information about part-of-speech or shallow syntactic features and returns a prosody annotation for each word or subword unit (e.g. Wightman et al. 1994; Syrdal et al. 2001; Chen et al. 2004; Ananthakrishnan and Narayanan 2008). These efforts have contributed greatly to the understanding of how prosody is encoded in the acoustic signal, but so far have not been successfully tested on spontaneous speech data.

17.1.5.4 *Assessing transcription reliability*

No corpus should be publicly released without at least two levels of quality validation. First, automatic verification using standard methods should be applied to any corpus prior to release. The energy of each waveform should be computed, in order to verify that every file in the distributed corpus contains speech. Transcription files should be spell-checked. The Linguistic Data Consortium (2004) recommends running a "syntax check" that searches transcription files for timestamps without text, illegal characters, ill-formed symbols (e.g. ill-formed foreign speech transcriptions or non-speech transcriptions), bad spacing around punctuation, and numerical utterances that are entered using digits rather than full orthographic words.

Second, any coding system that requires rater training (including phoneme, distinctive feature, and prosodic transcriptions) should be evaluated by measuring inter-transcriber agreement. It is usually impractical to duplicate transcriber effort for the entire corpus, but the general validity of the transcription system can be measured by assigning additional transcribers to re-code a small portion of the corpus. Cohen's kappa (Cohen 1960) or Fleiss's kappa (Fleiss 1971) are statistics that can be used to test the reliability of transcriptions across two or more transcribers.

17.1.6 Pronunciation dictionaries and lexica

As described in the preceding sections, it is possible to perform phonology research using a database of recorded speech, an orthographic transcription, and a pronunciation dictionary. It usually takes less time to write a dictionary than it would take to phonemically transcribe the entire corpus, but writing a dictionary is, itself, a time-consuming task. For this reason, until recently, the pronunciation dictionaries distributed with most speech technology applications (synthesizers and recognizers) were considered to be valuable pieces of intellectual property, protected by the full weight of international copyright law. Recently, encouraged by a few widely cited examples (Weide 1995), increasing numbers of dictionaries are being released to the public. These efforts are supported by the publication of open source licenses appropriate to the distribution of text data, e.g. the Creative Commons Share Alike license (Creative Commons 2009). The Creative Commons licenses allow users to add content to a published work, provided that, if the work is republished, it be republished under the same license with appropriate attribution; for example, Hasegawa-Johnson and Fleck have republished the dictionary for the Carnegie Melon University Pronouncing Dictionary (or *cmudict*, a machine-readable pronunciation dictionary for North American English) with added tags for syllabification, part of speech, and named entities, and with about 100,000 additional entries derived from other open sources (Hasegawa-Johnson and Fleck 2007).

Languages whose letter-to-sound mappings are more predictable than English may be well served by an orthographic dictionary. For example, Hussain et al. have published an Urdu pronouncing dictionary using pronunciation codes based on the traditional Urdu orthography plus vowels (Ijaz and Hussain 2007).

17.1.7 Statistical and computational methods for data analysis

After the researcher has obtained a speech corpus, created and assessed a transcription (if needed), and identified regions of interest within the corpus, data collection can begin. A wide variety of data may be extracted for the purpose of phonological investigation, depending on the researcher's specific interests. For instance, data may consist of acoustic measurements taken from the speech signal, articulatory measurements if they are available (e.g. Westbury 1994), measurements of lexical frequency or phonotactic probability, or properties of the phonological, syntactic, or discourse context in which a targeted phonological unit occurs. An important detail in coding the data is the assignment of a unique label to each data point which identifies the speech unit (e.g. word, phrase, or utterance) from where the measurement is extracted, and for ease of reference, that also identifies the speaker, file number, and any properties of the data or its context that will be considered in the analysis.

A benefit of corpus research in phonology is that it provides a ready training database for an analysis of the category structure of speech—a central concern of phonology. Statistical methods for classification analysis may be used to test how well the observed data can be classified into linguistically meaningful categories (e.g. voiced vs. voiceless stops, urban vs. rural dialect, phrase-final vs. phrase-medial position) based on one or more characteristics inherent in the items. There are many approaches to classification analysis, using linear or non-linear methods, e.g. regression, discriminant analysis, support vector machines, k-nearest neighbor, decision trees, neural networks, Bayesian models, Hidden Markov Models (e.g. Webb 1999; see also Chapter 22 this volume on statistical methods in laboratory phonology). Some of these methods are also used in machine learning to create computer algorithms that can automatically learn the distribution of the data items into linguistic categories (Mitchell 1997). These methods of classification analysis align with methods used for the creation of speech technologies, such as speech synthesis and automatic speech recognition, and many of the studies that employ these methods in the analysis of speech corpora simultaneously contribute to linguistic understanding and technology development (e.g. Chen et al. 2006; Liu et al. 2006; Hirschberg et al. 2007).

17.1.8 Summary

Speech corpora offer a valuable source of data for phonological investigation, and are arguably an essential resource for the study of sound patterns that arise in connected, casual speech. Relative to corpus-based research in other areas of linguistic inquiry, corpus phonology research is in its infancy, and there remains much to be learned from existing resources. But it is also true that the linguistic coverage of the existing corpora is limited to a fraction of the world's languages, and does not fully represent all the dialectal varieties and speech styles that are of phonological interest. Fortunately, the technology needed to construct a corpus, including recording equipment and digital storage, is fairly inexpensive and easily obtained. On the other hand, a corpus is only as good as its annotation, and the human resources needed to produce a reliable, quality transcription are considerable.

One of the distinguishing features of corpus-based research is the large volume of data that is available for analysis from even a medium-sized corpus, e.g. the Buckeye corpus (Pitt et al. 2007, comprising approximately twenty hours). On the other hand, even with a corpus of this size there may be a scarcity of examples of low-frequency phonological phenomena, reflecting the trade-off between the use of naturalistic speech materials drawn from a corpus and speech materials controlled by the experimenter and elicited in the laboratory.

17.2 Using the Internet for collecting phonological data

Dan Loehr and Linda Van Guilder

17.2.1 Introduction

While this chapter as a whole discusses corpus analysis, this section focuses on using the Internet for acquiring corpora. The Internet is a rich resource for collecting phonological data. One may search the Internet for data, or use the Internet as a platform for conducting experiments. We discuss each in turn.

17.2.2 Searching the Internet for phonological data

What does it mean to search the Internet for phonological data? There are two main types of search. In the first, one searches for existing phonological resources, i.e. linguistic data collected and transcribed by linguists, which has been made available on the Internet. In the second, one treats the Internet itself as a corpus, i.e. a source of raw linguistic data.

There exist a large and growing number of phonological resources on the Internet, which include both audio and transcribed data. Examples include the following.

- LDC (Linguist Data Consortium, <http://www.ldc.upenn.edu>) contains hundreds of speech corpora in a variety of languages, many with transcriptions.
- CLDC (Chinese Linguistic Data Consortium, <http://www.chineseldc.org/EN/index.htm>) contains corpora focused on Chinese.
- ELRA (European Language Resources Association, <http://www.elra.info>) and partner organization ELDA (Evaluations and Language Resources Distribution Agency, <http://www.elda.org>), contain corpora focused on European languages.
- UCLA (University of California at Los Angeles) Phonetics Laboratory Archive, <http://archive.phonetics.ucla.edu> contains recordings from hundreds of languages, and Phonetics Laboratory Data, <http://phonetics.ucla.edu>, originally "Sounds of the World's Languages," contains phonetic teaching material compiled by Peter Ladefoged.
- George Mason University's Speech Accent Archive, <http://accent.gmu.edu> contains recordings and IPA transcriptions of native and non-native speakers of English, representing hundreds of L1 languages and dialects, each reading an

elicitation containing most standard American English phones and clusters in various contexts.

- Berkeley's TELL (Turkish Electronic Living Lexicon, <http://linguistics.berkeley.edu/TELL>) provides orthographic and phonemic transcriptions for thousands of Turkish words.
- Boston University's National Center for Sign Language and Gesture Resources, <http://www.bu.edu/asllrp/cslgr> provides annotated videos of American Sign Language.
- TalkBank, <http://talkbank.org/data> contains a variety of transcribed conversations.
- CHILDES, <http://childes.psy.cmu.edu> contains corpora on child language, including the phonological collection PhonBank, <http://childes.psy.cmu.edu/phon>.
- SIDGrid (Social Informatics Data Grid, <http://sidgrid.ci.uchicago.edu>) is a portal for sharing corpora in the social sciences, including linguistics.
- VoxForge, <http://www.voxforge.org> provides freely available transcribed audio.
- OLAC (Open Language Archives Community, <http://www.language-archives.org>) contains pointers to other language corpora.
- ToBI (Tones and Break Indices, <http://www.ling.ohio-state.edu/~tobi>) is a framework for transcribing intonation and prosody which has been extended to over a dozen languages. Though ToBI is primarily a transcription standard, many language-specific ToBI websites contain transcribed data.

In addition to the above specific websites, there are genres of websites that can be useful to phonologists:

- Speech therapists compile word lists organized by phonological principles to aid patients with speech disorders. These can be located with an Internet search such as "speech pathology word lists." For example, Bowen (2008) provides an extensive list of English minimal pairs focusing on specific sounds.
- There are many other minimal pair collections, which can be found by simply searching on "minimal pair." For example, Higgins (2009) has collected several thousand, contrasting every pair of English phonemes for which minimal pairs have been found.
- Online dictionaries and language instruction websites typically provide pronunciations, sometimes with audio. Arguably the most comprehensive is the Oxford English Dictionary, <http://oed.com>. A dictionary commonly used in research is Carnegie Mellon University's downloadable Pronouncing Dictionary (Lenzo 2009), with over 100,000 English words transcribed in Arpabet (an ASCII-based

system for the phonemes of General American English). Spell-checking lists can also be useful, depending on how phonemic a language's orthography is.

One can also find phonological data by entering sample linguistic transcriptions into a search engine. For example, Googling the IPA characters *ðə* (*the*) will yield a variety of sites with IPA-transcribed English. Operating systems and browsers differ in the manner and ease of entering IPA characters. One fairly robust method is to copy and paste IPA characters from a webpage designed for this purpose, such as Ishida (2009). In addition to IPA, there exist other phonetic transcription schemes, mostly ASCII-based, which can be used in search terms to find transcribed data. These include Arpabet, TIMITBET, MRPA, SAMPA, X-SAMPA, WorldBet, and various transcriptions used by dictionary publishers (Phonetic Alphabets 2009).

As mentioned, the Internet can also be used as source of raw linguistic data. Viewing the Internet as a corpus has both benefits and drawbacks. On the positive side, the Internet is arguably the world's largest corpus, containing a vast amount and variety of recorded human language. Anyone hoping to download recordings of speech or videos of sign language needn't look far. Many sites maintain podcasts of audio. For sign language phonologists, vlogs (video blogs) also exist; an example site is the ASL Vlog and Video Directory, <www.aslvlog.net>.

Yet care is required when using the Internet as a source of linguistic data. Unlike traditional linguistic corpora, the Internet was not built by design, nor is it balanced. Corpus designers may, in fact, decline to label the Internet a corpus at all, but rather simply a data *collection*. Although the text portion of the Internet is largely indexed to facilitate search, little of the audio/video portion is indexed, hindering searches for data of interest. The researcher hoping to maintain experimental conditions has no control over the audio/video quality, or type of elicitation, or speaker's background. The data is often subject to intellectual property restrictions; unless permission is explicitly granted, researchers must restrict their use to "fair use." In short, linguistic data on the Internet is "found data," opportunistically acquired, with associated baggage. The situation is like the drunk looking under the lamp post for the lost car keys, though they could be anywhere, "because the light is better here." That is, one makes use of what is readily available, rather than searching in the dark for what one hopes to find.

Nevertheless, there is still much useful data under the lamp post. Assuming that some of it is suitable, a logical next step is to transcribe or annotate it for the researcher's purposes. Cole and Hasegawa-Johnson (this chapter) discuss transcription methods in more detail. This assumes the researcher is interested in audio recordings and their transcriptions. Some researchers, including Lavoie and Cohn (1999), Hayes and Zsuzsa (2006), and Zuraw (2007), have used written orthographic material as a source of phonological data.

A final word on phonological resources has to do with standards. Data annotated in a commonly recognized framework will be more useful to others, and more readily allow comparable analyses by different researchers on the same data. When

creating new data sets, therefore, researchers are urged to use standards where available for the phenomena of interest (e.g. IPA, or ToBI for prosody). In addition to using a standard *transcription*, researchers can also choose a standard *encoding type* for the transcription. A common encoding type is Unicode, which contains all IPA characters as well as most characters found in the world's writing systems. Unicode has variant implementations; a common choice is UTF-8. Finally, the choice of *file format* can also help or hinder data sharing. Microsoft Word and Excel are commonly used, though these are not universally compatible. More universal are "plain" text files, especially if they use UTF-8 encoding. For transcriptions making explicit reference to timelines, Annotation Graphs (Bird and Liberman 1999) provide a mechanism for capturing this temporal information. Schmidt et al. (2008) describe converters between Annotation Graphs and over a dozen annotation and analysis tools, including the widely used Praat, <http://www.praat.org>, enabling transcription interoperability.

17.2.3 Conducting phonological experiments on the Internet

Modern advances in audio processing and Internet technologies offer an unprecedented opportunity for expanding the empirical scope of laboratory phonology beyond the limitations imposed by traditional brick and mortar laboratories. Human speech can be faithfully captured and reproduced by commonly available audio recording and playback devices, and stored and transferred using high-fidelity audio formats. Ever-increasing bandwidth and data transfer speeds facilitate the deployment of these high-quality sound files in web-based speech applications. For laboratory phonology, the natural consequence of these advances is to deploy web-based speech experiments. The use of the World Wide Web as a vehicle for psycholinguistic experimentation promises access to a larger, more diverse and language-appropriate group of subjects and therefore a stronger empirical foundation for psycholinguistic analyses. Long-distance experiments allow subjects to participate at times and locations convenient to them, reducing the schedule coordination burden on the researcher while still offering a large degree of reproducibility in many facets of the experimental set-up. With this high-tech methodology, however, come a number of new considerations in experimental design. As large-scale, web-based experimentation emerges as a new paradigm in laboratory phonology, the community needs to define standards and best practices for conducting research at a distance. Among the questions to be addressed are:

- How to design a usable and effective Internet experiment?
- How to control for variations in the subjects' physical environments?
- How to collect, protect, and verify subject biographical data?
- How to factor in variations in subjects' computer set-ups?
- More fundamentally, how to determine when Internet-based experimentation is or is not appropriate?

The remainder of this section explores some of these issues, discusses technical challenges in deploying long-distance speech experiments, and highlights a few recent web-based psycholinguistic applications.

Web experimentation entails a computer-based, algorithmic approach to experimental design. Software-driven methods have both advantages and drawbacks. In an ideal laboratory-based context, computer-based experimentation provides precise and reproducible implementation of details such as the timing, presentation order, and volume of stimuli. Data collection and analysis can be automated and responses monitored in ways that are difficult for a human observer to achieve. Computers can accurately track details such as response timing, eye gaze, keyboard corrections, and mouse clicks. Experimental software can also provide a low-stress, face-saving way for subjects to generate feedback on various aspects of an experiment that they might not offer in face-to-face situations. Software-based experiments deployed over the Internet, however, are subject to additional limitations and considerations beyond those of computer-based set-ups conducted with the researcher present. The main detracting factors are lag time, which precludes time-based analyses such as eye-gaze tracking, and the lack of control over many facets of the subject's environment. In the web-based context, *environment* includes standard considerations such as background noise and external distractions as well as the technological details of the subject's computer set-up. These factors do not represent insurmountable obstacles to web-based experimentation, but they do need to be factored in to the design and presentation of long-distance experiments.

For web-based applications, the importance of managing user expectations and providing adequate documentation cannot be overstated. It is much easier for a subject to walk away from the experiment when there is nobody there monitoring the experiment and answering questions. In addition to the traditional task description provided at the beginning of an experiment, users should receive guidance on a number of matters that would be handled personally by the researcher under traditional conditions. Among the details that a web-based research introduction should provide are:

- How to operate the experimental software?
- Which environmental conditions should be sought or avoided?
- Should the experiment be completed in one session, or are multiple sessions allowed?
- Are rest breaks desired or to be avoided?
- Which combinations of hardware and software have been tested and are known to work?
- Can the experiment be performed easily using a slow Internet connection?
- What to do if the experiment is interrupted?
- Where to find help?
- Where to view annotation guidelines or Likert scale categories, if applicable?

As with lab-based experiments using computers, it can be helpful to familiarize the user with the software and the task using practice tests. In web-based experimentation, practice sessions have the additional benefit of offering subjects the opportunity to adjust their environment, for example, raising the speaker volume or switching browsers. Listing the computer and browser configurations that have been tested may encourage subjects to try alternative computer systems or browsers if they experience difficulties accessing or running the experimental software. Details of the user interface can be quite important as well, including easy access to relevant annotation guidelines and help documents. Visual clues such as progress bars can keep subjects engaged while files load or calculations are performed. By attending to such details, the research designer increases the likelihood that subjects will complete the experiment with minimal distractions or frustrations introduced by the experimental software itself. Pre-deployment trials of the experimental software on a range of operating systems and browsers allow researchers to address technical problems and enhance the user interface and online help by observing difficulties encountered by test subjects. Even with these precautions, long-distance experimentation is likely to require a certain amount of time allocated for email or telephone support.

Background noise and distractions are the most obvious challenges in a remote subject's environment. When participating at home or in an office, subjects may be disrupted by children, pets, the telephone, the boss, and countless mundane occurrences. To minimize the impact, the experimenter must take steps to educate subjects about the ideal environmental conditions for participating in the experiment. For example, if the research requires minimal background and white noise, the instructions might request that the subject perform the experiment in a quiet room with all extraneous electronic equipment turned off, and with the microphone or headphones as far from the computer's fan as possible. It can also be beneficial to provide mechanisms for reporting unexpected conditions that emerge during the course of participation. For example, Van Guilder (2007) reports that subjects used the optional text field provided with each stimulus-response pair in her cross-language perception experiment to report issues ranging from intermittent headphone problems to lexical interference. Background noise and interruptions are not a detriment to all experimental designs, however, and would be desirable conditions for researchers investigating speech in noisy or natural conditions. The key point for web-based research is to identify novel techniques for anticipating, detecting, mitigating, and tracking the same control conditions endemic to all empirical cognitive research.

The importance of anticipating less than perfect control conditions in long-distance experimentation extends to the collection of biographical data. In traditional laboratory phonology, researchers are present when collecting subject data such as language background and technical experience. They can therefore answer questions and provide guidance on issues such as what it means to be native,

fluent, or proficient in a language. In-person experimentation also allows the researcher to verify and personally assess certain facets of an individual's apparent versus self-reported data, including gender, age, dialect, and language and technical proficiency. Because these feedback modalities are unavailable in long-distance experiments, the researcher must be extra vigilant in designing biographical questionnaires. Technology offers potential innovations in this area, including the ability to automatically detect features such as the client browser's localization and language settings, country codes contained in email addresses and other clues indicating discrepancies from the self-reported control data. A benefit of Internet-based experiments is that they can collect data for extended periods of time at little to no additional cost. Consequently, if some subjects do not match the desired profile, the experiment can be kept open until completed by the target number of participants, barring external restrictions such as publication deadlines or expiration of review board approvals.

Web-based elicitation of biographical data requires new ways of handling certain administrative details. Internet-based experimentation offers increased potential for conducting anonymous experiments, if the experimental design allows for subject anonymity. If, however, an anonymous experiment is not possible or consent forms are required by an institution's experimental review board for any reason, then the consent must be signed electronically. In the Internet environment, the use of email addresses, passcodes, or digital signatures has become common practice for electronic signing. The web application must be designed to securely and anonymously store any subject data that should not be publicly exposed based on the conditions outlined in the consent form. Furthermore, as in traditional experiments, all instructions, consent forms, and other guide text must be offered to subjects in their native language if they require it, meaning that the web designer must be aware of localization features of each language. Relevant localization issues include the ability to properly display the writing system of each language by using an appropriate combination of encodings and fonts, as well as handling any special concerns pertaining to the calendar system of a specific locale. For detailed information on localization standards, see the Localization Industry Standards Association website, <http://www.lisa.org/>.

Variability in subjects' computing environments is inevitable. Participants will use dozens of permutations of operating systems, Internet browsers, and third-party plug-ins such as the Java Runtime or the Adobe Flash Player components. Their hardware will vary in terms of the type and age of their computers, as well as the quality of their sound cards, speakers, headphones, and microphones. The speed and quality of their Internet connections will range from those of dial-up modems to those of high-speed, high-fidelity fiber optics. Although these challenges must be addressed in the design of web-based experiments and questionnaires, they do not individually or collectively overwhelm the potential benefits of collecting statistically significant sample sets from a large, representative group of

subjects. The onus is on the researcher to implement experiments using technologies that balance functional requirements with the likelihood that the software is usable by a maximum number of subjects. Van Guilder (2007) deployed a cross-language perception experiment using a combination of Java Server Pages for the questionnaires and a Java applet for the main perception experiment. The applet allowed for fine-grained control over user interaction, but at the calculated expense of certain potential subjects being unable to run the experiment on unexpected combinations of operating systems, browsers, and Java Runtime plug-in versions. Although the applet was coded in a relatively old version of Java for maximum backward compatibility, some people interested in participating had even older versions of the Java plug-in running within outdated, unsupported browsers. In some cases problems resulted from non-standard, proprietary HTML syntax used by competing browsers. In the six weeks that the experiment was online, 119 participants started and one hundred completed it. More than thirty additional people tried to run the applet, but could not overcome the technical obstacles of updating browsers or plug-ins and ultimately opted not to participate. Of the nineteen participants who started but did not complete the experiment, six reported that their dial-up modems were too slow, while the remainder found the cross-language perception task too difficult or time-consuming.

The connection speed issue is one which should be carefully considered in determining whether web-based experimentation is appropriate for a given research goal. Transfer of data and responses between the subject's computer and the server hosting the experiment entails a delay, meaning that time-focused experiments are not suitable for Internet-based research. The latency or lag time of the responses is not constant, and can be affected by many factors, including the type of Internet service used by the subject and the amount of traffic on the Internet at different times of day. If response timing is not a key factor in the research, connection speed is still ranked high among a set of implementation details that should be addressed in a number of ways, including through user expectation management, efficient application design, and careful data preparation. For example, if an experiment involves high-volume file transfer to and from the subjects' computers, the researcher should provide a warning indicating that users with slow or unreliable connections may want to participate from an alternate location, including an estimate of the time required to complete the experiment at different connection speeds. In addition, care should be taken to reduce the size of pre-existing audio or video files as much as possible without clipping or distorting the speech signal. The easiest example is to trim silence from the edges of sound files to reduce both their size and the perceived lag time in playing them. Other alternatives for minimizing the impact of numerous or large stimulus files on the usability of the software include presenting them to subjects in smaller groups or loading the files asynchronously in the background as the user performs other tasks such as filling out biographical information or reading instructions.

Audio-visual and computer hardware, such as motherboards, sound cards, speakers, headsets and microphones, present a slightly different challenge, since the quality of this equipment theoretically has more potential to affect the results of speech-based experiments, particularly where older, faulty, or substandard equipment is concerned. Controlling the hardware used by remote subjects would be very difficult, but for many experiments merely tracking the equipment used represents an important control variable. The level of granularity does not necessarily need to be fine; for example, Scharinger (2007b) performed a web-based perception experiment which began by presenting users with a selection box to indicate the type of speakers used, with five fairly coarse-grained choices:

1. Headphones (small, one ear)
2. Headphones (standard, two ears)
3. Desktop loudspeakers (in screen)
4. Desktop loudspeakers (stand-alone)
5. High-quality loudspeakers

The web experiment was one component in a larger study examining vowel perception, which measured event-related potentials produced by subjects in response to stimuli beginning with the same consonant cluster onset, "st," followed by either a front-mid-round or a back-mid-round vowel. The goal was to determine average timing intervals required for subjects to "accurately discriminate between "st-back vowel" and "st-front vowel" stimuli, using several synthetically constructed gate sequences of 'st' followed by a vowel with 2 millisecond differences between each of the sequences" (Scharinger 2009, personal communication).

Since the perceptual divergence portion of study was not concerned with subjects' reaction times, it was possible to conduct the experiment online without concern over Internet connection speeds or loading time for HTML pages. Subjects were asked to judge whether or not pairs of inputs were identical. The speaker types were tracked to control for anomalies in the data, given that stimuli pairs could differ by intervals as short as two milliseconds. See Scharinger (2007a,2008) for more details. Information about subjects' equipment could correlate with otherwise inexplicable patterns in the data. For example, the use of laptop speakers might pattern with less accurate discrimination of certain phones or features. As with the other issues surrounding online experimentation, hardware configuration differences are not necessarily insurmountable obstacles; rather, researchers need to include these factors in the design of their experiment and in their determination of whether a web-based approach is appropriate.

As technology continues to advance, the appeal of conducting experiments in laboratory phonology over the Internet will grow. It is currently challenging to locate phonological experiments on the web, in part because they are less common than other types of online psycholinguistic research and in part because such experiments are transient, appropriately existing online only for the duration of

the research. A model for facilitating and tracking online linguistic experiments is offered by the Portal for Psychological Experiments on Language, a website sponsored jointly by the University of Edinburgh, the University of Glasgow, and the Universität des Saarlande (http://www.surf.to/experiments). This site hosts online experiments, lists past experiments, and offers researchers the possibility of submitting publications related to experiments conducted on the site. On the Portal site, as well as other online language experimentation sites such as the Max Planck Institute for Psycholinguistics (http://www.mpi.nl/) and the Language Cognition Laboratory (http://coglanglab.org/participate.html), the majority of online linguistic research explores semantic, pragmatic, or syntactic phenomena. A possible explanation for the relative scarcity of web-based phonology experiments could simply be that phonologists are not necessarily programmers or web developers, so unless the research is conducted by a team, the appropriate technical skills might be lacking. Another plausible factor is that the community is in the early stages of identifying and defining the types of phonological research that are appropriate for web-based experimentation. Categorical perception experiments seem to dominate this early stage (Eriksson 2007; Sharinger 2007a; Van Guilder 2007; Zeng 2009). By engaging in open dialog about the challenges in web-based experimentation, the laboratory phonology community will be able to define and standardize best practices for this emerging experimental paradigm.

17.3 Speech manipulation, synthesis, and automatic recognition in laboratory phonology

Henning Reetz

17.3.1 Introduction

Using natural speech in speech production or speech perception experiments is the ultimate goal in speech research. Unfortunately, even the largest corpus may not have all the specific sounds, variants, or contexts the researcher wants to investigate. For example, when investigating the influence of VOT durations (Voice Onset Time, Lisker and Abramson 1964) on the perception of voiced or voiceless stops, the desired range of VOTs may not be found in the data set. Well-controlled speech is needed to investigate such a question (e.g. Carney et al. 1977). Similar issues would arise for a variety of other research questions, including how perception is affected

by formant trajectories (Lindblom and Studdert-Kennedy 1967), or the shapes of intonation contours (Gussenhoven and Rietveld 1992), to name but a few topics. One possibility is to use natural speech signals and modify them for this type of perception research, e.g. by shortening their VOTs (see Section 17.3.2). Another approach is to analyze natural speech into a set of acoustic parameters, modify them appropriately, and re-synthesize a speech signal from them. Alternatively, a speech signal can be synthesized from a set of parameters directly (see Section 17.3.3, and Iverson, this volume) for use in perception experiments. Such synthesis can be based on an acoustic model of the human speaking apparatus (Klatt 1980; Markel and Gray 1976) or can try to mimic the behavior of the articulatory organs (Rubin et al. 1981). Although a (re-)synthesis allows very good control over properties of the generated speech signal, it can have a rather unnatural "metallic" or "synthetic" sound quality.

Studying speech production with large corpora of natural speech, rather than using only small sets of well-controlled laboratory speech, involves the transcription and analyses of large amounts of data. This process is extremely time-consuming and costly when performed by humans. To speed up this process (semi-)automatic speech recognition and labeling methods are often applied (e.g. Pluymaekers et al. 2010; see Cole and Hasegawa-Johnson, this chapter). The (semi-)automated approach uses automatic speech recognition (ASR) methods, discussed in Section 17.3.4. These ASR systems are based on stochastic models, which operate on principles that might be fundamentally different from human speech perception. Only a few systems have been developed to test a specific model of human speech perception in a computer simulation (see Section 17.3.5).

17.3.2 Manipulating natural speech

One of the easiest methods of manipulating natural speech is the splicing technique, where parts of a speech signal are cut out, repeated, or cross-spliced with another piece of the signal. An example is the manipulation of vowel length, closure duration, VOT, and other parameters to mark the contrast between voiced and voiceless (or lenis and fortis) stops. Van Dommelen (1983) found a complex interaction of duration and intensity properties guiding the perception of French plosives. This relation was found even when signal stretches *after* the target segment were manipulated. This regressive influence of duration manipulation on the perception of segments was also investigated by Repp et al. (1978) who manipulated the durations of silent pause and frication in the phrase DID ANYBODY SEE THE GRAY SHIP. They found that a shortened [ʃ] led to the perception of GRAY CHIP whereas lengthening the segment led to the perception of GREAT SHIP. These two experiments demonstrate that listeners integrate several durational and intensity cues into one phonetic percept. Replacing parts of a signal by other signals, for example by splicing a

vowel from an appropriate into an inappropriate context (Beddor et al. 2001) can lead to insights about the contribution of coarticulatory information in speech perception. Beddor et al. showed that participants could perceive the difference between oral and nasal vowels when they were in their appropriate oral and nasal contexts, but performance fell to chance level when the context was inappropriate (i.e. an oral vowel and a nasal context, and vice versa). They concluded that listeners compensate for the missing coarticulatory information. This observation can help explain historic sound change, where a ṼN is reinterpreted as a single nasal vowel, or cases like the cross-linguistic perception of a Ṽ as a VN sequence (e.g. American listeners hearing French *vin* 'wine' as *van*).

The gating technique is another form of natural speech signal manipulation often applied in psycholinguistic experiments, where parts of a speech signal are cut off, and incrementally more of the signal is presented to a listener. The participants in such an experiment have to write down the sounds or the words that they think they hear. For example, Lahiri and Marslen-Wilson (1991) presented /CVC/ and /CVN/ words to English speakers, and /CVC/, /CṼC/, and /CVN/ words to Bengali speakers. The underlying oral vowel in /CVN/ words is realized in both languages with coarticulated nasality as [CṼN]. But since Bengali has—unlike English—underlying nasal vowels, English and Bengali listeners in a gating experiment might respond differently to this vowel nasality. The CṼC stimuli in Bengali activated CṼC words as soon as the nasality of the vowel could be perceived. In Bengali and English, the oral stimuli activated mostly CVC and some CVN words, roughly proportionate to their distribution in the lexicon. The stimuli with nasalized vowels activated more CVN words in English due to the coarticulated nasality. In Bengali, on the other hand, these stimuli activated first mostly CṼC and few CVN words, until the nasal consonant was presented, when CVN words were perceived. Lahiri and Marslen-Wilson's conclusion was that the nasality, which was clearly perceived from the signal, activated only words with underlying nasality in Bengali in the mental lexicon (i.e. CṼC), and not forms with nasality on the surface (i.e. [CṼN]) (Ohala and Ohala 1995 found the same result with a slightly different methodology).

Another speech signal manipulation is the mixing of two signals. Two signals are usually mixed when a direct manipulation of the acoustic parameters with (re-)synthesized speech (see Section 17.3.3) does not lead to a natural sound quality or because the crucial acoustic parameters are not well understood. For example, Norris et al. (2003) mixed voiceless labial and alveolar fricatives with forty-one different amplitude relations (e.g. 20 percent [f] and 80 percent [s]). In the key experiment they found that exposure to the ambiguous fricative shifted the perceptional boundary, but—crucially—this occurred only for lexical items. They concluded that listeners can adjust rapidly to interpret ambiguous sounds by using their lexical knowledge.

While manipulation in the time domain by cutting, splicing, and inserting parts of a speech signal is a rather simple and straightforward procedure, manipulating articulation rate is a more complex operation. (*Articulation rate* refers to the number of produced segments per time unit whereas *speaking rate* includes pauses, hesitations, and the number of intended segments per time unit; for a discussion see Koreman 2006.) First, there is the simple fact that changing the speed of a reel-to-reel tape or overwriting the sampling rate of a file (*without* re-sampling it) always changes pitch *and* rate of a signal together. Second, a faster (or slower) articulation rate is not simply accomplished in natural speech by speeding up (or slowing down) the production of all segments in a homogeneous way. To speak faster, speakers reduce or delete segments (Koreman 2006), and vowels are compressed more than consonants (Max and Caruso 1997). Fundamental frequency, intensity, and vowel quality are further parameters that are subject to speaking and articulation rate (Gay 1978b).

To disentangle the physical dependency between speech rate and its pitch to manipulate intonation or articulation rate within a limited range, one can use the PSOLA technique (Pitch Synchronous OverLap Add; Moulines and Charpentier 1990). This method compresses or stretches glottal periods (to increase or decrease pitch) and replicates or removes them (to lengthen or shorten segments, or to compensate the length modifications of the pitch manipulations), and interpolates them with the original signal. The signal is thus a modified "original" signal, which usually gives a natural-sounding voice where the f0 contour or the articulation rate can be controlled. For example, Grabe et al. (2003) constructed eleven different intonation contours from a short English phrase with the PSOLA technique. They presented these phrases to British English, Iberian Spanish, and Mandarin Chinese speakers. They found the same grouping of the stimuli into High-Low and Low-High classes, although Chinese is a tone language, and English and Spanish have different stress accent patterns. They concluded that this result gives support for a universal auditory mechanism that is not language-specific.

The methods presented so far operate in the time domain, where a signal is directly manipulated (cut, replicated, spliced, amplified, or attenuated) or stretched/compressed with the PSOLA method. The outcome usually sounds quite natural since it is modified human speech fulfilling certain criteria (e.g. VOT durations, frication amplitudes, f0 contours). These methods have two limitations. First, spectral parameters cannot be manipulated directly (although it is possible to combine PSOLA with signal mixture to produce a transition between two sounds, e.g. from [n] to [m]; see Mitterer and McQueen 2009). Second, decisive control of acoustic parameters (e.g. formants) is not possible. Furthermore, the influence of articulatory movements on the produced speech signal cannot be investigated.

17.3.3 Synthesizing speech

Good control over properties of a speech signal can be achieved by synthesizing from a parameter set. Synthesis can be based on an articulatory or an acoustic model. An articulatory synthesizer uses geometric shapes of the vocal tract to generate a speech signal. An acoustic model uses formant frequencies and other acoustic parameters for the synthesis. The parameter sets themselves can be constructed from scratch, by simply writing them down or extracting them from a database. For an acoustic model, the parameter sets are often computed from an acoustic analysis of natural speech. They are then modified and a signal is eventually re-synthesized. This reduces the burden of finding the correct parameters and preserves the properties of a natural signal, as far as the analyzing/synthesizing procedure allows.

There are two main classes of acoustic synthesizers, the LPC and the formant synthesizers. The LPC *analysis* (Linear Predictive Coding; Markel and Gray 1976) converts speech into signal-processing parameters that can be transformed to phonetically interpretable data, namely formant frequencies, their bandwidths, signal amplitude, and f0 parameters. These parameters, which also can be generated from scratch, can be used to drive an LPC synthesizer to generate speech signals. The relatively small parameter set (usually twelve parameters for every 10 or 20 ms frame) is easy to manipulate and can be used to generate signals with well-defined formant trajectories or f0 contours (e.g. Kohler 1990b). Warner, Fountain, and Tucker (2009) manipulated signal intensity directly, duration with PSOLA, and the fourth formant with LPC resynthesis to investigate acoustic characteristics in a perception study of /t, d/ flaps in American English. They found that an intensity dip contributed the most to a flap perception, duration manipulation less, and F4 manipulation hardly at all, although the latter was clearly visible in the spectrographic analysis of their real speech data. They concluded that there must be additional cues to the perception of presence/absence of flapped /t, d/. One disadvantage of LPC synthesis is its somewhat "metallic" sound quality due to the glottal excitation by a pulse train (Sambur et al. 1978), which was also observed by Warner, Fountain, and Tucker.

A more sophisticated acoustic modeling of the speech signal is possible with a formant synthesizer (e.g. Klatt 1980), which typically models the vocal tract in much more detail, e.g. with separate oral and nasal tracts, "cascaded" formants (for the generation of vowels) and "parallel" formants (for consonants), and noise sources at different locations along the vocal tract. These parameter sets can be defined in small time steps if necessary. To produce high-quality speech requires fine-tuned parameter sets (e.g. 60 parameters in Klatt and Klatt 1990), which have to be adjusted in small temporal increments (5 ms or less). This leads to a large demand on the generation and manipulation of these parameters. To reduce the burden, the parameters can be estimated from LPC parameters (e.g. *copy synthesis*, Scheffers and Simpson 1995, or in Praat, Boersma 2009) which are used as input for

re-synthesis. For synthesis from scratch, the parameters are taken from a database or can be computed from higher-level parameters, which take care of dependencies between parameters. For example the *high-level synthesis* (HL, Stevens and Bickley 1991) uses ten "high-level" parameters to derive the "lower-level" parameters for the formant synthesizer. These higher-level parameters are intended to describe articulatory configurations of a speaker with information about, for example, the area of glottal opening, cross-sectional area of the velopharyngeal port opening, and intraoral pressure. Providing this information still requires the specification of considerable detail by the user. Another high-level approach to synthesis is the DELTA system (Hertz 1990), which was explicitly designed to test phonological and phonetic theories. A main objective in this high-level programming language was to provide access to a multi-tiered data structure to investigate the relationship between phonological and phonetic units. The system flexibly used user-specified tiers (from the phrase, word, morpheme, phoneme, CV, nucleus, syllable, and tone, down to duration, f0, and formant specifications) and allowed the implementation of rules referencing these tiers to generate data sets that eventually could be used to drive a speech synthesizer.

The speech manipulation and syntheses described so far are used to produce or manipulate speech with certain acoustic parameters, but are not well suited to testing or simulating speech production models. (The HL synthesis does not directly allow a specification like "move the tongue forward" but requires the specification of changing formant values that goes along with an articulation.) For the investigation of articulatory-acoustic processes, articulatory synthesis is used (e.g. Iskarous et al. 2003; see Iskarous, this volume). Such a synthesizer can be used to generate a speech signal according to articulatory specifications. Furthermore, it allows comparison of predictions about articulatory activities for phoneme sequences with observed movements, as they can be captured with articulatory measurements. For example, Kaburagi and Honda (1996) computed in their model the articulatory movements of jaw, lips, and tongue. Their movements are constrained by a cost function to restrict the "degree of freedom" problem in articulation-to-acoustic mapping: there are many possible articulations to produce one sound (Saltzman 1979) and restrictions must be given to choose only one articulatory movement. Kaburagi and Honda generated several vowel and VCV sequences and compared their simulations with articulatory measurements, finding reasonably good agreement between the data sets.

There are other possible approaches to speech synthesis. For example, Bangayan et al. (1996) used the program SPICE (Simulation Program with Integrated Circuit Emphasis), which was written to design loudspeaker filters, to model the acoustic properties of the vocal-tract filter. This gave them the capability to model details of the anti-formant behavior of retroflex consonants, which are hard to design in a conventional LPC-, formant-, or articulatory-based synthesis.

Speech synthesis (in a broader sense) in laboratory phonology research is used to modify natural speech, re-synthesize manipulated versions, or synthesize artificial speech sounds to test the influence of acoustic or articulatory parameters on speech perception. The methods for manipulating human speech usually give a more natural-sounding quality but are restricted to essentially temporal and intensity manipulations. Formant (re-)synthesis allows more—and naturally more complicated—control of parameters but often goes along with a more "synthetic" sound quality. Articulatory synthesis ideally would be the ultimate way to mimic (and study) speech production, but must simulate the whole complexity of the human speaking process, which is still a very complex challenge. Equally challenging is the other way of studying speech perception—by implementing models of (automatic) speech recognition in a machine. Some approaches in this area are described in the next section.

17.3.4 Automatic recognition of speech

Analyzing natural speech corpora to investigate speech production can be supported by automatic speech recognition (ASR) systems, which assist segmenting and labeling large amounts of speech (i.e. *align* it with an orthographic transcription), whose segment boundaries can then be adjusted by hand (e.g. Greenberg et al. 1996). Nowadays these methods are based on systems that work on the same principles as commercial ASR systems, which are solely based on stochastic Hidden Markov Models (HMMs, Rabiner 1989). There are many different implementations and even the HTK system (Hidden Markov Model Toolkit, Young et al. 2002), which is often used in speech research, is a toolbox to build HMM recognizers rather than a ready-made system.

In any HMM system, speech data and its transcription are used to train a recognizer, which is then used to transcribe unknown speech. Normally, there is much more training material than "recognized" speech. For example, Pluymaekers et al. (2010) used the ASR technology to segment 432 spoken Dutch words ending in *-igheid* /əxhɛɪt/ into phonemes. Their training set consisted of 13,328 phrases (not containing the words to recognize). They found that the duration of /xh/ was longer when *-igheid* was a single suffix than in cases where *-ig-* was a part of the stem and *-heid* a suffix. They argue that /xh/ was predictable in the first case from the preceding sequence of sounds. It was thus produced faster than in the latter case, where the suffix was added to a word stem, which was informative on its own. In their view, this result contradicts a prosodic structure hypothesis, where /xh/ should have been lengthened as a morphological boundary marker. Pluymaekers and his colleagues could have segmented the speech material by hand, probably faster than with the automated method, but they argue that the automatic method is not influenced by linguistic knowledge, that it is consistent, and that the

transcription can be equal to a human transcription (Pluymaekers et al. 2010: 519). Only a few ASR systems do not rely on stochastic principles claiming to mimic parts of the human speech perception process, for example the Acoustic Landmark model (Stevens 2002) and the FUL system (Lahiri and Reetz 2002, 2010), which are presented below in more detail.

Stevens (1992) transformed his LAFF (Lexical Access From Features) system into the Acoustic Landmarks model (Stevens 1994, 2002). This system uses acoustic landmarks as anchor points to locate information about segments. These landmarks are characteristic discontinuities at stop closures and releases, intensity peaks in syllable nuclei, and minima from glides as they can be observed in spectrograms and waveforms. The idea is that consonants are marked by sudden changes due to the building up of a constriction somewhere along the vocal tract. These rapid changes are often well demarcated in the spectrogram or waveform. In contrast, the mid-part of a vowel is reasonably uninfluenced by adjacent segments. Additionally, vowels, or more generally syllable nuclei, normally have high amplitude and are therefore robust against background noise, and serve as a starting point for formant estimations. Glides lie in between consonants and vowels and show a dip in the energy contour that does not appear in vowels. Segmental information is determined from the landmarks of these three sound classes and is filled into a table together with all features associated with the segments. Possible interactions with adjacent segments are then predicted by these feature sets, and possible assimilations, deletions, and insertions of segments are generated from the possible interactions. This leads to hypotheses about word forms, which are then tested against the spectral characteristics.

Another system that tries to mimic the access of words from the mental lexicon of humans is the Featurally Underspecified Lexicon system (FUL, Lahiri and Reetz 2002; Lahiri, this volume). This system assumes a sparse phonological representation of speech, to which the acoustic signal is mapped. Here abstract phonological features, rather than acoustically rich information, are the basis of representation in the mental lexicon. The acoustic front-end of this system computes online spectral characteristics, which are converted by rather broad acoustic specifications to phonological features (i.e. a first formant above 600 Hz triggers the feature [low]). The emerging feature sets are then mapped to the lexicon, which contains sequences of (underspecified) feature bundles to represent phonemes, with a ternary logic of *match* (a feature computed from the signal matches with a feature in the lexicon), *mismatch* (a computed feature is mutually exclusive, like [high] mismatches with [low]), or *no-mismatch* (a feature does neither match nor mismatch, e.g. a [low] does not mismatch with an underspecified tongue-height specification). Crucial to the system is that certain features (e.g. [coronal]) can be extracted from the signal but are not part of the underlying lexical representation. Hence, a sound like [n] will lead to the extraction of the feature [coronal], which is a *no-mismatch* with an underspecified place feature of an /n/, but it will *mismatch* with a [labial]

feature of an /m/ (i.e. [n] only activates an /n/ but excludes /m/). The other way round, [m] will lead to the extraction of [labial] features, which *matches* with the [labial] of /m/ but is a *no-mismatch* with /n/, which is not specified for place (i.e. an [m] activates /m/ but does not exclude /n/). This procedure should explain the asymmetry observed in assimilations (that coronal sounds can become labials and velars, but not the other way around) and the associated asymmetry in perception (i.e. that listeners accept a word like 'meam' as a variant of 'mean' but not 'crean' as a variant of 'cream').

Since the whole feature extraction and mapping process is envisioned as an online process, the system does not require segmentation of the signal into segments and performs without backtracking. In this respect it is surprisingly similar to HMM systems, although the latter store detailed acoustic information and rely on large, pre-compiled databases.

17.3.5 Conclusion

While two decades ago the manipulation, synthesis, and automatic labeling or recognition of speech was only possible with highly dedicated and expensive computer systems, usually requiring technical staff for its operation, any off-the-shelf laptop computer can now perform these tasks with affordable or even free programs. Although this has removed a technical burden from planning and performing experiments with speech, the quality of synthesized or automatically recognized speech is often still clearly distinct from human performance. (Synthetic stimuli can even lead to different brain activities than natural stimuli do; Lattner et al. 2003). Nevertheless, the relative ease of use of the available tools has the benefit that it helps the researcher concentrate more on the underlying theoretical questions rather than on dealing with technical obstacles.

17.4 Phonotactic patterns in lexical corpora

Stefan A. Frisch

17.4.1 Introduction

The final section of this chapter considers the use of lexical corpora, ideally in the form of an electronic dictionary, in studies of language phonology. Lexical corpora

may be created as stand-alone entities (e.g. the Hoosier Mental Lexicon based on the Webster's dictionary used below; Nusbaum et al. 1984) or derived from speech corpora by indexing all of the words used (e.g. the dictionaries created by the Linguistic Data Consortium based on the Callhome project).

The use of corpora and databases in laboratory phonology studies has opened up a new line of inquiry for researchers to examine phonological patterns in frequency and/or probability, and related factors of lexical neighborhood density and similarity. These lines of inquiry allow researchers to use statistical techniques to examine phonological patterns and investigate systematic non-categorical (gradient) phonological patterns. In many ways, this approach to laboratory phonology is a minimal departure from generative phonology, and a variety of generative implementations of frequency effects exist using Optimality Theory (e.g. Hayes 2000; Boersma and Hayes 2001; Albright 2002a; Frisch et al. 2004; Hammond 2004; Anttila 2008a; Hayes and Wilson 2008). Alternatively, some researchers have explored more holistic measures of similarity between words and phonemes to explore analogical approaches to phonological patterning (e.g. Bybee 2001, 2006; Pierrehumbert 2003). Many of these phonological studies have related work in the phonetic or psycholinguistic literature that use analogous approaches or that share quantitative metrics.

17.4.2 Frequency and similarity as tools in laboratory phonology research

17.4.2.1 *Frequency and probability*

Phonological frequency counts can consider either the token frequency or type frequency of the unit. Token frequency is the frequency of usage of the unit over a corpus of language usage. For example, multiple usages of the same word in a corpus will contribute to the token frequency of the constituents of the word (such as phonemes, onsets, and rimes). Type frequency is the frequency of the unit over the lexicon. In this case, each word contributes to the type frequency of its constituents only once. As an example, consider word-onset /ð/ in English. This onset is found in relatively few words in English, giving it a low type frequency, but many of these are function words such as *the*, *this*, *these*, *that*, *then*, *them*, so the token frequency of this phoneme in English is quite high.

As an example, this section will examine consonant combinations in a lexicon of English (Nusbaum et al. 1984). This particular lexicon includes the orthography, phonemic transcription, CV pattern, stress pattern, part of speech, frequency of occurrence, as well as lexical neighborhood information for about 20,000 words of English. Here, we will examine the distribution of onset and coda consonants in the 1,324 CVC words found in this lexicon. Since the lexicon contains a column with the CV pattern of the word, extracting all CVC words is easily done with a

script command (such as *grep* in Unix) or by sorting within a data table (such as in Microsoft Excel, which can load this entire dictionary in a single spreadsheet). Table 17.4.1 presents a typical tabulation of this type of data. Rows in the table show the different onset consonants, and columns show the coda consonants. The numbers show the number of CVC words in the lexicon with that combination of onset and rime consonant. A blank cell indicates no co-occurrences.[2] The row and column totals show the number of CVC words with each individual onset and coda consonant, respectively.

As can be seen, some consonants are more frequent than others. For example, /l/ is the most common onset consonant with 102 occurrences in distinct (non-homophonous) words.

Frequency counts will be affected by the size of the corpus that is being examined. Even for type frequencies, there are different sizes of lexical corpora (the CELEX dictionary, for example, is much larger than the one used here). Frequencies can be normalized in a variety of ways. One of the most common measures for data, such as in Table 17.4.1, is to use probability, rather than frequency. Probability of each unit (in this case a consonant) is the number of occurrences of the unit divided by the total number of occurrences of units of the same type. The right-hand column of Table 17.4.1 shows probabilities for the various onset consonants in CVC words in English. For example, for the onset /b/, there are 91 occurrences and 1,324 total onsets, giving a probability of 0.069.

$$probability = \frac{frequency\ (unit)}{frequency\ (total)}$$

Phonological patterning is the systematic combination or avoidance of a combination of phonological units. Table 17.4.1 shows the frequency of co-occurrence of onset and rime consonants. Cases where there are zero co-occurrences may be evidence for a phonological constraint against the combination. However, many constraints have the occasional exception. Conversely, cases where there are a very large number of co-occurrences may be evidence for a phonological constraint in favor of the combination. In this case it is much more difficult to decide how many co-occurrences indicate a constraint favoring co-occurrence without considering statistical measures.

Given the ability to compute probability for phonological units (such as onsets and codas in the example table), the expected probability of a combination can be computed. This is simply the product of the probabilities of the two (or more) units being combined. This probability represents the expected likelihood of the combination of the phonological units if units co-occurred with one another freely.

[2] Note that some entire rows and columns are blank, reflecting the absence of certain segments in word onset (/ʒ, ŋ/) or coda (/j, w, h/) position.

Table 17.4.1. Onset and coda consonant distribution in English CVC words

onset ↓	coda → b	p	f	v	m	θ	ð	t	d	s	z	n	l	r	ʃ	ʒ	tʃ	dʒ	j	k	g	ŋ	w	h	Total	p
b	3	1	2		4	4	1	12	8	4	4	8	8	4	3	1	5	2		9	6	2			91	0.069
p	1	9	1	2	1	2		10	4	6	4	9	10	5	2		7	2		9	3	2			89	0.067
f	2	1	2	1	3	2		7	5	2	6	9	11	5	1		1	1		2	3	1			65	0.049
v	1			1	1			3	1	4		3	5	1			2	1			2				25	0.019
m	1	3	2	2	3	4		8	7	9	1	7	9	5	3		2	2		6	1				75	0.057
θ			1	1	2			1	2			2	1				1			1	1	2			15	0.011
ð					1			1		2	2	3		1											10	0.008
t	3	6	3		7	2	2	8	2	2	1	9	9	6			2			11	3	3			79	0.060
d	4	4	2	2	8	2		6	6	4	5	10	8	4	3		2	3		6	3	1			83	0.063
s	2	7	2	4	4	2	3	10	5	3	2	8	6	4	1		2	4		7	1	4			81	0.061
z		1							1			1	1	1								1			6	0.005
n	5	4	1	2	3	1		10	4	4	3	5	6	2	1		3	1		5	1				61	0.046
l	3	5	6	4	6	2	3	8	8	8	3	8	4	6	3	1	3	3		10	5	3			102	0.077
r	4	6	4	5	7	4	2	10	8	4	5	6	6	3	2	1	4	2		7	4	4			98	0.074
ʃ		4	2	3	2	1	1	7	5		1	5	6	5						7	1				50	0.038
ʒ																									0	0
tʃ	1	5	3	1	2			4	2	3	3	4	2	4			1			7	1				43	0.032
dʒ	4	2		1	4			4	1	3	1	4	3	2	1			1		3	4				38	0.029
j		1			1	1		2		1	2	5	4	2						2		1			22	0.017
k	4	8	4	4	5	1		9	7	3	1	10	10	3	1		4	3		6	2	1			86	0.065
g	2	2	2	2	2	2		8	7	4	3	4	8	2	3			2		1	2	2			58	0.044
ŋ																									0	0
w	1	2	3	4	2	2		6	8	1	2	7	7	3	2		2	2		6	2	1			63	0.048
h	3	7	3	4	6	1		9	8	3	6	3	9	4	2		3	1		7	3	2			84	0.063
Total	44	78	43	43	74	33	12	143	99	70	55	130	133	72	28	3	44	30	0	112	48	30	0	0	1324	

$$\textit{expected probability(unit1, unit2)} = \textit{probability(unit 1)} \times \textit{probability(unit 2)}$$

The expected probability can be converted to an expected frequency of occurrence in the corpus by multiplying the expected probability by the total frequency of units of that type in the corpus. In the example in Table 17.4.1, expected frequency for each combination would be the expected probability for the combination multiplied by the total number of combinations (1,324 in this data set of CVC words). Table 17.4.2 shows the expected frequency of combinations of onset and coda consonants in CVC words in English.

$$\textit{expected frequency(unit)} = \textit{expected probability(unit)} \times \textit{frequency(total)}$$

Comparing the observed frequency in Table 17.4.1 with the expected frequency in Table 17.4.2 provides insight into whether combinations are found more or less frequently than would be expected by random combination. A metric to quantify this relative likelihood of co-occurrence is to divide the observed frequency by the expected frequency. Ratios near 1 indicate random co-occurrence. Ratios much below 1 indicate a constraint against co-occurrence. Ratios well over 1 indicate a constraint in favor of co-occurrence. This ratio is referred to as O/E (observed/expected; Pierrehumbert 1993).

$$\textit{O/E} = \textit{frequency(unit)/expected frequency(unit)}$$

Phonological constraints aren't usually based on single combinations of individual segments. The O/E measure can be used over larger groups of units, such as natural classes. Consistently low or high O/E values over a natural class would suggest a phonological constraint. In addition, the O/E measure can be easily aggregated for groups of combinations. For groups, the ratio is computed for the sum of observed frequencies of occurrence of each combination divided by the sum of expected frequencies of each combination (see Pierrehumbert 1993; McCarthy 1994; Frisch et al. 2004). For example, it has been noted that repeated similar consonants in English (e.g. /p/ and /b/ as in words like *beep* and *pub*) are not found as frequently across a vowel in English as would be expected, while repeated identical consonants are found commonly (e.g. *bob*, *babe*, *boob*, *pop*, *pipe*, *pup*, etc.). However, Wilson and Obdeyn (2009) argue that O/E values should be considered with some caution and that studying frequency of occurrence directly might be preferable.

While the example given above provides an expected frequency/probability for a combination of two units (onset and coda), in principle any number of units can be combined, with their probability of occurrence at random determined by the product of the probabilities of all of the units involved. Using this approach, any novel non-word can be evaluated for its cumulative phonotactic probability (see Coleman and Pierrehumbert 1997; Frisch et al. 2000).

Table 17.4.2. Expected frequency of onset and coda consonants in English CVC words given the data in Table 17.4.1

Onset ↓	coda → b	p	f	v	m	θ	ð	t	d	s	z	n	l	r	ʃ	ʒ	tʃ	dʒ	j	k	g	ŋ	w	h	Total
b	3	5.4	3	3	5.1	2.3	0.8	9.8	6.8	4.8	3.8	8.9	9.1	4.9	1.9	0.2	3	2.1	0	7.7	3.3	2.1	0	0	91
p	3	5.2	2.9	2.9	5	2.2	0.8	9.6	6.7	4.7	3.7	8.7	8.9	4.8	1.9	0.2	3	2	0	7.5	3.2	2	0	0	89
f	2.2	3.8	2.1	2.1	3.6	1.6	0.6	7	4.9	3.4	2.7	6.4	6.5	3.5	1.4	0.1	2.2	1.5	0	5.5	2.4	1.5	0	0	65
v	0.8	1.5	0.8	0.8	1.4	0.6	0.2	2.7	1.9	1.3	1	2.5	2.5	1.4	0.5	0.1	0.8	0.6	0	2.1	0.9	0.6	0	0	25
m	2.5	4.4	2.4	2.4	4.2	1.9	0.7	8.1	5.6	4	3.1	7.4	7.5	4.1	1.6	0.2	2.5	1.7	0	6.3	2.7	1.7	0	0	75
θ	0.5	0.9	0.5	0.5	0.8	0.4	0.1	1.6	1.1	0.8	0.6	1.5	1.5	0.8	0.3	0	0.5	0.3	0	1.3	0.5	0.3	0	0	15
ð	0.3	0.6	0.3	0.3	0.6	0.2	0.1	1.1	0.7	0.5	0.4	1	1	0.5	0.2	0	0.3	0.2	0	0.8	0.4	0.2	0	0	10
t	2.6	4.7	2.6	2.6	4.4	2	0.7	8.5	5.9	4.2	3.3	7.8	7.9	4.3	1.7	0.2	2.6	1.8	0	6.7	2.9	1.8	0	0	79
d	2.8	4.9	2.7	2.7	4.6	2.1	0.8	9	6.2	4.4	3.4	8.1	8.3	4.5	1.8	0.2	2.8	1.9	0	7	3	1.9	0	0	83
s	2.7	4.8	2.6	2.6	4.5	2	0.7	8.7	6.1	4.3	3.4	8	8.1	4.4	1.7	0.2	2.7	1.8	0	6.9	2.9	1.8	0	0	81
z	0.2	0.4	0.2	0.2	0.3	0.1	0.1	0.6	0.4	0.3	0.2	0.6	0.6	0.3	0.1	0	0.2	0.1	0	0.5	0.2	0.1	0	0	6
n	2	3.6	2	2	3.4	1.5	0.6	6.6	4.6	3.2	2.5	6	6.1	3.3	1.3	0.1	2	1.4	0	5.2	2.2	1.4	0	0	61
l	3.4	6	3.3	3.3	5.7	2.5	0.9	11	7.6	5.4	4.2	10	10	5.5	2.2	0.2	3.4	2.3	0	8.6	3.7	2.3	0	0	102
r	3.3	5.8	3.2	3.2	5.5	2.4	0.9	11	7.3	5.2	4.1	9.6	9.8	5.3	2.1	0.2	3.3	2.2	0	8.3	3.6	2.2	0	0	98
ʃ	1.7	2.9	1.6	1.6	2.8	1.2	0.5	5.4	3.7	2.6	2.1	4.9	5	2.7	1.1	0.1	1.7	1.1	0	4.2	1.8	1.1	0	0	50
ʒ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tʃ	1.4	2.5	1.4	1.4	2.4	1.1	0.4	4.6	3.2	2.3	1.8	4.2	4.3	2.3	0.9	0.1	1.4	1	0	3.6	1.6	1	0	0	43
dʒ	1.3	2.2	1.2	1.2	2.1	0.9	0.3	4.1	2.8	2	1.6	3.7	3.8	2.1	0.8	0.1	1.3	0.9	0	3.2	1.4	0.9	0	0	38
j	0.7	1.3	0.7	0.7	1.2	0.5	0.2	2.4	1.6	1.2	0.9	2.2	2.2	1.2	0.5	0	0.7	0.5	0	1.9	0.8	0.5	0	0	22
k	2.9	5.1	2.8	2.8	4.8	2.1	0.8	9.3	6.4	4.5	3.6	8.4	8.6	4.7	1.8	0.2	2.9	1.9	0	7.3	3.1	1.9	0	0	86
g	1.9	3.4	1.9	1.9	3.2	1.4	0.5	6.3	4.3	3.1	2.4	5.7	5.8	3.2	1.2	0.1	1.9	1.3	0	4.9	2.1	1.3	0	0	58
ŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
w	2.1	3.7	2	2	3.5	1.6	0.6	6.8	4.7	3.3	2.6	6.2	6.3	3.4	1.3	0.1	2.1	1.4	0	5.3	2.3	1.4	0	0	63
h	2.8	4.9	2.7	2.7	4.7	2.1	0.8	9.1	6.3	4.4	3.5	8.2	8.4	4.6	1.8	0.2	2.8	1.9	0	7.1	3	1.9	0	0	84
Total	44	78	43	43	74	33	12	143	99	70	55	130	133	72	28	3	44	30	0	112	48	30	0	0	1324

17.4.2.2 *Neighborhood density*

An approach to frequency effects in phonology that is conceptually very different from the approach of generative phonology can be found in theories that advocate analogy or similarity to existing forms as the basis for phonological generalization (e.g. Skousen et al. 2002; Bybee 2006; Vihman and Croft 2007). These studies also use lexical data or corpora, but may examine larger phonological units (e.g. entire words) and consider degrees of match or mismatch between groups of units. While some of these studies have not been quantitatively formalized to the extent that is found in the phonotactics literature using frequency and probability, there are measures of lexical and segmental similarity that are applicable to these types of studies, such as neighborhood density.

Neighborhood density was originally conceptualized as a factor that would influence spoken word recognition (Luce 1986). Luce proposed the Neighborhood Activation Model, which used the probability choice rule and measures of segmental confusion to predict words that would be easier or harder to identify due to lexical confusion. The original Neighborhood Activation Model had a highly gradient measure of lexical similarity, where individual segments in a word would contribute probabilistically to word confusion based on segmental confusion. The notion of neighborhood density was abstracted somewhat for experimental convenience in later studies, and became a more categorical measure of lexical confusability based on the number of highly similar confusable words. In this abstraction, the neighborhood density for a word (usually CVC) is taken to be the size of the set of words than can be created from the target word by a single substitution, addition, or deletion (cf. Greenberg and Jenkins 1964). For example, the word *plant* has lexical neighbors such as *planned* (subtitution of /d/ for /t/), *pant* (deletion of /l/), and *planter* (addition of /ɚ/). The concept of neighborhood density has been extended to longer words by considering a word to be a neighbor if it shares 2/3 of the phonemes of the target word, though for words beyond two syllables, the number of neighbors is minimal (Frisch et al. 2000). Words with a relatively large number of neighbors are more difficult to identify (greater number of errors) than words with a smaller number of neighbors (Luce and Pisoni 1998). By contrast, words with a relatively large number of neighbors are easier to produce (faster reaction time) than words with a smaller number of neighbors (Vitevitch and Luce 1998).

Neighborhood density for English words can be found in the lexical database used above (Nusbaum et al. 1994). Neighborhood density can also be applied to novel phonological items (non-words) and has been examined as a predictor of well-formedness judgments for non-words. Frisch et al. (2000) found that neighborhood density was a good predictor of wordlikeness judgments for relatively short and high-probability non-words, but could not differentiate long or low-probability non-words, which basically had no neighbors. In a study specifically designed to compare phonotactic probability and neighborhood density, Bailey and

Hahn (2001) found a stronger effect of neighborhood density on wordlikeness judgments. However, this finding was not replicated by Shademan (2006) who found a stronger role for phonotactic probability when both probability and neighborhood density were potentially relevant. From a theoretical perspective, there is no reason to conclude that only one of frequency or neighborhood density is relevant in phonology. The variation in findings so far may indicate generalizations at a variety of phonological levels as predicted by a phonotactic grammar grounded in the lexicon.

17.4.2.3 *Similarity*

Neighborhood density is related to the general concept of similarity, being a particular conceptualization and metric of similarity of words to other words in the lexicon. Influences of similarity in phonological processing and metalinguistic phonological tasks have also been investigated at a segmental level. For example, there is a long history in the literature examining the influence of similarity on speech error rates (e.g. Fromkin 1971) and even a study directly comparing different metrics of similarity for their predictiveness (van den Broeke and Goldstein 1980). More recently, similarity has been examined as a gradient phonological influence on co-occurrence, promoting dissimilarity between segments by the Obligatory Contour Principle (OCP, see Pierrehumbert 1993).

Metrics of segmental similarity generally use phonological features, with shared features promoting segmental similarity and non-shared features promoting dissimilarity. The influence of features on similarity may be modulated by their contrastiveness (see Hansson 2001; Frisch et al. 2004). However, there are a limited number studies that have presented evidence in favor of detailed influences of similarity on phonology or phonological processing (e.g. MacEachern 1999; Frisch et al. 2004). A comparison of segmental similarity effects on a variety of basic phonological processing tasks found that gross categorization in terms of place, manner, voicing, and sonority provided equally good predictions to more detailed similarity metrics (Bailey 2005).

$$similarity = \frac{shared\ properties\ (unit1,\ unit2)}{shared\ properties\ (unit1,\ unit2) + non\text{-}shared\ properties\ (unit1,\ unit2)}$$

The metrics of similarity discussed so far take similarity to be some intrinsic property of a phonological unit, defined by the internal characteristics (e.g. segments or features) of the phonological unit. Another approach to similarity which has seen very limited application thus far is to consider similarity in behavior or distribution. In general, two units can be considered phonologically similar if they share the same phonological constraints. Applying this notion to frequency/probability-based phonological analysis means that two units are phonologically similar if they share the same probability distribution: in other words, if they have similar

transitional probabilities with other units or have similar patterns of co-occurrence (Frisch et al. 1995).

17.4.3 Findings on frequency and similarity

Most of the research investigating phonotactic patterns using laboratory phonology techniques has focused on the lexicon as a data source. However, studies of the details of speech production (as reflected in acoustics) have shown frequency as a factor influencing hyperarticulation/reduction (Bell et al. 2003; Zhao and Jurafsky 2009). High-frequency words have a tendency to be reduced in comparison to low-frequency words. Diachronically, this tendency could lead to elision in phonological representation and eventual historical change. It has been argued that phonological patterns more crucially depend on the type frequencies found in dictionary studies rather than the token frequencies provided by a speech corpus (e.g. Bybee 2001), but it is likely that each type of frequency and level of generalization is relevant to different aspects of phonological/phonetic knowledge.

17.4.3.1 *Phonotactic probability as a phonological baseline*

Greenberg's (1950) study of consonant co-occurrence in the Arabic roots was the early foundation for research using lexical corpora and phonological statistics to examine phonotactic patterns. Like the CVC example given above, Greenberg (1950) examined observed frequency versus expected probability for consonant pairs in Arabic triliteral roots, providing a quantitative description of place of articulation-based co-occurrence restrictions that were well-known among Arabic scholars.

Pierrehumbert (1994) examined triconsonantal clusters in English (e.g. /ŋkl/ in *inkling*) with the express purpose of examining the extent to which expected probability can predict occurrence frequency. Pierrehumbert examined the combination of rime with following onset units as a predictor of occurrence for medial clusters. She found that that expected probability does provide a good baseline for predicting phonological variation in phonotactics. Pierrehumbert (1994) had two additional findings. First, there were phonological regularities that could not be accounted for by probability alone, one of which was a segmental OCP pattern, where triconsonantal clusters with identical first and third consonants were avoided. In addition, it was noted that extremely low-probability combinations were not found, even though a few would be expected over a database that is the size of the English lexicon. Thus, it would appear that there may be some sort of threshold or cut-off based on expected probability that prohibits very low-probability forms from occurring at all (see Frisch 1996 for discussion).

Coleman and Pierrehumbert (1997) conducted an experimental study of word-likeness judgments for novel English non-words and found expected probability to be a good predictor of participants' ratings. They found that words containing a low-probability constituent (in this case, an unattested consonant cluster) received relatively low ratings, but that the rating was also influenced by the probability of all other (attested) constituents in the word. Frisch et al. (2000) replicated and extended this finding, demonstrating phonotactic probability effects for well-formedness judgments for English non-words in both a wordlikeness task (with a 1–7 rating scale) and an acceptability judgment task (acceptable/unacceptable). Bréa-Spahn (2009) demonstrated similar expected probability influences on well-formedness judgments for Spanish speakers.

17.4.3.2 *Syllable and word structure*

Phonotactic probability has been investigated as a cause of the development of structure in phonological representations. For example, Treiman et al. (2000) examined the relative distributions of consonants and vowels in English to determine whether onset consonant-vowel combinations (e.g. CV) or vowel-coda consonant combinations (e.g. VC) exhibited more statistical regularities in co-occurrence. They found that CV combinations occurred nearly at chance levels, while many VC combinations occurred at levels well above chance or well below chance. In other words, transitions from the onset consonant to the vowel are relatively random, while transitions from the vowel to coda consonant are more systematic. They interpreted the non-random combinations of vowels with coda consonants to be evidence in support of onset-rime structure in English syllables (i.e. C-VC). Lee and Goldrick (2008) replicated this finding for English, and also demonstrated that Korean contains more statistical regularities for CV than VC combinations, supporting body-coda structure (CV-C) for Korean (see also Côté, this volume).

Analogous frequency and probability effects have also been observed for English morphology. In a series of corpus and experimental studies, Hay and colleagues have found a probabilistic component to word structure and the parsing of word constituents (Hay et al. 2004). They have found that frequency of occurrence of an affix (also relative to the stem) is one of several influences on processing complexity. Processing complexity effects can be seen in the productivity of use of affixes, both in experimental studies with novel morphologically complex words (Hay 2002) and in corpus studies of complex word use (Hay and Baayen 2002). Affix processing complexity has also been shown to influence affix ordering in multi-affix words (Hay and Plag 2004). Analogous to the baseline effects of probability for phonotactic combinations, Hay and Plag (2004) found that processing complexity provides a baseline prediction for possible affix combinations. The attested set of

affix combinations is further restricted by phonological, syntactic, and semantic selectional restrictions for specific affixes that are overlaid on top of the complexity constraints. However, forms that are otherwise unrestricted by affix-specific constraints are not found if they violate complexity constraints.

Rather than treating the statistics as a by-product of preordained phonological structure, several researchers have proposed that such statistical regularities can be the motivation for a language learner to develop phonological or morphological structures (e.g. Pierrehumbert 2003; Hay et al. 2004; Bybee 2006; Lee and Goldrick 2008). The emerging picture, then, is one in which frequency of occurrence of individual constituents provides information about the expected probability of combinations of constituents. The language learner can posit grammatical constraints on co-occurrence when the observed frequency of combinations significantly deviates from expected probability. Low co-occurrence points toward a grammatical generalization against co-occurrence, and high co-occurrence points toward a grammatical generalization requiring co-occurrence. These generalizations can be specific to phonological content (e.g. a constraint requiring nasal-stop sequences to be homorganic) or to more generally indicate phonological structure (e.g. words, affixes, syllables, onsets, rimes, and codas as constituents).

17.4.3.3 *Similarity: OCP and harmony*

Much of the work developing frequency, probability, and co-occurrence began in the study of co-occurrence restrictions for consonants in the Arabic triliteral roots (Greenberg 1950). The initial analysis was a restriction against co-occurrence of consonants with shared place of articulation (McCarthy 1988), but exceptions and subclassifications involved manner features and secondary place features as well (McCarthy 1994; Padgett 1995). This led to the hypothesis that similarity within place classes determined the degree to which a combination was avoided, making the prediction of co-occurrence a gradient rather than categorical phenomenon (Pierrehumbert 1993). Subsequent studies found analogous consonant co-occurrence patterns for place of articulation in a variety of languages (see Frisch et al. 2004; Coetzee and Pater 2008 for summaries) and for obstruent laryngeal features (MacEachern 1999). In the two languages that have been studied most thoroughly, Arabic and Muna, it has been shown that many manner, voicing, and stricture features affect co-occurrence within place classes quantitatively, providing support for the hypothesis that co-occurrence restrictions are the result of similarity avoidance. It has also been shown that these similarity avoidance constraints influence well-formedness judgments for novel non-words (Frisch and Zawaydeh 2001) and speech error rates (Rose and King 2007). This indicates that these constraints are psychologically real, and not merely a lexical pattern reflecting some sort of historical residue that is not relevant synchronically. In comparing between

languages, different non-place features appear to affect co-occurrence differently and the basis for these differences is unknown (see Wilson and Obdeyn 2009 for additional discussion, and additional detailed analyses of the Shona and Wargamay languages).

Similarity has also been shown to play a role in harmony systems for both vowels and consonants. Conceptually, both vowel and consonant harmony processes promote similarity between segments.[3] It has also been shown that consonant harmony is more likely to be triggered when consonants are similar to one another (Hansson 2001; Rose and Walker 2004). Analogous triggering effects have been found for vowel harmony, where vowel harmony is more likely to occur when vowels are similar to one another either featurally (Hare 1990) or in their structural context or proximity (Ringen and Heinämäki 1999). All of these harmony effects are parallel to OCP dissimilarity effects, suggesting that the two should be analyzed in comparable ways. Cole (2009) demonstrates that the asymmetries in transitional probabilities between vowels in harmonizing versus non-harmonizing groups can be learned by a connectionist network, resulting in emergent harmony class generalizations across the lexicon. Such a network approach might also succeed in creating emergent generalizations of consonant harmony or dissimilation classes.

17.4.4 Future directions

A rich literature on the effects of frequency, neighborhood density, and similarity has been developed within the laboratory phonology paradigm (e.g. Vitevitch and Luce 1998; Munson 2001; Storkel et al. 2006; Goldrick and Larson 2008; among many others). A fundamental issue for research in this area is to better define and identify the origin of specific effects. In many cases, the factors of frequency, neighborhood density and similarity overlap or are interrelated. For example, for monosyllabic words, neighborhood density, and phonotactic probability are difficult to distinguish. Neighborhood density itself is also a specific instance of similarity/analogy at the word level. Future research should investigate whether some particular levels of generalization are primary, from which others bootstrap or can be seen as emergent or epiphenomenal. If a primary level can be found, then much of the generative paradigm can remain intact, and integrating frequency and probability effects would merely be a quantitative extension of the paradigm (e.g. Coetzee and Pater 2008). On the other hand, it may be that generalizations can emerge at any level where regularity arises, with no *a priori* constraints on where this may be

[3] Resolving the apparent contradiction between OCP constraints that promote dissimilarity and harmony constraints that promote similarity is beyond the scope of this section, but is an interesting area for future research.

(e.g. Cole 2009). Clearly, broader generalizations are descriptively more powerful, and thus cognitively more useful, and so the majority of robust phonotactic patterns are likely to be broad (Pierrehumbert 2003). In addition, historical change may have a tendency to push quantitative generalizations toward these presumably simpler categorical end-points. Differentiating the true nature of quantitative generalizations in phonotactics and the phonology of the lexicon more generally will require careful research using the variety of data sources available to laboratory phonologists.

CHAPTER 18

ARTICULATORY ANALYSIS AND ACOUSTIC MODELING

ON THE ACOUSTICS AND AERODYNAMICS OF FRICATIVES

CHRISTINE H. SHADLE

The contributions in this chapter describe theory and research methods for articulatory, acoustic, and aerodynamic analysis of speech. Iskarous critically reviews four theories of how articulatory and/or acoustic factors define the system of contrasts used in the world's languages. Davidson discusses technical and experiment design considerations in using ultrasound as a research tool in linguistics. Hanson discusses methods of investigating speech aerodynamics and laryngeal function. Shadle discusses acoustic and aerodynamic analysis of fricatives.

18.1 Articulatory to acoustic modeling*

Khalil Iskarous

18.1.1 Introduction

A phonological description of a language usually begins with a description of the system of contrasts, expressed as specifications of distinctive features (Trubetzkoy 1939; Jakobson et al. 1952; Chomsky and Halle 1968; Dresher 2009). This first step is quite important, since a great deal of the rest of the phonological description, especially the description of lexical phonology, phonotactics, and post-lexical or allophonic systems in the language, will depend on the contrastive system posited (e.g. McMahon et al. 1994). Moreover, several theoretical aspects of a phonological theory like markedness, feature geometry, underspecification, and the phonetics-phonology interface are highly dependent on the nature of the underlying contrastive system. However, since a phonological system begins with the system of contrasts, an understanding of the origin of systems of contrasts has to come from external explanatory sources, like the cognitive or phonetic domains. That

* This work was supported by NIH NIDCD grant 02717.

is, phonological theory itself uses the notions of contrast and natural class, but does not itself define the possible contrasts or possible natural classes. This section provides a critical review of four theories that attempt to deductively explain the physical sources of distinctive features and the contrastive systems of the world's languages: Quantal Theory (Stevens 1972, 1989; Stevens and Keyser 2010), Theory of Adaptive Dispersion (Lindblom 1986; Diehl 2008), Dispersion-Focalization Theory (Schwartz et al. 1997b), and the Distinctive Region Model (Mrayati et al. 1988; Carré and Mrayati 1990). For further discussion of phonological categories, see Chapter 9 this volume. Quantal Theory (QT) deduces discrete natural classes from articulatory-acoustic relations, even though acoustics and articulation by themselves are continuous variables. The Theory of Adaptive Dispersion (TAD) deduces segment inventories from simultaneous application of principles of acoustic contrast maximization and articulatory effort minimization. Dispersion-Focalization Theory (DFT) and the Distinctive Region Model (DRM) combine ideas from both QT and TAD, but provide new insights of their own on the phonetic basis of phonological contrast. Unfortunately, despite the deep claims made by these theories about the nature of phonological contrast, they have been rarely used in the development of phonological descriptions of languages or of phonological theories (exceptions include Stevens and Keyser 1989; Halle 1992; Clements and Hume 1995; Padgett and Tabain 2005; Clements and Ridouane 2006; Dogil 2007; and Benus and Gafos 2007). The goal here is not to give a detailed introduction to the theories, but to give enough background to allow more researchers interested in laboratory phonology to access the literature in which the theories are developed. It is especially hoped that greater use of these theories in the development of phonological theory will itself assist the growth of the four theories of possible natural classes discussed here, since notions of phonological contrast are themselves changing (Cohn 2006; Scobbie and Stuart-Smith 2008)—indeed the very notion that discreteness is a requirement for contrast is now in question. It therefore seems that developments of the theory of contrast in phonology and the theory of possible natural classes in the field of articulatory-acoustic relations would have to develop jointly.

18.1.2 Quantal Theory

The basic idea behind QT is that even though acoustic and articulatory variables can be measured as continuous variables, the relation between them can serve as a source of discreteness. Specifically, Stevens (1972, 1989) identifies some articulatory variables for which continuous increase of the magnitude of that articulatory variable would first yield little change in some acoustic variable. However, further increase in the magnitude of the articulatory variable may yield a discontinuity in the acoustic variable, followed again by little change in the acoustic variable.

Therefore quantitative change in the articulatory variable yields two distinct qualitatively different acoustic behaviors. The suggestion is that distinctive features arise out of this qualitative differentiation in acoustic behavior.

The relation between change in location of constrictions in an acoustic tube and the consequent change in formants is one of Stevens's (1972, 1989) examples of the discretization of the vocal tract through articulatory-acoustic relations. Johnson (1997a) provides an accessible background to the necessary acoustic theory. Consider a tube 16 cm long closed at the back end and open at the front. If a 2 cm constriction is passed from back to front, and the resonance cavities of the front cavity (quarter wavelength resonances) and the back cavity (half wavelength resonances) are measured, we obtain the nomogram in Figure 18.1.1. As the constriction moves forward (moving rightward through Figure 18.1.1), the front cavity becomes shorter and shorter, leading to monotonic increase of its resonances (gray). At the same time, the back cavity is becoming longer and its (non-zero) resonances monotonically fall (black). If attention is paid to the lowest resonance frequency in this nomogram (which is a model of F2, not F1), it can be seen that until the front cavity is about 4 cm long, that lowest resonance monotonically tracks the shortening front cavity. However as the constriction continues to move frontward, the lowest resonance starts to monotonically track the increasing length of the back cavity, not the front cavity. At this point, therefore, there is a switch in the cavity affiliation of the lowest resonance from front cavity to back cavity. Two basic properties arise from this cavity-affiliation switch: (1) a stable plateau in the

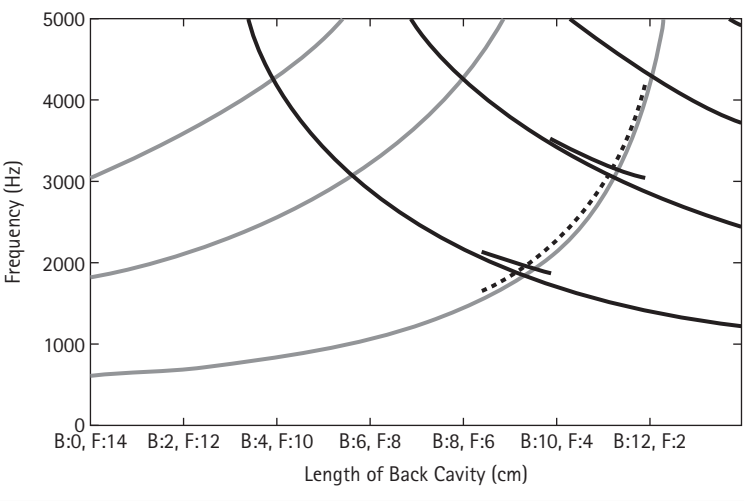

Figure 18.1.1. Nomogram for a 16 cm tube with a single constriction.

lowest resonance, (2) a focal point where two resonances come close together (seen when there is finite coupling between the front cavities, as in the dotted lines). The plateau is an acoustically stable region around which small changes in the position of the constriction induces little change in the resonance frequency. A constriction in the formant-cavity affiliation switch region would be a favored for a speech segment, according to this very simple model, since articulatory variability would be accompanied by acoustic invariability.

Stevens (1989) follows the reasoning presented in the previous paragraph to show that the vowels favored in the world's languages follow from stable regions of nomograms in 2-tube, 3-tube, and 4-tube models of the vocal tract for low vowels, non-low front vowels, and rounded vowels, respectively. That is, formant-cavity affiliation switch regions for different models are associated with different vowels in different regions in the vowel space. Stevens (1989) identifies several other quantal relations between articulation and acoustics and between acoustics and perception, which he argues are the source of various distinctive features used in phonological systems.

Stevens and Keyser (2010) argue that there are two types of quantal relations in speech, one associated with articulator-bound features and the other with articulator-free features (Halle 1992). Articulator-bound features refer to use of one of seven identified discrete articulator regions (e.g. [Round], [Distributed]), whereas articulator-free features define manner-type classes without reference to any particular articulator (e.g. [Sonorant], [Consonantal]). Quantal relations associated with articulator-bound features arise through motion of specific vocal tract articulators, which alters coupling between different vocal tract cavities, switching formant-cavity affiliations (as discussed earlier). Quantal relations associated with articulator-free features, on the other hand, are associated to aerodynamic-acoustic quantal relations that can be instigated by the motion of any articulator. Stevens and Keyser (2010) show that articulator-free quantal relations are a source of a temporal discretization in the speech signal around acoustic "landmarks." Briefly, vowels are associated with high energy in the low-frequency region, whereas obstruents are associated with low energy in that same region. Switching between these types of segments creates major landmark discontinuities in the speech signal that, they argue, are a main cue for changes in segment classes, and it is suggested that the generation of such landmarks is the motivation for their linguistic use by phonological systems.

The various versions of QT therefore propose several forms of discreteness which are proposed to play a role in the definition of possible natural classes of speech sounds: (1) discreteness in location or degree of constrictions around stable regions of articulatory-acoustic relations due to cavity coupling, (2) discreteness in time through changes in articulator-free features that introduce abrupt acoustic changes, (3) articulator discreteness for articulator-bound features.

To illustrate the use of QT (and articulatory-acoustic relations in general) to investigate phonological issues, two case studies will be briefly presented. Goldstein (1983) presented an articulatory-acoustic rationale for generalizations on vowel shifts described by Labov et al. (1972). The main generalizations in the latter work, which have been supported by later work (Labov 1994), are that vowels in the front and back series shift primarily in the vertical dimension, and additionally, back vowels can shift to front. Goldstein (1983) argued that natural speaker variability in speech production would lead to a distribution of possible constriction locations and degrees around each speaker's typical vowel constriction location and degree. However, due to articulatory-acoustic non-linearity, for the front non-low vowels, natural variability in location leads to little change in acoustics, whereas natural variability in degree leads to sizeable acoustic parameter variation. Articulatory-acoustic simulations showed that equal variability in constriction location and degree for the front vowels leads to little variability in F2, and a great deal more variability in F1. The argument is that natural variability in constriction location in the production of these vowels is not as acoustically apparent as is natural variability in constriction degree. Shifts occur primarily in the vertical dimension, since that is the dimension in which natural variability is communicated. This rationale is not available if acoustics and articulation are investigated by themselves. The key is their relation. Benus and Gafos (2007) and Gafos and Goldstein (this volume) use similar reasoning about non-low front vowels to derive a phonological pattern in vowel harmony systems. It is well known that some vowels can be transparent to backness harmony. Benus and Gafos (2007) argue that the transparent vowels are exactly the non-low front unround vowels that can articulatorily partially participate in the harmony process, but auditorily seem as if they were truly transparent. The reason is that for these vowels, partial participation in the harmony process leads to articulatory variability that is not acoustically registered, since these vowels are in a position where articulatory variability can be hidden. However, this argument crucially depends on the articulatory variability being in the position of the constriction and not in its degree.

18.1.3 Theory of Adaptive Dispersion (TAD)

Lindblom and Engstrand (1989) and Lindblom (2003) present a critique of QT which argues that the crucial phonetic property that needs to be modeled is not *stability*, but *contrast*. The maximization of contrast has a long history in phonology (Trubetzkoy 1939), and was given a numerical implementation to model the structure of vowel systems by Liljencrants and Lindblom (1972) and revised in Lindblom (1986). Crothers (1978) presented a set of seven implicational generalizations characterizing how vowel systems of various numbers of vowels are structured. These generalizations and later ones emerging from analysis of the UPSID database

(Schwartz et al. 1997a) relate the number of vowels in a system to qualitative features of systems, specifically (1) likelihood of peripheral vs. central vowels, (2) symmetric vs. asymmetric systems, (3) asymmetries between frontness and height as primary features. The Liljencrants and Lindblom (1972) model bases contrast maximization on a Euclidean metric measuring the distance between each two vowels in a mel-warped formant space, and choosing vowels that maximize a measure of the average distance between the vowels, in analogy to determining the positions of particles with mutually repulsive electric charge. This is a straightforward method of implementing the idea that segmental distributions should be chosen so that their acoustic behaviors are maximally distinct. Liljencrants and Lindblom (1972) and Crothers (1978) show that the theory accounts well for systems with three to six vowels, but is not as successful for larger systems. This is principally due to the prediction that there should be multiple non-peripheral high vowels, whereas most larger vowel systems do not contrast as many vowels as predicted by the model in this region. The model was revised by Crothers (1978) by implementing vowels not as points in a space but as finite regions. Lindblom (1986) and Lindblom (1990) revised it by replacing the formant parameterization of vowels by a spectral parametrization and by replacing the notion of *maximal* contrast with *sufficient* contrast, which factors in articulatory economy constraints. The revised system was shown to provide a better fit to larger vowel inventories and was then extended to predicting consonantal systems by Lindblom et al. (1984). However some empirical studies of closely similar language communities with different vowel systems have failed to confirm the predictions of TAD. Recasens and Espinosa (2006, 2009) studied eight dialects and subdialects of Catalan. When they compared systems with different numbers of vowels, they did find certain cases where more vowels in a system lead to greater dispersion in support of TAD, but they also found cases where increasing the number of vowels does not lead to any greater dispersion (even when the system is not overly crowded), which would not be predicted by TAD. Moreover, when they compared systems with the same number of vowels, they found different distances between vowels in different systems, against the predictions of TAD. More studies of such closely related languages are therefore necessary to determine the role of dispersion in determining the structure of segmental systems. As an example of a phonological study using TAD, Padgett (2003) argued that the alternation in Russian between the high front and centralized vowels before non-palatalized consonants is motivated by a diachronic constraint favoring the maximization of contrast between the two vowels. This reasoning can be used to motivate many historical changes; however, the problem is that neutralization is also quite prevalent in diachrony. Greater development of the theory is necessary to determine if it is possible to predict in which particular sound changes contrast maximization is available and when neutralization of various degrees is likely.

An important aspect of the theory that needs further exploration is the source of explanation. Liljencrants and Lindblom (1972) begin their exposition

by emphasizing that contrast maximization works with a predefined notion of what a possible vowel is and assumes the shape of the acoustic space, including asymmetries in height and backness. The notion of possible vowel and the shape of the space that is input to the contrast maximization procedure emerge from the structure of the articulatory model used to generate the acoustic space (Lindblom and Sundberg 1971). It is not clear how much of the predicted structure of vowel spaces, in Liljencrants and Lindblom (1972) as well as many further works based on TAD, is actually due to perceptual contrast maximization and how much is due to the structure of the assumed underlying articulatory model. This is especially the case in light of the study of dialects of the same language, such as Catalan, that have the same number of vowels, but different structures. The differences in inter-vowel distances in such dialects are evidence that there are factors other than the size of the system which determine the contrasts in it. The theory does allow economy of effort to interact with acoustic contrast maximization; however it is not clear how or why this principle can be invoked to derive the intricate facts of the dialects of a language like Catalan. Adding economy of effort to acoustic contrast maximization does indeed make the theory more powerful, but perhaps too powerful, especially when it is not clear what articulatory ease refers to (Pouplier 2003). Another critique that must be mentioned before TAD is used as a phonetic explanation of the phonological inventory is that it is not clear how phonemic vs. phonetic the various databases of vowel inventories are. This critique was made in an important paper by Lass (1984). If the labeling of vowels is already phonemic, then the distance between vowels, a quantity that is crucial to any theory that posits dispersion to be a crucial factor, has little meaning.

18.1.4 Dispersion-Focalization Theory

DFT combines two ideas: focalization *and* contrast maximization. Focalization is a property that emerges from acoustic model nomograms, as in Figure 18.1.1, and refers to points where constriction placement results in formants being close to each other (focal points). Based on work on the Center of Gravity effect in vowel perception by Chistovich et al. (1979), Escudier et al. (1985), and Schwartz and Escudier (1989) showed that focal vowels are auditorily more salient than non-focal vowels. Moreover Badin et al. (1991) showed that not all focal points are stable (in the QT sense). The DFT algorithm predicts vowel systems of different sizes by minimizing the sum of two energy terms: a dispersion term and a focalization term. Systems with well-dispersed focal vowels are optimal. Schwartz et al. (1997b) see DFT as an almost purely perceptual theory in contrast to QT and TAD. Schwartz et al. (1997a,b) discuss several generalizations about vowel spaces arising from an analysis of the UPSID database (Maddieson 1984), which TAD does not account for: (1) schwa can occur as the only non-peripheral vowel, (2) overprediction of high

non-peripheral vowels, (3) front rounded vowels are preferred to back unrounded vowels in non-peripheral systems, (4) symmetry in number between front and back vowels. Schwartz et al. (1997b) show that DFT accounts for most of these problems. It is not clear, however, how DFT would account for the difference in the vowel system structures of Catalan dialects investigated by Recasens and Espinosa (2006, 2009), especially for the fact that systems with same number of vowels show different inter-vowel distances.

An especially important contribution of DFT from a phonological point of view is that it addresses the introduction of secondary vowel contrasts (Schwartz et al. 1997a). The theory also distinguishes between independent and non-independent secondary contrasts. Independent secondary contrasts are secondary contrasts that do not interact with the primary vowel contrasts, while non-independent contrasts are secondary contrasts that affect the primary ones. Nasality, for instance, is an independent secondary contrast, since a nasalized vowel will have approximately same height and frontness as a non-nasalized one, whereas vowel length is non-independent, since a shorter vowel will often have a different quality than a longer one. Schwartz et al. (1997a) show that beyond about nine vowels, a system will become partitioned into a primary and a secondary system. Moreover, secondary systems, just like primary ones, tend to be symmetric and concentrated in the periphery. This work may have implications for more complex views of phonological contrast, especially the hierarchical view of contrast investigated by Dresher (2009).

18.1.5 Distinctive Region Model

The basic insight behind DRM (Mrayati et al. 1988; Carré et al. 1990) is that different regions of the vocal tract have uniform acoustic behavior. That is, formation or release of a constriction in a region raises some formants and lowers others. If only F1, F2, and F3 are considered, the vocal tract can be discretized into eight such regions, each with a distinctive acoustic behavior. Similarly, it is shown through acoustic arguments that there are three discrete modes of degree of constriction. Within each mode, the acoustic results of increasing the size of a constriction are qualitatively uniform and acoustically distinguish that mode from the other modes. It is then argued that region and mode discretizations of the vocal tract, based on distinctive acoustic behaviors, are the basis on which linguistic contrast builds and is the source for vocalic and consonantal contrasts in the world's languages.

The results of DRM are based on how formation and release of constrictions affect the formants at different locations within the vocal tract and DRM is a development based on the Perturbation Theory of Chiba and Kajiyama (1941). This theory calculates how each formant will be raised or lowered if a small constriction

is introduced, as a function of the location of that constriction, in a tube closed at one end and open at the other, a first approximation to a schwa.[1] The resulting functions relating constriction location and formant change are termed the Sensitivity Functions, since they express how acoustically sensitive each portion of the vocal tract is to the formation or release of a constriction.

The lower part of Figure 18.1.2 shows, schematically, the sensitivity functions for F1, F2, and F3 for a 17.5 cm acoustic tube, representing a generic schwa, indicated with a bold outline of the tube in Figure 18.1.2 (the cross-sectional area is 5 cm^2—the vertical axis scale refers to the value of the sensitivity functions, not the area function). A positive sensitivity value at a particular location means that a formant increases if the area function *increases*. For instance, F1 decreases due to an enlargement of the vocal tract (release of a constriction) at the glottis (left) and increases due to an enlargement at the lips (right). DRM arises out of the observation that each formant divides the vocal tract into a number of regions in which constriction anywhere in that region has the same *qualitative* effect on the formant. Within a region, it is possible for one formant to be positive and another point to be negative, but the crucial point leading to discretization is that within a region, each formant has uniform behavior, almost always positive or almost always negative, but not both. Therefore F1 divides the vocal tract into two halves: in the back half, constriction release lowers it, while release in the front half raises it. F2 divides the vocal tract into four regions of different lengths and F3 divides it into eight regions of different lengths.

The key insight is that even though the sensitivity functions are continuously varying functions of constriction position, the zero crossings of these functions define qualitatively distinct behaviors—in each region, certain formants raise and others lower. If only F1-F3 are considered, then the vocal tract is discretized into eight regions, each with a distinct qualitative behavior. The regions are divided by vertical lines in Figure 18.1.2, and the "distinctive" behavior of each discrete region is indicated with "+" or "−" for each formant at the top of each region in Figure 18.1.2. The name of each region is placed at the top of each region of the figure. The bars on regions in the posterior half of the vocal tract indicate that their acoustic behavior is opposite to the behavior of the corresponding regions in the front, as a result of the antisymmetry of the closed-open model. If a constriction release in one region is coupled with a constriction formation in a region with opposite behavior, the effect on formants is doubled, increasing potential for contrastiveness. Carré and Mrayati (1990) term this synergistic action.

At the middle of each region is a point, the constriction of which produces maximally distinct acoustic behaviors, when all three formants are considered and

[1] The vocal tract is at first assumed to be rigid, straight, lossless, but some of these assumptions can be relaxed as shown by Mrayati et al. (1988). They also show that the theory is applicable to perturbations of vowels other than schwa.

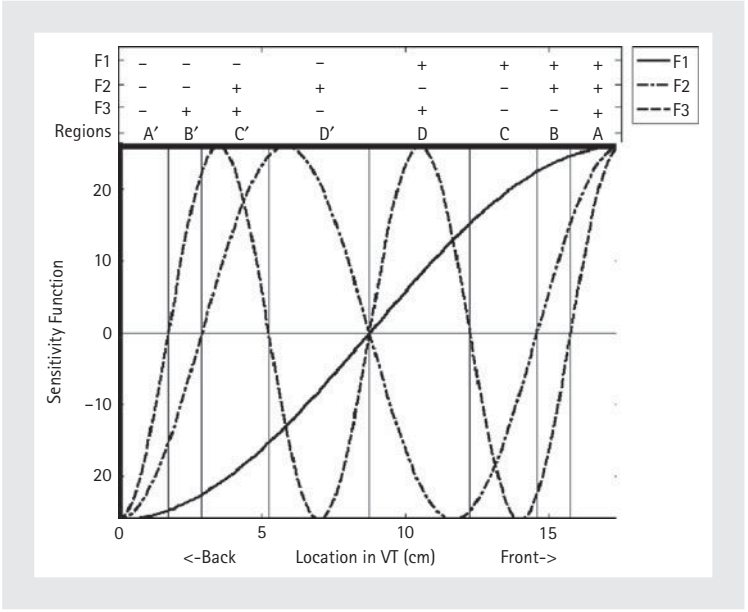

Figure 18.1.2. Discretization of the vocal tract based on distinctive formant behavior.

all regions are compared to each other. This is the basis on which Carré and Mrayati (1990) refer to the region formant behaviors as being pseudo-orthogonal. At the borders between regions, on the other hand, formant behavior is least contrastive amongst the regions. These regions are termed the *jittering* zones. A key idea of DRM is that linguistic contrast develops from forming and releasing constrictions at midpoints of regions to yield acoustic behavior that distinguishes that region from others, and to avoid the jittering zones, where constriction and release would yield acoustic results that are ambiguous about which region was constricted.

Mrayati et al. (1988) show that the sensitivity functions are valid as long as the constriction size is between about .8 cm^2 and 14 cm^2, if the neutral tract is 5 cm^2. This is termed the One Tract Mode (OTM), since different parts of the vocal tract are acoustically highly coupled and act as parts of one tract. If the constriction area is between 0 and .05 cm^2, the back and front cavities are practically uncoupled, yielding acoustic behavior that is qualitatively distinct from the OTM, which is termed the Two Tract Mode (TTM). If the constriction area is between about .05 cm^2 and .8 cm^2, a third mode of acoustic behavior is seen, and this mode is termed the Transitional Mode (TM). Both TTM and TM show a qualitatively different discretization of the vocal tract than OTM. Table 18.1.1 shows the discretization pattern that they share (based on Figure 17 of Carré and Mrayati 1990 and Figure 9 in Mrayati et al. 1988). Boxes where the TTM and TM behaviors are different from

Table 18.1.1. Vocal tract discretization based on mode behavior

Regions	A′	B′	C′	D′	D	C	B	A
F1	–	+	+	+	+	+	+	+
F2	–	+	+	+	+	+	+	+
F3	–	+	+	–	+	–	+	+

OTM are shaded. This region behavior is shared between TTM and TM; however, the two qualitatively differ in that the jitter zones are highly stable within TTM, but unstable in TM (see Mrayati et al. 1988).

The DRM theory of linguistic contrast is based on the pseudo-orthogonality of the discrete regions and on a dynamic articulatory-acoustic principle: segments are preferred that allow for the least motion from one segment to another, while maximizing contrast. For consonants, the main prediction is that the constriction location of each consonant will be in one of the discrete regions derived. Labials are articulated in region A, alveolars in region C, velars in D, uvulars in D′, and pharyngeals in C′. Complex consonants are then postulated to be articulated in multiple regions. The theory for vowels is more complex, since it involves a novel notion of minimal effort in switching between segments. This principle is different from the minimal articulatory effort principle invoked in TAD. The principle invoked by Carré et al. (1995) and Carré (2004) is rather that vowel constriction location and degree are chosen by languages to allow minimum switching effort between vowels. This principle invokes the ease of VV transitions, even though vowel hiatus is a relatively rare phenomenon. However, Öhman (1966) has argued that VCV transitions are fundamentally VV transitions with a superimposed consonantal perturbation, making inferences based on VV switching applicable to VCV switching. The DRM theory is partially supported by Iskarous (2005b), which showed that transitions between two lingual segments occur in a highly organized way, regardless of the segments in the transition. Specifically, when the constriction locations of two segments in a transition are not contiguous, the vocal tract is discretized into two locations, where area function change is concentrated, separated by a functional pivot. However, when the constriction locations are contiguous, the change in area function is not discretized.

18.1.6 Comparison of theories

The main difference between the theories is in how they particulate or discretize phonological space (Studdert-Kennedy and Goldstein 2003). QT proposes a spatial discretization based on stability, a temporal discretization based on variation

of articulator-free features, and an articulator discretization based on vocal tract physiology. DRM proposes a spatial discretization based on pseudo-orthogonality of the acoustic behavior of different vocal tract regions and an efficiency criterion for VV switching, while DFT proposes articulatory-acoustic focalization and its consequent perceptual stability as a basis of differentiation and distinctiveness. TAD assumes discretization, but rather focuses on the modeling of contrast maximization, without explicitly differentiating between articulatory factors that shape a vowel space (e.g. asymmetry of tongue back and front and difference in the anatomical boundary conditions at the top and bottom of the vowel space), from the perceptual factors.

Each of the theories uses its basic principle(s) to explain some aspect of segmental systems, but it is not clear how easily extendible each is to other aspects of contrast. QT focuses on proving stability of certain features based on assumed models. However, there are many possible models that would also have stable points, but whose stable vowels are not chosen by languages of the world as vowels or consonants. For instance, Stevens (1989) proposes three different types of models (2-tube, 3-tube, and 4-tube models) for three regions of the vowel space. But there are many more models that vary minimally from these three models obtainable by changing constriction parameters, number of cavities, shape of tube parameters, etc., each of which would have its own stable focal points. If each such model were a possible model for a different part of the vowel space, then there would be many more vowels than the assumed thirty-seven vowels. TAD accomplishes a great deal through perceptual contrast minimization and does assume minimization of articulatory effort, but does not ascribe any explanatory role for the underlying articulatory model and its structure (differentiation between organs, jaw-tongue synergy, etc.), except that articulatory effort needs to be minimized. Articulation involves a great deal of articulator dependency and other forms of structure that could be invoked explanatorily. Even if certain aspects of segmental inventories can be explained by perceptual principles, it cannot be concluded that it must be the perceptual principles that give rise to the contrasts; it should also have to be shown that articulatory principles and articulatory-acoustic principles are unable to motivate the same generalizations, since it is possible that perceptual effects follow from articulatory ones. DFT and DRM improve a great deal on QT and TAD and provide their own insight into the systems, but it is not clear if DFT can also provide a rationale for consonant systems or if the discretization insights that DRM brings to consonant systems apply to vocalic systems.

We believe that each of these theories provides some deep insight into the origin of phonological contrasts. Further development of each of the theories and new theories built upon their insights, in combination with new insight into how phonological contrast actually functions, promises to provide a new basis for understanding of phonological contrast, and how the physical basis of contrast delimits possible generalizations in lexical and post-lexical phonology.

18.2 Ultrasound as a tool for speech research

Lisa Davidson

18.2.1 Introduction

Ultrasound imaging in speech research is becoming a popular tool for investigating a wide range of phonetic, phonological, and sociolinguistic questions. The appeal of ultrasound is that it is relatively inexpensive, non-invasive, and often portable (compared to electromagnetic midsagittal articulography (EMMA), magnetic resonance imaging (MRI), or electropalatography (EPG), for example). This makes ultrasound suitable not only for traditional laboratory research, but also for fieldwork with less-studied populations (Gick 2002; Gick et al. 2005; Miller 2008), sociolinguistic research that aims to elicit vernacular speech and reproduce more naturalistic conditions (Lawson et al. 2008), and studies with small children and clinical populations (Bernhardt et al. 2003; Bressmann et al. 2005a, 2005b; Davidson et al. 2007; Ménard et al. 2007). In this section, a brief overview of considerations that are of particular interest to laboratory phonologists is presented. An in-depth explanation of the technical details of ultrasound in speech research can be found in Stone (2005).

As an articulatory technique, the main advantage of ultrasound is that the whole tongue can be imaged. Typically, it is possible to obtain a midsagittal image that extends from the tongue blade to the beginning of the tongue root, since the air in the sublingual space and/or the jaw tends to prevent the ultrasound machine from adequately imaging the tip, and the hyoid bone can block the posterior portion of the tongue root. However, depending on the specifics of both the ultrasound system and the speaker's anatomy, it is sometimes possible for the image to also include a larger portion of the tongue tip, allowing researchers to examine changes in the tip, blade, body, and root simultaneously. Other technologies, such as EMMA, are good techniques for tracking individual points on the tongue, but may not always allow for the reconstruction of the whole tongue surface. EPG is appropriate for examining tongue-palate contact, but does not provide information about tongue shape. MRI provides excellent spatial resolution of the whole tongue (and vocal tract anatomy in general), but the time resolution is not as good as ultrasound and the MRI machine is expensive, stationary, and does not allow for good simultaneous audio recording because of the noise of the machine. The real-time MRI technique developed by Narayanan and Byrd and colleagues has a frame rate of 21 frames per second (Bresch et al. 2008; Byrd et al. 2009), which is slightly less than ultrasound, but it has the other disadvantages of conventional MRI as just noted. Thus, for

researchers who have questions that can be answered by examining whole tongue shapes, ultrasound is a practical and convenient tool.

Examples of two ultrasound images are presented in Figure 18.2.1. The white line representing the midsagittal image of the tongue is created as a result of the transducer of the ultrasound machine emitting high-frequency sound that travels toward the tongue. When the sound wave hits the tongue, it sends an echo back to the transducer, which reconstructs the density change between the tongue surface and the air above it. In each of these images, the tongue tip is on the right. Figure 18.2.1a illustrates an adult Russian speaker producing the closure portion of palatalized [tʲ]. Using the acoustic recording, the frame corresponding to the midpoint of the closure was selected. Although there is no palate information in this image, it can be assumed that the tongue tip is in contact with the alveolar ridge, and there is a raising of the tongue body that would be expected for a palatalized consonant (cf. Kedrova et al. 2008). In Figure 18.2.1b, a 7-year-old American English speaker is producing a bunched articulation of [ɝ] in the nonsense sequence [ɝd]. Both a coronal constriction and a tongue root retraction can be clearly seen in this image (cf. Tiede et al. 2004).

Despite the considerable advantages of ultrasound for articulatory data collection, there are also drawbacks that all researchers should be aware of. First, the time resolution of ultrasound imaging is typically lower than it is for methodologies like EMMA or EPG. For researchers wishing to stream directly from the ultrasound machine to a video camera, VHS, or a capture card on a computer, the NTSC or PAL recording standards impose a frame rate of 29.97 or 25 frames per second, respectively. Some ultrasound machines allow a faster internal frame rate, but this can only be captured using the hard drive of the ultrasound machine, which has two main limitations: the length of the video is constrained by the available space on the

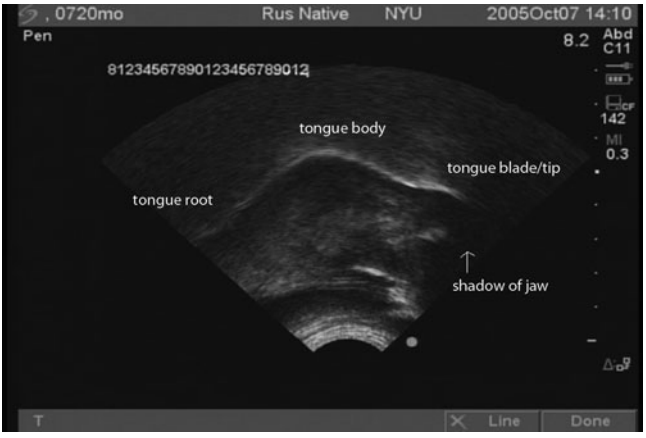

Figure 18.2.1a. The closure portion of /tʲ/ for an adult Russian speaker.

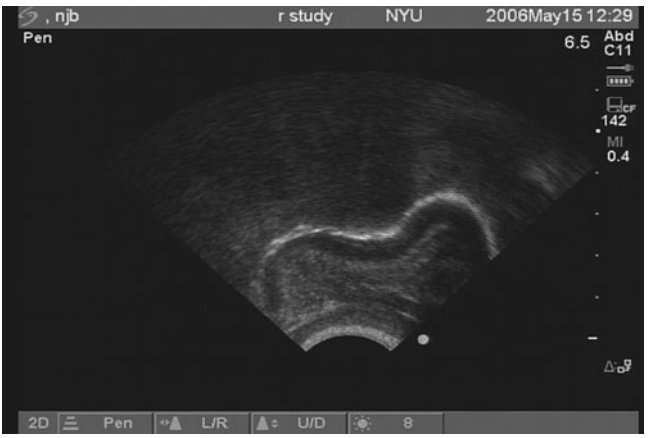

Figure 18.2.1b. A bunched /r/ for a normally developing 7-year-old American English speaker.

hard drive or the flash card, and if researchers wish to align audio to the video, it must be done post hoc. Audio-visual synchrony is not a trivial issue, since there is not usually an identifiable articulatory event that can be used as a reliable marker for aligning the audio. However, researchers have developed some techniques for audio-video alignment after data collection, either by simultaneously aligning multiple articulatory events to their presumed acoustic outcomes (Miller 2008), or by using a "Brightup" unit that produces both a flash on the video screen and a pulse in the audio that can later be matched up (Miller and Finch 2011; Wrench and Scobbie 2008).

Another concern when interpreting ultrasound images is the presence of scanning artifacts in the image. There are two main sources of artifacts. The first is called scan lag, and it is due to the fact that the crystals in the transducer fire sequentially, not in parallel. Consequently, when the echoes are reconstructed into the image displayed on the ultrasound machine and captured to video, the line being imaged will not contain information collected simultaneously along the length of the curve. That is, in any given frame, the front of the curve may represent a tongue shape that occurred up to 24 ms before the shape taken by the back of the tongue. During relatively long sounds like vowels, scan lag is not problematic, but an image that corresponds to a shorter sound can actually be a composite of the tongue shapes for more than one sound. The second type of artifact occurs when streaming the output of the ultrasound machine to a video capture card on a computer, which can result in mismatches between the video standard and the ultrasound's internal frame rate. Wrench and Scobbie (2006) provide a technical explanation of the problem, and show that frames may contain information from more than one sweep of the transducer. This is illustrated by the scanning discontinuity shown in Figure 18.2.2. One potential solution to this

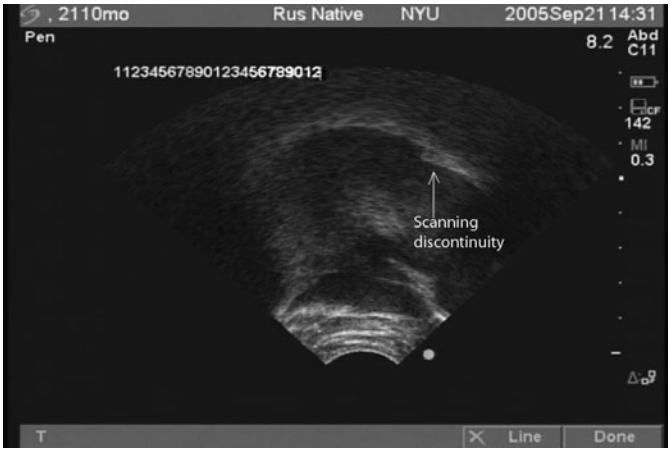

Figure 18.2.2. A scanning artifact in which two different scans of the ultrasound transducer are combined into one image during the production of /n/ by a Russian speaker.

problem is to use an ultrasound machine that saves "cineloop" data, which are frames containing data from only a single sweep of the ultrasound transducer, directly to the hard drive of the ultrasound machine (Wrench and Scobbie 2008). Another way to address occasional artifacts is to collect many repetitions of the data so that they can either be averaged, or so that an individual token can be discarded without jeopardizing the number of data points available.

Ultimately, while researchers must remain vigilant about a number of technical issues, ultrasound is still a non-invasive and relatively quick method for collecting articulatory data. In the next section, further considerations for data collection and analysis are reviewed.

18.2.2 Methodological considerations in ultrasound data collection and analysis

18.2.2.1 *Head and transducer stabilization*

In order to carry out quantitative analyses of ultrasound data, the transducer must be kept in the same position relative to the tongue at all times (see Figure 18.2.3). Stabilization ensures that tongue shapes can be compared across trials for the purposes of phonetic and phonological research. Certain types of movements, such as rotation or translation of the transducer, can be corrected as long as the transducer stays in the same imaging plane throughout data collection. That is, transducer movement can be rectified if it tilts or slides forward or backward, but if the transducer moves in a side-to-side motion, then it is no longer imaging the

Figure 18.2.3. The head and transducer stabilization set-up at New York University. The ultrasound transducer is restrained with a Magic Arm and the head is held with a moldable head stabilizer.

same slice of tongue tissue. The data from this type of error cannot be salvaged. Researchers have developed two main techniques to avoid such movement: head and transducer immobilization, or a headset system that affixes the transducer to the jaw while the head is allowed to move more freely.

A number of head and transducer immobilization systems have been developed, ranging from those that provide fine control of the set-up to those that are "low-tech" and portable. The Head and Transducer Support system (HATS) designed at the University of Maryland consists of a metal ring with adjustable padded screws that can accommodate any head size and a metal support for the transducer that is controlled by a joystick to ensure accurate placement (Stone and Davis 1995). The Comfortable Head Anchor for Sonographic Examinations (CHASE) developed at the University of Toronto was designed after an ophthalmic chair, where the participant leans his head into a forehead rest and the transducer is affixed under the chin using a specially-designed holder (see <http://www.slp.utoronto.ca/aboutus/rlabs/vrlab/vrhome/Research.htm>, accessed 7/21/2011). An inexpensive but effective system using a moldable head stabilizer (Comfort Company, intended for people with low head-and-neck tone) and a Magic Arm (Manfrotto by Bogen Imaging) for the transducer is currently in use at New York University (NYU) and the University of Rochester (see Figure 18.2.3). The consistency of the stabilization was tested at NYU. It was found that once speakers settled into a comfortable position during the first block of trials, vertical and horizontal head movement was at most 2 mm (Davidson and De Decker 2005).

One drawback of immobilizing the head and transducer is that typically, speakers move their heads while talking, so preventing them from doing so may affect the

naturalness of their speech. In addition, affixing the transducer under the jaw may restrict natural jaw movement during speech to some extent. Some researchers have addressed this issue by using a headset system that allows the transducer to move with the jaw. Articulate Assistant is a headset for use with ultrasound imaging which has been commercially developed by Articulate Instruments and tested at Queen Margaret University (Scobbie and Lawson 2008). Another method that does not require head and transducer stabilization is the HOCUS system developed at Haskins Laboratories (Whalen et al. 2005). HOCUS is agnostic as to how the transducer should be affixed to ensure that the same slice of tongue is imaged continuously (Haskins researchers have held the transducer by hand or with elastic bands), but any translation or rotation movement can be corrected for by using Optotrak to track infrared emitting diodes (IREDs) placed in three locations: the transducer, a pair of goggles worn by the speaker, and the articulators. A similar system called Palatron developed at the University of Arizona is a low-tech, portable alternative that uses sticks attached to goggles and to the transducer with dots that can be videotaped and later superimposed onto the ultrasound image to correct for any unwanted movement (Mielke et al. 2005). Miller (2008, see also Miller and Finch 2011) combined both the Articulate Assistant helmet and the Palatron alignment technique when collecting data on clicks in IsiXhosa.

Another advantage of stabilization is that it allows a researcher to collect a tracing of the palate which can later be superimposed on images containing the tongue contour. This is useful for research that is enhanced by information about tongue-palate contact. The palate cannot normally be seen in an ultrasound image because the ultrasound beam reflects off the tongue-air boundary, which prevents the beam from reaching the palate. However, when the vocal tract is filled with water or some other substance, the beam can image the boundary between the liquid and palatine bone. Epstein and Stone (2005) explain techniques for collecting palate data by imaging participants during various kinds of swallows (e.g. wet, dry, different size boluses). Traces of the palate images can be extracted using the same software used to extract the tongue edge (see Section 18.2.2.2) and then combined with tongue images to examine tongue-palate contact patterns.

18.2.2.2 *Edge extraction and statistical analysis*

As a precursor to either quantitative or qualitative analysis of ultrasound tongue images, many researchers begin by extracting the tongue contours into a set of numerical values that can be plotted and submitted to statistical analysis. A number of labs have developed semi-automatic tracking procedures that detect the boundary between the surface of the tongue and the air using snakes, an edge-detection technique in which a curve changes shape over time until it determines the best edge in an image (Iskarous 2005b). This is the principle behind ultrasound analysis software packages like EdgeTrak (University of Maryland, Li et al. 2005)

and Ultra-CATS (University of Toronto, Bressmann et al. 2005b). Articulate Assistant (Articulate Instruments), which has also been developed for tracking tongue edges in ultrasound, uses a different algorithm. In most software packages, the user specifies a few points on the ultrasound image corresponding to the tongue surface, and the program then employs these points to further define the tongue curve. Programs that have semi-automatic tracking capabilities use the information provided for the first tongue curve to track the shapes of subsequent frames, making it easier for researchers to process larger data sets. An example frame from EdgeTrak is shown in Figure 18.2.4, demonstrating how the edge can be tracked in order to be extracted in a numerical format.

Once the tongue edge is extracted, it can be analyzed in a variety of ways. One of the simplest techniques is to use a fixed line for measurement or a fan overlay. These are transducer-centric measures which allow the researcher to superimpose a line or a grid on top of the ultrasound image and then measure changes along fixed lines radiating out from the transducer, as illustrated in Figure 18.2.5 (Bressmann et al. 2005; Gick, Pulleyblank, Campbell, and Mutaka 2006; Benus and Gafos 2007; Vasquez-Alvarez and Hewlett 2007; Bressmann 2008). It should be noted that this method does not allow for tracking fleshpoints, but as long as the head and transducer remain stabilized, conclusions about how the tongue passes through a point of interest in the oral cavity can be made.

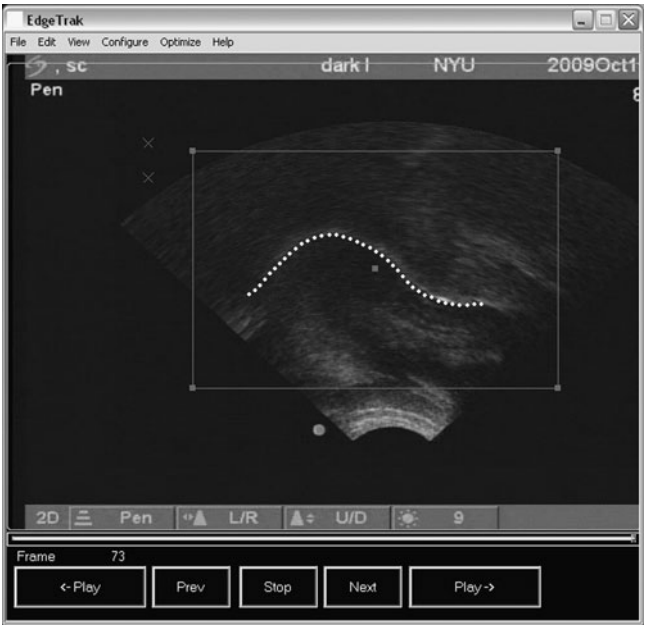

Figure 18.2.4. An EdgeTrak screenshot of an English speaker producing dark /ɫ/.

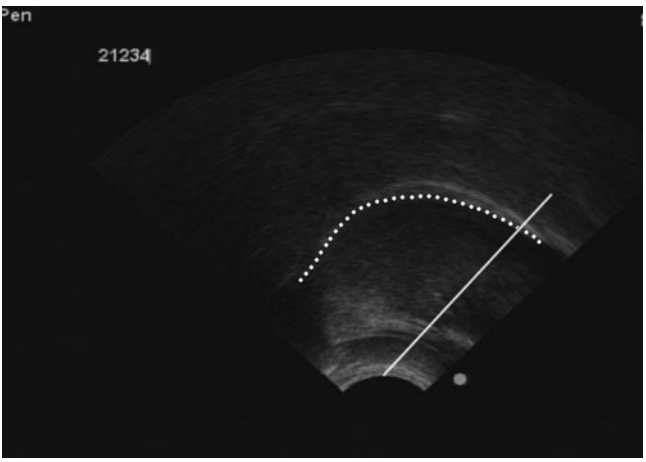

Figure 18.2.5. An example of a measurement along a fixed line for the tongue tip position of /n/ produced by a Russian speaker.

One limitation of measurement along fixed lines is that it only provides values for a discrete number of points and does not necessarily take advantage of the whole tongue curve. Two methods that do operate over the whole tongue curve are root mean squared (RMS) difference and smoothing spline analysis of variance (SS ANOVA). To obtain RMS differences, first squared distances are calculated between specified points on two tongue curves, then those squared values are averaged, and then the square root of the average is taken (Stone 2005; Buchwald et al. 2007; Davidson 2007a). RMS differences can be submitted to statistical tests such as ANOVAs, t-tests, or sign tests. This value provides a global difference measure to indicate how different two tongue curves are (see also Zharkova and Hewlett 2009 for a similar technique using nearest neighbor calculations).

The SS ANOVA method is a more holistic method for comparing tongue curve data, but it also allows researchers to divide the tongue quasi-arbitrarily into articulators such as tongue blade, body, and root (Davidson 2006a). For the SS ANOVA, multiple repetitions of tongue curves for the linguistic element being investigated are first fit by a smoothing spline (e.g. Wahba 1990). Subsequently, an analysis of variance in which each component of the model is estimated with a smoothing spline is carried out (e.g. Gu 2002). The SS ANOVA does not return an F value; instead, the smoothing parameters of the components of the equation are compared to determine their relative contributions. To determine specifically where differences occur, 95 percent Bayesian confidence intervals can be constructed around the smoothing splines that best fit the data. Where the confidence intervals do not overlap, there is a significant difference among tongue shapes. This is demonstrated in Figure 18.2.6a and 18.2.6b, which is a comparison of the tongue shapes for the

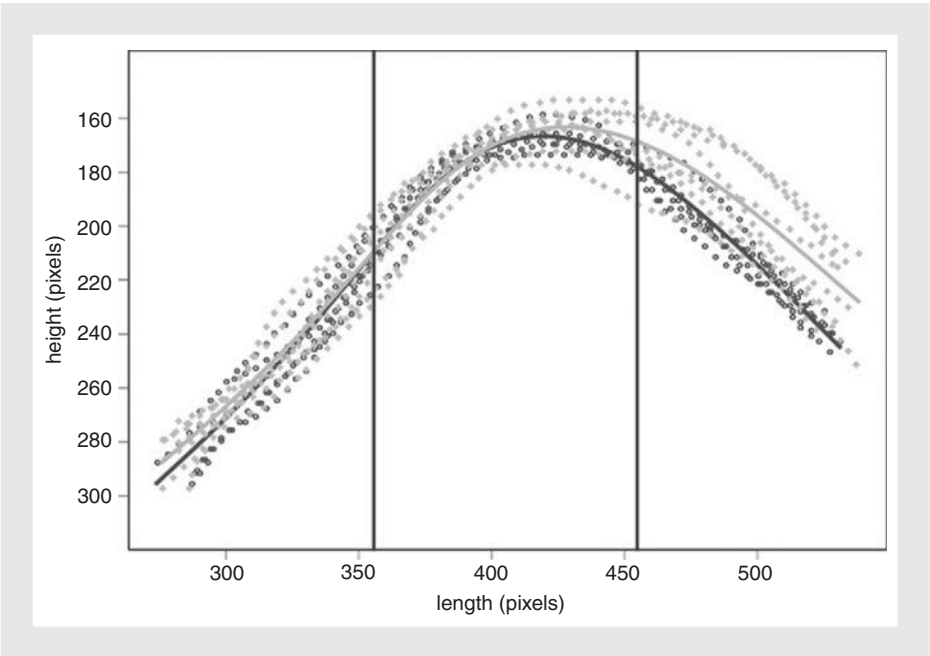

Figure 18.2.6a. Eight repetitions of the closure of the [g] in *bag dazzled* (dark "o") and *Baghdad* (light "+"), each fitted with a smoothing spline.

closure for [g] in *bag dazzled* and *Baghdad*. The three divisions of the tongue roughly correspond to tip/blade (right), body (mid), and root (left). The confidence intervals around the splines, represented by the dotted lines around the solid lines of the smoothing spline in Figure 18.2.6b, show that there is a difference in the tip/blade (right side of the image), but not in the body or root (left side of the image).

18.2.3 Applications of ultrasound to laboratory phonology

Initially, the use of ultrasound imaging in speech research focused primarily on studying the tongue shapes associated with individual consonants and vowels and the relationship between the tongue, jaw, and palate in speech (e.g. Stone et al. 1992; Parush and Ostry 1993; Stone 1995; Stone and Vatikiotis-Bateson 1995; Stone and Lundberg 1996). More recently, researchers have used the technology to address specifically phonological questions, which are the focus of this section.

Since ultrasound imaging is particularly well suited to examining questions about differences in tongue shape, it is natural that a number of studies have focused on the phonology and phonetics of vowels. Benus and Gafos (2007) used

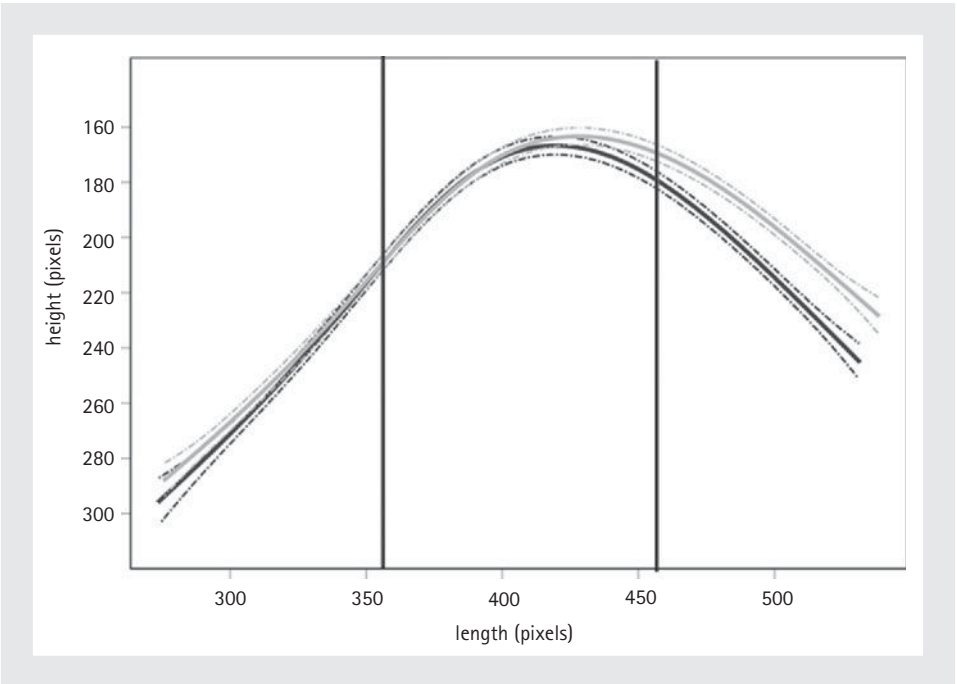

Figure 18.2.6b. Smoothing spline estimate (solid lines) and 95% Bayesian confidence interval (dotted lines) for comparison of the mean curves for /g/. The vertical lines roughly divide the tongue into the blade (right), body (middle), and root (left). The x-axis is the length of the tongue, and the y-axis is the height of the tongue. The scales correspond to the pixels of the original JPEGs, where 1 mm = 2.63 pixels and the origin is in the top left corner (accounting for why the values on the y-axis increase).

ultrasound (as well as EMMA) to investigate transparent vowels in Hungarian. Usually in Hungarian, suffix vowels take on the same [±back] feature as the last vowel in the stem. However, vowels like [i] in the stem are called transparent vowels because they can be followed by either front or back vowels in the suffix. Taking measurements along fixed lines placed at the tongue root, Benus and Gafos show that transparent vowels found in a back harmony context have a significantly more retracted tongue body posture than the same vowel in front harmony contexts. They argue that the choice of suffix vowel after the transparent vowel is not arbitrary, but rather correlates with the horizontal position of the tongue. Gick, Pulleyblank, Campbell, and Mutaka (2006) investigated whether a cross-height harmony feature in Kinande involves advanced tongue root, and whether the tongue root advancement is also found on low vowels, which has been explicitly claimed to not be possible. Results from measurement along a fixed line placed at the tongue root

indicated that the articulation of vowels in Kinande is consistent with a [±ATR] feature, and that the same articulations are also present for low vowels.

Researchers have also used ultrasound to examine the status of excrescent schwas in various phonological environments. Gick and Wilson (2006) looked at the articulation of words like *heel* or *hire*, which have sometimes been claimed to be disyllabic ([hi.əl], [haɪ.ɚ]). Gick and Wilson argued that the percept of a schwa in these types of words is not due to the production of a phonological vowel target, but rather is an articulatory by-product of the tongue passing through a "schwa space" as the tongue root moves from an advanced position to a velarized or pharyngeal constriction. Inspection of tongue shape changes showed that the tongue took on a schwa-like shape in the middle of the trajectory from the high vowel to the liquid. A similar claim was also made for Nuu-chah-nulth (Wilson 2007). Davidson (2005) examined a different type of excrescent schwa found in the production of non-native consonant sequences by English speakers. Davidson hypothesized that when English speakers repair sequences like [zg] in word-initial position by inserting a schwa, they are not epenthesizing a vowel but rather failing to accurately coordinate the consonant gestures such that the constrictions of the gestures do not sufficiently overlap. This would result in a period of open vocal tract between the constrictions, giving rise to the percept of a schwa. Davidson compared the pronunciation of words like *succumb* and *scum* to the "repaired" production of non-words like *zgomu* (pronounced as [zəgomu]). Findings from L_2 norm distance measures (Horn and Johnson 1990) showed that for most speakers, tongue shape changes over time were more similar to native #CC- than #CəC- words, and that there was no movement toward a schwa gesture between the two consonants in the non-native utterances. Buchwald et al. (2007) investigated a similar question in an aphasic speaker who exhibited errors in spoken language production, including the insertion of vowels (e.g. *bleed* → [bəlid]). Results showed that both the speaker's lexical and inserted schwas had the same acoustic and articulatory characteristics, indicating that her errors were consistent with phonological epenthesis, not a failure to accurately coordinate consonantal gestures.

Tongue shape in consonant articulations has also been studied for various goals: to shed light on the possible phonological specifications of phonemes, to study the phonetic underpinnings of natural classes, and to investigate patterns of coarticulation. Miller and colleagues have examined the articulation of clicks in languages such as N|uu (Miller et al. 2009), Khoekhoe (Miller et al. 2007), and IsiXhosa (Miller 2008). Using ultrasound data, Miller claims that languages with clicks are not fundamentally different systems than languages without them, but rather that linguo-pulmonic consonants differ from lingual consonants only with respect to the airstream of the release. Mielke (2005) applies measurements from acoustic recordings, nasal and oral airflow, and ultrasound to a simulation that attempts to use the raw phonetic data to discover natural classes of consonants (see also Mielke, this volume). Wodzinski et al. (2007) examined the coarticulation of velar

consonants with the following vowel in English, and showed that the angle, or position, of the velar closure is well correlated with the F2 of the following vowel, but that while some speakers seem to prefer a distinct closure position for front vowels versus back vowels, the closure location is more continuous for other speakers. Zharkova and Hewlett (2009) developed a method for quantifying coarticulation using a nearest neighbor metric that finds the closest points between two curves in order to provide a distance measure. They use the nearest neighbor technique to examine the coarticulatory effect of /i/ and /a/ on the production of /t/. The same measure was also used to investigate differences in the degree of coarticulation among /ʃV/ sequences in children and adults (Zharkova et al. 2008). Results showed that for these sequences, children showed significantly more anticipatory lingual coarticulation than adults did.

One specific consonant that has received considerable attention is the approximant /r/ found in several varieties of English. Mielke et al. (2006) examined the production of /r/ in American English, both preceding and following a large variety of consonants and vowels. They were interested in whether retroflexed and bunched varieties of /r/ are conditioned by context, or whether individual speakers produced only one or the other variant of /r/. After classifying tongue shapes as retroflexes, bunching, or other, Mielke et al. found that many of their speakers did produce a combination of both retroflexes and other possible tongue shapes for /r/ (Delattre and Freeman 1968), and that average retroflexion rates are highest before vowels and /l/. They argue that there is less retroflexion next to articulations that are incompatible with the tongue shape necessary for retroflexion. Campbell et al. (2010) used ultrasound to study the relative timing of the independent gestures necessary to produce Canadian English /r/, namely the lips, the tongue body, and the root. Just to discuss one result, they found a strictly front-to-back ordering of gestural timing in syllable-initial position. This is not entirely consistent with previous studies that hypothesized that the gestures comprising /r/ should be produced simultaneously in syllable-initial position, but Campbell et al. discussed how Browman and Goldstein's (1995) proposal that constriction width predicts gestural timing patterns could be extended to account for Canadian English. Finally, Lawson et al. (2008) examined the spread of de-rhoticization among young male working-class speakers in West Lothian, Scotland (see also Scobbie and Stuart-Smith, this volume). In addition to the classification of auditory data demonstrating that de-rhoticization seems to be conditioned by unstressed syllables and utterance-final position, they also provided ultrasound evidence that there is a delayed tongue-raising gesture in apparently non-rhotic tokens that occurs only after the voicing for /r/ has trailed off.

Another application of ultrasound imaging is collection of articulatory data on speech errors. Pouplier (2008) confirmed previous findings that speech errors are often composed of both the target and the intruding gesture (e.g. both tongue dorsum and unexpected tongue tip raising during the production of the [k] in *top*

cop, giving rise to the perception of [t]) (Goldstein et al. 2007), and extended prior work to demonstrate that intrusion errors are more likely to occur when the stimuli contain codas than when they do not. Pouplier discusses the ramifications of these findings for theories of gestural organization.

In sum, there is a growing body of work using ultrasound to examine questions central to laboratory phonology, such as transparent vowels, vowel harmony, syllable structure, and consonant-vowel coarticulation, to name a few. Not only can researchers use ultrasound to test hypotheses about how phonological processes should affect tongue shapes, but it is a non-invasive, portable, and relatively inexpensive technology that makes collection of articulatory data ultrasound practical both for the laboratory and for use in the field. Though limitations of the technology should be kept in mind, methodological and measurement techniques currently being developed may help overcome some of the drawbacks of ultrasound. For example, Miller and Finch (2011) have developed a technique for recording ultrasound images at a much higher-than-typical frame rate (124 frames per second) that is suitable for recording short clips of speech even in the field. Ultimately, the type of ultrasound set-up that an investigator prefers will be determined by the particular research question being asked. Those researchers interested in examining the fine temporal details of either a short sound or a complex articulation may prefer a high-speed system. Researchers who tailor their questions to sounds that are relatively unchanging over a longer timescale, such as vowels or stop closures, may find that a more conventional ultrasound that captures 30 frames per second is sufficient. Another area that will benefit from further advances is statistical analysis. Current techniques such as SS ANOVA or fixed line measurements are adequate for single frame comparisons, but further work is necessary to provide meaningful analyses of tongue shape changes over several sequential frames. Just as partnerships between linguists and engineers have led to many of the advances detailed in this section, future collaborations with statisticians will greatly benefit the use of ultrasound for speech research.

18.3 Methodologies used to investigate laryngeal function and aerodynamic properties of speech

Helen M. Hanson

Most techniques described in this section originated in clinical settings. They have been borrowed, and sometimes adapted, by linguists, speech scientists, and en-

gineers who want to better understand the links between speech physiology and the acoustic speech waveform. I first describe the commonly used aerodynamic measures, and then turn to a description of techniques for studying laryngeal function, which plays a role at both segmental and prosodic levels, in the form of voice quality and fundamental frequency.

18.3.1 Measuring aerodynamic properties

18.3.1.1 *Oral and nasal flow*

It is through the control of airflow in the vocal tract that we produce speech sounds in all their variety. Different rates of flow, different types of flow, and different pathways of flow produce distinct classes of sounds, for example, obstruent vs. sonorant sounds. There is a need, then, for methods of measuring flow through the oral and nasal cavities.

The volume velocity at the mouth, or oral airflow, is most often measured using a pneumotachograph mask (e.g. Rothenberg 1973; see also Shadle, this chapter). The mask is like an anesthesia mask, but it has holes around its circumference, close to the mouth (Figure 18.3.1). These holes are covered with a fine wire mesh, which allows air to flow freely, but also presents a small resistance to the flow. This resistance results in a small pressure differential between the air inside and outside the mask. The pressure drop is measured using a pressure transducer, which converts it to a voltage signal. Because the pressure differential is directly proportional to the oral airflow, one can derive volume velocity from this voltage signal, using appropriate calibration data. A pneumotachograph mask can cover both the nose and mouth, or only one or the other.

The mask does affect the volume velocity, and the nature of this distortion depends on the design of the mask used. The Rothenberg mask, for example, minimizes the time delay and mask resonance because the wire-covered holes are close to the mouth. If the holes were at the far end of the mask, distortion would be greater. Nevertheless, the frequency response of the mask is limited to below 29000 Hz (Rothenberg 1973; Hertegård and Gauffin 1992). Details about the varieties and use of pneumotachograph masks can be found in Baken (1996) and Rammage et al. (2001).

An example of the use of the oral volume velocity in research is an early study by Hixon (1966) in which airflow rates associated with [s, ʃ] were observed. Oral volume velocity can also be used to derive other measures. Koenig (2000) used the oral airflow signal to label closures, releases, and voice onset times (VOT) for stop consonants. Likewise, Slifka (2006) used oral airflow (along with lung pressure) to estimate glottal area during vowel production. Oral airflow is commonly recorded as a means to estimating glottal flow, or the volume velocity through the glottis (see 18.3.2.2).

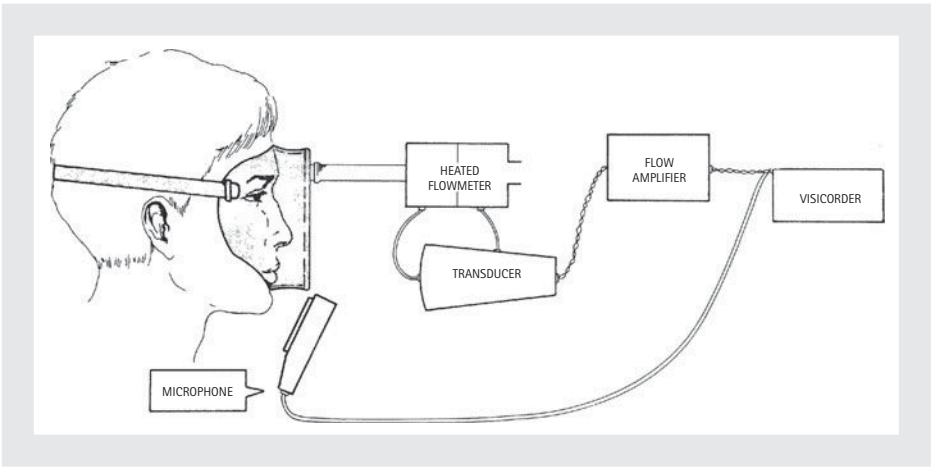

Figure 18.3.1. A set-up for recording airflow using a pneumotachograph mask (reprinted with permission from Warren 1982).

18.3.1.2 *Lung volume*

In general, lung volume increases rapidly during inhalation and decreases slowly as a speech utterance is produced. Small variations in lung volume during this exhalation phase can be linked to speech events (Ohala 1990a). In addition, lung volume data may be necessary for estimation of other aerodynamic signals (e.g. Section 18.3.1.3 on estimating lung pressure from esophageal pressure). There are two methods suitable for deriving lung volume variations during speech production.

Respiratory inductive plethysmography, also known as Respitrace, is based on the assumption that movements of the ribcage and diaphragm contribute independently to lung volume (Konno and Mead 1967). Elastic bands are fitted around the ribcage and abdomen of a subject (Figure 18.3.2). As the subject inhales and exhales, the bands change size according to whether the ribcage and abdomen are expanding or contracting. The wires attached to the bands are stretched, changing their inductances. These changes in inductance can be measured and related to the cross-sectional areas of the ribcage and abdomen. Upon completion of data collection, the cross-sectional area data are calibrated and combined to obtain lung volume. A good discussion of this process can be found in Baken (1996) or Slifka (2000).

A limitation of the Respitrace method is that non-respiratory movements of the subject may also contribute to changes in the cross-sectional areas of the ribcage and abdomen—even movements as small as shifting an arm can have an effect. Therefore, the subject must be very still during data collection, which is difficult if the protocol is lengthy. Seating the subject in a comfortable seat, with both foot

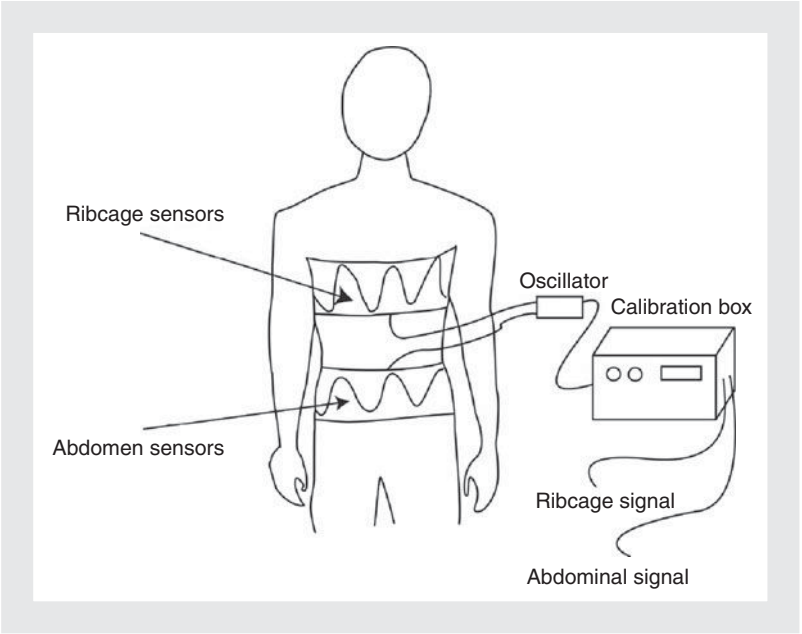

Figure 18.3.2. Example of a respiratory inductive plethysmography system.

and head supports, is best. Because of these restrictions on subject movement, the Respitrace method works best for relatively short, scripted protocols.

However, studies such as those by Slifka (2000, 2006) and Hanson et al. (2005, 2007) on the physiology of prosody suggest a need for measures of lung volume during more unscripted, spontaneous speech, such as that elicited using the map-task method (Anderson et al. 1991). These methods would seem to require more activity on the part of the subject, which could be allowed by the full-body plethysmograph (Figure 18.3.3). One of the earliest of these devices was described by Mead (1969). Today, there are several types of full-body plethysmographs (Goldman et al. 2005). These are largely used in clinical settings to measure respiratory volumes such as vital capacity. However speech researchers such as Ohala (1977) have made use of them to, for example, relate lung volume changes to oral airflow rates. Some full-body plethysmographs completely encase the body and the subject may breathe through a pneumotachograph, allowing collection of airflow data as well. As the subject inhales and exhales, their body volume increases and decreases accordingly. Correspondingly, the volume of the air in the chamber decreases or increases, resulting in a pressure increase or decrease, respectively. These pressure changes are tracked, and can be mapped to lung volume changes given the appropriate calibration data.

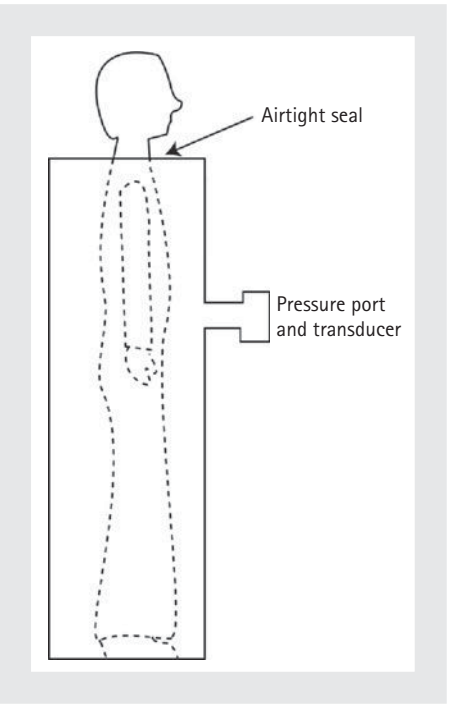

Figure 18.3.3. Schematic of a full-body plethysmograph in which the head is not encased in the device.

To observe volume changes during speech production, it may be preferable to use a version in which the subject's head is not enclosed in the chamber (Baken 1996), or in which the subject speaks through an opening in the wall of the chamber (Ohala 1977). In either case, a tight seal must be maintained at the neck or face, respectively, to prevent air leakage into or out of the chamber. A seal at the face can be provided by a facemask, such as an anesthesiology mask.

18.3.1.3 *Subglottal pressure*

Subglottal pressure is considered to be the energy source for speech production; hence the interest in recording it. In tracheal puncture, a hypodermic needle is used to puncture the trachea (Figure 18.3.4). A tube is then inserted and connected to a pressure transducer. Although the procedure is relatively simple and the results are accurate, it is not commonly employed because it is not appealing to subjects. In addition, there is some risk to the procedure and it requires the presence of a physician. When it has been used, the study usually involves a single subject,

typically one of the experimenters (e.g. Isshiki 1964; Hertegård et al. 1995). The main contribution of this method has been to verify other methods, described below.

For most speech sounds lung pressure, or alveolar pressure, can be considered to be the same as subglottal pressure. Although we cannot measure lung pressure directly, we can measure esophageal pressure using a thin latex balloon that is passed through the nasal cavity and into the esophagus (van den Berg 1956) (Figure 18.3.5). If lung volume is measured simultaneously and appropriate calibration data are collected, lung pressure can be estimated from esophageal pressure (Kunze 1964). The calibration process is somewhat involved, as it requires an estimation of static recoil of the lungs for a range of lung volumes.

While still invasive, the esophageal pressure method is not quite as unappealing to subjects as tracheal puncture, meaning that studies can be based on data from more than one subject (Slifka 2000, 2006; Hanson et al. 2005, 2007). There are several potential sources of problems when collecting esophageal pressure data.

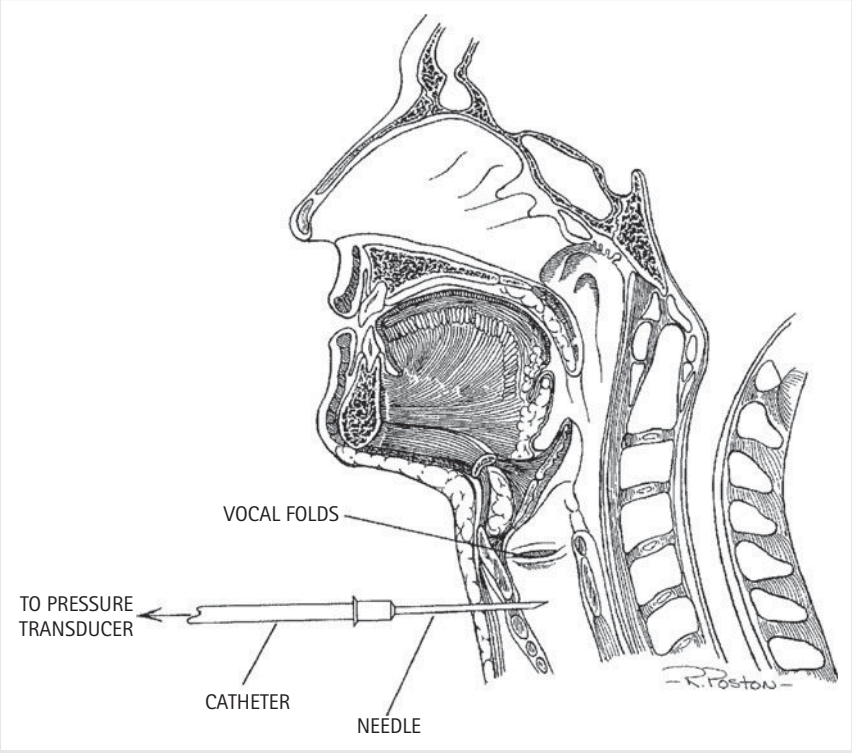

Figure 18.3.4. Subglottal pressure being measured by tracheal puncture (reprinted with permission from Warren 1982).

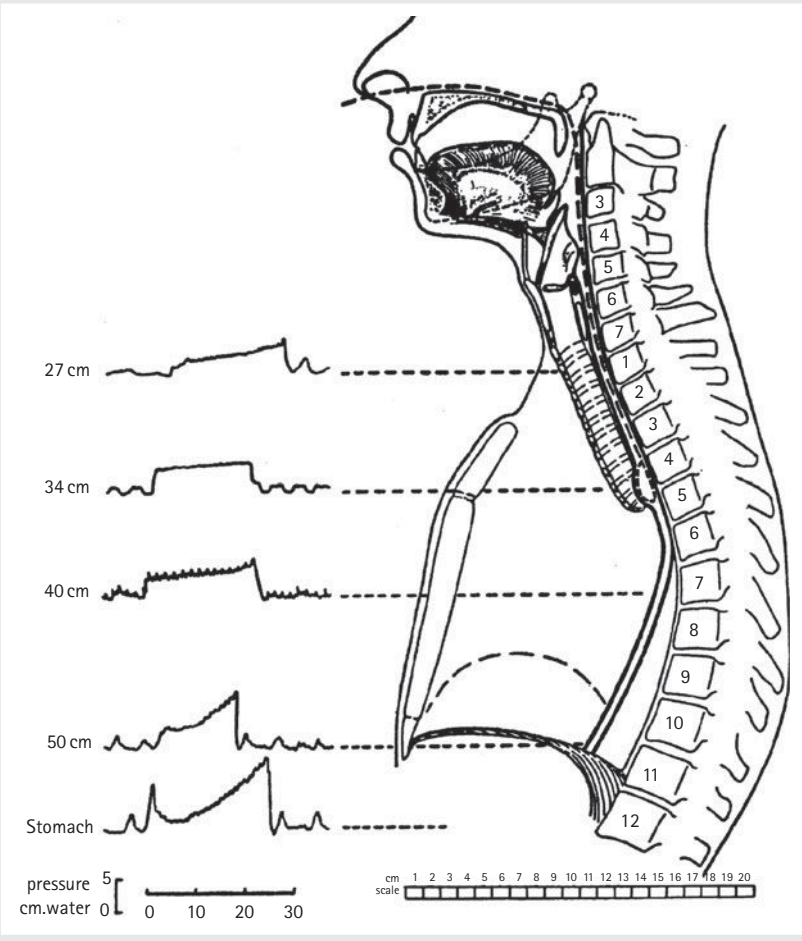

Figure 18.3.5. Showing the placement of an esophageal balloon, through the nasal cavity and into the esophagus (reprinted from Ladefoged 1962, with permission from DeGruyter Mouton).

For one, the balloon must be carefully placed, so that it is below the trachea, but does not stray into the diaphragm. Proper placement cannot be confirmed by sight, so the experimenter(s) must rely on characteristics of the pressure waveforms as a guide during placement. For another, the subject must be aware of their swallowing and warn the experimenter(s) when they do so, because swallows result in peristaltic waves in the esophageal pressure data. In addition, the subject's heartbeat can interfere with the data. Slifka (2000) provides an excellent, thorough discussion of the data collection process, including calibration.

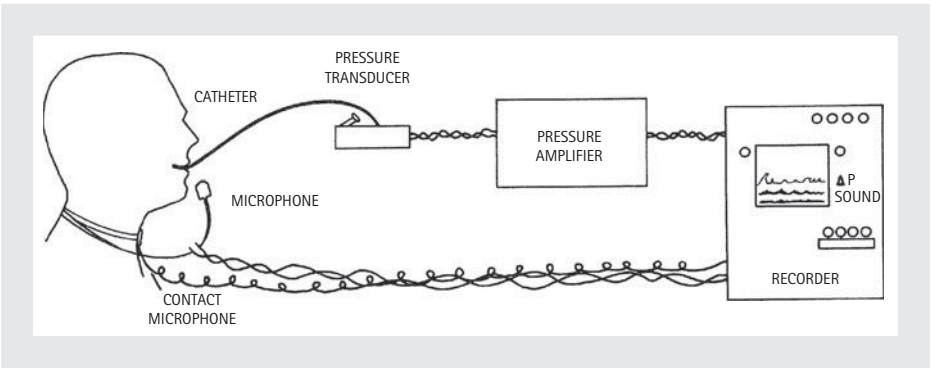

Figure 18.3.6. A set-up to record intraoral pressure (reprinted with permission from Warren 1982).

The easiest method of estimating subglottal pressure is to measure the oral pressure during voiceless stop-consonant occlusions. When the vocal folds are spread and the oral cavity is completely occluded, as occurs for voiceless stops, oral pressure quickly builds up to equal subglottal pressure. Intraoral pressure is measured by placing one end of a tube in the mouth (Figure 18.3.6; see also Shadle, this chapter). It is positioned so that it doesn't interfere with normal articulation of speech sounds and the tongue doesn't obstruct its opening. The other end of the tube is outside the mouth, connected to a differential pressure transducer. The transducer senses the difference between the oral and atmospheric pressures and converts it to a voltage. The voltage signal can be converted to a pressure signal using appropriate calibration data. (See Baken 1996 and Rammage et al. 2001 for more details.)

Hertegård et al. (1995) collected oral pressure simultaneously with subglottal pressure measured through a tracheal puncture, and verified the high accuracy of this method. However, the speech sounds that can be articulated without interfering with the tube inserted into the oral cavity are limited. Typical stimuli are strings of /pV/ syllables (Holmberg et al. 1988). The combination of a labial stop consonant with an open vowel is less likely to be perturbed by the tube placed in the oral cavity. Reiterant speech using /pæ/ and /pə/ syllables has also been used in an effort to simulate more natural speaking conditions (Hanson 1997b). Another disadvantage is that subglottal pressure is not being estimated continuously; subglottal pressure during the vowels is assumed to be an average of the values estimated in the neighboring stop consonants. Still, this simple method could be used to collect pilot data or verify calibration of esophageal pressure (Slifka 2000).

Despite its invasiveness and somewhat complex calibration process, the use of esophageal pressure to estimate lung pressure seems to be the preferred method for obtaining subglottal pressure. It has been consistently used since the early work by Ladefoged and his colleagues (summarized in Ladefoged 1962), is not as invasive as

tracheal puncture, and provides a good, continuous estimate. With current interests in the physiology of prosody, it should continue to be a useful method.

18.3.2 Methodologies for studying laryngeal properties

18.3.2.1 *Glottal area*

Glottal area or glottal width can be most accurately estimated using images of the vocal folds obtained through ultra-high speed photography (e.g. Timcke et al. 1958; Metz et al. 1980), schematized in Figure 18.3.7. Computation of the glottal area from such images is labor-intensive, but fortunately it has been shown that glottal width, which is far easier to compute, is a good correlate of glottal area (Koike and Hirano 1973). The presence of the laryngeal mirror in the oral cavity, however, limits the possible speech sounds studied to be sustained vowels, thus limiting the value of this method to linguists.

One alternative is fiberscopy, in which the camera lens and light source are threaded through a nostril and the nasal cavities into the pharynx, until they sit just above the vocal folds, at about the level of the epiglottis. The oral cavity is unobstructed and subjects can speak more or less naturally. Video can be used to capture the laryngeal activity, but measures made on them are not very precise. Recently, however, advances have been made in digital imaging via fiberscopy. A solid-state image sensor captures images of the vocal folds as they vibrate and these images are stored in memory for later analysis (e.g. Hirose 1997). This method appears to be the best alternative to ultra-high-speed photography.

A semi-invasive alternative is photoglottography. In this method, a bright light is shined on the neck just below the vocal folds, and a flexible fiberscope is used to position a photosensor above the vocal folds (or vice versa). The amount of light that passes through the vocal folds and reaches the photosensor will depend on the glottal area: during the closed phase of a vibratory cycle, less light will be passed through and sensed than will be during the open phase. Thus, a voltage signal will be obtained, the amplitude of which reflects glottal area. Photoglottography has been compared to high-speed photography by Harden (1975) and Baer et al. (1983), and was found to provide comparable estimates of glottal area functions (Baken 1996). An example of its use is a study of geminate and singleton stops in Berber (Ridouane 2003).

Electroglottography (EGG) produces a voltage that is believed to reflect vocal-fold contact area, not glottal area. Two electrodes are attached on either side of the front of the throat, at the vertical level of the vocal folds. A small current is applied. Because human tissues are good conductors, while air is a good insulator, the applied current will be conducted most effectively when the vocal folds are approximated and least effectively when they are spread. During phonation,

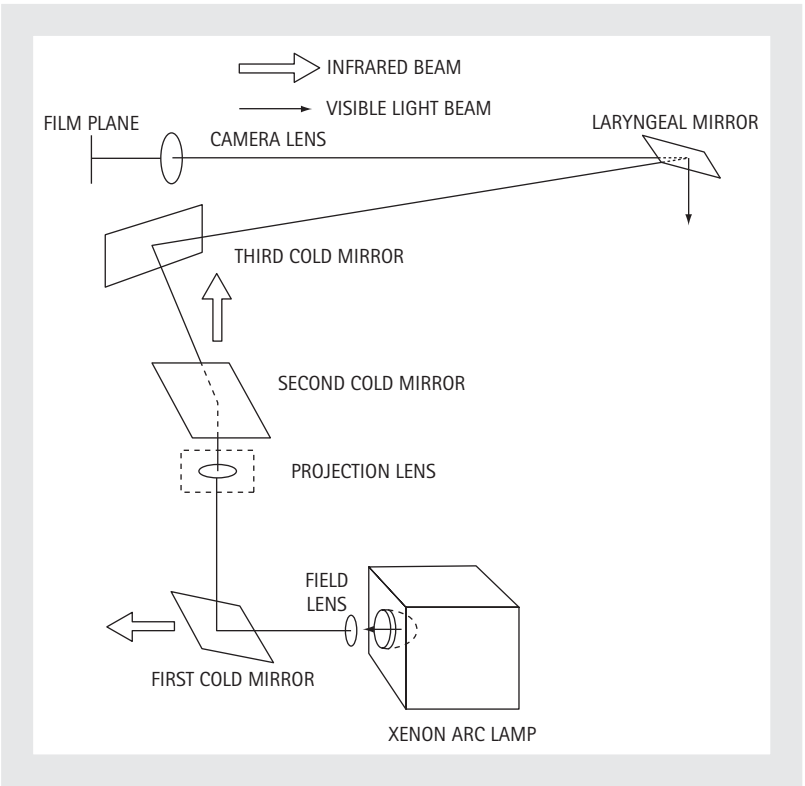

Figure 18.3.7. Schematic of a high-speed photography system (reprinted with permission from Metz et al. 1980).

therefore, the output voltage will be high during the open phase and low during the closed phase of vocal-fold vibration. While this method is non-invasive and relatively easy to employ, interpretation of the resulting data is not straightforward, and much research has been aimed at figuring out just what the EGG waveform can tell us (see Baken 1996 for a summary). Childers et al. (1990) carefully recorded simultaneous EGG and high-speed photography data. Measures were made on the EGG waveforms and the high-speed films. The results suggest that some EGG features can be associated with aspects of the glottal area waveform, although these features (e.g. f0, loudness) can be heard by clinicians and measured with other methods. On the minus side, the EGG does not reflect some important features of the glottal area function, for example, vibratory events that do not involve vocal-fold contact, and it has not been found to correlate with voice quality. However, Baer et al. (1983) found that EGG data combined with photoglottography data can provide much of the information available from high-speed films. Therefore, the

value of EGG data to both research and clinical applications may be uncertain unless it is combined with other data.

18.3.2.2 *Glottal waveform*

The glottal volume-velocity waveform cannot be measured directly. What we can measure is its filtered form, after it has been shaped by the vocal-tract resonators. If we assume that the system is linear and time-invariant, the glottal waveform can be recovered via inverse filtering. That is, we take either the acoustic pressure signal or the oral airflow signal and we remove the effects of the vocal-tract filter to reveal the glottal waveform. Once the glottal waveform has been obtained, we can make measures on it to obtain properties reflecting laryngeal function, such as the open quotient. Here we discuss three methods for extracting the glottal waveform or its properties. We refer to these jointly as inverse filtering techniques.

The first method is inverse filtering of the sound pressure waveform. A high-quality microphone must be used for making the recordings. In addition, special care should be taken to prevent interference from other sources of sound in the environment, e.g. recordings should be made in a sound-attenuated booth or an anechoic chamber. Once the data are recorded, the effects of the formants are removed by applying an inverse filter. Ideally this filter will have zeros that exactly line up with the poles of the original vocal-tract filter, and the bandwidths of the zeros will match the bandwidths of the filter. There are several methods of setting the parameters of the inverse filter. One is to manually tune the frequencies and bandwidths of the filter. Manual tuning, however, is quite time-consuming and limits the amount of data that can be analyzed for any given study. An alternative is to use automated, semi-automated, or fixed-filter methods for setting the filter parameters. For example, Javkin et al. (1987) developed algorithms for automatic inverse filtering. Quality of the results varies depending on factors such as the fundamental frequency range of the utterance (better results for lower f0), or whether the voice is deviant in some way.

Deriving glottal waveforms from acoustic sound pressure signals is not without its problems. For one, the recordings are sensitive to low-frequency noise. In addition, the acoustic signal actually represents the excitation as the derivative of the glottal waveform. The signal could be integrated to obtain the volume velocity, but any DC component of the glottal waveform[2] will be lost. Finally, it is difficult to calibrate the signal.

For these reasons, oral flow inverse filtering was introduced by Rothenberg (1973). The oral airflow is recorded using techniques described in 18.3.1.1. This

[2] The DC (direct current) component of the volume velocity is a constant flow that sometimes occurs even during the closed phase of the vibratory cycle. It is due to a lack of complete closure of the vocal folds, usually (but not always) at the cartilaginous portion of the vocal folds. The time-varying component of the volume velocity is superimposed on the DC component.

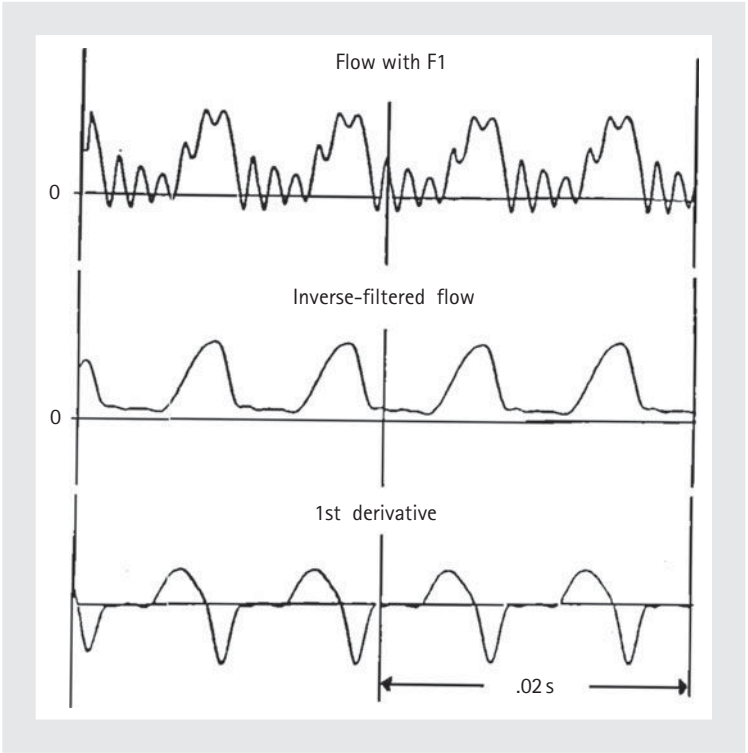

Figure 18.3.8. Example of glottal waveforms obtained through flow inverse filtering. Row 1 shows oral flow after it has been highpass-filtered. Row 2 shows the inverse-filtered waveform, corresponding to glottal flow. Row 3 shows the first derivative of the waveform in Row 2, which is sometimes considered to be the effective glottal source (Fant 1979). (Reprinted with permission from Holmberg et al. 1988, c1988, Acoustical Society of America.)

signal can be calibrated, is not subject to distortion by low-frequency noise, and preserves the DC flow. It is inverse filtered as above. This method has been used to study laryngeal characteristics of normal voice (e.g. Holmberg et al. 1988, 1995) and disordered voice (e.g. Hillman et al. 1989, 1990). Figure 18.3.8 shows examples of waveforms obtained by flow inverse filtering.

Although it has the above-mentioned advantages over pressure inverse filtering, it has several disadvantages, too. The frequency range of the mask is limited to about 1.6 kHz (see 18.3.1.1 above and also Hertegård and Gauffin 1992). As a consequence, very rapid changes in glottal flow, as might occur at the moment of glottal closure, are lost. Because the instance of glottal closure provides the main excitation of the vocal tract during phonation (Fant 1979), this side effect is particularly unfortunate.

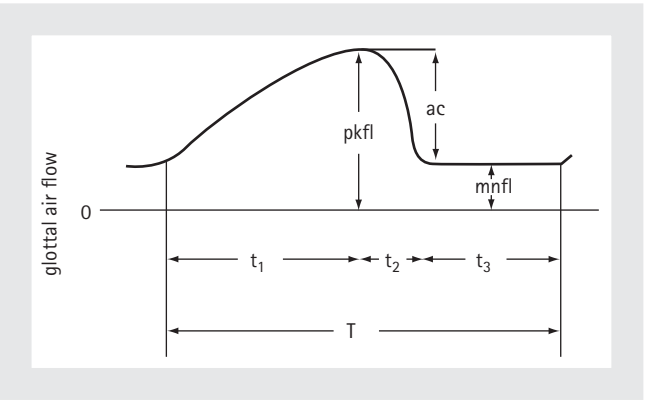

Figure 18.3.9. Example of measures made on the glottal waveform obtained through flow inverse filtering. (Reprinted with permission from Holmberg et al. 1988, c1988, Acoustical Society of America.)

Another problem is that it is difficult to ensure that the seal between the mask and the subject's face is airtight; air leaks can have major effects on the recordings.

Once the volume-velocity waveform has been obtained by inverse filtering, one of two methods is used to obtain measures of laryngeal function. Figure 18.3.9 illustrates how glottal characteristics can be measured directly from the waveform, as was done by Holmberg et al. (e.g. 1988, 1995). Or, a glottal waveform model can be fit to the measured data; the parameter values resulting in the best fit are assumed to be characteristic of the glottal waveform. This method has been used by, for example, Ní Chasaide and Gobl (1993), who used both flow inverse filtering and pressure inverse filtering to obtain the volume velocity waveform *U(t)* and *dU(t)/dt*, respectively. The resulting glottal pulses were fit to a model of *dU(t)/dt* (Fant et al. 1985).

When the volume-velocity waveform itself is not needed, acoustic measures that reflect or indicate the glottal parameters can be made on speech spectra or waveforms. For example, it is common to see the measure H1-H2 (relative amplitudes of the first two harmonics) used to indicate the open quotient of a glottal pulse (e.g. Klatt and Klatt 1990). Some typical measures are illustrated in Figure 18.3.10. But making this measure on the speech spectrum is not quite correct because the magnitude of the harmonics of the glottal source will be influenced by the frequency response of the vocal tract, particularly if the harmonic in question is close to a formant frequency. Thus, attempts have been made to correct the harmonic magnitudes to remove the boosting effects of the formants. In fact, Fant (1982) refers to this method as frequency-domain inverse filtering (FDIF) and cites Mártony (1965) as having used it. More recently, Holmberg et al. (1995) used such "correction factors" on H1-H2; Chen (1995, 1997) used them on the amplitudes of

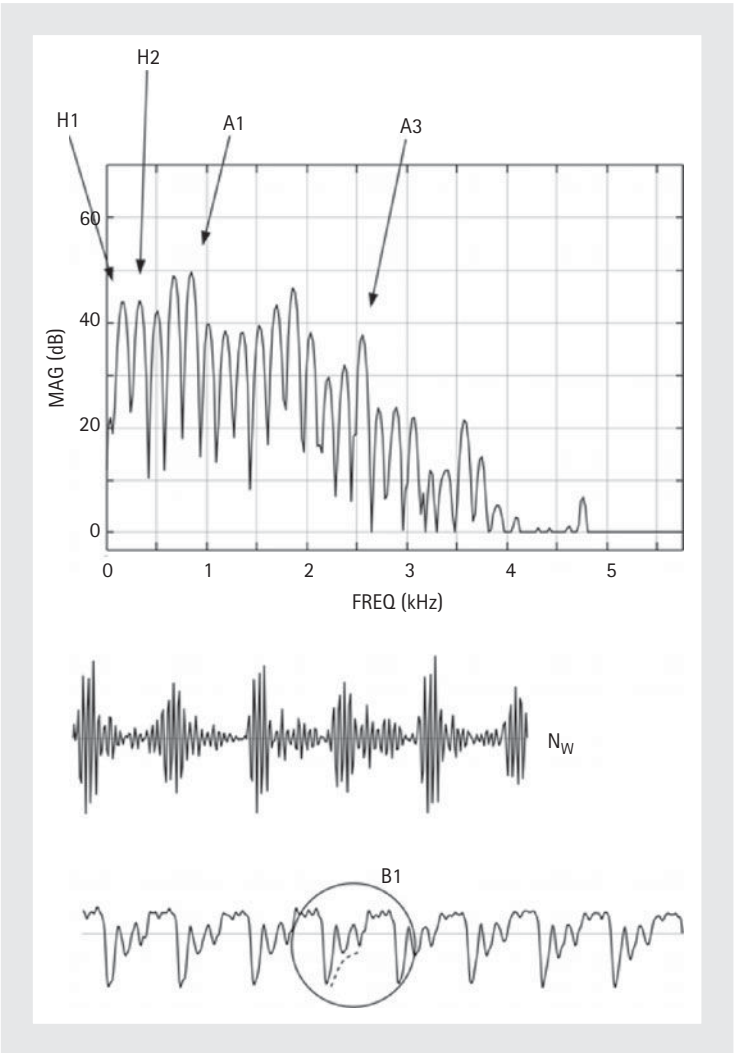

Figure 18.3.10. Illustrating measures made on the speech spectrum that correspond to measures made on the glottal waveform. H1 and H2 are the amplitudes of the first two harmonics, respectively. A1 is the amplitude of the strongest harmonic of the first formant peak. A3 is the amplitude of the strongest harmonic in the third formant peak. The waveform labeled N_W was obtained by bandpass-filtering a vowel waveform in the F3 region, and is an example of waveforms used to make noise ratings. The bottom waveform illustrates the decay of the first-formant oscillation, from which one can estimate bandwidth B1. (Note that each example is from a different female speaker.) (Reprinted with permission from Hanson and Chuang 1999, c1999, Acoustical Society of America.)

nasal poles; and Hanson (1995, 1997a) used them on H1-H2 and H1-H3 (where the latter measure reflects spectral tilt of the periodic source spectrum). These three experimenters used simplified correction factors based on the assumption that the bandwidths of the resonances of the all-pole filter are zero.[3] The rationale for ignoring bandwidth is that it is difficult to estimate, unlike formant frequencies. However, a limitation of this simplified correction factor is that it is not accurate for vowels in which the harmonics of interest are close in frequency to the formant frequency. In general, this assumption means that (1) only low, and possibly mid, vowels can be used; and (2) fundamental frequency cannot be too high.

Recently, Iseli and Alwan (2004) found that bandwidth estimates do not have to be precise, and therefore a more general form of the correction formula, which takes bandwidth into account, can be used with relative ease, opening up the possibility of using FDIF for almost any combination of formant frequencies and fundamental frequency. Okobi (2006) used this technique in his study of the acoustic correlates of stress in American English, and we hope to see more examples of its use in the future. Shue et al. (2009) have developed MATLAB code to automatically calculate measures such as H1-H2 and H1-H3, both with and without the correction factors.

Other acoustic measures reflecting laryngeal configuration can be made in the time domain. For example, the bandwidth of the first formant (B1) not only reflects losses at the vocal-tract walls, but also losses at the glottis. Hanson (1995, 1997a) estimated B1 from the speech waveform after it had been bandpass-filtered in the vicinity of the first-formant frequency, and linked these estimates to other measures of laryngeal configuration. Another acoustic characteristic related to glottal characteristics is aspiration noise generated during vowel production. Klatt and Klatt (1990) made noise ratings on the acoustic waveform. Hanson (1995, 1997a) did likewise and also included a similar rating based on observations of the spectrum in the third-formant region. Again these can be linked to other measures of laryngeal function and to perceptions of voice quality.

Although these measures on the speech spectrum and waveform are not direct measures on the glottal waveform or derivative, several of them have been found to correlate with glottal source parameters. For example, Holmberg et al. (1995) found a strong correlation between adduction quotient (1-OQ) and the measure H1-H2.

Analysis by synthesis is another technique with which one can estimate source parameters without first inverse-filtering the speech waveform. Copy synthesis is used to duplicate a natural voiced speech waveform. The synthesis parameters related to the voice source are then taken to be measures of the glottal waveform (e.g. Alwan et al. 1999). One problem with this approach is that sometimes two or more synthesis parameters can be used to change an acoustic parameter. For example, in the Klatt synthesizer (Klatt and Klatt 1990), the amplitude of F3 can

[3] The equations are given in Hanson (1997a). Derivations of these equations can be found in Hanson (1995).

be adjusted using either B3 (third-formant bandwidth) or TL (source spectral tilt). Therefore, results based on analysis by synthesis should be interpreted with care.

18.3.3 Summary

Given the importance of subglottal pressure and laryngeal function to the production of speech, methodologies such as those discussed in this section will continue to be useful, although some will be more useful than others. Using esophageal pressure to estimate subglottal pressure should become more common as interest in the production aspects of prosody increases. Intraoral pressure as an estimate of subglottal pressure will continue in situations where subglottal pressure is not required for all segments, or where reiterant speech will suffice. Inverse filtering of oral airflow signals, combined with waveform model fitting, will continue to be useful for obtaining glottal waveform parameters. In addition, acoustic measures made on speech spectra or waveforms are gaining in popularity and tools such as those provided by Shue et al. (2009) will make it easier to apply these measures.

18.4 On the acoustics and aerodynamics of fricatives

Christine H. Shadle

18.4.1 Acoustic analysis of fricatives

18.4.1.1 *Acoustic properties of fricatives*

Fricatives are noisy sounds, which means that they need to be analyzed differently from periodic sounds such as vowels. Even voiced fricatives, with periodic as well as noisy components, need to be analyzed differently. Unlike stops, fricatives are more continuous and tend to be thought of as having a steady-state region, but such a region in which the acoustic properties are nearly stationary does not always exist, and there is evidence that the transition regions carry important cues to the identity of the fricative. However, analyzing these intervals of rapidly changing acoustic properties is more challenging.

The acoustic signals of voiceless fricatives and of the noise component of voiced fricatives are considerably lower in amplitude than those of sonorants. However, the

frequency range over which significant energy is produced is much higher. Different recording and analysis techniques are needed in order to capture the fricative sound accurately.

Phonologically, all fricatives have a noise source; voiced fricatives also have a voicing source. Phonetically, either source can disappear in some contexts: voiced fricatives can devoice, either mid-fricative or for the entire "voiced" fricative, and the noise source can disappear, particularly for non-sibilants. When both sources are present, the voicing source often modulates the noise source, resulting in the noise amplitude varying at the same rate as the fundamental frequency, with these variations visible in the time waveform or spectrogram.

Because the noise source occurs somewhere along the vocal tract rather than mainly at the glottis, it excites anti-resonances as well as resonances of the vocal tract. As a result, the spectral shape of fricatives has a more complex relationship to the vocal tract configuration than for vowels. Resonances can be excited that are then canceled, or nearly canceled, because the noise source location results in anti-resonances at nearly the same frequency. Such cancelations produce regions in the spectrum that are flat or have only a small ripple. However, where resonances and anti-resonances do not cancel each other, deep troughs and high peaks can occur, especially for sibilants, leading to a difference of amplitude across the spectrum of 40 dB or more. Regions of high energy in the spectrum may be due to several formants clustering together; a slightly different set of formants may cluster for another speaker. The spectral tilt calculated across the entire frequency range (i.e. up to 20 kHz) is meaningful aerodynamically, as will be discussed below, and can be related to specific segmental and suprasegmental properties.

18.4.1.2 *Recording techniques*

Given the acoustic properties of fricatives described above, the ideal recording environment has low ambient noise, and a sensitive microphone with a frequency response to 20 kHz. This requires a sampling frequency of 44 kHz or more, and an anti-aliasing low-pass filter that cuts off at 20 kHz, or slightly less than half the sampling rate. Such an ideal environment is not always available or necessary, however. The ambient noise should be recorded and analyzed; its time-averaged power spectrum can be compared to that of the fricative signals recorded in the same session to determine whether the low-amplitude parts of the fricative spectrum are getting lost in the ambient noise, and whether the choices of microphone gain and microphone location relative to the speaker are optimal (e.g. Shadle et al. 2008). The signal can be downsampled if the frequency range of the original recording is not needed, but filtering with an anti-aliasing filter set to half the new sampling frequency is needed before downsampling.

If signals are to be compared across speakers or recording sessions, or if variations in the sound pressure amplitude are to be compared to other measurements such as volume velocity, intraoral sound pressure, or degree of constriction, it is important to record the absolute sound pressure. A simple way to do this is to use a sound-level meter in the acoustic far-field, and record the distance from the subject's mouth to the meter. Another way is to record a calibration signal and all relevant parameters such as distance from subject to microphone, angle with respect to the subject's mouth, and all gain settings. Then it is possible both to determine the absolute sound pressure levels for a given recording and to compensate for different gains or microphone positions between recordings (Beranek 1954: 91–115, 1988: 177–92).

In some experiments it is not possible to make good-quality recordings. Scanning sequences for Magnetic Resonance Imaging (MRI) generate very loud noises inside the scanner, and condenser microphones cannot be used because of the strong magnetic field. Commercial systems exist that use optical microphones and adaptive noise-canceling technology, allowing recordings to be made of the subject's speech while being scanned, but for fricatives especially these are limited: the sampling frequency is low (e.g. 8 kHz) and the system can overadapt to the point where it removes fricative noise as well as scanner noise from the recorded signal. Such recordings are useful to monitor the subject's speech, but a separate recording in a quiet environment should be made if detailed acoustic analysis is planned. Other experimental environments can also pose problems: ultrasound scanners typically have loud fans. A condenser microphone should not be used within the field of an EMA system. For both of these cases, a directional microphone helps. A close-talk head-mounted microphone is useful in any situation with inescapable ambient noise, but the microphone is then in the acoustic near field which would make computation of an equivalent source at the lips more complex, and requires much greater precision if comparison of amplitudes across subjects or sessions is desired. A directional microphone used at a distance is a good choice for an EMA experiment and is less problematic, but its distinct directional characteristic may require care if comparison with signals recorded using a different omnidirectional microphone is desired.

Corpus design should be considered in terms of the other experimental constraints, and also in terms of the acoustic analysis planned. As we will discuss in the next section, some form of spectral averaging is needed to analyze noisy sounds. For MRI, speech needs to be sustained, so time averaging which is based on an assumption of a stationary signal is a natural choice. The corpus design for all sessions generating speech data to be compared to the MRI session data should include sustained fricatives with symmetric phonetic contexts. For experimental set-ups in which more natural speech is possible and desired, ensemble averaging may be a better choice; the corpus design should then include enough repetitions to form an ensemble, in which the signals need not be stationary but are assumed to contain

the same order of events that may affect the production. For ensemble averaging (see Section 18.4.1.3), where segmentation is crucial in defining the alignment points for the ensemble of signals, recording a second channel of an electroglottograph (EGG) signal will make both voicing detection and segmentation of the speech signal easier.

18.4.1.3 *Analysis techniques*

In the transition from a vowel to a fricative, two articulatory events occur with aerodynamic and acoustic consequences. The articulators move to form the main constriction, causing changes in the vowel formants. The vocal folds abduct, reducing the amplitude of the voicing signal and, for voiceless fricatives, leading to complete cessation of voicing. When the constriction is small enough and velocity of the air through it is high enough, noise production begins. The acoustic events do not necessarily occur in the same order; noise can begin either before or after voicing has died out, for instance. Segmentation based on the acoustic signal alone is thus not straightforward. For voiced fricatives it is even more problematic, since the voicing does not in general cease and the noise component is generally weaker. Techniques that have been found useful include filtering the speech signal with a high-pass filter at 3.5 or 3.9 kHz; the fricative onset can be defined as the time at which the high-passed signal increases above an amplitude threshold, as significant noise is being generated. Other signals recorded in the same experiment can also be used: cessation of low formants as visible on a spectrogram, decrease of an aerodynamically derived constriction area estimate below 0.2 cm^2, peaks in the oral airflow, or in the amount of anterior contact shown by electropalatography can all be used, but are likely to define slightly different frication intervals (Scully et al. 1991, 1992). The important thing is to use a consistent set for a given data set.

When the EGG signal is recorded, it provides a cleaner way of determining when voicing begins to decrease in amplitude, and when it ceases altogether. It can aid segmentation and also, obviously, any study of devoicing. Many automatic measures of voicing based on the speech signal only, or on the EGG signal, have been devised; see Jesus and Shadle (2003) for a review of these, a description of one algorithm, and an account of how that algorithm compares to manual measurement.

Spectral analysis of a fricative is properly thought of as forming an estimate of the spectrum. Analysis appropriate for vowels is a poor choice for fricatives, or any noise-excited sound. Computing a Discrete Fourier Transform (DFT) from a single windowed interval will result in a spectral estimate with a large error: at any frequency the error is proportional to the mean amplitude at that frequency. Increasing the length of the window will not change the error, counter-intuitively;

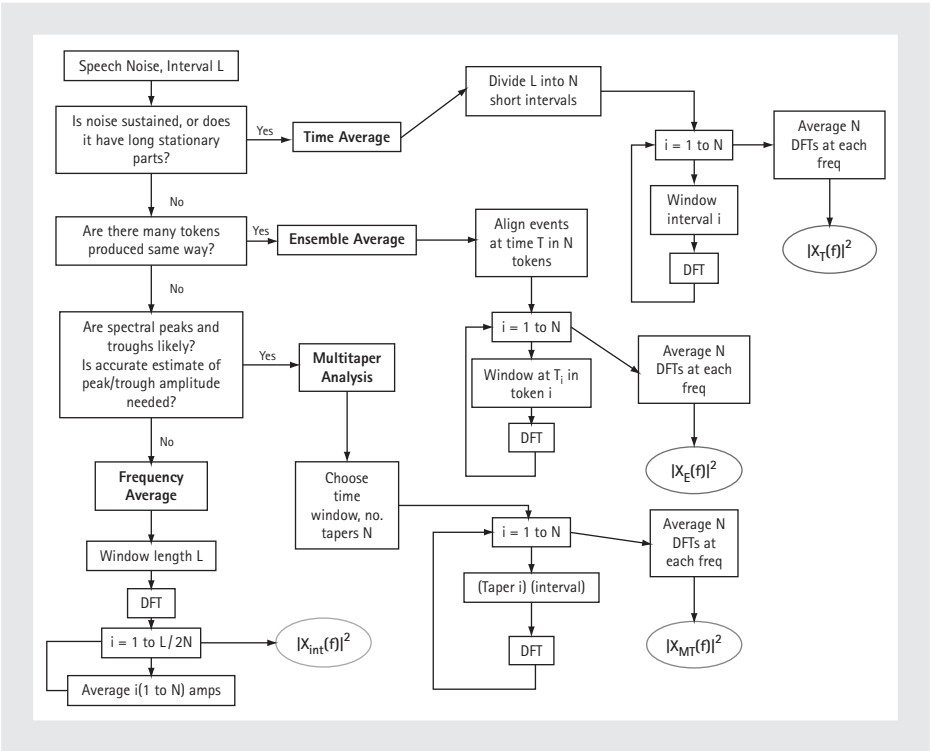

Figure 18.4.1. Diagram illustrating the basis for choosing a spectral averaging method for noisy sounds, and the general procedure for each method. The output in each case is the magnitude-squared spectrum.

some form of spectral averaging is needed. The particular form of averaging to be used depends on the nature of the signal, its likely spectral shape, and whether multiple tokens exist that were produced under the same conditions. Figure 18.4.1 is a diagram of the decision process and the main steps involved in each type of averaging; we now cover these in turn.

In *time averaging*, a single long interval is cut up into many shorter intervals; each of these is windowed and then the DFT is computed. The DFTs are then averaged (the magnitude-squared amplitudes, not the dB values, are averaged at each frequency). The windows may overlap by up to 50 percent; the error is reduced proportional to N, the number of DFTs averaged. The frequency resolution is inversely proportional to the length of each window. Time averaging is based on the assumption that the signal properties are stationary during the long interval, and therefore the short windows represent independent samples of the same random process. The accuracy of the estimate will depend on the extent to which this assumption is justified (see Jesus and Shadle 2002: 443–4).

When the fricative is very short, or when its time-varying properties are of interest, *ensemble averaging* may be a better choice. In this method, the same event is identified in several signals, which form an ensemble. The event obviously cannot be determined from spectral properties of the signal; it could be the fricative onset or offset, as determined from EGG or a combination of aerodynamic and articulatory signals. A DFT is computed of a windowed interval at that event in each signal and the spectra are averaged. As with time averaging, the error is proportional to N, the number of signals in the ensemble; the frequency resolution is related to the window length used on each signal. The underlying assumption is that each member of the ensemble is produced in the same way, so that the signal properties are the same at the event identified in each signal (see Jesus and Shadle 2002: 443–5).

A third method is *frequency averaging*. A single DFT is computed from a windowed signal, giving a spectral estimate with relatively high-frequency resolution but also high error. Then adjacent spectral amplitudes are averaged together so that every N amplitudes create a single new amplitude, reducing the frequency resolution but also decreasing the error. This does not require a long stationary interval, but it does introduce bias in the spectral estimate, particularly at peaks and troughs; that is, the spectral estimate will not converge to the true value as N, the number of points averaged together, increases. (See Shadle 2006 for more details.)

A fourth method uses *multitaper analysis*. Here, a single interval is chosen, and the speech signal is then multiplied by a set of different taper functions that are orthogonal to each other. DFTs are then computed for the products of the signal by each of the tapers, and averaged. The result gives a small error with good time and frequency resolution; the number of tapers used, N, and the interval length determine the resolution. This method does not require an ensemble nor an assumption of stationarity. Gray-scale multitapergrams can be plotted similar to spectrograms; see Blacklock (2004) for a complete exposition and examples of such plots of fricatives.

If a voiced fricative is averaged using any of these techniques, the fundamental and its harmonics will be averaged along with the noise. If f0 is relatively steady during the interval L, or similar across the members of the ensemble, the harmonics will reinforce each other and show up in the final averaged power spectrum. In frequency regions with steady f0 and some noise, the more averaging is done (the larger N is) the more the harmonics should stand out from the noise. If f0 changes during the analyzed interval, the frequency of a harmonic in the averaged spectrum will be blurred and the amplitude at any one frequency will decrease, with higher harmonics blurring more. A useful way around this is to estimate the spectrum twice, with a relatively low N (e.g. $N = 4$) and a high N ($N = 8$ or 10). The error for the noise components will be higher in the former, but the harmonics will be

clearer. If the harmonics are not much changed in the latter, then the estimate with higher *N* should be used.

Once a good spectral estimate has been computed, the problem remains to parameterize it to enable quantitative comparisons. Many different systems have been used, none of which is entirely satisfactory. The systems can be roughly divided into two types: model-based parameters, in which the acoustic parameters are related to articulatory or aerodynamic production parameters, and descriptive parameters, in which gross aspects of the acoustic spectrum are described. Model-based parameters include peak and trough frequencies (related to vocal tract resonances and anti-resonances, and thus to front cavity dimensions in particular), the lower-frequency limit of the high-energy region (as used by Strevens 1960), and spectral tilt (related to noise source properties, and thus to sibilance, effort level, and stress). Descriptive parameters include spectral moments (Forrest et al. 1988) and locus equations (Sussman and Shore 1996).

Shadle and Mair (1996) used both moments and model-based parameters on a large and varied corpus of two speakers. Neither set of parameters was sufficient to distinguish fricatives by place, though the model-based parameters did vary as expected with effort level, source type (whether sibilant or not), and location within fricative. A revised set of these parameters used on European Portuguese also captured source characteristics as predicted (Jesus and Shadle 2002).

Though Forrest et al. (1988) found that their use of spectral moments on a limited corpus resulted in /f/ and /θ/ being completely overlapped, Jongman et al. (2000) found a small but statistically significant difference between them. The degree of overlap in the distributions means that the moments are insufficient for distinguishing place, but can be useful for assessing differences due to phonetic context, dialect difference, and so on. In general, any parameter set distinguishes normal productions of /s/ and /ʃ/ within subject, and typically within gender; Matthies et al. (1994) used the first two moments in a longitudinal study of /s/-/ʃ/ distinction following a cochlear implant. Because of the gross descriptive nature of spectral moments, any difference in moments can be difficult to attribute to a particular articulatory or aerodynamic cause. They also must be computed with care: the same frequency range and amplitude floor must be used for all moments to be compared. Blacklock (2004) discusses these and other factors, and shows moments computed throughout intervocalic fricatives in English words, demonstrating token-to-token variation, cross-subject differences, and vowel context effects.

Finally, methods of decomposing a signal into harmonic and anharmonic components are worth mentioning because of their use in analysis of voiced fricatives. The different techniques used are reviewed in Jackson and Shadle (2001).

18.4.2 Aerodynamic analysis of fricatives

18.4.2.1 *Aerodynamic properties of fricatives*

18.4.2.1.1 *Flow noise sources*

The mean velocity in a constriction and that constriction's diameter determine the Reynolds number; when that is above a critical value, it indicates that the flow will become turbulent and can thus generate turbulence noise. The sound power generated by the turbulence noise depends on the velocity as V^n; the value of n depends on the type of flow noise source, which is determined largely by the shape of the tract downstream of the constriction (Shadle 1991, 1997, 2010). Knowing the type and location of sources is essential for speech synthesis, but is also important for understanding the factors that affect the sound generated. How loud a fricative is; how fast its noise increases at onset, when flow rate and constriction area are changing rapidly; how localized the source is, and therefore structures and dimensions that may be critical to its sound, are all consequences of the type of noise sources that predominate.

For a given configuration, the pressure drop across the constriction can be related to the velocity in the constriction through the orifice equation. The pressure drop can then be used to determine the source strength, and predict the radiated sound pressure. Stevens (1971) showed how the pressure drops across the glottis and the supraglottal constriction trade off, explaining why voiced fricatives tend to have weaker noise sources than their voiceless equivalents.

18.4.2.1.2 *Interaction of noise sources with vocal tract*

Classic speech models are based on an assumption that the source and filter do not interact. In a literal sense, there is interaction for fricatives, since the walls of the vocal tract downstream of the constriction define the filter, but also can affect noise generation and therefore the source. However, when it is possible to separate the two experimentally, by, for instance, moving the constriction and obstacle to different positions within a duct, it has been shown that the same source model can be used with different filters according to the position within the duct, to predict the far-field sound equally well (Shadle 1990). Thus, there is no true interaction for most instances of voiceless fricatives.

Whistles, however, cannot be modeled by independent source and filter; the whistle arises when an aerodynamic instability is reinforced by positive feedback. The "source" couples into resonances of the surrounding cavities, exhibiting hysteresis and other non-linear behavior. Whistles occur fairly commonly in fricative production, often in sibilants in rounded vowel context (Shadle and Scully 1995) but not necessarily restricted to that (Shadle et al. 2008). The high-amplitude, narrow-bandwidth peaks that are characteristic of whistling can coexist with turbulent rumble, leading to whistly fricatives that are distinctive acoustically yet apparently present no perceptual difficulties.

A third type of source interaction occurs in voiced fricatives. It has long been noted that the voicing source modulates the noise (Fant 1960; Flanagan 1972). It appears that the harmonic sound propagating from the glottis interferes with the turbulent jet in the front cavity, inducing a phase delay between the maxima of the harmonic and anharmonic components of the speech that is proportional to the length of the front cavity. The phase change occurs rapidly at fricative onset, and may be an additional cue to the place of the fricative, of particular use for intrinsically weak fricatives (Jackson and Shadle 2000, 2001).

18.4.2.2 *Measurement techniques in humans*

In mechanical equipment there are many ways to measure flow velocity and pressure. However, many of these are too bulky, too fragile, or both, to be used in human speakers. Volume velocity can be measured by use of the Rothenberg mask, a lightweight mask with screened vents that forms a seal around nose and mouth. The mask has resonances which should be considered when analyzing the acoustic signals (Badin et al. 1990), but it distorts much less than a typical oral-nasal mask without screening does. A pressure transducer mounted in one of the ports measures the pressure difference across the mask; this, with the known mechanical resistance of the screening, can be used to estimate the volume velocity through the mask. This can be used as an estimate of the volume velocity through the lips as long as the velum is sealed and jaw movements are minor.

Intraoral pressure can be measured by using the other port in the mask. A short airtube inserted through the port and between the lips, and with a pressure transducer on the end of the tube outside the mask, can measure the pressure during bilabial closure. A longer tube wrapped behind the molars can detect pressure upstream of an alveolar or palato-alveolar constriction. The air in the tube has its own inertia, and so effectively low-pass-filters the pressure readings. This can be avoided by using a thin pressure catheter so that the sensor itself is located where the intraoral pressure is desired.

The volume velocity at the lips and the intraoral pressure can be used together to estimate the area of the constriction (see Scully et al. 1992; Shadle 1997). The intraoral pressure during /p/ closure can also be used to estimate the subglottal pressure. The subglottal pressure can be measured more directly by using a balloon to measure the esophageal pressure, or by tracheal puncture, but these methods are more invasive and have a limited frequency range (up to 300 Hz). A different method using miniature pressure transducers suspended in the pharynx and through the glottis in the trachea can give pressure measurements with increased frequency range (up to 1 kHz) and, by combining measurements from two sensors, allow estimation of the volume velocity in the pharynx (Cranen and Boves 1985, 1988). Though these measurements were developed in order to study glottal flow, they hold promise for the study of fricatives with more posterior constrictions.

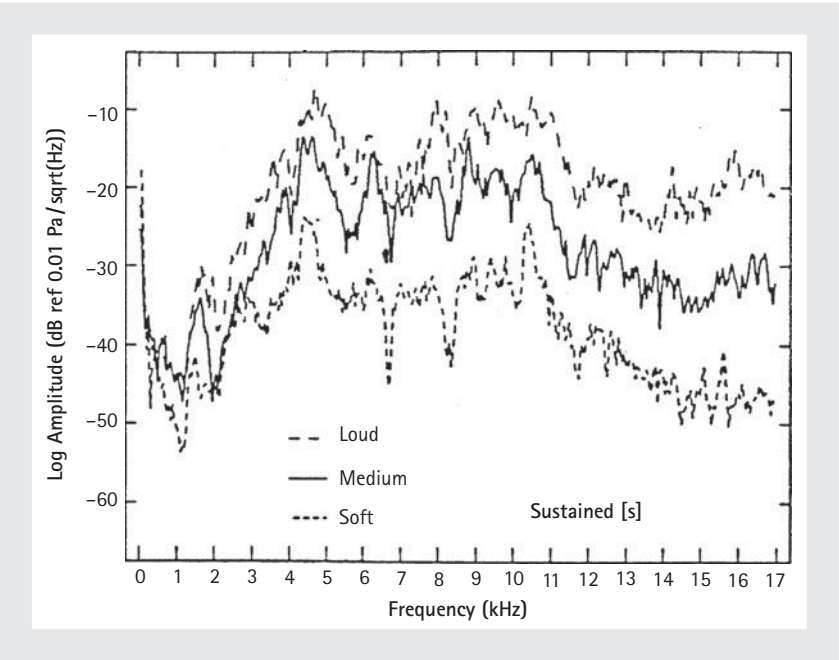

Figure 18.4.2. Time-averaged power spectra of three sustained [s] tokens, each at a different effort level, by an adult male speaker. The analysis used 25 20-ms Hanning windows, placed with 50% overlap, so that 25 DFT spectra were averaged at each frequency. (After Badin et al. 1994: Figure 3.)

18.4.3 Current findings

A typical aerodynamic sequence of events in a vowel-fricative-vowel sequence is well described by Scully et al. (1991, 1992; Shadle 1997: 45). Glottal abduction allows volume velocity, U, to rise; formation of the main constriction with the tongue, or lip and teeth, causes the intraoral pressure, P_{oral}, to rise. In voiceless fricatives the glottal abduction gesture is large and occurs more quickly than the constriction formation, leading to two peaks in U at onset and offset. These are not as large in voiced fricatives, and in fact the second peak may even be missing; P_{oral} does not increase as fast or reach as high a pressure. As a result, the constriction area estimated from U and P_{oral} has differently shaped contours: in voiceless fricatives A_c decreases rapidly, plateaus, then increases more gradually. In voiced fricatives there is no real plateau once A_c reaches a minimum. Corresponding to this, voiceless fricatives have longer durations on average than voiced fricatives, though the transitions are longer proportionately for voiced fricatives (Scully 1979; Mair and Shadle 1996; similar results for German are reported by Fuchs et al. 2007).

Scully found that noise generation occurred beginning when A_c dropped below 0.2 cm^2, though the amplitude of the noise continued to increase. This additional noise generation is visible in spectrograms or multitaper-grams: often the main peak is excited by noise first, and thereafter noise appears at higher and higher frequencies. Averaged power spectra computed at beginning, middle, and end of fricative tokens show the amplitude at the main peak increases from beginning to mid-fricative, but the amplitude at high frequencies increases more, consistent with an increase in flow velocity in the constriction (Shadle et al. 1996). This could be the result of the constriction area decreasing, volume velocity increasing, or a combination of the two. In sibilants, this could also result from the jaw rising, bringing the lower teeth to a more optimal position for noise production. A similar high-frequency boost to the spectrum occurs with increased effort level of a sustained fricative; an example is shown in Figure 18.4.2, in which the main peak at 4.5 kHz increases by 10 dB from Soft to Medium, and by 6 dB from Medium to Loud, but high frequencies (e.g. from 14–16 kHz) increase by approximately 15 and 12 dB, respectively. When voiced and voiceless fricatives are compared at the same effort level, a similar effect is shown, with the voiceless fricative having a greater amplitude and the largest amplitude difference occurring at the highest frequencies (Jesus and Shadle 2002).

In addition to changing the source strength, changes in area of the constriction during a fricative can also affect the transfer function that is excited by the noise source. As discussed by Badin (1989), when the area of the constriction is small enough relative to the area of the cavities on either side, those cavities are acoustically decoupled and the back-cavity resonances are canceled by anti-resonances. If the area of the constriction increases, back-cavity resonances increase in amplitude because they are no longer completely canceled (the poles and zeros move apart in frequency). The back-cavity resonances will be most apparent below the frequency of the main peak. Coupling can also increase if the constriction becomes shorter, or if the entrance or exit of the constriction becomes more tapered. Back-cavity resonances are often observable in the spectrum at the beginning and end of an intervocalic fricative, presumably when the constriction area is small enough to generate turbulence but still large enough to allow some coupling. An example is shown in Figure 18.4.3 of ensemble-averaged spectra analyzed at the beginning, middle and end of the [s] in repeated tokens of /pasa/. While all three spectra have a peak between 5 and 6 kHz, the peak at 2.8 kHz is significant only in the beginning spectrum, and drops by 10 dB from the beginning to the middle spectrum; this is very likely a back-cavity resonance that is incompletely canceled at the beginning of the [s]. Note that in the beginning spectrum, there is also energy in low-frequency peaks suggesting a fundamental and a few harmonics, indicating that voicing continues even after noise generation has begun. In the middle spectrum, in addition to the drop in energy at 2.8 kHz, the first main peak at 5 kHz has increased by approximately 6 dB, and the amplitude remains high for all frequencies above

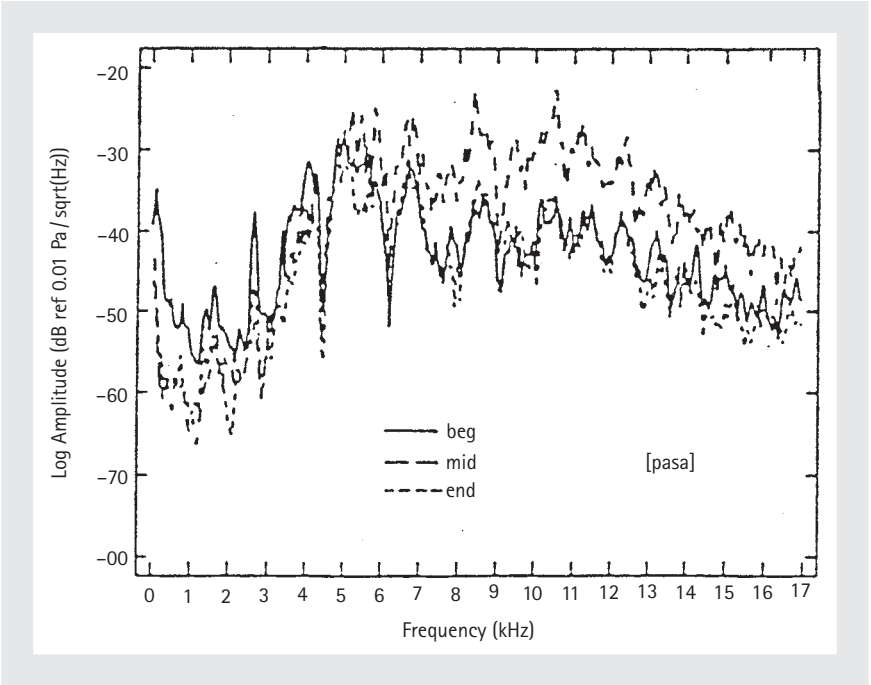

Figure 18.4.3. Ensemble-averaged power spectra of eight repeated tokens of /pasa/ spoken by an adult male speaker, with 20-ms Hanning windows located at the beginning, middle, and end of the [s] in each token used to generate the "beg, mid, end" spectra shown here. Eight DFT spectra were averaged at each frequency for each of the spectra shown. (After Shadle et al. 1996.)

5 kHz, indicating more efficient noise generation. In the end spectrum, the high-frequency amplitudes are similar to those at the beginning, but the low-frequency amplitudes remain low, indicating that noise generation has decreased but the constriction is still small enough to keep the front and back cavities decoupled. Apart from differences in constriction area, other differences in constriction shape may account for some of the acoustic differences observed token-to-token, in different contexts, or across subjects.

In general, the back cavity resonances contribute little to the spectral shape of a fricative. In sibilants, the lowest front-cavity resonance is emphasized by an anti-resonance at an even lower frequency, which results in a large dynamic range. For a more posterior place, the front cavity is longer, so its resonances are lower in frequency. This would seem to lead to two predictions, neither of which is borne out: automatically distinguishing between the place of fricatives should be a simple matter of determining the frequency of the lowest spectral peak, and men and women should have similar fricative spectra for a given place since pharynx length

does not affect the most salient acoustic characteristics. In fact, the constriction location of a given "place" can vary widely between subjects, leading to spectral peaks that vary widely in frequency. Even within a subject, phonetic context can cause enough variation in acoustic spectra to obscure what might seem to be a simple set of distinguishing characteristics. Regarding gender difference, women tend to have higher frequency ranges for spectral peaks. It is not understood why, though there is evidence that listeners' perceptual boundaries between /s/ and /ʃ/ shift to higher frequencies when the adjacent vowel has formants indicative of a shorter vocal tract (May 1976).

The spectra of non-sibilant fricatives are even more varied than those of sibilants. Simple models predict that for a constriction and therefore source location at the lip end of the vocal tract, there should be no front-cavity resonances; the entire transfer function should be made up of pole-zero pairs canceling each other, producing a low-amplitude, relatively flat spectrum. Figure 18.4.4 contrasts ensemble-averaged spectra of /θ/ in /piθi/ and /s/ in /pisi/ context. The [s]-spectra have an overall amplitude range of 45 dB, and an obvious range of peaks from 5–10 kHz. The [θ]-spectra have no obvious peaks, but are not completely flat; the beginning and middle spectra have an amplitude range of 30 dB, with a trough at low frequencies similar to that seen in [s]. Thus, while the simple model for non-sibilants is roughly true, such models do not include the lip horn, which is in essence a very short front cavity. This may account for the presence in many [f] tokens of a peak at approximately 11–12 kHz; a low-frequency trough in some tokens of [f] and [θ]; and evidence that the spectral shapes of these fricatives vary more with vowel context than do the sibilants (Shadle et al. 1996). Figure 18.4.5 shows spectra for [f] in /pufi/. As the spectra progress through the fricative, the peak at 1.3 kHz moves to 1.8, then 2.4 kHz, reflecting the changes in lip rounding and tongue position. As the peak frequency increases, the low-frequency trough deepens, and the amplitude increases from 8–11 kHz especially. The transitions appear to be more important than in sibilants (Harris 1958), but there are apparently cues to place in the frequencies above 10 kHz as well (Tabain 1998).

Three recent studies are aimed at establishing a firmer theoretical basis for determining the noise source characteristics for fricatives, and use somewhat different approaches. Zhang et al. (2002) determined the relative strengths of dipole and quadrupole sources generated by an orifice plate with a centered circular constriction in a duct. Their spectral decomposition method does not require an assumption of plane-wave sound propagation and provides a more general source-filter model with the means of testing whether interaction exists. Howe and McGowan (2005) used a theoretical approach to study whether the position of the upper and lower incisors in [s] contribute to the noise source. They concluded that the diffraction provides a high-frequency boost to the source spectrum, and found good agreement with existing speech data by Shadle (1991) and Badin (1989). They also determined the relationship between the overall sound pressure level in the

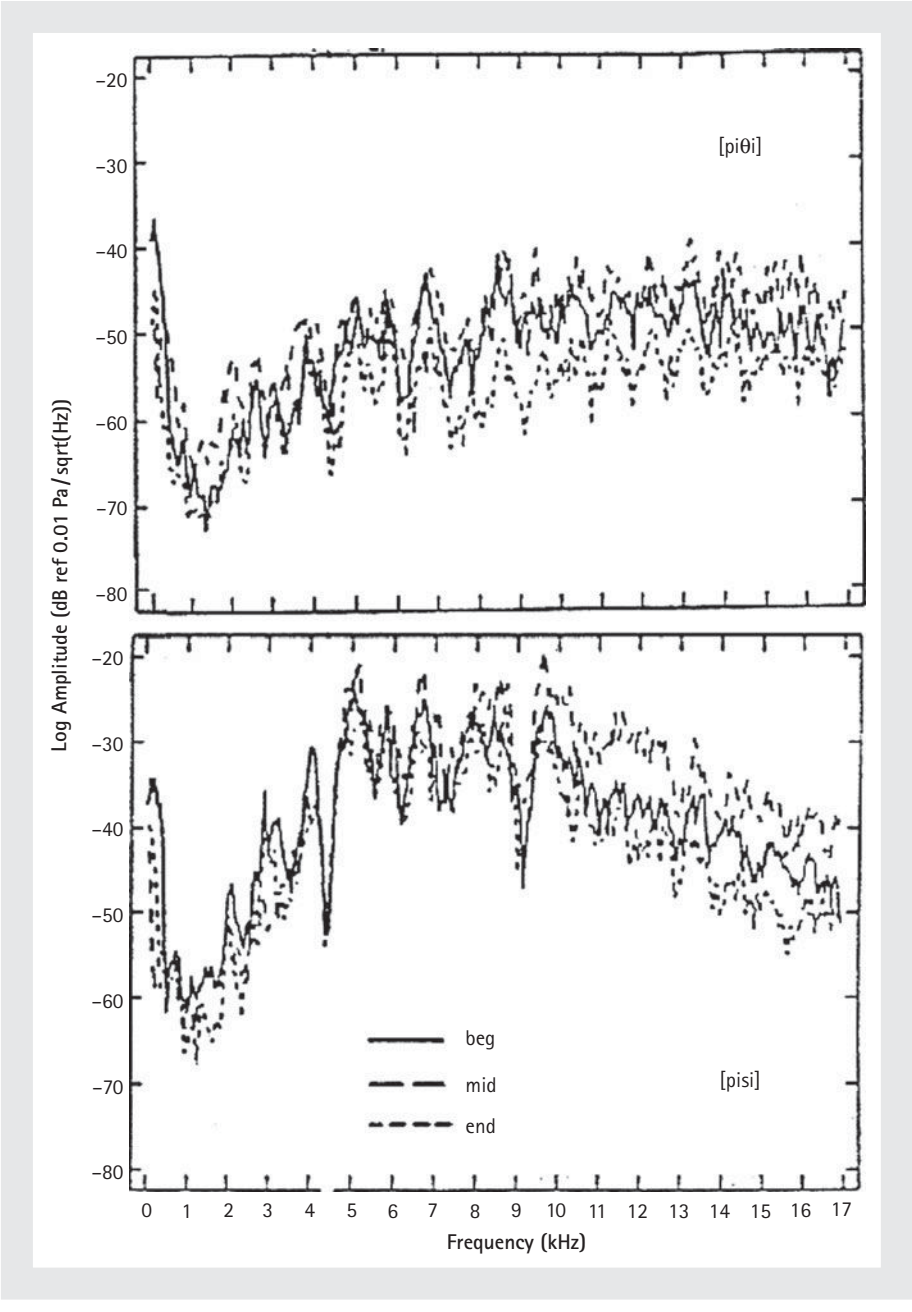

Figure 18.4.4. Ensemble-averaged power spectra of eight repeated tokens of /piθi/ (top) and /pisi/ (bottom) spoken by an adult male speaker, with 20-ms Hanning windows located at the beginning, middle, and end of the fricative in each token used to generate the "beg, mid, end" spectra shown here. Eight DFT spectra were averaged at each frequency for each of the spectra shown. (After Shadle et al. 1996.)

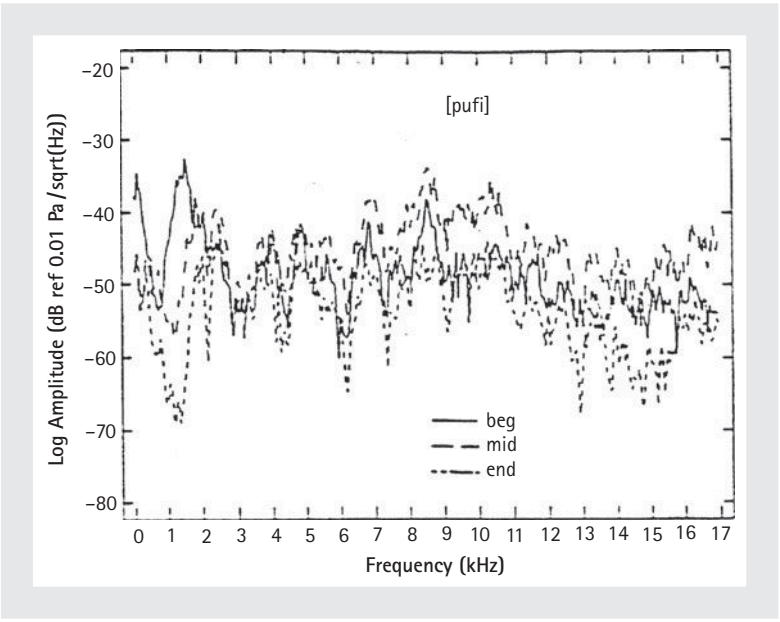

Figure 18.4.5. Ensemble-averaged power spectra of eight repeated tokens of /pufi/ spoken by an adult male speaker, with 20-ms Hanning windows located at the beginning, middle, and end of the [f] in each token used to generate the "beg, mid, end" spectra shown here. Eight DFT spectra were averaged at each frequency for each of the spectra shown. (After Shadle et al. 1996.)

far field and the intraoral pressure. Krane (2005) also used a theoretical approach to identify functions pertaining to the jet vorticity and the vocal tract shape, and explained how the relative bandwidth of these functions determines whether the output spectrum is harmonic—whistle-like—or broadband. He found that when vorticity encounters an area discontinuity, that is enough to generate turbulence noise; an obstacle per se is not needed. All three of these studies immediately suggest further work in terms of more realistic and more varied "vocal tract" shapes; by modeling the various contributions to the acoustic output, they aid in developing model-based ways to describe the acoustic spectra of fricatives.

Studies of human subjects encounter the difficulties of complex, at best only partially known, vocal tract shapes, and the natural variability that occurs in speech production. Koenig et al. (2008) used functional data analysis of airflow signals to separate variability due to amplitude and timing, allowing a more detailed comparison of production of [h, s, z] across subjects and age groups. This approach allowed them to define developmental stages in airflow management, and also demonstrate individual differences among adults in the variability of the three fricatives. Fuchs and Koenig (2009) measured palate contact via EPG and intraoral

pressure simultaneously during German clusters. Of the various contact parameters computed, they found that percentage contact was more strongly correlated with intraoral pressure than a measure related to place of articulation, as might be expected, but that the slope of the regression line—the exact relation between amount of contact and intraoral pressure—varied with speaker. Their data and the measures they used add to the sparse literature on articulatory-aerodynamic relationships in human production of obstruents, and suggest measurement and analysis methods that should facilitate such studies. It seems clear that although theoretical and mechanical model studies offer more control and understanding of the underlying physical processes, they are not so well suited to exploring the extent of, and reasons for, the variability evident in articulatory, aerodynamic, and acoustic studies of human fricative production.

CHAPTER 19

PROSODIC ANALYSIS

EXPERIMENTAL METHODS AND PARADIGMS FOR PROSODIC ANALYSIS

PILAR PRIETO

DATA COLLECTION FOR PROSODIC ANALYSIS OF CONTINUOUS SPEECH AND DIALECTAL VARIATION

BRECHTJE POST AND FRANCIS NOLAN

The contributions in this chapter present a detailed overview of research methods investigating prosody. Prieto reviews experimental approaches to prosodic analysis. Post and Nolan describe design and analysis of prosodic corpora of both naturalistic and controlled speech.

19.1 Experimental methods and paradigms for prosodic analysis*

Pilar Prieto

19.1.1 Introduction

There is a long tradition of experimental research in the field of prosody, as different aspects of speech production and perception related to prosody have often been part of traditional laboratory and phonetics investigation. However, in recent years, the development of a set of laboratory tools to investigate language performance and its neurocognitive basis has prompted a new experimental approach to prosody research, in addition to the linguistic approaches that have traditionally been used. This new approach encompasses a wide range of methodological paradigms such as acoustic analysis of speech productions, direct measurement of articulator movements, judgments and reaction times obtained during identification and discrimination tasks, measurements of brain activity and patterns of attention in babies.

In this section we provide an overview of some of the experimental methods and paradigms that are currently used for the phonetic and phonological analysis of prosody. Given the large amount of literature on prosody research from the linguistics, speech, and psycholinguistic communities it would be impossible to provide an exhaustive list of relevant research reports. Instead, selected examples of such work and the methodological paradigms used are provided. Importantly, we show how these methodological advances have contributed in various ways to our understanding of a large range of issues in the field of speech prosody, as experimental findings have been able to empirically test the various predictions posed by different models of prosody.

19.1.2 Acoustic analysis

Production studies have been widely used for phonetic and phonological analyses of prosody. There is a long tradition of using acoustic analysis of speech productions under various elicitation conditions in the field or in the laboratory. Though most such studies work with laboratory speech and within the strict demands of corpus design and experimental control, other studies are increasingly working with large

* I am grateful to G. Elordieta, S. Frota, M. Grice, C. Gussenhoven, D. Mücke, B. Post, M. Swerts, and to the editors A. Cohn, C. Fougeron, and M. Huffman, for their comments on an earlier version of this manuscript. This research has been funded by projects FFI2009-07648/FILO and CONSOLIDER-INGENIO 2010 CSD2007-00012 (awarded by the Ministerio de Ciencia e Innovación) and by project 2009 SGR 701 (awarded by the Generalitat de Catalunya).

databases of read and spontaneous speech (Post and Nolan, this chapter; Cole and Hasegawa-Johnson, this volume and Warner, this volume).

In prosody research, acoustic parameters such as fundamental frequency (f0), duration patterns, and intensity or amplitude patterns have been widely investigated. In the last two decades, the increasing availability of freeware to carry out acoustic analyses, in particular the Praat program (Boersma and Weenink 2009), has made it much easier to carry out acoustic analyses even without access to a specialized phonetics laboratory. In most of these production experiments, the analysis of the above-mentioned acoustic parameters is manually performed. However, corpus analyses increasingly resort to automatic procedures such as automatic segmentation procedures (for the analysis of duration), automatic detection of f0 turning points, and even automatic prosodic labeling—see Cole and Hasegawa-Johnson (this volume).

There is a long tradition of research that has been concerned with the acoustic characteristics of diverse prosodic phenomena such as stress prominence or prosodic boundary phenomena (see the classical experiments by Fry 1955, 1958 on word stress in English). Another issue that has been the focus of both production and perception research is intonational meaning. One of the crucial issues in intonational phonology is how phonetic elements encode intonational contrasts. Since the development of the Autosegmental Metrical approach to intonation (Pierrehumbert 1980; Beckman and Pierrehumbert 1986; and others), tonal alignment has been shown to play a central role in encoding intonational contrasts. For a thorough review about work done on tonal alignment and tonal association, see Arvaniti and D'Imperio (this volume). Detailed acoustic studies have also served to develop predictive models of phonetic realization. For example, an important goal of intonation research has been to develop predictive models of tonal alignment (Silverman and Pierrehumbert 1990 for English; Prieto et al. 1995 for Spanish). Overall, this work has shown that L and H targets are independently aligned relative to the syllable and that the accentual rise is neither of fixed slope nor of fixed duration, that is, the fixed rise-time hypothesis cannot be maintained (Arvaniti et al. 1998; Prieto et al. 1995; among others).

19.1.3 Articulatory analysis

An important line of investigation within the Articulatory Phonology framework has used kinematic data of articulator gestures obtained using electromagnetic midsagittal articulography (EMMA) to study the intragestural dynamics of boundary-adjacent lengthening phenomena (Byrd and Saltzman 1998). This work interprets boundary-adjacent lengthening as a local slowing of the gestures in the immediate vicinity of sufficiently strong prosodic boundaries at multiple levels. Thus, just as the syllable edges influence intergestural timing, other types of

prosodic boundaries have been shown to influence it too (for a review, see Byrd 1996a, Krakow 1999, and Turk, this volume).

In recent work, researchers have started paying attention to the coordination between tonal gestures (measured as f0 turning points) and oral constriction gestures. Recent work by D'Imperio, Espesser, Loevenbruck, Menezes, Nguyen, and Welby (2007), Mücke et al. (2006), Mücke, Grice, Becker, and Hermes (2009), Prieto et al. (2007), and Mücke, Nam, Prieto, and Goldstein (2009) has investigated tonal-oral constriction alignment patterns for three different languages (Italian, German, Catalan respectively) by using the magnetometer (EMMA). This work shows that the temporal coordination between pitch movements and articulatory gestures is in many cases stronger than that between acoustic events and f0. Interestingly, there is some variation as to the articulatory landmark which serves as an anchor for the tonal target. For example, in German nuclear LH accents, the H peaks co-occurred with the intervocalic C target, whereas in pre-nuclear accents peaks co-occurred with the target for the following vowel (accent shift, Mücke, Nam, Prieto, and Goldstein 2009). In Catalan rising-pitch accents it was the consonantal peak velocity rather than the maximum constriction for the consonant which served as the landmark (Prieto et al. 2007). Such an apparently small alignment difference in the articulatory anchor type may be used by speakers to make phonological distinctions, as in Neapolitan, where H in L* + H (questions) aligns with the maximum constriction, and H in L + H* (statements) with peak velocity (see D'Imperio, Espesser, Loevenbruck, Menezes, Nguyen, and Welby (2007)). Recently, Gao (2008) has formalized the coordination patterns between tonal and vocalic gestures within the Articulatory Phonology framework, proposing that the different alignment patterns can be analyzed as differences in phasing between supraglottal and tonal events.

Another research area in which facial and gestural articulatory analysis has been performed is that of visual prosody. In recent years researchers have analyzed quantitatively the head and facial movements that accompany speech, and analyzed the visual correlates of prominence and focus, question intonation, as well as the audiovisual expression of affective functions such as uncertainty. For example, Cavé et al. (1996) analyzed the production of eyebrow movements and their association with tonal rises in French. They found that in only 71 percent of the cases was there an association, while 38 percent of the eyebrow movements occurred while the subject was not speaking. Thus, eyebrow movements may serve as back-channel signals and play a role in turn-taking during conversation.

19.1.4 Categorization: Identification and discrimination tasks

In the past few decades there has been a significant increase in the number of studies that take a psycholinguistic or cognitive approach to prosody research.

Common research paradigms consist of behavioral experiments in which subjects are presented with stimuli and asked to make conscious decisions about them. Such experiments can take the form of an identification test, a discrimination test, a gating task, etc.

One method that comes from the study of consonantal contrasts and which has been applied to the study of tonal and intonational contrasts across languages is the well-known Categorical Perception Paradigm (CP paradigm; Liberman et al. 1957; and see Iverson, this volume for a thorough review of this paradigm). The CP paradigm involves, first, an identification/classification task in which the listeners have to categorize stimuli taken from a continuum, and secondly, a discrimination task in which listeners are asked to judge pairs of stimuli as being either the same or different. The solid lines in Figure 19.1.1 show the ideal S-shape functions of responses for the identification task, i.e. an abrupt shift from one category to the other. The dashed line shows the ideal function for the discrimination task: if perception is categorical, discrimination between stimuli should be more accurate between categories (where a peak of discrimination is obtained) than within them.

The CP paradigm has been applied to both tonal languages (Francis et al. 2003; Francis and Ciocca 2003) and intonational languages, both for *boundary tones* (Remijsen and van Heuven 1999; Post 2000; Schneider and Linftert 2003; Falé and Faria 2006; Prieto et al. 2008) and for *pitch accents*, in terms of either differences in peak alignment (Kohler 1987; D'Imperio and House 1997; Chen 2003; Gili Fivela 2009; Grice and Savino 2011; Dilley submitted) or differences in pitch height (Ladd and Morton 1997; Vanrell 2007; Prieto et al. 2008). Yet the application

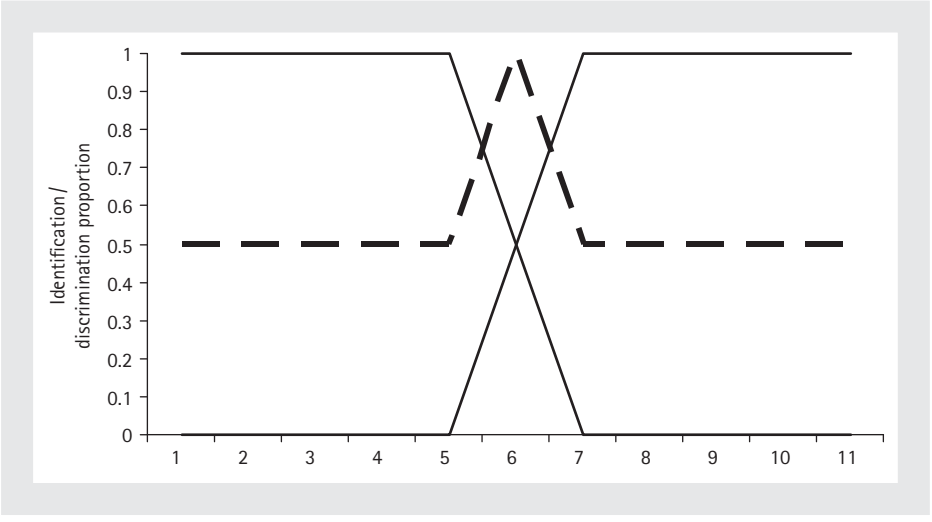

Figure 19.1.1. Idealized identification and discrimination functions—solid lines and dashed line respectively.

of this paradigm to intonation research has met with mixed success and there is still a need to test the convergence and degrees of adequacy of this particular experimental method (see Ladd and Morton 1997; Chen 2003; Gussenhoven 2004, 2006). Although claims have been made of categorical perception for a particular contrast, in the majority of cases no discrimination peaks appear in the crossover of categories revealed by the identification test (Remijsen and van Heuven 1999; Falé and Faria 2006; Prieto et al. 2008). Only three studies present clear evidence of categorical perception, with a clear discrimination peak in the expected position (Kohler 1987; Schneider and Linftert 2003; Vanrell 2007). Chen (2003) claims that the use of reaction time (RT) measures is essential in conjunction with the results of identification tasks to help interpret whether tonal categorical perception effects are linguistically real. Mean RTs are longer for across-category stimuli (that is, ambiguous stimuli) and shorter when the stimulus corresponds to an unambiguous category. The importance of the reaction time data has been confirmed in a number of studies (Falé and Faria 2005; Vanrell 2007; Savino and Grice 2008; Prieto et al. 2008, among others). Other tasks have also been proven to be successful to discover whether two pitch contours are phonologically distinct. In the imitation task, subjects are asked to imitate or try to reproduce the target contours (e.g. Pierrehumbert and Steele 1989). As an example of a semantic task, Gussenhoven and Rietveld (2000) undertook a perception experiment with the high-rise pitch configuration (H* HH%) and the low-rise configuration (L* HH%) in Dutch. Listeners were asked to rate "perceived surprise" with a continuum of stimuli with different pitch ranges. Listeners showed a different pattern of responses for both categories, lending support to the hypothesis that the two pitch contours are categorically distinct in Dutch.

19.1.4.1 *The Gating Paradigm*

The Gating Paradigm methodology was initially developed to study online word recognition (Grosjean 1980, 1996). The goal of this online speech processing task is to study the speakers' online performance with an identification task when only part of the speech signal is available. In this experimental paradigm, target sequences are cut into smaller pieces. These gated stimuli are presented to subjects in a sequential order of increasing duration. Subjects are asked to identify the target unit and rate their level of confidence in their own answers. Two independent measures are important, namely, the isolation point (or the location where correct identification is achieved and maintained over fragments), and the recognition point. The recognition point will be determined by the rating of the confidence level given by subjects, which will be reached when a stimulus is first rated as "sure" and this rating is maintained throughout the sentence.

In prosody research, gating tasks have been used to study the contribution of stress information in spoken word recognition or word spotting. For example,

Lindfield et al. (1999) tested the role of stress on word recognition in English by using a word onset gating technique that allowed subjects to hear only the full prosodic pattern of a word (number of syllables and syllabic stress), deprived of segmental information beyond that contained in the onset gate. Their results suggest that word prosody in English is represented in the mental lexicon and is indeed used by listeners in spoken word recognition.

In the last decade, the gating paradigm has been used to investigate the contribution of early intonational features in the processing of interrogative meaning (Face 2005 for Spanish; Falé and Faria 2005 for European Portuguese; Vio and Colas 2006 for French; Petrone 2008 for Neapolitan Italian). In these studies, listeners had to identify sentence type when presented with auditory speech stimuli that were gated in specific sentence locations. In general, findings indicate that listeners are very accurate in identifying the sentence type early in the utterance, and that the first post-tonic syllable tends to be a good isolation point. These studies demonstrate that early prosodic information provides listeners with enough cues to access and recognize sentence type well before the end of the utterance.

19.1.4.2 *The Priming Paradigm*

Psychologists have developed several ways of probing for association among representations in memory, and one of them is priming tasks (see Schiller, this volume). This procedure allows the comparison of reactions to a target stimulus when the presentation of the target is primed or is not primed by the immediately prior presentation of another stimulus. For example, when adult speakers of English are asked to perform a self-paced reading task (that is, to read aloud a series of words presented one by one on a video screen), they respond more quickly if they hear a semantically related word just before seeing the target word. Psychologists interpret priming results as evidence that representations of semantically related words are associated in memory.

The cross-modal priming paradigm allows moment-by-moment activation of word meanings to be accurately tracked during ongoing spoken sentence comprehension as subjects are exposed to visual cues. Cross-modal priming has been employed in research to test the effects of lexical stress on spoken language processing and word recognition. For example, Cooper et al. (2002) conducted a study in which English-speaking participants were asked to perform a lexical decision task. They were presented with a visual target (e.g. a card reading "MUSIC") after a one-syllable-long fragment. Responses were faster when the prime's stress matched the target word (e.g. *mu* from *music*) than when it did not (e.g. *mu* from *museum*).

Recent research has demonstrated that prosodic perception is very sensitive to multimodal dimensions, that is, audiovisual integration. One of the most compelling demonstrations of multimodal speech perception is known as the McGurk effect (McGurk and MacDonald 1976). This effect has been extensively replicated

across conditions: the experiments involve discrepant audible and visible utterances which are dubbed so as to be synchronously produced. The results show that what hearers perceive is strongly influenced by the visual component, demonstrating that many times the visual component overrides the audio component. A particular area which has used priming techniques is that of visual prosody. Swerts and Krahmer (2008) investigated the importance of visual cues to prominence. Their perception experiment probed the relation between auditory and visual cues by means of testing reaction time to congruent and incongruent stimuli, that is, stimuli where auditory and visual cues to prominence were occurring on the same word (congruent) or on different words (incongruent). Their results showed that participants can more easily determine prominence when the visual cue occurs on the same word as the auditory cue, while displaced visual cues hinder prominence perception.

The facial affect decision task is one of the latest approaches to gathering new information on how emotional prosody is processed for meaning and then integrated with other communicative events. This technique analyzes emotional congruity priming effects in perceptual processing. For example, Pell (2005) undertook a study with this task to see how implicit activations of prosodic stimuli representing an emotion were related to specific facial expressions. The subjects had to listen to auditory stimuli over headphones and at the same time judge a face on a computer screen. Auditory stimuli were spoken in different emotional tones and facial stimuli were portraits of an actor conveying either "true" emotional expressions (e.g. happy or sad) or "false" emotional expressions (resembling a "grimace" or facial expression that does not represent basic emotions). The results indicated that the prime-target relationship induced both error and latency responses to facial expressions, revealing that emotional prosody biases online processing of "emotional faces."

19.1.4.3 *The Eye-Tracking Paradigm*

Monitoring eye movements while a person is silently reading written text is a procedure that was originally used to infer the processes that underlie skilled reading. There is growing evidence that in silent reading, readers tend to project a default prosodic structure (what Fodor 2002 calls the implicit/silent prosody hypothesis) onto the written words that can then influence syntactic processing, just as "real" prosody does when we listen to speech. In these experiments, eye movements are tracked by two miniature cameras that record the size of the pupil and the corneal reflection for each eye. By computing the position of the eyes with respect to the position of the markers, one can infer, in real time, the direction and fixations of the participant's eye gaze on a predefined area (see Speer, this volume for detailed review).

A rapidly expanding community of psycholinguists is now using the eye-tracking paradigm to study the role of prosody in spoken language comprehension. Dahan et al. (2002) study provides a good example of investigation of the role accentuation

and deaccentuation play in the processing of information structure. They used the action-based version of the so-called visual world paradigm (Tanenhaus et al. 1995), which contains a visual display consisting of four black and white line drawings representing four distinct objects, arranged within a grid. Participants had to perform a word recognition task and pick up one of the competing objects (for example, *candle* vs. *candy*) while their eye movements were monitored. They were prompted by a series of instruction sequences such as *Put the necklace below the candle. Now put the CANDLE above the square.* The target in the second instruction was either accented or unaccented, and referred to either the picture mentioned in the first sentence or to a previously unmentioned picture. The pattern of eye fixations to the target objects demonstrated that deaccented nouns were initially biased toward a given and anaphoric (mentioned) entity, whereas accented nouns were biased toward a non-anaphoric new entity.

More recently, other studies have successfully used the eye-tracking paradigm to study the role of pitch accent type in online processing of information structure. Using the same technique as Dahan et al. (2002), Chen et al. (2007) found that rise-falls create a strong bias towards newness, whereas rises and also deaccentuation create a strong bias towards givenness. Watson et al. (2008) used the same paradigm to investigate whether pitch accent type in English (namely L+H* vs. H*) can bias listeners toward interpreting a temporarily ambiguous noun as referring to a discourse-given or discourse-new entity. Their results show that the interpretive domains of both pitch accents overlap, that is, L+H* creates a strong bias toward contrast referents, whereas H* is compatible with both new and contrast referents. For a review of the eye-tracking paradigm applied to prosody research, see Watson et al. (2006, 2008), and see also the examples given in Speer (this volume).

19.1.4.4 *The Head-Turn Preference Procedure*

The Head-Turn Preference Procedure (HPP) has been shown to be a valuable technique for testing attentional patterns in babies (see classic reference Kemler et al. 1995; Maye, this volume for a review of this paradigm). This procedure records the summary patterns of the babies' eye fixations and does not require very detailed eye-tracking technology. Though the procedure is of limited applicability, since infants have not yet learned the lexicon of their language, it can indirectly tell us significant facts about the acquisition of prosody.

Several types of cues to word and phrase recognition have been experimentally studied, such as phonotactic cues, stress patterns, prosodic boundary cues, with many studies showing that infants at a very early age are sensitive to these prosodic cues. For example, Jusczyk, Cutler, and Redanz (1993) used the HPP procedure to examine the potential role that sensitivity to predominant stress patterns of words might play in lexical development. In English, the majority of words have stressed initial syllables, and the authors demonstrated that by 9 months of age American

infants listen significantly longer to words with strong/weak stress patterns than to words with weak/strong stress patterns.

19.1.5 Neurophysiological and neurobehavioral methods

With the advent of modern neuroimaging techniques such as event-related potentials (ERPs) and functional magnetic resonance imaging (fMRI), there has been increasing interest in investigating the neural mechanisms involved in the processing of prosody. These techniques have also been brought to bear on many of the questions addressed above.

19.1.5.1 *Event Related Potentials (ERPs)*

ERPs can be reliably measured using electroencephalography (EEG), a procedure which measures electrical activity of the brain and which allows for the non-invasive measuring of brain activity during cognitive processing—see Schiller (this volume) and Idsardi and Poeppel (this volume) for a thorough review of this technique. One important and robust response used with event-related potentials is the so-called magnetic mismatch negativity (MMN), which is elicited when the auditory perceptual system detects a mismatch between a neural representation of a frequently repeated stimulus (the standard) and a stimulus deviating in at least one parameter (the deviant). This is the so-called mismatch paradigm. Since its discovery, this well-known MMN index of automatic acoustic change detection has also been found to be a sensitive indicator of long-term memory traces for native language sounds (e.g. phonemes, syllables) and native lexical words. When comparing MMNs to words and meaningless pseudowords, researchers have detected larger amplitudes for words than for meaningless items. This is interpreted as a neurophysiological signature of word-specific memory circuits/cell assemblies activated in the human brain in a largely automatic and attention-independent fashion. Other studies have found evidence that the MMN reflects automatic syntactic and semantic processing commencing as early as ~100 ms after relevant information becomes available in the acoustic input; see Shtyrov and Pulvermuller (2007) for a review of ERP work related to language functions.

A number of ERP studies have focused on various aspects of prosody processing. For example, some studies have investigated the neural correlates of intonational phrase boundaries, which elicit a specific component in event-related brain potentials, the so-called closure positive shift. Although there is general agreement on the disambiguating role played by intonational phrase (IP) boundaries, the role of lower phrase boundaries seems to be still an open issue, as does the role of word boundaries. Li and Yang (2009), on the basis of EEG measures, investigated whether

prosodic boundaries at different levels could evoke the closure positive shift reflecting prosodic boundary perception: they found that listeners were very sensitive to both intonational phrase boundaries and phonological phrase boundaries.

19.1.5.2 *Brain imaging techniques (fMRI)*

Functional Magnetic Resonance Imaging (fMRI) is a type of MRI scan that measures the hemodynamic response to neural activity in the brain (see Idsardi and Poeppel, this volume for a thorough review of this technique). Since the early 1990s, fMRI has come to dominate the brain-mapping field due to its low invasiveness and lack of radiation exposure. One of the advantages of this technique is that it provides a high spatial resolution, but on the other hand it has poor temporal resolution compared with EEG.

By using the same mismatch paradigm, fMRI studies have been recently applied to the analysis of brain activation localization patterns found during tonal and intonation processing. Gandour et al. (2003) examined neural responses to the discrimination of differences in illocutionary force (questions vs. statements) and emotional valence (happy vs. angry vs. sad) in Chinese utterances in Chinese and English speakers. In both groups of subjects, discrimination of illocutionary force compared to a passive listening baseline led to widespread increased neurological signal in both hemispheres, suggesting bihemispheric processing of intonation. In general, recent studies confirm a right hemisphere dominance during the processing of intonation contrasts (Gandour et al. 2003; Friederici and Alter 2004; Meyer et al. 2004; Fournier et al. 2010).

19.1.6 Conclusion

In the last two decades, the field of prosody has witnessed a significant growth in interdisciplinary research that integrates behavioral experimentation with neurophysiological and neuroimaging studies. As we have seen, a wide range of methodological paradigms are now available for prosody research, including acoustic and articulatory analyses of speech productions, judgments and reaction times obtained during identification, discrimination, gating and priming tasks, and measurements of brain activity, eye movements, and infant attention patterns. Many laboratory phonologists have started to use these diverse and complementary methods to address questions in prosody research, instead of relying on only a small range of methods. As noted earlier, this section has attempted to illustrate a selection of methods that are representative of experimental approaches to the study of prosody. We have selected some representative articles for each paradigm in order to provide a glimpse into some of the issues of current interest in the field, such as the categorical or gradient perception of intonation, the target

vs. movements approach to intonation, and the use of prosodic information in the recognition and processing of lexical representations, syntax, and discourse. We believe that the full exploitation of these methodological advances will provide important answers and will most likely lead to even more improved experimental paradigms.

19.2 Data collection for prosodic analysis of continuous speech and dialectal variation*

Brechtje Post and Francis Nolan

19.2.1 Introduction

A better understanding of the functions and realizations of prosody in continuous speech requires detailed studies of prosodic phenomena in context, in existing speech corpora as well as in more controlled purpose-built data sets. In areas of research in which we expect to observe prosodic variation for which the sources are still unknown, for instance in dialects or second-language learners, large-scale analyses of such data could be particularly useful in improving our understanding of those sources (see Gut 2007). Such analyses are time-consuming and error-prone, since they still require much hand-labeling of the data, but with the increasing availability of (semi-)automatic techniques for prosodic analysis (e.g. Momel/INTSINT: Hirst and Espesser 1993; Hirst et al. 2000; Prosogram: Mertens 2004), large-scale studies are rapidly becoming more feasible (e.g. Govender et al. 2007; see Cole and Hasegawa-Johnson, this volume).

In this section, we provide an overview of the main considerations that will determine the design of prosodic speech production studies carried out in a laboratory phonology framework. Since this section is primarily concerned with research on prosodic phenomena in continuous speech, the overview will focus on studies in which speech corpora were specifically developed for this purpose. A speech corpus is understood here as a collection of computer-readable speech data with associated annotation and documentation that allow the data to be used for research purposes (Gibbon et al. 1997). By this definition, an important difference between a purpose-

* We would like to thank the editors, Pilar Prieto, David House, and an anonymous reviewer for their very helpful comments and suggestions. Preparation of this contribution was supported by grants from the ESRC (RES-061-25-0347) and the European Community (MRTN-CT-2006-035561).

built controlled production experiment and a speech corpus is that experimental data are not normally made available to the wider academic community. However, the same basic principles govern the collection of both types of data, and more-controlled experiments will be referred to where relevant in order to illustrate the discussion—if only because speech corpora that were specifically created to investigate prosodic phenomena are quite rare.

In Section 19.2.2, the role of the research focus in determining the design of the study is discussed, and an outline given of areas of research and types of prosodic phenomena that have been studied in a laboratory phonology framework. In Section 19.2.3, different empirical approaches that have been prevalent in prosodic corpus research are reviewed, and in Section 19.2.4, we evaluate the advantages and disadvantages of different elicitation techniques that can be used to obtain more or less naturalistic data.

19.2.2 Research focus: Defining the scope of the corpus

Prosody is studied from many different perspectives within the framework of laboratory phonology. The perspective determines to a large extent which phenomena are investigated, what methodologies are used, which technical constraints apply, and what variables are introduced into the design of the study. For instance—as the examples discussed below will illustrate—variables of interest to a sociolinguist or language typologist will tend to include speaker characteristics such as age, socioeconomic, educational, and regional background, as well as contextual factors like speaking style (see Docherty and Mendoza-Denton, this volume). However, unlike the sociolinguist, the typologist is likely to pursue cross-linguistic comparisons of features which are hypothesized to differ in theoretically relevant ways or can be considered as universal. Different language backgrounds may also be of interest in developmental studies of prosody in first- and second-language acquisition, but here, comparisons of speaker groups at different stages of acquisition will be a main factor of interest. By contrast, prosodic research in speech pathology is quite unlikely to rely on large samples of directly comparable data, since there will be more between-speaker variation in the smaller sampling populations involved.

The prosodic phenomena which most commonly form the object of study in laboratory phonology research are prosodic phrasing, intonation, focus and accentuation, and rhythm (see Chapter 11 this volume). Prosodic phrasing is the chunking of speech into linguistically relevant units like words, phrases, and utterances. In the framework of Prosodic Phonology, the constituents are hierarchically organized at different levels of structure (Selkirk 1986; Nespor and Vogel 1986; see Truckenbrodt 2007b). The edges of constituents are cued by juncture phenomena like final lengthening (e.g. Wightman et al. 1992), changes in pitch (e.g. Streeter

1978), and laryngealization and glottalization (see Ernestus, this volume and Turk, this volume).

An example of a speech corpus which was specifically developed to investigate phrasing is the Romance Languages Database (RLD; Elordieta et al. 2003; D'Imperio et al. 2005). The aims of the RLD project were to investigate (1) patterns of placement of intonational boundaries, (2) the influence of syntactic and prosodic factors on boundary placement, and (3) the phonetics and phonology of the intonational boundaries in a set of typologically related languages (Central Catalan, Standard and Northern European Portuguese, Neapolitan Italian, and Central Peninsular Spanish were included in the corpus). In a cross-linguistic typological study of this type, direct comparability of the data between the languages is of paramount importance (e.g. Grabe and Post 2004). In the RLD, comparability was achieved by eliciting a set of sentences which were cross-linguistically matched for syntactic structure and length. By systematically varying these factors of interest in the same way across the languages, their relative contribution to phrasing could be compared directly, revealing cross-linguistic differences in the preferences for grouping of subjects with following verb-object constructions, but also a clear cross-linguistic similarity in the role of length, which was found to play a larger part than syntactic branching in determining boundary placement (as found in French by Post 1999).

By focusing on intonational phrase boundary marking, the RLD project investigated one particular function of intonation. This particular use of intonation is grammatical, since it can affect the propositional content of the utterance. Intonation can also be used grammatically when the choice of contour affects the pragmatic interpretation of an utterance in its context; for instance when a rise is used to signal that an utterance is intended as an interrogative. The grammatical use of intonation is also referred to as its linguistic use, in that it reflects formal structure in language, and there is no one-to-one relation between the forms and functions it expresses (Gussenhoven 2004). Two further uses of intonation can be identified (House 2006): an indexical function, when intonation conveys meanings that are paralinguistic and speaker-oriented; and a discourse function, when it signals the structure of a discourse (e.g. the marking of a new topic, Wichmann 2000) or when it signals functions in conversational interactions like turn-taking and floor-holding (e.g. Couper-Kuhlen and Selting 1996).[1] In many languages, intonation can also convey aspects of information structure like focus distribution.

These different functions of intonation are all investigated in the Spontal project (Edlund et al. 2010), in which a multimodal spontaneous speech database is specifically developed for research on talk-in-interaction. The data consist of audio and

[1] Another communicative function of intonation that falls outside the scope of this section is that it can help separate voices in adverse listening conditions by providing cues to auditory continuity (the cocktail party effect; Brokx and Nooteboom 1982).

video recordings of speakers engaging in different dialogue functions, with the aim of improving our understanding of the visual and acoustic properties of grammatical functions like prominence, grouping, and phrasing; of indexical functions like speaker and listener attitude and emotion; and of dialogue functions like turn-taking and floor-holding. Five percent of the recordings also make use of a motion capture system for recording facial gestures in 3D, as well as body and head gestures. The corpus makes it possible to examine multimodal aspects of conversational interaction such as the timing relationships between speech signals and facial and body gestures. The research findings could also find a practical application in the development of an animated talking agent (talking head) for conversational spoken language systems. However, direct comparability between speech samples is considered to be of less importance than spontaneity in the productions (see discussion in Section 19.2.4), with the consequence that no variables are specifically introduced or controlled in the data. Also, although the annotation of the data is partly automated, it is still quite time-consuming, and limited in scope, since only orthographic transcriptions are provided for the audio signal, and head positions for the video signal (see Edlund et al. 2010 for details).

Rhythm is closely intertwined with phrasing, intonation, and accentuation, since the marking of prosodic edges and prominent syllables through variations in duration, timing, pitch, and other spectral properties all are important contributors to the percept of rhythm in speech. An example of a corpus-based study of rhythm which illustrates laboratory phonology work in the area of language acquisition is the APriL project (Acquisition of Prosody in L1; Payne et al. forthcoming; Astruc et al. forthcoming; Post et al. forthcoming; http://april-project.info/). The main objectives of this cross-linguistic project were (1) to clarify differences in L1 acquisition of rhythm in what are traditionally referred to as stress-timed and syllable-timed languages; (2) to investigate the contribution of a number of phonological factors to cross-linguistic differences in speech rhythm in adults; (3) to investigate intonational development cross-linguistically; and (4) to explore the prosodic properties of the children's speech input (i.e. child-directed and adult-directed speech produced by the caregivers). At least three variables had to be included in the design of the corpus in order to meet these objectives. Language was the first variable, with three levels representing languages that have been reported to belong to different rhythmic classes (English: "stress-timed," Spanish: "syllable-timed," Catalan: "intermediate"). The second was child age (2-, 4- and 6-year-olds) to allow for a cross-sectional analysis of prosodic development in young children. The third variable was speech type (child speech, child-directed speech, and adult-directed speech). The relatively complex design had the advantage that any effects that were found in the data could be attributed quite precisely to a specific source, but the number of data points in each cell of the design was relatively low, which weakened power in the statistical analyses (see Payne et al. forthcoming). However, this issue was a consequence of the quantitative approach that was adopted

in the study, which implied that the data points had to be directly comparable within cells.

19.2.3 Empirical approach: Qualitative versus quantitative methods

Until recently, most corpus-based studies of prosody took a qualitative approach, and there is a particularly rich tradition of qualitative corpus work in the areas of conversational interaction and discourse analysis (also e.g. Brown et al. 1980). The qualitative method could be defined as a "nonmathematical process of interpretation, carried out for the purpose of discovering concepts and relationships in raw data and then organizing these into a theoretical explanatory scheme" (Strauss and Corbin 1998: 1). In prosodic research, the data are typically naturally produced speech in radio broadcasts, telephone conversations, or recorded interviews (see Section 19.2.4 for a discussion of data elicitation). Qualitative data analysis contrasts with a quantitative approach in that it does not involve the measurements and quantification which make the data amenable to statistical analysis, and instead any findings are typically based on the in-depth analysis of a small number of cases, leading to rich and precise descriptions of specific prosodic phenomena in context. For instance, Ogden (2006) impressionistically analyzed assessment sequences in the CALLHOME corpus (a corpus of phone calls from Americans to friends and family abroad), and he established that agreement versus disagreement in turns can be signaled by changes in pitch span, loudness, degree of stricture in articulations, and more or less dynamic pitch movements on accented syllables, where the way in which these cues combine depends on the sequential environment of the turn in the conversational interaction.

Although, originally, qualitative corpus work relied almost exclusively on auditory impressionistic data, it is often supplemented by acoustic analyses in current qualitative research on prosody. Thus, Ogden (2006) measured pitch range and rate of articulation to confirm his impressionistic observations acoustically, but the study is nevertheless qualitative in nature, because none of the data were quantified over cases at any point. This implies, of necessity, that the generalizability of the findings to other speakers or contexts is limited.

The 1980s saw a gradual shift towards quantification in corpus-based research on prosody. The development of off-the-shelf speech processing software for more widely available platforms, combined with the expanding memory capacities of desktop computers, made acoustic analysis practicable for larger volumes of data, and they helped to make auditory analysis more reliable and efficient by allowing the researcher to verify observations in spectra, spectrograms, and fundamental frequency traces of the speech signal. These developments also gave a wider community of researchers access to speech analysis tools, which led to a sharp increase

in experimental work, and an attendant increase in quantifiable data (see Johns-Lewis 1986 for early corpus-based work).

For data to be quantifiable, any variation that is extraneous to the question at hand needs to be factored out or controlled for. The Intonational Variation in English (IViE) corpus provides an example of a speech corpus designed to investigate prosodic variation using a quantitative approach (Grabe et al. 2001a, 2001b; Grabe 2004; for a similar project on Irish see Dalton and Ní Chasaide 2006). The corpus was designed to contain directly comparable data on cross-dialectal and stylistic variation in British English intonation to investigate the implications of this variation for intonational theory. Quantifiability and comparability necessitated that near-homogenous groups of speakers who were representative of the local dialect produce the recordings, using the same stimulus materials. A purpose-built annotation system was developed to compare multiple recordings of the materials across speaking styles and dialects (Grabe et al. 2001a). This resulted in a large set of directly comparable audio recordings with machine-readable linguistic analyses of the intonation patterns produced (e.g. fourteen sets of controlled sentences eliciting different types of statements and questions in nine varieties of English). The comparability of the data across conditions allowed for statistically valid generalizations about the types of prosodic structures encountered, as well as their phonetic realizations (Grabe and Post 2004). Discovering frequency distributions of particular phenomena also provided information about the scope of variation, which can inform the development of an adequate model of intonational phonological knowledge (see Docherty and Foulkes 1999). The potential for quantification of the IViE data also facilitated the development of statistical models that could be applied in speech technology (Grabe et al. 2007).

Current empirical work in prosody often takes a mixed methods approach, where qualitative and quantitative analysis complement each other; this approach has a long history in the social sciences (e.g. Campbell and Fiske 1959). For instance, a qualitative analysis deductively provides a hypothesis that is based on a limited number of observations, and then tested quantitatively, as in Post (2000). This study of French intonation proposes a tonal grammar which is based on an auditory analysis of a fairy tale read by four speakers. No statistically valid generalizations could be made on the basis of such a small data set, but clear hypotheses could be developed about which tonal configurations can occur in French. A subsequent categorical perception experiment, which was combined with a semantic rating task, tested some of the more tenuous hypotheses about Intonation Phrase final rises. Alternatively, the methods are mixed when particular variables or hypotheses emerge inductively through a quantitative analysis in which large numbers of directly comparable cases are compared for a particular feature (or features), and a subsequent qualitative analysis is used to shed further light on the precise properties of the phenomenon at issue.

Taking a quantitative approach has the advantages of replicability and robustness, but only if appropriate sampling and statistical techniques are used (see Chapter 22 this volume). Since our knowledge about prosody is still rather limited, relevant variables may well be overlooked during data collection and analysis, introducing a bias towards certain types of sampling in specific subsets of speech data (see Warren and Hay, this volume). However, such a bias could also affect qualitative work, so regardless of the approach chosen, the data must be maximally representative of the phenomena under investigation (see McEnery and Wilson 2001; Biber 2004), and the basis of any conclusions that are drawn from patterns in the sample should always be carefully considered before the conclusions are generalized to other contexts, for instance from the laboratory to natural speech.

19.2.4 Data elicitation: Ecological validity and experimental control

Audio and audio-visual data can be collected by means of a wide range of materials and elicitation procedures (see also Chapter 21 this volume). In general, a balance must be struck between the "ecological validity" of the elicitation, that is, how close recordings are to natural speech communication, and how tightly controlled the elicitation needs to be to induce the speaker to produce the required prosodic events sufficiently frequently and in experimentally comparable ways. The latter, of course, can only be a goal if enough is known about the prosodic phenomenon at issue to make it possible to target it specifically. Below, we briefly exemplify some of the points on the continuum from ecological validity to control.

The most ecologically valid data come from corpora which have recorded natural spoken communication. There is no reason why, to be useful for prosodic research, a corpus should be recorded with a particular research question in mind, since corpora merely contain language as it is spoken in defined circumstances. With natural corpora there is no elicitation as such, beyond obtaining the consent of individuals to contribute their speech. Such corpora provide a substantial resource for the preliminary analysis of prosody—providing a tractable but extensive sample from which to get the feel of the prosody of a dialect—but they may also facilitate more specific studies. Wichmann and Cauldwell (2003), for instance, in an experiment exploring the role of discourse context in affect judgments on questions, presented to listeners a number of wh-questions from the ICE-GB corpus (the British English subset of the International Corpus of English), recording varieties of English worldwide and covering a selection of speaking styles from conversation to scripted broadcasts (see http://ice-corpora.net/ice/). Another example of prosodic research using "natural" data is the prosodically labeled Boston database of FM radio news speech collected at Boston University (Ostendorf et al. 1995). Although the data are scripted speech produced by professional news readers, they are natural

in the sense that the speech content and production setting were not controlled or manipulated for the purpose of the investigation. The corpus was used by Dilley and colleagues (Dilley et al. 1996) to investigate the role of prosodic structure in the glottalization of word-initial vowels. Since a large enough number of tokens was available, generalizations across contexts were possible in spite of the relative lack of experimental control over the data.

Scripted speech, however, reflects the structures of written text, unlike spontaneous speech which reflects the structuring of talk-in-interaction (see e.g. Biber et al. 1998), as well as a speaker's cognitive processing (Wichmann 2008). Spontaneous corpus data can prove very useful in uncovering unknown prosodic correlates of functions in conversational interactions, such as the cues to agreement and disagreement in assessments found in the CALLHOME corpus by Ogden (2006; see also Section 19.2.3).

Other types of spontaneous interactive data are interviews, and free or guided conversations and discussions. In the Spontal corpus (Edlund et al. 2010), the spontaneity of the conversations was considered of paramount importance, although an elicitation method was used which could elicit comparable speech samples between different conversations. The participants were asked to open a wooden box containing objects whose identity or function is not immediately obvious, and were invited to talk about them. To enhance the likelihood of obtaining spontaneous productions, all interlocutors interacted with a dialogue partner whom they had met before, and they were told that they were allowed to talk about absolutely anything, and not obliged to talk about the contents of the box.

The advantage of interactive data is that the participants concentrate on communicating successfully rather than on their speech production, which increases the likelihood of eliciting naturally occurring spontaneous data. However, the participants are unlikely to produce long or complex sentences; they will produce overlapping speech, and many utterances will be incomplete. Also, many participants are likely to resort to a relatively narrow repertoire of communication strategies, which may result in the elicitation of a more limited set of prosodic phenomena than in scripted speech.

Greater experimental control over prosody can be achieved with other non-scripted tasks, often involving descriptions of scenes or events depicted with pictures or film clips. Swerts et al. (2002), in a study on Dutch and Italian, devised an interactive card-matching game using cards with geometric shapes in different colors to elicit four focus conditions. This technique was adopted by Hellmuth (2005a) in a study of Arabic, and she extended the technique using a "murder scenario trio of person, weapon and place", eliciting triads with appropriate focus such as "the manager, with the POISON, in the kitchen".

In the APriL corpus introduced above (Section 19.2.2), the main goal was to elicit utterances that were as comparable as possible in terms of segmental and metrical structure, lexical targets, and syntactic constructions across the three types

of speech (adult-directed speech, child-directed speech, and child speech). This was necessary to ensure that any cross-linguistic differences that were observed could be attributed to the developmental nature of the speech rather than random differences in the data. Short interactive dialogues were elicited by means of two structured games, based on animated clips eliciting simple noun-verb-object utterances (*She's blowing bubbles*). Picture scenes in a second task elicited lexical words with different metrical structures. The scenes represented familiar objects and easily describable events in order to increase the likelihood of obtaining directly comparable speech samples across the age groups, even though children as young as 2 participated in the study.

Such tasks have the advantage that targeted prosodic phenomena can be elicited in a controlled fashion, while the resulting data are still spontaneous in the sense that they are unscripted. They are probably more "cost-effective" for prosodic research than the much-used "map task" (Anderson et al. 1991), where one participant tries to describe a route on a map to an interlocutor, who, unknown to either of them, holds a version of the map with different landmarks. The map task is a good way to elicit names and words specified on the maps, but it tends to yield a lot of minimal utterances (including back-channel utterances) which may not exemplify much prosodic variation, or contain identifiably distinct speech acts (Shobbrook and House 2003).

Clearly, the most control can be exercised over the content and structure of data when the speech is scripted. Arguably the most naturalistic way to do this is to incorporate dialogue into a read story. Asu (2004), for instance, placed members of Estonian three-way quantity triplets in sentence-final position in quoted sentences, both questions and statements, within a constructed story. Similarly, Post (2011) embedded target items in which morphosyntactic structure and constituent length (in syllables) were systematically varied to investigate the interaction between various constraints on pitch accent distribution and prosodic phrasing in French.

Read utterances have been widely used in prosodic research as they allow complete control over parameters relevant to discourse status (mainly statement versus question) and number of pitch accents, down to the segmental constituency of relevant syllables, and they facilitate auditory and acoustic analysis. In the classic experiment of Liberman and Pierrehumbert (1984), speakers read lists of between two and five berry names in a downstepping series, and provided data supporting a "local" model of this kind of downtrend, with accents scaled relative to the preceding accent rather than determined by a global utterance contour. As is not uncommon this experiment relied not only on the production of specified segmental strings, but also on the speakers rendering them with one (or more than one) desired prosodic pattern. At the more microprosodic level, recent research into the alignment of pitch targets relative to segmental material has relied on carefully controlled read materials (see Arvaniti, this volume and D'Imperio, this volume for a review on tone association and alignment). Arvaniti et al. (1998) defined the

notion (and subsequent research paradigm) of "segmental anchoring," whereby H and L targets align with segmentally defined events, on the basis of an experiment on Modern Greek. The relevant pre-nuclear accent was by design on a long or short syllable, and the same number of syllables intervened before the next accent. With this level of control it was possible to show that the LH pitch accent aligned consistently with the L around the start of the accented syllable, and the H just after the start of the first post-accentual vowel; concomitantly the slope and duration of the rise varied according to the length of the accented syllable.

The most extreme forms of experimental control are probably stimuli that elicit reiterant speech, i.e. speech in which the same segmental material is repeated to form an utterance, like *mamamama* (Liberman and Streeter 1978; Sluijter and Van Heuven 1996). Reiterant speech has been particularly useful in uncovering the variations in the acoustic correlates of stress in different positions in the speech chunk (equivalent to a word or a phrase), allowing precise measurements that can be abstracted away from confounding factors such as the segmental make-up of the materials.

19.2.5 Conclusion

The development of databases for the study of prosody in connected speech is still in its infancy, but the progress that is currently being made in the development of (semi-)automatic techniques for the annotation and analysis of large amounts of speech data makes possible rapid advances in the field. Quantitative analyses of spontaneous data will become increasingly feasible, and they will provide a better understanding of prosodic cues and the way in which they interact in conveying different communicative functions (see Post et al. 2007). The development of adequate models of prosody crucially depends on this.

CHAPTER 20

ENCODING, DECODING, AND ACQUISITION

EYE MOVEMENTS AS A DEPENDENT MEASURE IN RESEARCH ON SPOKEN LANGUAGE

SHARI R. SPEER

NEUROPHYSIOLOGICAL TECHNIQUES IN LABORATORY PHONOLOGY

WILLIAM IDSARDI AND
DAVID POEPPEL

The contributions in this chapter introduce experimental methods used to study the encoding, decoding, and acquisition of speech. These methods, primarily developed in the fields of psycholinguistics and most recently neurolinguistics, offer a tool set that provides different windows on the nature of speech in these varied functions. Maye reviews methods used to study the acquisition of phonetics and phonology. Schiller discusses methods to study the encoding of speech. Iverson reviews methods for studying speech perception in adults. The final two pieces survey newer methodologies, with Speer's discussion of eye tracking and Idsardi and Poeppel's review of recent developments in neuroimaging.

20.1 Studying the acquisition of a receptive phonetic/phonological system

Jessica Maye

20.1.1 Introduction

In the past few decades remarkable strides made in infant testing methodology have dramatically altered what we know about the development of a receptive phonetic and phonological system. Earlier research into phonological development focused primarily on children's utterances, noting the difference between child and adult pronunciations, and looking for patterns in how children alter the pronunciation of a word. This line of research is still alive and well; however, through new methodologies we have discovered that infants' phonological *knowledge* is greatly advanced compared to the phonological complexity of their own productions (see Munson et al., this volume). We now know that although infants do not begin speaking until around 12 months of age, by 5 months they are beginning to recognize familiar word forms (Mandel et al. 1995); by 6 months they have already begun tuning in to the native language phonetic properties (Kuhl et al. 1992); by 8 months they have begun to selectively discriminate native language (and not foreign) phonetic contrasts (Werker and Tees 1984a; Polka and Werker 1994; Bosch and Sebastián-Gallés 2003); and by 9 months they are familiar with the typical phonotactic patterns of the native language (Friederici and Wessels 1993; Jusczyk et al. 1994). This section discusses the primary methods currently used to test infants' early perceptual, phonetic, and phonological development.

20.1.2 Designing the study: Dependent measures and testing paradigms

Testing the receptive linguistic knowledge of pre-verbal infants can be tricky, but researchers have capitalized on infants' keen senses of sight and hearing. Most early language research utilizes infants' looking and/or listening preferences to infer the linguistic knowledge of these young participants.

20.1.2.1 *Dependent measures*

In infant research, the dependent measures available for testing differ based on the age of the infant. It is possible, in fact, to test infants pre-natally, using infant heart

rate as the dependent measure. Because the auditory system is fully developed by the third trimester of gestation (Fifer and Moon 2003), infants begin learning about their linguistic environment during their last 2–3 months in the womb. Near-term fetuses can differentiate between different talkers' voices (Lecanuet et al. 1993) and react to their mother's voice (Lecanuet et al. 1991; Masakowski and Fifer 1994).

After birth, although heart rate remains a viable measure for neonates, a more commonly used measure is high-amplitude sucking (HAS). Infants suck on a pacifier attached to a pressure transducer measuring the strength of each suck. When the infant sucks harder than usual (producing a high-amplitude suck) an auditory stimulus is presented. Pairing the stimulus this way provides a way for infants to demonstrate their preference for one stimulus over another: infants produce more high-amplitude sucks when they hear stimuli that they prefer, presumably because doing so causes the auditory stimulus to continue playing. This methodology has been used to demonstrate that newborns can differentiate between their native language and a foreign language (Moon et al. 1993), specifically when the two languages fall into different rhythmic classes (Nazzi et al. 1998). The most well-known study utilizing HAS to test phonological development is perhaps the classic study by Peter Eimas and colleagues (1971) who used HAS in a habituation paradigm to determine whether infants perceive voice onset time categorically. HAS can also be used in a preference paradigm to determine whether infants prefer one type of sound stimulus over another.

After about 4–6 months of age, heart and sucking rates are less robust measures of infants' interest, so studies focusing on older infants typically use looking-time and head-turn response as dependent measures. Infant eye gaze can be used in two different ways to make inferences about language processing. In some procedures (such as the Intermodal Preferential Looking and Anticipatory Eye Movement procedures) researchers track the proportion of time infants spend looking at one visual stimulus versus another. For example, an infant may hear "Look at the bear!" while they see images of a bear and a dog and a researcher tracks which picture the infant spends more time looking at. Although it is possible to use a remote eye-tracker to track infants' gaze, many researchers have had success with the "poor man's eye-tracker" method, which involves videotaping the infant during an experiment and having experimenters manually code infant eye gaze offline based on the video (e.g. Swingley et al. 1998; Golinkoff et al. 1987). Since infant studies typically depict a limited number of items in the visual scene (often only two pictures side-by-side) it is possible for an experimenter to accurately judge the direction of an infant's gaze.

A second set of procedures (such as the Head-Turn Preference and Visual Fixation Procedures) uses infant gaze to infer how much attention the infant is paying to what they hear. The infant is given one visual stimulus to look at (a flashing light or simple image), and allowed to look at it while they hear an auditory stimulus. If the infant looks longer while they are listening to one type of stimulus

(e.g. a grammatical sentence) than another type (e.g. an ungrammatical sentence), this longer looking-time is taken to indicate that they paid more attention to, and thus preferred, that stimulus type.

20.1.2.2 *Testing paradigms*

Three basic types of testing paradigms are used in infant language processing research: habituation, preference, and conditioned response. Each of these testing paradigms can be conducted with a variety of dependent measures.

The term *habituation* refers to the extinction of a novelty response (Kagan and Lewis 1965). If an infant hears the syllable [ba] they will initially be interested in it. Over time, if the same syllable repeats many times, the infant's interest will wane as they become *habituated* to the stimulus. Habituation can be used to test an infant's discrimination of two items. For example, once the infant has habituated to the stimulus [ba], a change stimulus [pa] can be presented. If the infant perceives the difference between [ba] and [pa], presentation of the change stimulus should re-engage their interest, resulting in *dishabituation*.

Eimas et al. (1971) used high-amplitude sucking in a habituation paradigm to study categorical perception in 1–4-month-old infants. Each infant's sucking rate was measured over windows of one minute while they listened to a repeating stimulus (e.g. [ba, ba, ba. . .]), and the number of high-amplitude sucks they produced was compared to the immediately preceding minute. When the sucking rate decreased by 20 percent over two consecutive minutes the infant was considered to be habituated to the background stimulus, triggering four minutes of the change stimulus (e.g. [pa, pa, pa. . .]). This study found that infants were most likely to dishabituate if the background and change stimuli came from opposite sides of the English voicing boundary. Habituation is also used in studies employing looking-time as a dependent measure. See Hoben and Gilmore (2004) for additional discussion of habituation in infant studies.

The most common paradigm in infant language studies is the preference paradigm, which tests whether infants show a greater response to one type of stimulus over another. In this paradigm an infant is presented with two types of stimuli (e.g. phonotactically legal vs. illegal words), and the researcher measures which stimulus holds the infant's attention the longest. The dependent measure is most often looking-time, but HAS and heart rate can also be used in a preference paradigm: infants will suck harder or faster to maintain the presentation of a preferred auditory stimulus, and increases or decreases in heart rate are indicative of enhanced interest or calmness.

The preference paradigm has been used to measure infants' naturally developing preferences as they acquire their language. For example, newborns (using HAS) and near-term fetuses (using heart rate) have been shown to prefer listening to their own mother's voice over that of an unfamiliar woman (DeCasper and Fifer 1980;

Kisilevsky et al. 2003); and newborns (using HAS) prefer hearing their own native language over a foreign language with different rhythmic characteristics (Nazzi et al. 1998). By 5 months (using looking-time) infants prefer to hear their own name over that of a different name with the same stress pattern (Mandel et al. 1995), and by 9 months they show the same preference even when the names are partially masked by background noise (Newman 2005).

The preference paradigm can also be used to measure discrimination. While the ability to discriminate two items does not necessarily entail that one will prefer one item over another, having a preference for one item *does* entail that the two items are discriminable. For example, Friederici and Wessels (1993) found that 9-month-old infants preferred to listen to phonotactically legal sequences over illegal sequences, from which we can infer that by 9 months infants have learned which sequences occur in their native language. Moreover, a preference can be induced in infants using the preference paradigm to test discrimination of other types of stimuli. For example, although infants are unlikely to have an inherent preference for one of two test syllables (e.g. [ba]-[pa]), they may prefer to hear strings of alternating syllables over strings of repeated syllables; this method is known as the Stimulus Alternation Procedure; Best and Jones (1998). This issue is discussed in greater depth by Houston and colleagues (2007) in their evaluation of several similar methods of testing phonetic discrimination using infant preference.

The third type of infant testing paradigm is that of conditioned response. In this paradigm the infant is trained to perform some explicit behavior in response to the test stimulus. Two procedures that utilize conditioned response are the Conditioned Head-Turn (CHT) and Anticipatory Eye Movement (Anti-EM) procedures (discussed at fuller length below). In the CHT procedure (Werker, Polka, and Pegg 1998) there is a repeating background stimulus (e.g. multiple tokens of the syllable [ba]), and the infant is trained to turn their head to the side whenever that background stimulus changes to an oddball stimulus (e.g. tokens of the syllable [da]). In the Anti-EM procedure (McMurray and Aslin 2004) the infant is implicitly trained to look to the right side of a display when they hear one type of stimulus, and to look to the left side when they hear a different type of stimulus.

A final issue regarding experimental design is that of familiarization. A familiarization component may be added to the beginning of any of the described experimental paradigms. The term *familiarization* simply refers to some sort of pre-exposure to relevant stimuli before an infant is tested. One example of this type of method is the Pattern Induction paradigm (Saffran and Thiessen 2003). Familiarization is also used to test infants' ability to process aspects of natural language. For example, studies of infants' ability to segment word forms out of fluent speech have familiarized infants to a set of sentences containing a particular word (e.g. several sentences that all contain the word *cup*), followed by test trials in which infants hear the familiarized target word (*cup, cup, cup*. . .) versus non-familiarized words (e.g. *dog, dog, dog*. . .; Jusczyk and Aslin 1995).

20.1.2.3 *Specific procedures*

A variety of specific infant testing methodologies have crystallized into named procedures. This list of procedures is by no means exhaustive, but reference to any of the named procedures provides a shorthand description of several aspects of the methodology.

20.1.2.3.1 *Head-Turn Preference Procedure (HPP)*

The HPP was developed in the 1980s (e.g. Fernald 1985; Hirsh-Pasek et al. 1987; Kemler Nelson et al. 1989), and has become a staple in infant language research. The infant sits on a caretaker's lap inside a booth equipped with three flashing lights: one directly in front of them and one on each side (see Figure 20.1.1). The experimenter sits outside the booth, watching the infant via video camera, and manually keys in which direction the infant is looking at any point during the experiment. At the beginning of each trial the center light flashes, which brings the infant's attention to the center. When the infant is facing forward, one of the side lights begins flashing. The experimenter initiates a sound stimulus when the infant looks toward the flashing side light, and the sound continues until the infant looks away. Thus, the infant controls how long they hear the sound. Some wiggle-room is built into the criterion for what constitutes a "look-away" to ensure that the trial is not prematurely terminated if the infant's eyes briefly dart away from the light and then back.

The HPP has been used with infants as young as 3 months (e.g. Hayes and Slater 2008), as well as toddlers up to at least 19 months of age (e.g. Höhle et al. 2006), but it appears to be easiest to use at ages 6-12 months. With the HPP, researchers have demonstrated that infants as young as 3 months prefer alliterative over non-alliterative lists of words, providing some evidence for preference of segmental constancy across syllables from a very early age (Hayes and Slater 2008; Jusczyk, Goodman, and Bauman 1999).[1]

The HPP has also been used to demonstrate that infants' memory of words also contains extralinguistic information: infants show better memory of familiar words spoken by the same talker (Houston and Jusczyk 2003) and in the same affect (Singh et al. 2004) as those they originally heard. Using the HPP, Chambers et al. (2003) demonstrated that infants quickly learn phonotactic regularities in a laboratory setting. The HPP has been used to demonstrate that infants' phonological knowledge about their language (phonotactic probabilities: Mattys and Jusczyk 2001; typical stress pattern: Jusczyk, Cutler, and Redanz 1993) guides their word segmentation strategies.

[1] These studies focused specifically on consonantal constancy. However, other researchers, most notably Marean et al. (1992) utilizing a different methodology, have demonstrated vocalic constancy by 2–3 months of age.

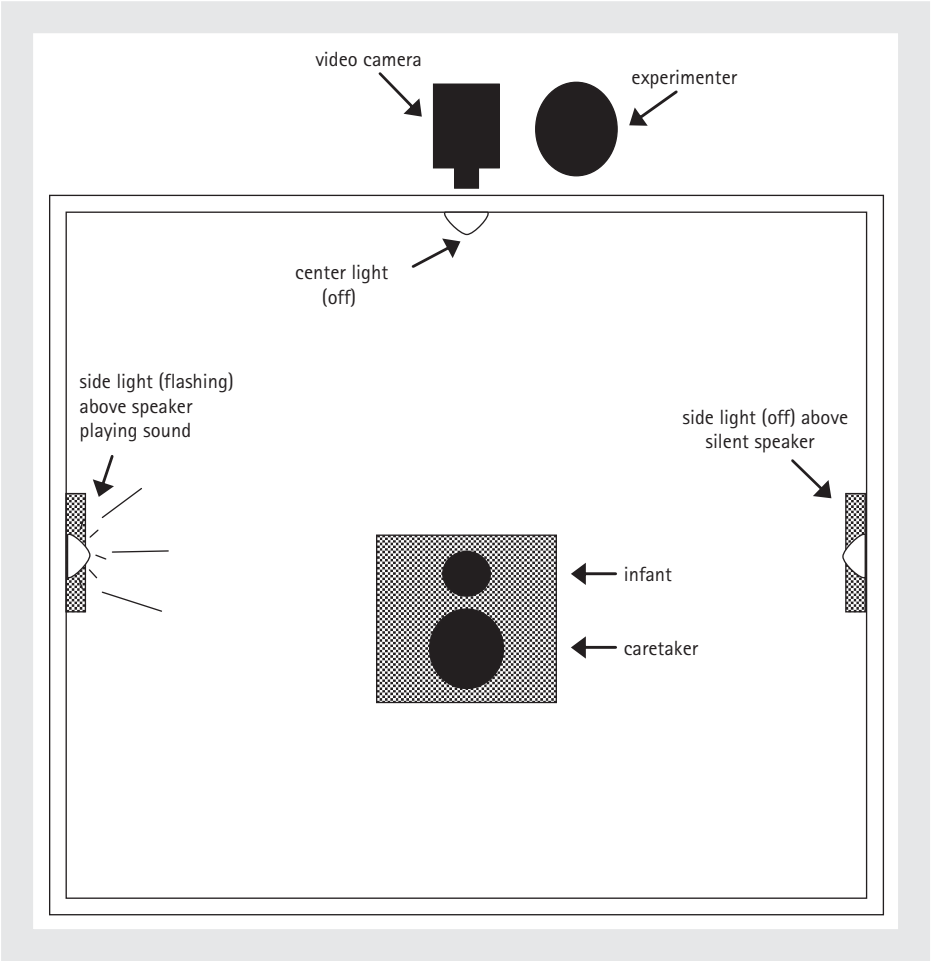

Figure 20.1.1. Testing booth layout for the Head-Turn Preference Procedure (redrawn from Kemler Nelson et al. 1995), depicting a trial on which the stimulus is presented on the infant's left side.

20.1.2.3.2 *Visual Fixation Procedure (VFP)*

The Visual Fixation Procedure is similar to the HPP, in that infants control the duration of a trial by looking at a visual stimulus while they are listening to an auditory stimulus. The difference is that while in the HPP the visual stimuli are flashing lights located to the sides of the infant, in the VFP the visual stimuli are images on a monitor or screen located in front of the infant. Because the VFP does not entail the infant turning their head to the side it may place fewer processing demands on infants, making it particularly well suited to younger infants (under 12 months), while the HPP is best suited to somewhat older infants (over 8 months).

Like the HPP, the VFP can be used to measure infants' preference for one type of stimulus over another. The VFP is also often used in a habituation paradigm: a habituation stimulus can be presented repeatedly until an infant's looking-time decreases to some pre-set criterion, followed by the presentation of test trials.

The VFP has been used to examine infants' speech perception and the development of linguistic perceptual biases that reflect the phonetic properties of the native language. For instance, Best and colleagues used the VFP to demonstrate that although infants lose sensitivity to some native-language phonetic contrasts during infancy, other phonetic contrasts (that are not assimilable to the native phonetic system, such as English speakers' perception of Zulu clicks) continue to be discriminated well throughout life (Best et al. 1995; Holt, this volume). Horn et al. (2007) have used this procedure to examine the discrimination abilities of children with cochlear implants. My own research has used the VFP to demonstrate that infants' phonetic discrimination is affected by the statistical distribution with which speech sounds occur in the infant's linguistic input (Maye et al. 2002; Maye et al. 2008).

20.1.2.3.3 *Intermodal Preferential Looking Procedure (IPLP)*

In the IPLP the direction of an infant's eye gaze is used to make inferences about their language processing. This method has been used in infant studies for three decades (Spelke 1979; Golinkoff et al. 1987) and continues to be very fruitful for assessing aspects of phonological (e.g. Swingley and Aslin 2000), lexical (e.g. Fernald et al. 1998), morphological (e.g. Zangl and Fernald 2007), and syntactic development (e.g. Golinkoff et al. 1987). In the IPLP infants see two images or moving videos presented side-by-side while they hear a single sound file. For example, they may see a picture of a ball on the left, and a doll on the right, and hear "Look at the ball!" This paradigm compares the amount of time spent looking towards the target image (in this case, the ball) to the amount of time looking at the competitor item (the doll).

Swingley and Aslin (2000) have used the IPLP to test infants' degree of phonological specificity of the words in their lexicons. In this study infants saw images of two familiar objects (e.g. a baby and a car) and heard either the correct name of one of the two items ("Look at the baby!") or a one-feature mispronunciation of one of the items ("Look at the vaby!"). Infants looked longer at the matching (*baby*) or near-matching (*vaby*) picture than at the distractor item (*car*), indicating that they recognized the similarity between the mispronunciation and the correct pronunciation of the target item. But they looked toward the target item *faster* when the word was pronounced correctly, demonstrating that early phonological representations of familiar words are fairly specific, rather than loosely specified, as had been previously argued (e.g. Charles-Luce and Luce 1990).

20.1.2.3.4 *Switch Procedure*

The Switch Procedure is designed to study word learning (Werker, Cohen, Lloyd, Casasola, and Stager 1998). In the Switch Procedure infants are trained to match a sound file with a particular image. This training occurs by habituating infants to a sequence of trials on which two objects are shown, each consistently paired with one of two labels. These training trials continue until the infant habituates, as indicated by a decrease in looking-time. Following habituation, the infant receives two test trials to test their ability to remember the mapping between word and object. One test trial is called the "Same" trial, and it is identical to one of the habituation trials (e.g. Object 1 paired with Word 1). The other is called the "Switch" trial, where there is a mismatch between the word and the object (e.g. Object 1 paired with Word 2). If infants have successfully learned the word-object pairings presented during the habituation phase, they should show an increase in looking-time on the Switch trial (i.e. a dishabituation to the novel stimulus) compared to the Same trial.

The Switch Procedure has been utilized to examine the interaction between speech perception and word learning. Although infants show a remarkable ability to discriminate fine phonetic detail in speech perception tasks, Stager and Werker (1997) found that at 14 months infants learn novel word-object pairings only if the word forms are maximally distinct. If the novel words form a minimal pair, such as *bih* [bɪ] vs. *dih* [dɪ], infants fail to notice the mismatched word-object pairing until the age of 18 months.[2]

Subsequent studies have suggested that this failure to learn minimal pairs at 14 months is due to processing demands. If processing demands are reduced by allowing infants to become familiar with either the two objects or the two word forms prior to learning the word-object pairings in the Switch task, 14-month-olds succeed at learning minimal word pairs (Fennell and Werker 2004). Presenting word forms within a grammatical sentence context rather than in isolation also facilitates 14-month-olds' discrimination of minimal pairs (Fennell and Waxman 2010).

20.1.2.3.5 *Conditioned Head-Turn Procedure (CHT)*

The Conditioned Head-Turn Procedure (Werker, Polka, and Pegg 1998) tests speech perception by conditioning infants to turn their head whenever they hear a change from a repeating background stimulus (e.g. the syllable [ba]) to a new stimulus (e.g. [da]). The CHT is conducted in a sound booth equipped with one or more smoked plexiglass boxes, each containing a light and an animatronic toy. Because the plexiglass is darkened, the infant cannot see inside the box unless the light inside

[2] Pater et al. (2004) subsequently replicated this result using the phonotactically permissible stimuli *bin* [bɪn] vs. *din* [dɪn].

is turned on. When the light turns on, the animatronic toy also begins to move and make noise, attracting the infant's attention and reinforcing the infant's head-turn. The infant sits on a caretaker's lap, facing a member of the research team ("the Assistant"), who is showing the infant quiet toys to keep them facing forward. The plexiglass boxes are to one side, so that the infant must turn their head to look at them. Both the Assistant and the caretaker listen to masking music or noise through headphones throughout the procedure so that they cannot hear any of the sounds presented to the infant. A second member of the research team ("the Experimenter") remains outside the testing booth and watches the infant through a window or via video camera. The Experimenter controls the experiment by pressing buttons to initiate test trials and to record infant head-turns.

There are multiple variations of the CHT, but in each the infant is conditioned to produce a head-turn when the background stimulus (presented repeatedly throughout the procedure: e.g. [ba, ba, ba, . . .]) is replaced by a change stimulus (e.g. [da, da, da, . . .]; see Werker, Cohen, Lloyd, Casasola, and Stager 1998, and Tsao et al. 2006 for two examples of how to train infants in this procedure). Once infants have been conditioned to perform this task, it can be used to test their discrimination. One strength of the CHT is that unlike many other infant testing methods, in the CHT infants are actively rewarded for performing well. Once they have learned to perform the task, the *only* time infants get to see the fun animatronic toys is if they turn their head when the sound changes. Thus, infants are highly motivated to discriminate the sounds in question.

While some studies report percent correct as a dependent measure, signal detection measures such as d' (Macmillan and Creelman 1991) or A' (Grier 1971) are more common. Both d' and A' involve calculation of hits vs. false alarms. A head-turn produced when there has been a sound change is called a "hit," and a head-turn when there has been no sound change is called a "false alarm." An infant may look toward the reinforcer not because they perceive a sound change but simply because they have a personal bias to continually check and see if the reinforcer is turned on. Thus, to control for response bias it is important to examine both hits and false alarms.

The CHT procedure was used in a landmark study by Werker and Tees (1984a) who showed that infants become selectively attuned to native language phonetic contrasts between the ages of 6–12 months. This study found that infants from English-speaking homes could discriminate the Hindi dental-retroflex contrast ([da]-[ɖa]) and the Nthlakampx velar-uvular ejective contrast ([k'a]-[q'a]) at 6–8 months, but by 10–12 months could not. The CHT procedure has been used in many of the studies that have documented infants' phonetic discrimination abilities between the ages of 4–12 months (e.g. Eilers et al. 1977; Kuhl et al. 2006; Polka et al. 2001), and thus is largely responsible for what we know about infants' early, language-general discrimination ability, as well as their later, language-specific discrimination patterns.

20.1.2.3.6 *Anticipatory Eye Movement Procedure (AEM or Anti-EM)*

The Anti-EM procedure was developed by McMurray and Aslin (2004) to test infants' categorization of two types of stimuli. In the Anti-EM procedure, like CHT, infants are conditioned to perform a particular behavioral response on the basis of what they hear. However, the Anti-EM procedure may place lighter processing demands on infants since the only motor movement required of the infant is that they move their eyes, while in the CHT infants must turn their head and/or their body to the side. In the Anti-EM procedure, infants are seated in front of a large monitor on which they see a T-shaped occluder (i.e. an object that things can go behind, shaped like a capital "T"; see Figure 20.1.2). On each trial, an image such as a small geometric shape or the face of a familiar cartoon character appears at the bottom of the monitor, below the occluder. It then travels upward, disappearing behind the occluder, and after a short delay it reappears from one of the two sides of the top of the "T," seeming to have traveled upwards behind the occluder and made a 90 degree turn to one of the two sides. When the object disappears behind the occluder, one of two types of sound stimuli plays. The two types of sound stimuli indicate which side of the occluder the object will emerge from: if one stimulus type plays (e.g. the syllable [pa]), the object will emerge from the left side, if the other type of stimulus plays (e.g. the syllable [ba]), the object will emerge from the right side. Over subsequent trials, infants begin to pick up on this consistent pattern, which is evident by their production of anticipatory eye movements. That is, while the object is hidden behind the occluder, infants begin to look to the side where they expect the object to reappear.

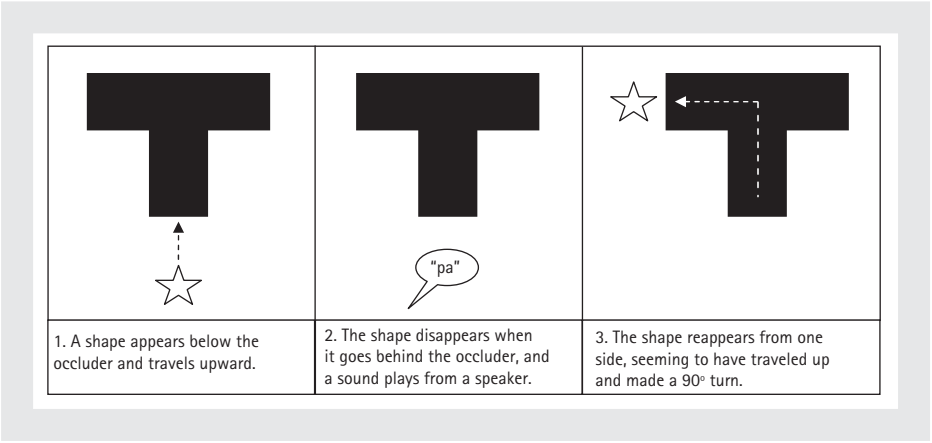

Figure 20.1.2. Example display used in the Anticipatory Eye Movement procedure. Once infants have learned the pattern (e.g. that when they hear the syllable /pa/ the shape will appear from the left side of the occluder), they begin to look toward the anticipated side *before* the shape reappears.

This methodology is similar to the CHT in that infants are trained to look in a particular direction, based on what they have heard. One advantage of the Anti-EM procedure, though, is that there are two sides to look at, so it is possible to have infants categorize *two* types of sound stimuli, rather than just one.

Anti-EM has been conducted both using a remote eye-tracker device (McMurray and Aslin 2004), as well as with experimenter judgments of infant eye gaze (Albareda-Castellot et al. 2011). This methodology is still new and has not been widely used, but it is likely to prove very fruitful in the future.

20.1.3 What have we learned?

The development of these infant methodologies has led to tremendous advances in our knowledge of early phonetic and phonological development (see also Munson et al., this volume; Demuth and Song, this volume). We now know that infants begin to learn about language while still in the womb. The sound from the outside world that reaches a fetus's ears is low-pass filtered through the mother's body and uterine environment (Richards et al. 1992), allowing only the lower frequencies to pass through. However, low frequencies carry prosodic information such as pitch and rhythm that infants can use to begin to parse the speech stream into smaller chunks (Gleitman and Wanner 1982) and learn the rhythmic patterns of the native language (e.g. Mehler et al. 1988; Nazzi et al. 1998).

After birth, infants begin rapidly learning additional aspects of their native language. In addition to fine-tuning their prosodic knowledge (Bosch and Sebastian-Galles 1997; Nazzi et al. 2000), infants begin to learn the prototypical phonetic properties of native-language vowel sounds by the age of 6 months (Kuhl et al. 1992). In their first few months of life, infants discriminate many phonetic contrasts, regardless of their phonemic status in the native language (e.g. Trehub 1976; Werker et al. 1981). But by 12 months, infants stop discriminating phonetic contrasts that correspond to a single phoneme category in the native language (e.g. Werker and Tees 1984a; Kuhl et al. 2006; Tsao et al. 2006; the shift is even earlier for vowel contrasts than consonant contrasts: Polka and Werker 1994; Bosch and Sebastián-Gallés 2003). And discrimination of native-language phonemic contrasts improves at around the same age (Kuhl et al. 2006; Tsao et al. 2006).

Nine-month-old infants differentiate between strings that conform to native phonotactic constraints versus strings that violate them (Friederici and Wessels 1993; Jusczyk, Cutler, and Redanz 1993), and between phonotactic patterns that occur frequently versus infrequently in the native language (Jusczyk et al. 1994). This knowledge helps infants parse word forms from fluent speech (Mattys and Jusczyk 2001).

Also by 9 months, infants prefer to listen to words that follow the typical stress pattern of their native language (Jusczyk, Cutler, and Redanz 1993). They use this

stress pattern to help parse words out of fluent speech: English-learning infants parse strong-weak syllable sequences (like *hamlet*) as words, even when to do so is erroneous (e.g. incorrectly segmenting the *taris* sequence out of the phrase "*the guitar is. . .*"; Jusczyk, Houston, and Newsome 1999).

The development of novel methodologies has also enabled researchers to ask questions about the mechanisms of language acquisition by exposing infants to artificial languages and testing to see which aspects of the language are learned. This technique has revealed that infants use statistical cues to begin segmenting potential word forms from fluent speech (Saffran et al. 1996). Statistical cues also influence infants' discrimination of speech sounds: infants' perception of a phonetic contrast is enhanced if they are exposed to a statistical distribution of sounds indicative of a phonemic contrast, but suppressed if the distribution of sounds suggests a single category (Maye et al. 2002, 2008). Perhaps the most remarkable finding from these studies is the incredible speed with which this learning occurs: infants are typically only exposed to the artificial language stimuli for about 2.5 minutes prior to testing.

And finally, we have learned that these early measures of language development correlate with later language development. Newman et al. (2006) found that infants' performance on word segmentation tasks at 12 months of age correlated with vocabulary size at 24 months, and with language abilities at 4–6 years of age. Tsao et al. (2004) found that infants who require fewer trials to learn the CHT task at 6 months (discriminating a foreign phonetic contrast) have larger receptive and productive vocabularies as toddlers. Furthermore, Kuhl et al. (2005) found that better discrimination of native phonetic contrasts at 7 months of age predicted advanced vocabulary and syntax development throughout toddlerhood, while better discrimination of foreign phonetic contrasts at 7 months predicted poorer development. This finding suggests that there may be some cost associated with maintaining the ability to discriminate a non-phonemic contrast, such that infants who tune out these contrasts earlier are facilitated in their language processing.

Although great advances have been made in infant testing methodology in the past few decades, there is still room for additional development. One of interest is the development of methods that will enable the documentation of individual differences between infants. Current methods are largely focused on group performance at particular ages. This is due to the fact that there are so many sources of variability when any infant is being tested, as well as the fact that each infant can only provide a small number of trials. Thus, it is very difficult to draw strong conclusions based on the performance of any one infant. However, some researchers are beginning to propose new methods that will enable reliable testing of individual infants. For example, Houston et al. (2007) have proposed a method of analysis (utilizing autoregression) that computes the significance level of a single infant's preference across a series of trials. Methodological advances of this nature will make it possible to diagnose infants with language problems far earlier than we

are currently able, and will enable us to develop more finessed theories of language development that take into account differences in processing style and different routes to learning.

20.2 Experimental methods and designs to investigate phonological encoding of spoken language*

Niels O. Schiller

20.2.1 Introduction

In recent decades, within psycholinguistics many experimental methods and innovative designs have been developed to investigate the encoding of speech. This section gives an overview of the most important paradigms employed to investigate word-form encoding in spoken language production. The section focuses on form aspects of the verbal signal, i.e. the so-called *word form*, leaving aside other aspects of language production such as syntactic and semantic processing.

Experimental methodology is discussed, illustrating investigations of two important theoretical questions in the field. The first is the *time course of word-form encoding*. When speakers have retrieved a word from their mental lexicon, do they have all the necessary information to produce the word (e.g. segments, stress pattern, syllable structure) at their disposal at once, or is there a temporal progression of encoding, such that information from the beginning of words is available earlier than that from later parts of the words? Evidence from speech errors (e.g. ***h****eft* ***l****emisphere* instead of ***l****eft* ***h****emisphere*, Fromkin 1971), suggests that words are not stored as whole, undividable units in our mental lexicon. However, whether or not the process of encoding the individual units (e.g. segments, lexical stress, lexical tones, etc.) is a parallel or incremental process is an independent processing issue.

The second theoretical question concerns the role of syllabic units in word-form encoding. After having been previously judged unimportant to phonological

* Niels O. Schiller is currently supported as a Fellow in Residence 2010/11 at the Netherlands Institute for Advanced Study (NIAS) in the Humanities and Social Sciences, Wassenaar, The Netherlands.

descriptions (Chomsky and Halle 1968), the syllable now has a firm place as a basic phonological unit in most phonological theories (though see Côté, this volume). However, the role of the syllable in psycholinguistics, especially word-form encoding during speech production, is still ambiguous (see overview in Schiller 2008). Some of the data discussed in this section bear on the status of the syllable as a unit of representation in the mental lexicon.

20.2.2 Language production

A widely assumed view of phonological encoding during language production involves the construction of a phonological word, i.e. prosodically fully specified word form, including all segmental and suprasegmental information necessary to pronounce the word (for an overview, see e.g. Levelt et al. 1999). After the selection of a word form from the mental lexicon, its constituent phonemes and metrical frame are retrieved in parallel (e.g. Roelofs and Meyer 1998). The metrical frame of a word specifies at least the number of syllables and the stress pattern of the word (e.g. Schiller 2006; Schiller et al. 2006). Whether or not CV structure, i.e. the consonant/vowel status of segments (phonemes), is also specified at this level is a matter of debate (see e.g. Dell 1986, 1988; Sevald et al. 1995; Meijer 1996; Costa and Sebastián-Gallés 1998; Cholin et al. 2004; but see also Roelofs and Meyer 1998). Depending on the language, other information may also be part of the metrical frame, for instance lexical tone (see Chen et al. 2002; Zhang and Yang 2007; Zhang et al. 2007; Zhang and Damian 2009).

Once the ordered segments and the metrical frame of a word have been retrieved, the segments can be inserted into metrical frames specifying syllable onset, nucleus, and coda positions (Shattuck-Hufnagel 1979; Dell 1988), a process called segment-to-frame association (Levelt and Wheeldon 1994). The insertion of phonemic segments into these slots proceeds in a piecemeal fashion from the beginning to the end of words. During this insertion process, the word form is assumed to be syllabified. The argument put forward by Levelt et al. (1999) for this later metrical framing (rather than having syllables stored in the lexicon) is as follows: word forms in the lexicon (such as *pen* or *and*) do not carry information about syllable boundaries because syllabification can transcend lexical boundaries—depending on the phonological context in which words occur. In the phrase *a pen and a pencil*, the syllable boundaries in (informal) fluent natural speech are *a.pe.nan.da.pen.cil* (syllable boundaries indicated by dots) and not *a.pen.and.a.pen.cil* due to a strong cross-linguistic tendency to maximize onsets (Pulgram 1970; Kahn 1976). This argument is supported cross-linguistically by the absence of syllable priming effects in speech production, whereas there is ample evidence for segmental priming effects during phonological encoding (e.g. Schiller 1998, 2000; Schiller and Costa 2006; but see Ferrand et al. 1996).

20.2.3 Methods for investigating phonological and phonetic encoding in the laboratory

This section focuses on online methods to investigate phonological encoding in speech production. That is, analyses such as offline speech error data (e.g. Nooteboom 1969; MacKay 1970; Fromkin 1971; Shattuck-Hufnagel 1979, 1987, 1992; Stemberger 1982; see Meyer 1992 for an overview) and speech error analyses of the speech of aphasic patients (e.g. Blumstein 1973; Dell et al. 1993; Dell et al. 1997; Romani and Calabrese 1998; Gordon 2002; Wilshire and Nespoulous 2003) are not considered. Behavioral and neurocognitive methods are treated separately.

20.2.3.1 *Behavioral methods*

Participants who take part in a psycholinguistic study are usually instructed to react as quickly and accurately as possible to a stimulus, i.e. a sound, a letter, a word, a phrase, or a sentence. Two so-called *dependent variables* can be measured: the time it takes participants to react to the stimulus, the so-called *reaction time*, and the number of errors. The reaction time (or RT) can be derived from a button-press response in case of a lexical decision ("is the stimulus a word in the language or not?") or a semantic classification ("is the stimulus referring to an animate or inanimate object?"). However, in case of an overt verbal response it can also be a *naming* (or *production*) *latency*, measured by means of a microphone and a voice key.

Reaction times are collected and averaged per condition to compare them statistically. Errors participants make while carrying out the task are coded as well on the basis of *a priori* criteria. For each condition to be compared statistically an *error rate* is computed. Usually, more difficult conditions yield longer reaction times and higher error rates than less difficult conditions. When faster reaction times are accompanied by higher error rates, participants have apparently reacted faster but at the same time less accurately. This phenomenon is called *speed-accuracy trade-off*. In such cases, participants applied an unknown strategy to solve the task, making the reaction times difficult, if not impossible to interpret. This basic paradigm is used in a variety of tasks; examples are outlined below.

20.2.3.1.1 *Priming (implicit priming/preparation vs. explicit priming)*

One way to investigate form encoding processes in language production is to study implicit priming, also known as the form preparation paradigm, introduced by Meyer (1990, 1991). She had Dutch participants produce sets of words that either overlapped in the onset phoneme (***h**ut*, 'tent'; ***h**eks*, 'witch'; ***h**iel*, 'heel'), or in the first two phonemes (***ha**mer*, 'hammer'; ***ha**ring*, 'herring'; ***ha**gel*, 'hail'), or in the first three phonemes (***hav**er*, 'oats'; ***hav**en*, 'haven'; ***hav**ik*, 'hawk'), or in the final phonemes (*h**aard***, 'stove'; *p**aard***, 'horse'; *k**aard***, 'map'). These were the so-called

homogeneous conditions, which were compared to so-called heterogeneous conditions in which the words did not overlap at all (*hut*, 'tent'; *dans*, 'ballet'; *klip*, 'cliff'). Reaction times were found to be faster in homogeneous compared to heterogeneous conditions, presumably due to response preparation when the beginning of the target words within a set overlapped and could thus be planned in advance, whereas in the heterogeneous condition the response could not be prepared since words within a set did not overlap in form. Note that this preparation effect does not occur when the final, but not the initial, part of the word could be prepared. The argument is that when participants cannot prepare the initial segment(s), there is no advantage of preparing the final segment(s) of a word because the production latencies will be determined by the latency of producing the initial segment. If this segment cannot be prepared, potential preparation of final segment(s) does not have any effect on the response times.

For the overlap of the initial segments of words, the magnitude of the preparation effect depended on the size of the string that could be prepared, i.e. the more segments overlapped among the words within a homogeneous set, the larger the preparation effect. Importantly, this was only true for begin-overlap. This has been taken to suggest that the phonological (and/or phonetic) planning of words is a strictly sequential process, i.e. proceeding in a left-to-right fashion from the beginning of words to their end. The form preparation paradigm has been employed not only to investigate phonological encoding but also morphological encoding (Janssen et al. 2002) and it has been applied to a variety of languages including Dutch, English (Damian and Bowers 2003), and Chinese (Chen et al. 2002).

Another method that has been used to investigate phonological encoding during speech production is *explicit priming*. In language and speech production research, it is important to ensure that participants are running through the entire production process, from conceptualization to articulation. Reading aloud, for instance, also includes components of speech production since the to-be-read words are uttered overtly, but it is unclear whether this task necessarily taps into processes such as conceptual preparation and lexical selection, since many words can be read by applying non-lexical grapheme-to-phoneme conversion procedures (see Coltheart et al. 2001 for a review). Furthermore, reading aloud also involves components of language comprehension, e.g. visual word recognition, and it is unclear how much these comprehension processes exactly interact with production processes. Therefore, one of the most challenging tasks of language production researchers is to design experimental paradigms to elicit target words in a controlled way without presenting the to-be-elicited verbal stimulus itself to participants—otherwise language comprehension is involved as well. Often, this is done by having participants name pictures. Pictures are non-linguistic stimuli, which can easily be recognized by the visual-perceptual system and subsequently named. It is generally assumed that after the visual recognition of a pictorial stimulus, this stimulus leads to the conceptualization of the meaning of the picture, which can afterwards be verbally

encoded by activating and retrieving its corresponding lexical item from the participant's mental lexicon.

One widely used paradigm employed in speech production research is the so-called picture-word interference (PWI) paradigm (see Glaser 1992 for an overview). In PWI, a picture of an object (or an action) is presented as a line drawing or a photograph, and at some stimulus onset asynchrony (SOA) relative to the onset of the presentation of the pictorial stimulus a verbal stimulus is presented visually or auditorily. Participants are usually told to name the picture and ignore the verbal stimulus they perceive. However, it has been demonstrated that participants nevertheless process the verbal stimulus, and therefore the influence of the verbal stimulus, also called the distractor, on the speech production process of the picture name can be measured, for instance by measuring voice onset latencies using a microphone connected to a voice key.

The PWI paradigm has proven to be very useful in language production research because distractor words can be manipulated in various dimensions, e.g. semantically, lexically (language type, word frequency, grammatical gender, etc.), phonologically, etc., which allows researchers to investigate the effects of distractor word processing on different levels of the production process, e.g. conceptual-semantic processing, lexical selection, or word-form encoding. Furthermore, the task is popular among researchers in the field of language production because it reflects relatively automatic production processes—participants' task is to produce (a phrase involving) the picture name while simply ignoring the distractor word. Whether language production processes using PWI are free of strategic processes, however, is a matter of debate. Generally, researchers try to keep the proportion of related trials as low as possible by incorporating as many filler trials as the practical constraints of an experiment allow.

Schriefers et al. (1990) used the PWI task to investigate the time course of speech production. They asked Dutch participants in the laboratory to name pictures while presenting them with auditory distractor words simultaneously (SOA = 0 ms, i.e. the onset of picture and distractor coincided). When the distractor words were phonologically related to the picture name (e.g. *harp*, 'harp'), the naming of the target picture name *hark* 'rake' was faster than in the control condition consisting of an unrelated word matched in frequency and length. The phonological facilitation was accounted for by assuming that the auditory presentation of the phonological distractor *harp* pre-activated segments (phonemes) in the production network necessary to encode the picture name. The segments that are shared between distractor and target (/h/, /a/, /r/) can presumably be selected faster, due to pre-activation by the distractor, when the target picture name *hark* is phonologically encoded. However, this effect disappeared when the phonologically related distractor words were presented before picture onset, i.e. at a slightly negative SOA (SOA = −150 ms), presumably because the production system of the speaker was not yet involved in the process of phonological encoding, i.e. retrieving the target segments needed

for overt pronunciation of the picture name (see also Meyer and Schriefers 1991). At a slightly later SOA of +150 ms, however, the effect was still demonstrated. This has been taken by some researchers to show that phonological encoding in speech production is a rather late process in the production cycle.

The phonological facilitation effect can also be observed when distractors are presented visually. Upon the visual presentation of a word, the phonological representation is activated only a couple of milliseconds after the orthographic code has been activated (Ferrand and Grainger 1992, 1993, 1994), and well within 50 ms after visual onset of the target. Also, phonologically related non-words or parts of words (e.g. segments or syllables; see Schiller 1998, 2000; Starreveld 2000) can be employed to investigate phonological encoding, which gives researchers additional degrees of freedom when creating the phonologically related distractors. Segmental (phonemic) facilitation effects are extremely stable and reliable, and therefore are often used as a kind of litmus test when manipulating the relation between the target picture name and the distractors on another dimension. However, other phonological units, such as syllables (Schiller 1998, 2000; Schiller et al. 2002; Schiller and Costa 2006; but see Ferrand et al. 1996; Ferrand et al. 1997) or prosodic properties like lexical stress (Schiller et al. 2004), have not been implicated using the PWI paradigm.

As mentioned above, analyses of naturalistic speech errors are not discussed in this section because it is considered offline data (though they have been a valuable source of evidence for psycholinguistic theories). However, speech errors can also be elicited in more controlled ways in the laboratory, e.g. by using the SLIP (Spoonerisms of Laboratory-Induced Predisposition) technique to elicit spoonerisms. This method was first introduced by Baars et al. (1975) and has since then successfully been used to tackle important theoretical questions such as the issue of the lexical bias effect in naturalistic speech error data, i.e. the fact that more segment substitutions result in existing words than in non-existing, pseudo-words than would be statistically expected by chance (see also Hartsuiker et al. 2005; Nooteboom and Quené 2008; Oppenheim and Dell 2008; though see Pouplier and Goldstein 2005; Gafos and Goldstein, this volume).

20.2.3.1.2 *Monitoring (phoneme, syllable, and metrical stress monitoring)*

Another behavioral task that has been used in the past to investigate phonological encoding in speech production is verbal self-monitoring. In a monitoring paradigm, participants are asked to monitor their speech plan for certain targets, for instance, a segment or a syllable, during the implicit production of another stimulus, sometimes called the carrier. If the target is contained in the carrier, the participant is often required to indicate this by pressing a button, and the button press latencies are measured and taken to reflect the time of phonological encoding of the carrier relative to a control condition.

Wheeldon and Levelt (1995), for example, required bilingual Dutch–English participants to internally generate Dutch translations to English prompt words, which

were presented via headphones. However, participants did not overtly produce the Dutch translation words but self-monitored them internally for previously specified target segments. For example, participants would hear the English prompt word *hitchhiker* and were asked to press a button on a button box in front of them as fast as possible if the Dutch translation (lifter) contained the phoneme /t/. Thus, for *hitchhiker* participants would press the button, whereas for *cream cheese* (*roomkaas*) they would not. Wheeldon and Levelt found that the button press latencies varied as a function of the position of the target phoneme in the translation word. That is, participants were faster when the pre-specified phoneme (e.g. /t/; typeset in bold print) was in word onset position (e.g. *garden wall—**t**uinmuur*) than when it occurred in the middle (e.g. *hitchhiker—lif**t**er*) or at the end of the translation word (e.g. *napkin—serve**t***). The earlier the target phoneme occurred in the Dutch word, the shorter the decision latencies. The authors interpreted these data as support for the claim of rightward incremental phonological encoding during speech production. The fact that the location of metrical stress influenced listener responses shows that these effects are localized at the phonological word level, but not at an earlier (lexical) or later (phonetic) level of word production planning (for arguments, see Wheeldon and Levelt 1995). Moreover, Wheeldon and Levelt (1995) observed a significant increase in monitoring times when two segments were separated by a syllable boundary. One possible explanation is that the monitoring difference between the target segments at the syllable boundary (e.g. *fie**t**.ser* vs. *lif.**t**er*) might be due to the existence of a marked syllable boundary or a syllabification process that slows down the encoding of the second syllable. These results have been replicated for English (Wheeldon and Morgan 2002; Morgan and Wheeldon 2003) and Dutch (Schiller 2005). Furthermore, it has been shown that morphological boundaries do not influence segmental monitoring latencies (Schiller 2005), presumably because morphological encoding of words precedes phonological encoding and therefore does not exert any influence on phonological encoding in the monitoring paradigm.

More recently, the self-monitoring paradigm has also been used to investigate the phonological encoding of lexical or metrical stress. For instance, Schiller et al. (2006) asked Dutch speakers to indicate by a button-press if a word had initial or final stress, respectively. They presented line drawings to their participants. However, instead of naming the pictures overtly, participants were requested to decide on the location of the metrical stress, e.g. initial or final stress. It should be noted that this metalinguistic task may be too difficult for some participants. Furthermore, it is necessary to include control experiments to make sure that any findings of the stress monitoring could not be due to stimulus-set-related factors such as visual differences between the pictures or differences in the articulatory properties, especially the word onset phonemes, of the corresponding picture names (see also Schiller 2006 for recommendations).

The button-press latencies of the stress-monitoring experiment indicated that participants were significantly faster in deciding that a picture name had initial stress than final stress. This fits the hypothesis that not only segmental but also suprasegmental encoding of word forms follows an incremental time course, with earlier parts of a word being phonologically encoded before final parts. However, to exclude the alternative account that the stress-monitoring result reflected the default metrical stress distribution in Germanic languages such as Dutch (the majority of bisyllabic words have initial stress in languages such as Dutch, Quené 1992; Levelt and Schiller 1998; and English, Cutler and Carter 1987); Schiller et al. (2006) also tested trisyllabic targets in which both second and third syllable stress are deviant from the default initial stress pattern in Dutch. The results, were similar to the bisyllabic targets, i.e. participants were faster in indicating that a picture name had second syllable stress compared to third syllable stress. Together, these results were taken to indicate that speakers not only encode the segments of word forms incrementally but also their corresponding metrical frames.

20.2.3.2 *Neurocognitive methods*

Useful as they may be, behavioral methods are indirect. During the neurocognitive revolution in the 1990s, neurocognitive methods such as electroencephalography (EEG)—and derived from that event-related potentials (ERPs)—as well as positron emission tomography (PET), transcranial magnetic stimulation (TMS), and functional magnetic resonance imaging (fMRI), to name just a few, were also introduced into the areas of psycho- and neurolinguistics, allowing more direct observations about processing (see Idsardi and Poeppel, this chapter). Here, I describe how the EEG/ERP method can be applied to research on phonological encoding in speech production and summarize some recent findings.

20.2.3.2.1 *Electroencephalography (EEG)/Event-Related Potential (ERP)*

Electroencephalography (EEG) is a well-established, non-invasive method to measure neuronal activity of certain cells in the cortex. Certain requirements must be fulfilled before neuronal activity can be picked up by surface electrodes on the skull (see Kutas and Van Petten 1994 and Kutas et al. 2006 for overviews), and the activity that is generated by these neurons is relatively weak. However, neuronal activity related to the cognitive processing of a particular stimulus, for instance a word, can be visualized by EEG activity employing a relatively simple trick. When the same (type of) stimulus is presented often enough, EEG signals related to the same (class of) stimuli or the same experimental condition can be averaged. As a result, the random background EEG activity that normally masks the EEG activity specific to the stimulus is filtered out and what remains is the neuronal activity specifically related to the (class of) stimuli under investigation, represented as the event-related potential (ERP), sometimes also called event-related brain potential.

Van Turennout et al. (1997) used the so-called lateralized readiness potential (LRP) with the EEG methodology to investigate phonological encoding during speech production planning. These authors presented pictures which participants wearing an electrode cap were required to name overtly. However, on some trials, a frame was presented around the picture indicating that participants should not name the picture but make a phonological decision about the target picture name. More specifically, they pressed a left or right button in case of certain target phonemes ("go trials") and refrained from button-pressing in case of certain other phonemes ("no-go trials"). Van Turennout et al. (1997) varied the position of the target phoneme in the picture name, i.e. it could occur in the word onset (e.g. the /t/ in *tafel*, 'table' or *tijger*, 'tiger') or in the word offset (e.g. the /n/ in *schoen*, 'shoe' or *spin*, 'spider'). Decision latencies showed that participants were significantly faster in detecting target phonemes that occurred in the word onset than in the offset (replicating findings by Wheeldon and Levelt 1995; see above). Furthermore, this behavioral result has an electrophysiological counterpart, i.e. the LRP waveform on no-go trials, i.e. trials on which participants did not press the button, went back to zero baseline earlier when the phoneme occurred in word onset position than in offset condition. This indicates that participants stopped preparing a motor response earlier when the target phoneme appeared in word onset position rather than word offset position.

Schiller et al. (2003) used a related task to investigate the relative time course of two phonological encoding processes, i.e. metrical encoding and syllabification. Metrical encoding involves the retrieval of the stress pattern of a word, whereas syllabification is construction of the syllabic structure of a word. However, the relative timing of these two processes is unknown. Schiller et al. (2003) employed an implicit picture-naming task and recorded event-related brain potentials to obtain fine-grained temporal information about metrical encoding and syllabification. For instance, in the metrical task, participants saw line drawings and were required to indicate by a yes/no-response whether the picture name had initial stress (e.g. *kano*, 'canoe') or final stress (e.g. *kanon*, 'cannon'). In the syllabification task, the same participants made decisions about the syllable affiliations of certain target phonemes. For instance, does the intervocalic consonant, e.g. the /n/ in this case, in 'ka-non' or in 'ka-no' belong to the initial or final syllable? Instead of the LRP, these authors measured the N200, an ERP component related to response inhibition visible on frontal electrodes. The N200 effect is the difference between the N200 component for no-go trials and go trials. Results revealed that both tasks generated effects that fall approximately within the 275–450 ms time window generally assumed for phonological encoding (Indefrey and Levelt 2004). However, there was no timing difference between the two effects. The observed ERP effects for both tasks fell within the 250–350 ms time window and were therefore interpreted as showing parallel processing of metrical and syllabic encoding.

Schiller 2006 employed the ERP technique to investigate the time course of metrical stress encoding in Dutch. He used the same task and similar materials that were used in Schiller et al. (2006), but in addition to reaction times, event-related brain potentials were also measured. The monitoring latencies obtained by Schiller et al. (2006) were much longer than the picture-naming latencies for the same pictures, presumably because the monitoring task is more complex and indirect—due to the additional decision component in this task—compared to simple picture naming. However, Schiller 2006 aimed to find out more about the stress-encoding component during phonological encoding in speech production and obtain more precise information about when metrical stress is encoded. The N200 effect may help answer this question. The behavioral results were similar to the earlier, purely behavioral study (Schiller et al. 2006)—there was a significant monitoring advantage of 86 ms for initial stress over final stress in bisyllabic words. The difference between the two N200 peaks was 58 ms, i.e. a similar temporal difference between decisions to initial and final stress, though the N200 peak latencies were much earlier than the button-press responses. The N200 peak is taken as indicating when information the response is based on must be available. Therefore, it gives an upper boundary about the temporal availability of this information, but it does not say anything about the earliest point when the information must be available. Indefrey and Levelt (2004) estimate phonological encoding in speech production to take place between approximately 275 ms and 450 ms post picture onset. Therefore, the N200 peak latencies obtained in Schiller (2006), which were taken as upper boundaries for information availability, were interpreted to be in accordance with this estimation.

The studies by Van Turennout et al. (1997), Schiller et al. (2003), and Schiller 2006 employed tacit or implicit picture naming because it was reasoned for a long time that overt articulation of picture names, for instance, would cause speech-motor artifacts that would make any interpretation of the EEG signal impossible (see also Schmitt et al. 2000; Zhang et al. 2007; Zhang and Yang 2007; Zhang and Damian 2009). Motor movements, even as small as eye blinks, are much larger in magnitude than the ERP waveforms generated by cognitive processes, such as speech planning for instance. However, recently, researchers nevertheless started to use overt speech production tasks in EEG/ERP research. In a typical naming experiment, e.g. picture naming, participants need to plan their speech production response and typically do not start to move their articulators until 500-600 ms after picture onset, and it has been shown that until that point in time approximately, the EEG/ERP signal is relatively free of motor artifacts (see Schmitt et al. 2002; Guo and Peng 2006; Christoffels et al. 2007; Hirschfeld et al. 2008; Koester and Schiller 2008).

As mentioned above, there are other neurocognitive methods, even electrophysiological methods such as magnetoencephalography (MEG; see for instance Levelt et al. 1998; Maess et al. 2002), but also hemodynamic methods such as PET and

fMRI, as well as brain stimulation methods (e.g. TMS; Schuhmann et al. 2009) that have been successfully employed to investigate the speech production process (see overview in Indefrey and Levelt 2004).

20.2.4 Future directions

This section is meant to give the reader a sample of some of the available experimental methods and designs to investigate phonological encoding of spoken language. Especially in the area of neuroimaging methodology, there are many new developments, which have not been mentioned in this section, for instance, functional Near-InfraRed Spectroscopy (fNIRS), a non-invasive optical measurement method to monitor brain activity in awake participants (including young children) by determining the amount and oxygen content of hemoglobin in the blood through the degree of absorption of near-infrared light (see e.g. Kovelman et al. 2008). However, research in the area of word-form encoding during language production will have to combine behavioral and neuroimaging methods to be able to answer important theoretical questions about the time course of the processes involved and responsible for word-form encoding and the architecture and representation of the underlying (neuro)cognitive systems. The goal of a comprehensive theoretical account of the neurocognitive mechanisms underlying word-form encoding can only be achieved by joining forces between behavioral and neurocognitive research on the one hand, and research on healthy participants and brain-damaged individuals on the other hand.

20.3 Measuring phonetic perception in adults

Paul Iverson

20.3.1 Introduction

The task of measuring phonetic perception in typical adults may seem straightforward in comparison to studying perception in children. You could play your participants natural recordings and ask them to identify what they heard, play pairs of synthetic stimuli and ask whether they were the same or different, or choose a variety of other methods. However, phonetic perception relies on many underlying processes, and individual differences in measures of phonetic abilities are often

poorly correlated with each other. For example, identification accuracy is often only moderately correlated with production accuracy for second-language learners (e.g. Flege et al. 1997; Bradlow, Akahane-Yamada, Pisoni, and Tohkura 1999; Hattori and Iverson 2009; Iverson et al. 2011), and it is likewise not unusual for individuals to have patterns of discrimination that do not correspond to their identification performance (e.g. Iverson et al. 2006; Heeren and Schouten 2008). Different levels of phonetic processing can interact in complex ways. For example, the difficulty that Japanese adults have in learning the English /r/-/l/ category may be due to their lower-level perceptual sensitivities for these phonemes which makes irrelevant acoustic variation more salient than the critical differences in F3 (Iverson et al. 2003). It is also possible that the similarity of their native-language flap interferes with learning these English categories, and it may be that articulatory, rather than acoustic, similarity between categories is more important (Aoyama et al. 2004; Best and Tyler 2007; cf. Hattori and Iverson 2009). The choice of methods to examine phonetic perception depends on which of these types of processes you wish to assess.

The methodological choice also becomes complex if one wants to analytically investigate the use of different acoustic cues. One of the current trends in the literature is to go beyond investigations of primary acoustic cues, and investigate finer-grained phonetic variation in more naturalistic connected speech materials (e.g. Pitt et al. 2005; Van Engen et al. 2010). This type of investigation can be done with acoustic measurements of read speech, but it can be much harder to test whether this phonetic variation is relevant to perception; it is difficult to control this variation in stimuli yet still retain naturalistic complexity. Methodologies that better allow for multidimensional phonetic variation can be used in these situations, but they are more difficult to implement than traditional identification and discrimination tasks.

The following is a review of some of the methods for assessing phonetic perception in adults, beginning with a discussion of synthetic, processed, and edited speech (Section 20.3.2), and followed by a discussion of techniques to measure categorization (e.g. identification ability, use of acoustic cues, 20.3.3) and sensitivity (i.e. ability to discriminate acoustic differences along phonetic dimensions, 20.3.4). The question of why such investigations of phonetic contrasts are interesting scientifically is mostly left to other chapters (e.g. Holt, this volume; Nguyen, this volume).

20.3.2 Synthetic, processed, and edited speech

The use of natural speech recordings has the clear advantage of having all of the complex and subtle phonetic variability of real speech, but it has the disadvantage of not giving the researcher much control over what is varying. Synthetic speech often sounds unnatural and has limited acoustic variation, but the complexity and naturalness of synthetic speech is more of a limitation of the researcher than it is of

the technology (see Reetz, this volume for a presentation of synthesis techniques). That is, there is no reason, in principle, why synthetic speech cannot fully match natural speech, although it is difficult to know exactly what the synthetic speech ought to reproduce. The suitability of synthetic speech can be assessed by using both natural and synthetic speech within the same study, as part of a larger test battery. If the more analytic results generated from experiments using synthesized speech correlate with identification results from natural recordings, this can help verify that the synthetic speech reproduced the major acoustic cues that are relevant in more natural materials (e.g. Iverson and Evans 2007; Iverson et al. 2008; Hattori and Iverson 2009).

Synthetic speech can be generated fully parametrically, such as within Klatt synthesis (Klatt and Klatt 1990). In these synthesizers, listeners specify a range of acoustic parameters (e.g. formant frequencies, fundamental frequency, voicing amplitude) at different points in time, and the synthesizer generates stimuli based on these values. Although one could use values based on average acoustic measurements for these parameters (e.g. specifying the formant frequencies for an average male /i/ vowel), this usually produces stimuli that sound unnatural. A more effective approach is to attempt a copy synthesis, where one starts with a single natural recording of a particular item (e.g. an adult saying the word /bid/) and then the researcher attempts to choose parameter values that makes the synthetic output match this natural recording. Acoustic measurements of the recording can be used to generate an initial set of values that are input into the synthesizer, which can be done automatically (e.g. Boersma and Weenink 2009). The output of the synthesizer can then be compared to the natural recording so that the parameter values can be further adjusted. By iteratively adjusting the parameters in this fashion, it is possible to generate synthetic speech that is a close match to the original. Once this copy synthesis has been generated, the parameters can be manipulated as necessary to create additional stimuli with variation in the desired acoustic dimensions.

It is also possible to signal-process natural recordings of speech to manipulate key acoustic cues. For example, overlap-add methods (e.g. Moulines and Charpentier 1990) can be used very effectively to manipulate pitch and duration (e.g. alter a pitch contour, or manipulate the duration of a vowel). Formant frequencies can be manipulated using techniques that model the source and filter of the original speech signal (e.g. LPC, Markel and Gray 1976; STRAIGHT, Kawahara et al. 1999). Using these techniques, one can, for example, take a natural recording of a syllable, decompose it into the acoustic characteristics of the source (e.g. glottal spectrum) and filter (i.e. vocal tract frequency response), manipulate the filter function to change the formant frequencies, and then recombine the source and filter to create a new stimulus (e.g. Liu and Kewley-Port 2004; Iverson et al. 2005). These techniques can be used to morph one syllable into another (e.g. Slaney et al. 1996; Kawahara et al. 1999; Stephens and Holt submitted), creating stimulus continua that can

sound highly natural, although they may not offer the full control over acoustic parameters that is possible with fully parameterized synthetic speech.

Finally, the acoustic waveforms of natural speech can be directly edited. For example, samples can be removed or added to alter voice onset time (e.g. Newman 2003) or the salience of short vowels (Dupoux et al. 1999). Fricative noises can be mixed directly without needing to decompose the speech into sources and filters (Norris et al. 2003; Repp 1981b). Critical aspects of the speech signal can also be replaced with silence (e.g. Strange et al. 1983). These editing techniques are useful in a fairly narrow range of situations, but they also can be effective for producing natural-sounding stimuli.

20.3.3 Measurements of categorization

20.3.3.1 *Identification tasks*

A basic way of assessing phonetic perception is to play individuals natural recordings of syllables or minimal-pair words and have them give a forced-choice response indicating what they heard. One purpose of such an experiment is to provide a baseline assessment of how accurately a particular group can recognize a pair of phonemes (e.g. the identification of English /r/-/l/ by Japanese adults), which is a useful control within a larger suite of more analytic tests.

The identification of larger sets of stimuli, such as a set of all of the consonants used in a particular language, can lend itself to more detailed analyses. Such experiments generate a confusion matrix, which is a table that lists how often each phoneme was identified as each response (e.g. Miller and Nicely 1955). The structure of the responses can be revealing. A researcher can start by inspecting which phonemes were confused and make inferences based on the error patterns. To better quantify the patterns of errors, the confusion matrix can be analyzed using statistical techniques such as hierarchical cluster analysis or multidimensional scaling. Hierarchical cluster analysis generates nested trees to describe what groups of phonemes sound similar (i.e. are more frequently confused with each other; e.g. Shepard 1972; see Clopper, this volume). Multidimensional scaling places the phonemes in a graphical space where similar phonemes are plotted close together and dissimilar phonemes are plotted far apart; researchers can then inspect this plot to infer which perceptual dimensions were most important to listeners (e.g. Shepard 1972). One can also analyze confusion matrices using information transfer analysis (Wang and Bilger 1973), which can be used to examine how well individuals perceived specific phonetic properties (e.g. voicing).

A typical identification experiment with synthetic speech uses a series of stimuli that varies along a one-dimensional continuum. This continuum could vary a single acoustic cue (e.g. voice onset time along a /b/-/p/ continuum), or could covary

multiple acoustic cues simultaneously (e.g. voice onset time, F1 cutback, onset f0, and burst amplitude). The latter continuum is still considered to be geometrically one-dimensional because the multiple cues covary rather than being manipulated independently (e.g. all cues vary together from /b/ to /p/). The results can generate a labeling graph that plots the identification proportions for each stimulus (see Figure 20.3.1 of Prieto, this volume for an illustration). This will display, for example, a curve that shows where along a /b/-/p/ continuum the listener stops hearing the stimuli as /b/ and begins to hear it as /p/ (i.e. the identification boundary location, corresponding to the point where each phoneme has a 50 percent chance of being identified) as well as the steepness of the transition between the two phonemes. The location of the category boundary can be important for examining, for example, how listeners of different languages perceive a particular distinction, such as voicing in stop consonants (e.g. Lisker and Abramson 1970). The steepness of the boundary can indicate, for example, how consistently a group of listeners use a particular acoustic cue (e.g. Simon and Fourcin 1978).

20.3.3.2 *Multiple dimensions and goodness-rating tasks*

One limitation of typical identification tasks with synthetic speech is that the variation must be constrained to a small number of dimensions (e.g. a number of acoustic parameters that are covaried along a one-dimensional continuum), but natural speech varies on many independent dimensions. It is possible to use stimuli that vary along multiple dimensions (e.g. vowel spaces independently varying F1, F2, and/or duration; Johnson, Flemming, and Wright 1993; Morrison 2006), but the number of possible stimuli grows rapidly when multiple dimensions are used (i.e. one must synthesize many combinations of the various acoustic cues) and it is hard to calculate or visualize identification boundaries that cut through high-dimensional spaces. Thus, one is effectively limited to manipulating one or two dimensions in identification tasks.

An alternative approach for higher-dimensional stimulus spaces is to map best exemplars using a goodness-rating task. Individual stimuli are played to subjects and they rate whether they sound like good or poor examples of the stimulus. One advantage to this approach is that the best exemplar can be represented as a single point (i.e. stimulus location) regardless of the number of dimensions. In contrast, identification boundaries are single points when the stimulus space has one dimension, lines when there are two dimensions, planes when there are three dimensions, etc. Moreover, these complex identification boundaries can have curves that are hard to model. In higher-dimensional spaces, the large number of possible stimuli is still a problem if one measures best exemplars, but it is possible to use adaptive algorithms that change the stimulus selection based on the listener's previous responses (e.g. Iverson and Evans 2007; Oglesbee and de Jong 2007). That is, computational methods can be used to efficiently guide the search, such that the

listener must make judgments on only a small number of stimuli rather than having to hear every possible combination of acoustic cue values.

Best exemplar and identification boundary locations are similar in that they both assess how variation in acoustic information affects perceived categorization. Best exemplars may be more sensitive to secondary acoustic cues that have relatively minor effects on identification accuracy. For example, the identification of English /r/-/l/ by native speakers is almost entirely determined by F3, but best exemplar maps can reveal that listeners are sensitive to additional acoustic cues that often vary between these phonemes in production (e.g. transition duration is shorter in /l/ than /r/; Hattori and Iverson 2009). Identification and discrimination measures most clearly differ in terms of whether central (best exemplars) or peripheral (identification) areas of the category are being assessed. Identification boundary locations essentially depend most on how listeners categorize the most ambiguous items in a stimulus set, and such judgments on ambiguous stimuli can be more affected by biases than are best exemplars (e.g. due to lexical status, Allen and Miller 2001). That being said, identification boundaries can be more reliable to measure, particularly for consonants. Listeners tend to have high perceptual sensitivity near category boundaries (e.g. Liberman et al. 1957) which sharpens the identification function (i.e. small acoustic differences can produce large differences in categorization). However, listeners tend to have poor sensitivity near best exemplars (e.g. Iverson and Kuhl 1995), which means that there can be a large region of "best exemplars" that essentially sound the same.

20.3.3.3 *Eye tracking*

Speech recognition does not necessarily need to be assessed by pressing a key. For example, eye fixations can be recorded using an eye tracker in a visual world paradigm (Tanenhaus et al. 1995; Weber 2008); listeners are shown several objects then played a speech instruction to manipulate the objects (e.g. "click on the beetle"). The eye fixations can be used to track how the ambiguity of the speech signal changes over time (e.g. how long it takes for individuals to look at the correct object more often than the other objects), and this temporal evaluation of the speech recognition process can be important for evaluating how differences in phonetic information affect other linguistic processes (see Speer, this chapter).

20.3.4 Measurements of perceptual sensitivity

20.3.4.1 *Discrimination judgments*

The simplest discrimination judgment, sometimes referred to as AX or Same-Different, involves playing two stimuli on one trial and having subjects judge

whether they are same or different; about half of the trials are the same stimulus played twice and the others are two different stimuli. This task is typically applied to a synthetic stimulus continuum, allowing the experimenter to assess how well a listener can discern acoustic differences between neighboring stimuli both within and across phonetic categories. Statistics, such as d' within the framework of Detection Theory (e.g. Macmillan and Creelman 1991), can be calculated to assess sensitivity by comparing the proportion of "*hits*" (i.e. listeners correctly judging that the stimuli were different) to those of the "*false alarms*" (i.e. listeners erroneously judging that the stimuli were different when they were actually the same). Comparing same and different trials in this manner is important because it controls for response bias. For example, if a listener is conservative (i.e. reluctant to press the "different" button unless they are absolutely sure that the stimuli are different) this would reduce the number of hits but also reduce the false alarm errors. On the other hand, a listener who is more willing to press the "different" button would have an increased number of hits but also more false alarms. Comparing the results for same and different trials can thus control for this kind of variability in bias between subjects, allowing more direct comparison of how well these listeners could discern the acoustic differences between sounds.

Many other tasks can also be used to assess perceptual sensitivity. For example, oddity tasks can be used where, for example, listeners hear three stimuli with two being the same and one being different; listeners need to decide which is the different one. AXB tasks can similarly be used, in which listeners hear three stimuli on each trial and judge whether the middle stimulus sounds more like the first stimulus or the second one. One advantage of these kinds of tasks is that listeners are forced to judge which stimulus was different on each trial, so this factors out their individual level of willingness to declare that they heard a difference. Same trials are thus unnecessary, which reduces the number of trials that must be presented and makes it easier to employ adaptive techniques that change the acoustic difference between stimuli to find a discrimination threshold (Levitt 1971).

20.3.4.2 *Category discrimination judgments*

The discrimination judgments described above are generally appropriate only for synthesized stimuli; careful control of the stimuli is necessary to make sure that listeners can detect an acoustic difference only along the phonetic dimension that is of interest, rather than because of some irrelevant stimulus artifact. However, natural stimuli can be used in category discrimination tasks that use acoustically variable stimuli and listeners have to judge which stimuli were the same phoneme, rather than judge which stimuli were acoustically identical (e.g. Best et al. 1988; Hojen and Flege 2006). For example, in an oddity task as described above, listeners could hear three stimuli recorded by three different talkers, with two stimuli having the same phoneme and one being different (e.g. /pi-pi-bi/). Listeners make their

judgments based on paying attention to whatever phonetic dimensions they think are relevant for these types of phonemes and ignoring irrelevant variation.

This task has obvious appeal in that fully natural stimuli can be used. However, the interpretation of the results is somewhat ambiguous. There is a sense in which this is essentially an identification task, except that it does not require listeners to give explicit response labels. That is, listeners who have well-formed categories for these phonemes, or similar phonetic categories in their first language, could perform this task by covertly identifying each stimulus and judging which belonged to the different category. However, there is also a sense in which this task assesses perceptual sensitivity along phonetic dimensions. That is, listeners who are unable to identify these stimuli could still perform this task by judging which stimulus sounds the most different from the others, as long as they are able to ignore irrelevant variation due to talker differences. If a listener has poor performance on this task, it indicates both that the listener has trouble categorizing the stimuli and attending to differences along the relevant phonetic dimensions, because either would have allowed the listener to perform the task well. It does not, however, demonstrate that listeners are completely unable to discriminate the acoustic differences between stimuli; ignoring irrelevant variation is as important as hearing the important differences. If a listener is accurate with this task, it demonstrates that listeners can discern differences between the stimuli along the relevant phonetic dimensions, but it is unclear whether or not the listener has well-formed categories for these stimuli.

20.3.4.3 *Multidimensional scaling*

There are at least two limitations with assessing perceptual sensitivity using discrimination experiments. First, discrimination judgments are only useful with stimuli that cannot be discriminated with 100 percent accuracy; the relative similarity of easy-to-discriminate items cannot be easily measured due to ceiling effects. Second, discrimination judgments are most feasible for measuring variation along a single continuum, for the same reasons given above for identification experiments, but mapping perceptual sensitivity in multiple dimensions can give you a broader view of how the different dimensions interact (e.g. see Iverson et al. 2003, 2008).

One solution is to play pairs of stimuli to subjects, ask them to rate on a continuous scale whether they sound similar or dissimilar, and analyze these ratings using multidimensional scaling. For example, one could start with a set of 18 /r/-/l/ stimuli that vary orthogonally in F1 and F2 frequencies, play subjects every possible pair of these stimuli to collect similarity judgments, and analyze the resulting similarity matrix with multidimensional scaling to generate a perceptual space for these stimuli that reveals how listeners perceive the physical distances between the stimuli (e.g. Iverson et al. 2003). Multidimensional scaling is effectively limited by the number of dimensions that can be visualized in a graph (i.e. 1–3) but it can still give a broader

view of the perceptual space than can be seen by examining individual pairs in a discrimination experiment. One could imagine that these sorts of subjective ratings are less reliable than more objective measures such as discrimination accuracy, but studies have revealed that the two measures correspond quite well (e.g. Iverson and Kuhl 1995).

20.3.5 Conclusion

Much of the basics of what we know about the relationships between speech acoustics and perception stem from work at Haskins Laboratories in the 1950s, with most of this relying on identification and discrimination judgments along synthetic one-dimensional continua (e.g. Liberman et al. 1957). Although the experimental techniques used in these studies are still valuable, gains in our current understanding of speech perception can be made if we start to make more connections between this kind of work to speech perception under more naturalistic conditions, assessed with a wider range of populations and communicative situations, and with a broader range of measures (e.g. neuroimaging). Towards this goal, it seems particularly important to begin to incorporate more naturalistic multidimensional acoustic variation in studies of speech perception, either using some of the techniques described above or by developing new methods.

20.4 Eye movements as a dependent measure in research on spoken language

Shari R. Speer

20.4.1 Introduction

When people speak and listen, they tend to look at locations and objects in their immediate environment that are relevant to their conversation. That these looks might serve as a window to underlying cognitive processes was suggested in the psycholinguistic literature as early as 1974, when Cooper tracked the eyes of participants who saw pictures as they listened to stories. His findings foreshadow many in the current literature, such as the fact that listeners looked to the picture of a mentioned object even before its name was completely pronounced, and that upon hearing the verb *wormed*, they looked to a semantically and visually similar item, a snake. Contemporary widespread use of eye-tracking technology to study

speech processing began with Tanenhaus et al.'s (1995) "visual world" experimental paradigm (and the availability of less expensive eye-tracking systems, many with data reduction software). Here, a participant typically follows spoken instructions to manipulate real-world or screen-displayed objects. Inferences are drawn about the listener's language processing based on the timing and overall pattern of *saccades* (ballistic movements of the eye) and *fixations* (stable periods of visual information intake) to the visual objects when they are potential referents in the speech stream. A smaller number of studies have used eye movements to study language production (e.g. Meyer et al. 1998; Griffin and Bock 2000; Gleitman et al. 2007). As discussed below, the eye-tracking method is particularly well-suited to theoretical questions involving changes in linguistic representation over the time course of language processing. Depending on the experimental task, the method is also a good choice when an implicit measure (one not susceptible to conscious attentional or response strategies) is required.

The following brief overview of eye-tracking methodology includes initial discussion of the advantages, applications, and assumptions necessary for its use (Sections 20.4.2–4). These are followed by examples of experiments illustrating the breadth of approaches to data collection and the range of questions addressed, with discussion of factors that influence the linking assumptions governing the interpretation of the data (Section 20.4.5). The examples chosen explicate certain methodological choices and concerns as well as basic theoretical and empirical questions relevant for laboratory phonology. Next is a discussion of methodological concerns and data analysis issues, including the choice of dependent variables available from eye-tracking, assumptions about the time-locking of the eyes to the spoken signal, issues of synchronization between the auditory signal and the eyes, and implications of how the data are displayed (Sections 20.4.6–7). The final section (20.4.8) presents current issues in statistical analysis, and issues of data interpretation and analysis that remain under debate.

20.4.2 Advantages

Eye tracking has many advantages for the study of spoken language processing (for an extensive list of these, see Tanenhaus and Trueswell 2005). It provides a continuous, non-intrusive and implicit measure of processing difficulty. It allows data collection throughout the time course of listeners' perception and interpretation of speech, or speakers' apprehension of the gist of a situation and their description of it. Eye data can be reliably time-locked to spoken language data. Eye movements during language processing are largely unconscious, and thus less susceptible to the development of response strategies. Participants wearing lightweight head gear (as with the ASL6000, SMI Eyelink, and ISCAN head-mounted systems), or seated near a tabletop or screen-mounted eye camera (as with Tobii or Senso-Motor

Instruments, ASL and ISCAN remote systems), can speak, manipulate objects, and move about during data collection.

20.4.3 Applications to spoken language research

Eye tracking can be used during natural tasks, with varying levels of complexity in the spoken signal and the visual display. As exemplified below, it has been used to study speech perception, word recognition, sentence parsing, and discourse processing. Visual stimuli may be presented on a computer screen as words, pictured objects or scenes, or video clips. A language processing context can be evoked through non-linguistic stimuli in the visual scene, increasing complexity without increasing memory load. Researchers have used real-world objects and instructed cooperative tasks to address questions about discourse-level and pragmatic factors that were not illuminated by more traditional techniques requiring lists of text or sentences of a particular form. Eye movement monitoring may be the only experimental method that provides a continuous implicit record of cognitive processes as they unfold over time in unscripted conversation. Example research includes the study of the generation of noun phrase forms as speakers follow a recipe and cook together (Hanna and Tanenhaus 2004), and the study of speakers' intonation as they work together to decorate a holiday tree. Figure 20.4.1 shows the laboratory set up for the holiday tree task (Ito and Speer 2008).

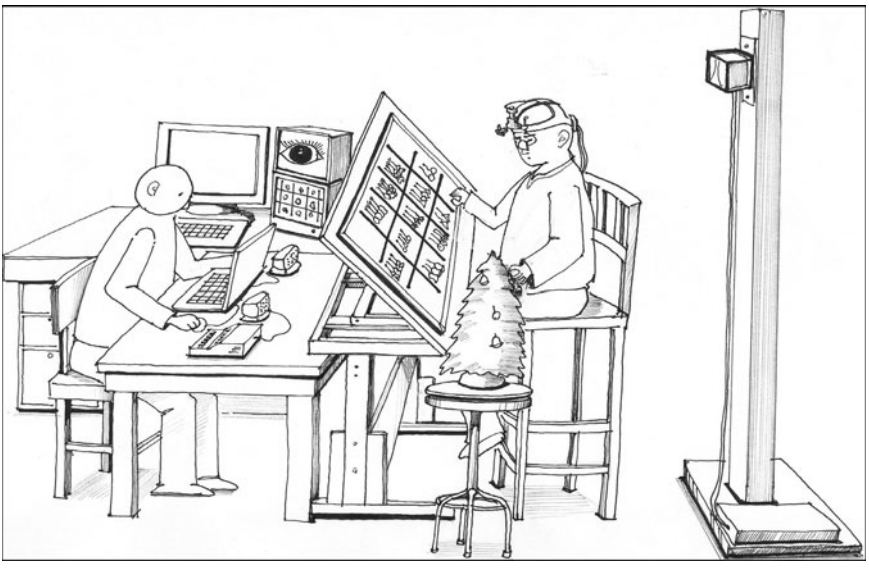

Figure 20.4.1. Experimental set-up for Ito and Speer (2008) holiday tree experiments. Originally Figure 5 from K. Ito and S. R. Speer (2008).

20.4.4 Necessary linking assumptions

Successful experimental design and useful interpretation of eye movement monitoring data depend critically on specification of the assumptions used to link performance in the visual task with the underlying attentional and linguistic processes. Many of these "linking assumptions" rely on background knowledge from the study of visual perception, including assumptions about the execution and timing of eye movements (see Allopenna et al. 1998; Altmann and Kamide 2004; compare Viviani 1990; Matin et al. 1993) and their interface with visual cognition (for a review see Castelhano and Rayner forthcoming), visual processing in object and scene perception (for a review see Henderson and Ferreira 2004), and visual attention and search (see review of Findlay 2004; and Huettig and McQueen's 2007 discussion of factors influencing language-mediated visual search). For each experiment, a tight set of linking assumptions should specify the cognitive processes and representations that must intervene between the presentation of a sound and the observation of a saccadic eye movement and subsequent fixation to a visually available object. This requires careful consideration of the experimental context, including (but not limited to) the relative complexity of the visual and auditory stimuli, the nature of the task, and the linguistic and non-linguistic representations that may mediate between the critical sound onset and fixation to the visual target. For example, in an experiment to explore spoken word identification, visual targets might be line drawings, photographs, or text. The underlying representations involved in processing might lead the researcher to reject the use of text since, when seen, text could evoke a phonological representation of the word in memory before the experimental speech sound is heard, thus biasing its subsequent perception in the spoken form (McQueen and Viebahn 2007).

20.4.5 Example experiments

20.4.5.1 *Spoken word recognition*

Allopenna et al. (1998) examined the relationship of eye movements to the recognition of speech sounds in instructions such as *Pick up the beaker*. Their task used a screen array of line drawings arranged in a grid (an example from their Experiment 1 is shown in Figure 20.4.2).

These included four movable objects that changed from trial to trial (a target, e.g. *beaker*, a cohort competitor, e.g. *beetle*, a rhyme distractor, e.g. *speaker*, and an irrelevant distractor, e.g. *stroller)*, and four stationary geometric figures that were the same across trials. Drawings were present for three seconds before the experimenter spoke the instruction, (*Pick up the beaker. Now put it above the triangle.*) Participants used a mouse to "move" objects. This experiment demonstrated the usefulness of head-mounted eye tracking for the study of speech—listeners

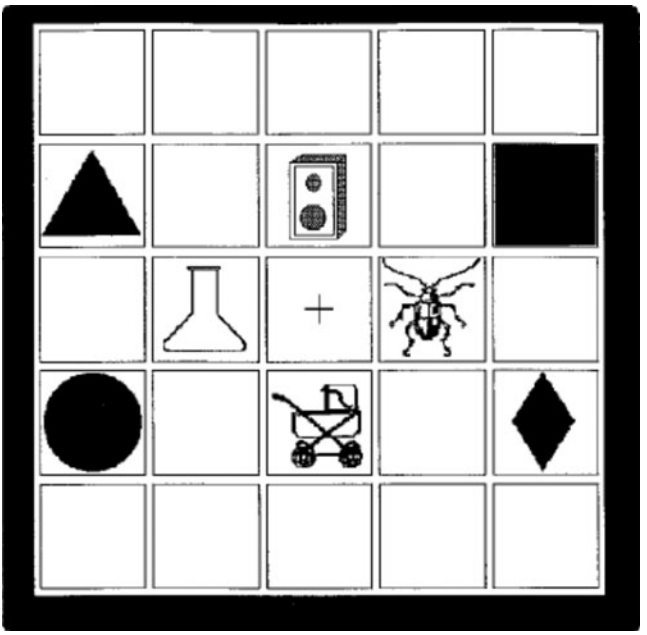

Figure 20.4.2. Example screen-based stimulus display from Experiment 1, Allopenna et al. (1998). Originally Figure 3 from Allopenna et al. (1998).

distinguished phonemes in spoken context while engaged in a language processing task that did not require interruption of the sound signal, participation in an unnatural laboratory task, or metalinguistic reflection on the identity of individual speech sounds. Results showed that eye movements are time-locked to the speech signal, and could be used to study the time course of lexical competition during spoken word recognition.

Figure 20.4.3 shows fixation probabilities over time for the four moveable objects in this experiment. Spoken target words averaged 375 ms in duration. Listeners began to fixate the target and competitor objects (*beaker* and *beetle*) more than the distractors about 200 ms after word onset, with looks leaving the competitor and settling on the target at about 400 ms, while looks to rhyme competitors (*speaker*) did not begin until about 300 ms after the onset. These results, taken together with an estimate of the time it takes to plan and execute a saccade (at least 150 ms, Matin et al. 1993), indicate that listeners were able to respond on the basis of the first 50–75 milliseconds of the word, and suggest that both word onsets and rimes are active while words are recognized.

Allopenna et al. (1998) also discuss a possible concern for interpretation of their results, the idea that "the presence of a circumscribed visual world interacts with the word recognition process" (p. 438). This is the "closed set" problem, which

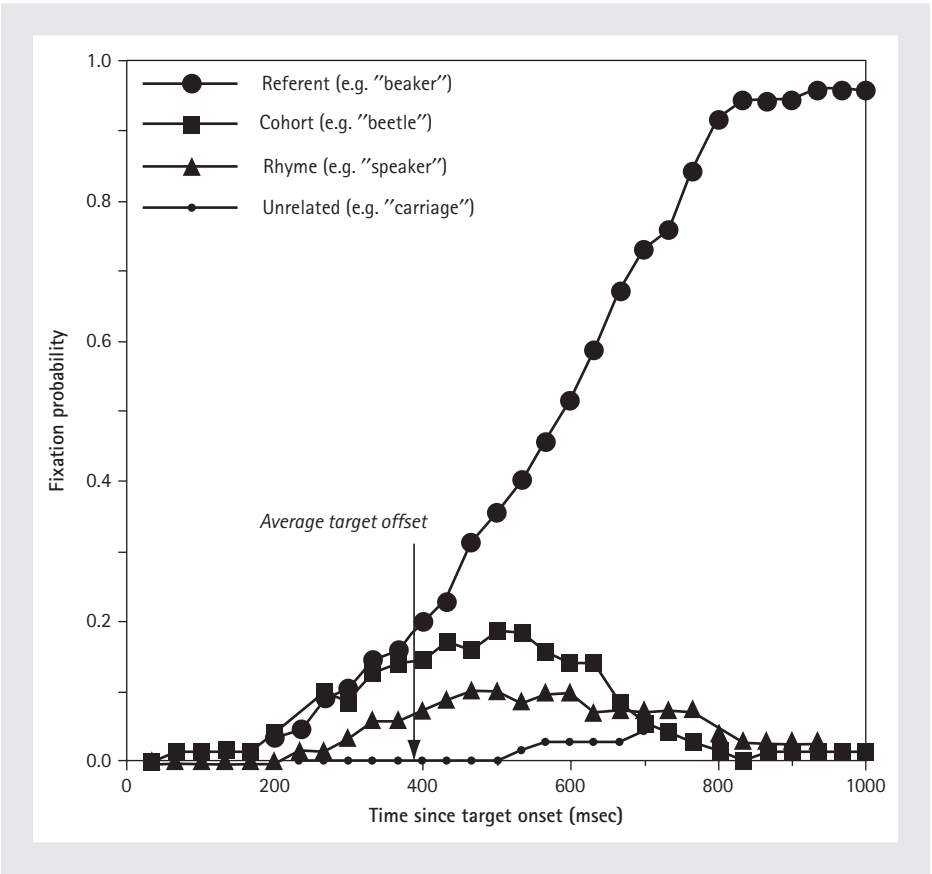

Figure 20.4.3. Fixation probabilities for four object types during spoken word recognition, from Experiment 1, Allopenna et al. (1998). Originally Figure 4. Allopenna et al. (1998).

potentially constrains the generality of the results from visual world studies. That is, to what extent does the visual presence of a small set of relevant objects "preactivate" their lexical entries? Allopenna et al. had participants name the pictured objects before beginning the experiment, and the structure of the trials required that one of the four changing pictures would be mentioned. The possibility that object names were available from both the visual and the auditory input complicates the interpretation of the rhyme effect. Did listeners look to the rhyme competitor during the second half of the spoken word because their attention was drawn to a picture consistent with the ongoing sound, or because the associated lexical entry was already active before the sound was heard (and thus without the visual input, no rhyme activation would occur)? The answers to these questions remain open.

(For an empirical approach, see Dahan, Magnuson, and Tanenhaus 2001; for recent discussion of these and related issues, see Ferreira and Tanenhaus 2007–8.)

20.4.5.2 *Speech production*

Griffin and Bock (2000) monitored speakers' eye movements as they described simple events, comparing extemporaneous description (speaking in the presence of the visual stimuli) to prepared speech (speaking after the pictures disappeared). Visual stimuli were black and white line drawings of events involving two characters (e.g. a mouse squirting a turtle with a squirt gun). Dependent measures were the location and duration of looks to the characters and actions, and the produced speech. Speakers showed two stages of looking behavior. First they extracted information about the event (during this time, eye movement patterns did not differ from those in a condition where no utterance was formulated). Then they moved their eyes to pictured event elements just under a second before speaking their names. The order of mention followed the order of fixation, regardless of the syntactic structure the speaker chose, the location of objects on the screen, or their event roles.

When speaking extemporaneously, speakers took about a second and a half to begin speaking, and looked to pictured elements of the events on average just over 900 ms before naming them. Prepared speakers began talking after about four seconds, but their eye movement patterns were comparable to those of the extemporaneous speakers. The study demonstrates that aspects of the speech production process are available from and closely time-locked to the eye movement record. In addition, it suggests that spoken sentence formulation begins with event apprehension and message generation, and then proceeds incrementally through the generated sentence components. Although these results showed no effect of the structure of the visual image or the location on the image that drew speakers' initial attention and fixation, more recent work (Gleitman et al. 2007) has shown that explicit manipulations of visual attention (via an attention capture technique involving a sudden, but not consciously detectable, onset of the image) can influence utterance form. Concepts that are the focus of a speaker's attention, and thus their initial eye movements, are more likely to be grammatically encoded as subjects.

20.4.5.3 *Sentence processing*

Altmann and colleagues examined eye movements during the use of visual context during auditory sentence comprehension (Altmann and Kamide 1999, 2007; Kamide et al. 2003; Altmann 2004; see Altmann and Kamide 2004 for a summary of results and issues). They tested whether listeners could combine information from the verb (e.g. *eat* in *The boy will eat the cake*) and a set of visually available

objects to make anticipatory looks to a target before hearing its name. They used clip art drawings to create a pseudo-scene that included the target (the only edible object, a cake), the subject (the boy), a woman, and a newspaper. Participants were not given a particular task, but were told that the sentences described the pictures. Results showed early looks to the cake—before it was pronounced, and during the processing of the word *eat*. This effect was not due to just the simple association between the words *eat* and *cake*; analogous effects were shown when listeners saw a scene with a girl, a carousel, a man, and a motorcycle. When listeners heard *The man will ride*, they made anticipatory looks to the motorcycle, but when they heard *The girl will ride*, they looked to the carousel, indicating that eye movements can reflect complex integration of information from the visual display, the subject and verb of the spoken sentence, and pragmatic knowledge of the world.

Interestingly, anticipatory eye movements are made during sentence comprehension even in the absence of a concurrent visual display. When the experiment using *The boy will eat the cake* was repeated, but altered so that the spoken sentence occurred after the picture had been shown and then removed for 2.5 seconds, participants looked to the locations on the blank screen where the named objects had been, again looking to the previous location of the *cake* while hearing *eat*. The blank screen manipulation was interpreted to indicate that eye movements reflect not only attention to objects that match what is being referred to in the auditory input, but also attention to the cognitive representation of those objects in memory.

20.4.5.4 *Intonation and discourse*

Ito and Speer (2008) investigated whether the felicitous use of English pitch accents to mark contrast during discourse comprehension could produce anticipatory looks to a target object. Their experiment used a large array of real-world objects, and a relatively complex instructed visual search task—decorating miniature holiday trees with small ornaments (Figure 20.4.1 above shows the experimental setting). They followed pre-recorded instructions, such as *Hang the green ball. Next, hang the BLUE ball.* (Upper case indicates the use of a salient L+H* accent.) Use of this somewhat familiar but manually tricky task absorbed the attention of participants, who suspected they were completing a memory test. Anticipatory looks to the most recently mentioned ornament type were found during the accented adjective for felicitous accent sequences (*green ball→BLUE ball*) but not sequences with accentual patterns that did not mark contrast intonationally (*green ball→blue ball*).

In addition, the study showed an intonational "garden-path" effect: Participants incorrectly fixated cells containing the previously mentioned ornament type when they heard infelicitous L+H* (*red angel → GREEN drum* produced initial fixations to the green angel). Such incorrect initial fixations were not observed when the

instruction had the felicitous accentual pattern (H* !H* for *red angel → green drum*). This experimental method takes a different approach from that of Altmann and colleagues. Here, the information in the speech signal is relatively simple, with a phonological manipulation in direct instructions. In contrast, the visual display is relatively complex, so that upcoming referents, though continuously present, are less predictable. Thus in this case, anticipatory looks can be linked somewhat more closely to the effect of intonation on attention, but still must be considered in the context of the stimuli and task.

20.4.6 Methodological issues

When designing the spoken language and visual displays for eye-tracking experiments, there are some basic questions to keep in mind. First, given a particular visual scene and spoken input, where will people look and why? That is, what specific linking hypotheses connect the linguistic input, attention, and the visual scene? The use of eye movements as a dependent measure for speech processing does not require that eye movements themselves directly indicate underlying cognitive states, or that a listener's attention necessarily be directed to a location throughout the time the gaze is fixated there. Most linking hypotheses employ the general assumption that the probability of looks to a visual target at a specific time is a direct function of the probability that the target is the referent of the speech, and where the likelihood of looking at the target is determined by the activation of its mental representation (e.g. lexical representation, or representation in discourse structure) relative to the activation of the other potential targets (the other objects in the current or previous display). The activation of this representation, in turn, increases the likelihood that attention will be directed to the matching target. However, it is important to note here that attention need not be explicit or conscious to affect eye movements (see Hannula and Charan 2009 for a recent demonstration).

Second, what affects the salience or availability of a visual target? Factors include the complexity of the display (the number, color, and size of objects, the presence or absence of scene-like structure and background), the spoken materials, the task and the interaction of these factors. A closely related design issue is the timing and duration of exposure—how long should the visual display be available before, during, and after the sound is heard? Many aspects of these questions remain the subject of research. For example, Huettig and McQueen (2007) explored what they referred to as the "tug of war" between phonological, semantic, and shape information over the time-course of word recognition. Participants heard critical words in neutral sentences while viewing displays containing a distractor and three types of competitor: phonological, visual shape, and semantic. When the display was available from sentence onset, fixations to phonological competitors preceded

fixations to shape and semantic competitors, but when the display appeared 200 ms before the critical word, participants looked first to shape and then to semantic competitors. Research continues to address the "closed set" problem mentioned above: Are the names of visual targets pre-activated in the lexicon due to their presence in the visual field given the task and the extent of preview? Recent work (Dahan and Gaskell 2007) suggests that abstract shape or object representations are more likely than picture names to attract looks during initial search, with linguistic input having a slightly later effect, and that this relationship interacts with preview duration (see also Dahan et al. 2007).

20.4.7 Data analysis and display

The analysis of eye movement data presents a challenge. Depending on the speed of the camera, eye position can be sampled from 33 to 250 times a second, before, during, and after the language signal, producing an embarrassment of riches in data points. Issues in analysis include how to define the dependent measure, how to align data from critical stretches of speech of different durations, and how to aggregate information from eye movement records to identify responses to the auditory input of interest.

By far the most common dependent measure in the visual world literature is fixation probability over time, displayed as a time-course graph (a summary plot of the mean likelihood of looking to a visual target at successive points in time, usually aligned from the onset of a critical word in the speech signal. See Figures 20.4.3 and 20.4.4). Fixation latency, or time from the spoken onset of the target to the initiation of the first saccade (analogous to reaction time measures), is less commonly used, perhaps because the inclusion of saccade planning inflates the resulting durations, or because it is redundant with full time-course measures. Researchers calculate proportion of looks in a variety of ways; some code fixation onset times from the time the eye stabilizes its position on the target, and others from the onset of the preceding saccade, as planning the ballistic saccade movement implies knowledge of the target location and content (see discussion in Altmann and Kamide 2004).

The numerator and denominator of the plotted proportion also vary across the literature; some groups use a denominator that is the total possible looks at a given point in time (e.g. the number of trials per condition × the number of participants), while others use the number of looks to coded regions in the display (i.e. total possible looks minus data lost to blinking, equipment failure, or looks away from the display). Note that the latter method has the disadvantage of changing reliability of the probability from timepoint to timepoint and between included participants, with increasing unreliability as the data become more sparse.

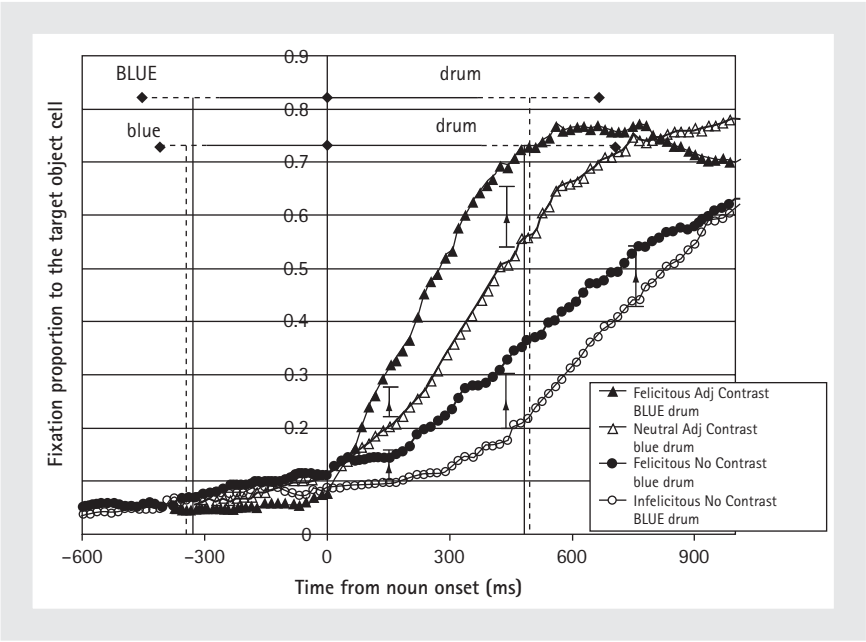

Figure 20.4.4. Example data display from Experiment 2, Speer and Ito (2008). Fixation proportions are aligned at noun onset. Spoken word durations for adjectives and nouns with two different pitch accent patterns are displayed: vertical lines indicate mean durations, horizontal lines show maximum (dotted) and minimum (solid). Floating bars indicate confidence intervals for time regions where mean differences reached statistical significance. Originally Figure 5 from Ito and Speer (2008).

Another concern has been finding an optimal way to align and plot fixation proportion data over time, averaged over multiple trials where the duration of words in different items differs (e.g. compare spoken nouns *kit* and *strengthener*). A single alignment point at the beginning of a spoken word would seem to misrepresent processing that occurs after the first syllable. Different alignment strategies may be appropriate for different research questions. For example, Ito and Speer (2008) aligned the data at the onset of the noun for their adjective-noun pairs (e.g. *BLUE drum* and *blue drum*), because this word singled out the referent to be mapped onto the real-world object. An example display from their Experiment 2 is shown in Figure 20.4.4.

Note that this creates a "backward alignment" of the eye data observed during the adjective, highlighting anticipatory effects of the pitch accent that occurred toward the end of adjective in felicitous trials.

In contrast, researchers interested in the effects of successive words in a sentence have examined data aligned at multiple onsets based on the average duration of

words across items (e.g. Snedecker and Trueswell 2003). Altmann and Kamide (2004) argue that eye data from sentence-processing studies should be collapsed over the duration of each constituent of interest (e.g. subject, verb), producing a single point that shows how often participants fixated the target during that constituent (see their appendix II for a comparison of the merits of four types of analysis and visual display). Although this method is appropriate when the research question concerns the effect of a word or phrase, it is not effective for research that queries the point *within* a critical word where an effect begins. A parallel concern is "window size," the amount of time to include when comparing conditions for statistical analysis, discussed more below.

20.4.8 Statistical analysis

Although simple inspection of fixation proportion functions can convey a great deal about an experiment's results, researchers wish to specify when measured differences reach statistical significance. The bulk of published eye movement studies in psycholinguistics before 2008 use a series of repeated measures ANOVAs to compare mean fixation proportions between conditions, calculating two parallel analyses, one with participants and the other with items as the random variable. A "window size" is set, and means are calculated over item or participant scores within successive windows.

As with data alignment, the appropriate window size and the number of windows analyzed must depend on the research question. For example, sentence-level analyses have been conducted for an initial series of relatively large windows (200–500 ms), with regions of particular interest reanalyzed using a smaller (100 ms) window (see e.g. Snedeker and Trueswell 2003). Specifics of the chosen window size, alignment point and underlying assumptions about the time it takes to plan and initiate a saccade can interact when results are interpreted—sometimes producing a quandary. For instance, suppose the data are aligned at the onset of the spoken word and analyzed in successive 300 ms windows, and we assume saccade planning takes 200 ms. If fixation proportion functions in the 0-300 ms window begin to diverge at 200 ms, averaging over the window may obscure early effects. On the other hand, the cautious researcher would want to avoid determining window size based on a post-hoc observation of the area most likely to produce a statistical effect.

At the time of this review, statistical methods in the social sciences are undergoing a shift toward the more sophisticated and powerful techniques available with the use of multilevel mixed-effects linear modeling (see e.g. Baayen 2008, this volume). The analysis approach described above suffers from over-reliance on ANOVA, and inherent "oversampling" of the data—while individual eye movements are discrete events, the possibility of their occurrence is evaluated with an arbitrary frequency

in time that corresponds to the shutter rate on the system camera, so that many "observations" are made of a single event. Consequences include (1) the inflation of alpha error due to conducting multiple analyses on successive (and even the same) time windows, where data cannot be assumed to be independent; (2) time, an inherently continuous variable, is artificially segmented into categorical "bins" to create analysis regions of manageable statistical power (given the wealth of observations); and (3) fixation proportions are treated as a continuous dependent variable, even though they are generated by an experimental event that is in essence categorical—participants are either looking at the target or not. Such data are binomially (rather than normally) distributed and more amenable to logistic regression (for discussion see Barr 2008). One of the most promising options for new statistical approaches includes the use of mixed-effects models for continuous dependent variables and random effects; this is most appropriate for traditional psycholinguistics designs, and allows the researcher to examine simultaneously crossed random effects for items and subjects (Baayen et al. 2008). Particularly promising for eye movement data is the use of multilevel logistic regression, with time properly analyzed as a continuous variable, and gaze location as categorical (Barr 2008); another approach is growth curve analysis, which models the data using additional parameters such as inflection points and rates of change (Mirman et al. 2008).

20.4.9 Summary

Eye tracking in the visual world paradigm is particularly useful for the study of spoken language processing. The technique is adaptable and appropriate for a broad range of theoretical and empirical questions, but is most apt for research questions that require observation of the changing state of linguistic representations over time (as during comprehension or production in linguistic and visual contexts), and for those that require an implicit measure, uncontaminated by metalinguistic judgments or conscious response strategies. Current eye movement monitoring systems, portable and with lightweight or no headgear, allow ever-increasing flexibility in experimental design, so that we might expect to see future use in the field, and with children and disabled populations. The discussion here emphasized the linking assumptions that connect eye movements to underlying language processing, and therefore govern the interpretation of results. Similarly important are the choice of dependent variable(s), assumptions about timing and synchronization between the auditory signal and the eyes, and decisions about data display and statistical analysis. The continuing development of these aspects of eye-tracking methodology should increase its usefulness for the precise and rigorous modeling of an increasing range of theoretical and empirical questions in laboratory phonology.

20.5 Neurophysiological techniques in laboratory phonology

William Idsardi and David Poeppel

20.5.1 Neurophysiological and neuroimaging techniques: general advantages and disadvantages

Until the late 1980s, the principal evidence regarding neuronal implementation of speech derived from patient studies (neuropsychological deficit-lesion correlations). These studies are theoretically well motivated and experimentally nuanced, but the neuronal information is necessarily restricted to spatial information (lesion location), and even those data remain rather coarse, with no meaningful temporal information. The aggressive development of non-invasive recording using both electrophysiological (EEG, MEG) and hemodynamic (fMRI, PET, NIRS) techniques has made available a wealth of new data, at spatial and temporal resolutions that are permitting researchers to develop increasingly sophisticated hypotheses about the neuronal basis of perception and cognition. Notwithstanding the excitement surrounding the widespread availability of these new empirical approaches, robust skepticism is indicated. It is, indeed, not always obvious precisely how such data can enrich our understanding of speech and language processing. There is great promise, but the hurdles are significant.

Phillips (2001) provides an excellent review of the many issues in applying neurophysiological techniques to the investigation of speech perception. As outlined there, researchers face an enormous number of issues in unpacking the mapping from the input (a continuous acoustic waveform) to the ultimate language percepts (features, sounds, words, sentences, etc.). For each available experimental technology, we face significant challenges at every level of analysis. One major advantage of neurophysiological techniques is that they offer measures related to, but separately collected from, the usual behavioral measures such as response accuracy and response reaction times, and thus provide quasi-independent verification of those measures. In addition, for some of the measures (EEG, MEG) the observable brain responses are hundreds of milliseconds earlier than the behavioral responses (which typically occur later than 500 ms). The advantage gained by observing earlier responses is that we can begin to potentially dissect the complex computations involved in tasks like word recognition. A response at 100 ms is unlikely to be directly modulated by long-term memory representations (for example the lexicon) and thus provides evidence for the nature of phonetic and phonological processing that occurs in advance of lexical access.

The first portion of this section briefly reviews some of the available technologies. The second part reviews some studies linking brain measures with properties of interest to laboratory phonology researchers. We focus here on the use of these techniques for studying speech perception, although it is also possible to study some neural aspects of speech production with some of these techniques. We conclude with some pragmatic recommendations based on the currently available technologies.

20.5.1.1 *Available technologies*

The available non-invasive recording techniques of brain functions and their anatomical correlates (Nolte 2009) trade off spatial resolution, temporal resolution, directness of the measure of brain activity, invasiveness, convenience, and cost. Here we consider only gross characteristics of passive recording techniques; those based on blood flow (fMRI, PET, NIRS) are inherently slower than those based on the electromagnetic fields generated by neural activity (EEG, MEG). Simultaneous detailed resolution in time and space is possible only with intra-cranial recording, included here for comparison as this technique is restricted to surgical populations (see Besle et al. 2008 and Boatman 2004 for examples of its use in investigating speech perception). The trade-off is then fairly direct—time versus space. The electromagnetic measures (EEG, MEG) have millisecond-level temporal resolution, but resolve brain areas relatively poorly (though well enough to distinguish visual cortex from auditory cortex, say, or to detect hemispheric differences). The blood-flow techniques can resolve well spatially within specific brain regions, but at the expense of temporal resolution. Given the rapidity of many phonetic changes, resolution at the level of 1–10 seconds will not impress laboratory phonology researchers accustomed to 44.1 kHz sampling rates for audio.

There are also active techniques such as surgical or pharmacological interventions which involve momentarily or permanently changing brain function; the only one considered here is transcranial magnetic stimulation (TMS, see 20.5.1.7). We also disregard lesion studies and post-surgical populations. Pictures of the various machines and representative data plots can be found on Wikipedia (e.g. http://en.wikipedia.org/wiki/Electroencephalography).

In summary, for laboratory phonology researchers, the space-time trade-off is the most relevant one, as summarized and simplified for the major technologies in the following table.

20.5.1.2 *Electroencephalography (EEG)*

Electroencephalography (EEG) directly measures the electrical mass activity, using from one to hundreds of sensors attached to the scalp, generated by the coordinated activity of large groups of neurons in the brain (10^5–10^6 neurons). The source

Table 20.5.1. Trade-offs between time and space resolution in brain-imaging technologies

Space		Time			
		0.001 s	0.01 s	0.1 s	1 s
1 mm^3	1 uL	(Intra-cranial)			fMRI
1 cm^3	1 mL	MEG		fMRI (low res)	NIRS, PET
1 dm^3	1 L	EEG			

of the signal is hypothesized to be the post-synaptic current flow in the apical dendrites of pyramidal cells in cortex (not action potentials). EEG has a number of significant advantages. Having been in use for more than a century (Swartz 1998), the equipment is now tried and true, and in this group of technologies, relatively inexpensive. Portable EEG systems are available, and this technique does not require extensive magnetic shielding to prevent interference from nearby electrical equipment. The sensors must be placed on the scalp; to better increase the conductance to the sensors, they are applied along with special gel or saline solutions (which do make the process somewhat messy). Although there is considerable debate on the ideal number and types of sensors, in practice researchers often report single channel data (when focusing on temporal aspects of the recorded signal), or data averaged over groups of sensors, rendering moot many of the theoretical issues (especially those involving localization). It is possible to combine EEG with fMRI (see 20.5.1.4) to improve the spatial localization (Bagshaw et al. 2006) either with simultaneous recording or by using separately collected fMRI scans to constrain the EEG localization solutions.

There are a number of ways to use EEG to get reliable brain measures. One of the most common is to measure the brain response arising from particular events controlled by the experimenter—event-related potentials (ERPs, Luck 2005), a particularly common technique in many psycholinguistic experiments. Because of the large amount of uncorrelated electrical noise in individual trials, the responses to a series of replicated trials are averaged together to increase the signal to noise ratio. Other measures include steady-state responses such as the auditory steady-state response (aSSR, Burkard 2009) which will entrain to certain amplitude or frequency-modulated frequencies in auditory stimuli and other changes in the endogenous brain rhythms (Buzsaki 2006) and which thereby show frequency matching with the external stimuli, such as following the pitch of an utterance (Patel and Balaban 2001, 2004). A particularly useful technique is mismatch negativity (MMN, Pulvermüller and Shtyrov 2006; Näätänen et al. 2007), an automatic response when an "oddball" is detected within a series of "standards." The flexibility of MMN designs makes it a very useful technique, although the necessarily high ratio of

standards to deviants (7:1 or more) makes this technique less efficient than ERP-style evoked responses.

20.5.1.3 *Magnetoencephalography (MEG)*

Magnetoencephalography (MEG, Hämäläinen et al. 1993) measures the magnetic field generated by the electrical activity of the brain from just outside the scalp, also using from one to hundreds of sensors. The equipment is much more expensive than EEG, both in the initial cost and in the ongoing supply costs. The measures obtained in MEG are largely analogous to those obtained with EEG: they can be event-related, steady-state, related to endogenous rhythms or mismatch fields. At the detailed level of the generation of electrical signals in the brain there are differences between the techniques, but for many high-level cognition experiments these differences are negligible. The main advantage of MEG is the relative simplicity of the source localization algorithms (due to the relative magnetic transparency of the intervening materials), which allows all of the temporal resolution of EEG plus reasonable spatial resolution ($\sim$ 1 cm^3), especially when combined with structural magnetic resonance scans. However, because the magnetic fields are tiny, the equipment requires extensive magnetic shielding and, as a practical matter for experiment design, a larger number of replicated trials to achieve reasonable signal to noise ratios. How MEG is used in studies of audition and speech perception is reviewed in detail by Lütkenhöner and Poeppel (2011).

20.5.1.4 *Functional Magnetic Resonanace Imaging (fMRI)*

Magnetic resonance imaging (MRI) is functionally like an X-ray machine; it detects differences in the resonance given off in response to extremely strong magnetic fields ($>$ 1 Tesla, about 100,000 times the strength of the Earth's magnetic field) applied to body tissues. While this technology is useful for viewing structural details and diagnosing tumors, we are more interested in the brain in action, and can use fMRI to study the difference between oxygenated (arterial) and de-oxygenated (venal) blood. Oversimplifying wildly, when a brain area is working hard (though cf. Sirotin and Das 2009) it requires more oxygenated blood, and consequently we can measure the blood-oxygenation-level-dependent (BOLD) resonance properties as they change over time. Unfortunately, changes in blood flow are relatively slow (the *hemodynamic response*, and its *lag*, peaking roughly 4–10 seconds after an "event" of interest), and are only indirectly related to the electrical activity of the neurons. Thus, as described in the table, fMRI offers excellent spatial resolution on the order of 1 mm^3 but poor temporal resolution ($\sim$ 1s). Clever experimental designs can overcome some of the temporal limitations of the technology. The equipment is expensive, and requires specialized training to operate safely. However, there are many more fMRI installations than MEG centers due to the extensive

clinical applications of magnetic resonance technologies. New analysis techniques, such as diffusion tensor imaging (DTI) allow for anatomical pathways of functional neural connections to be inferred from MRI data. There are as yet few such studies applied to language, but one notable one (Saur et al. 2008) offers support for separate dorsal and ventral pathways in language perception (Hickok and Poeppel 2004, 2007).

However, laboratory phonology researchers are not only interested in measuring the brain. MRI also offers significant potential as an imaging technique for the vocal tract. If the scanner is held in a mid-sagittal orientation, then low-resolution images can be collected at between 10 and 20 frames per second (i.e. 50–100 ms per frame). This speed is sufficient to study some aspects of speech production, such as velopharyngeal movements, tongue position, lip opening, and larynx height at spatial resolutions equal to or better than ultrasound techniques (Davidson, this chapter). As yet, MRI remains an unusual technology for studying speech production, and it is not yet clear what the overall advantages and disadvantages are of MRI for studying speech production. MRI offers two obvious advantages over the classic X-ray studies: (1) it is safer, and (2) it is more available than X-ray microbeam facilities. However MRI machines have a major disadvantage in that during the data collection phase (where the polarity of the magnet is rapidly reversed) the noise of the machine is quite loud (typically $>$ 80 dB SPL) which limits the ability to collect clean spontaneous speech samples during MRI scanning. MRI speech production studies are relatively scarce. No relevant review article exists, but the reader is referred to Park and Iverson 2009; Martins et al. 2008; Vampola et al. 2008; Mády and Beer 2007; Serrurier and Badin 2008; Story 2008; Clément et al. 2007; Mohammad et al. 2006; Kim et al. 2005; Kim 2004; Narayanan et al. 2004; Engwall 2003; Magen et al. 2003; Ettema et al. 2002; Nissenbaum et al. 2002; Stone et al. 2001; Fitch and Giedd 1999; Moore 1992; Baer et al. 1991.

20.5.1.5 *Positron Emission Tomography (PET)*

Another spatial resolution brain-imaging technology is positron emission tomography (PET). Subjects are given a dose of a radioactive compound which migrates to the brain tissue. As the radioactive substance decays, it releases positrons which when they encounter electrons annihilate each other, and in the process release two photons which travel in opposite directions. The photons are detected outside the brain, and the position of the collision is inferred. For our purposes, there is no significant difference between the information that can be obtained using PET and fMRI (Feng et al. 2004) though there are substantial differences when studying neurochemistry and neuropharmacology. As emphasized by a reviewer, PET does continue to offer certain advantages over fMRI for speech studies. First, the PET scanning process is quiet, and thus offers a clear environment for speech production and perception by the subjects. Secondly, PET offers better imaging of certain brain

areas (for example the anterior temporal lobe), with less distortion and less incidence of motion artifacts. However, as fMRI technology is improving *rapidly*, most of the imaging advantages are disappearing, leaving only the quiet environment as the major advantage of PET scanning, while there is a significant disadvantage: the exposure of subjects to a radioactive tracer. For older but still relevant reviews of speech studies using PET see Poeppel (1996) and, in reply, Démonet et al. (2002).

20.5.1.6 *Repetitive Transcranial Magnetic Stimulation (rTMS)*

In addition to recording electrical, magnetic, or hemodynamic correlates of brain activity, it is also possible to disrupt normal brain activity with electrical stimulation (as in electroconvulsive therapy), with moderate cooling of brain tissue (Malhotra and Lomber 2007), and, more usefully, with transcranial magnetic stimulation (TMS, Pascual-Leone et al. 2002). While such techniques were recently approved by the FDA for treating depression, their long-term effects are currently unknown, making this a relatively unlikely technique for laboratory phonology researchers. It is the case that the susceptibility of the motor areas to this stimulation would allow for testing motor theories of speech perception (Liberman and Mattingly 1985) and mirror neuron conjectures (Arbib 2006), and a few studies supporting motor involvement in speech perception have been published (Iacoboni 2008; Roy et al. 2008), but Lotto et al. (2009) detail the limitations of such studies in addressing the exact nature of the connection between perceptual and motor areas.

20.5.1.7 *Summary*

For most speech perception questions, better time resolution is more important than better spatial resolution. MEG, especially when combined with structural MRI scans, offers excellent temporal resolution while maintaining a good compromise for spatial resolution. However, EEG is a viable alternative in many circumstances, and has much lower set-up and maintenance costs. Naturally, the choice is ultimately dictated by the question at hand, so if the hypotheses require an answer in terms of anatomic information, fMRI is the most available and most appropriate technique; if the research centers on processing models or any issue requiring a temporal answer, MEG and EEG are optimal. It is worth noting that for acquisition studies, even with infants, Near Infra-Red Spectroscopy (NIRS or optical tomography, the little cousin of fMRI) and EEG can be used effectively. These techniques are less susceptible to movement artifacts, are silent, and generate the types of data that permit evaluation of hypotheses regarding the processing of speech information in learners. For a recent example testing newborns, see Telkemeyer et al. (2009).

20.5.2 A brief survey of EEG and MEG findings relevant for laboratory phonology

We now turn to some particular experimental paradigms using EEG and MEG. We concentrate on these two techniques because of their relative availability and suitability for testing questions involving the time course of speech perception and phonological processing. The majority of studies examine either the first prominent evoked responses (N100/M100) or mismatch (oddball detection) responses (MMN/MMF). Phillips (2001) reviews the same material and provides much useful phonological and psycholinguistic context for the general nature of these studies; consequently this section primarily updates that review with more recent publications.

20.5.2.1 *N1/N100/N1m/M100*

Any auditory stimulus with a well-defined onset will elicit a characteristic pattern of brain responses. Among these responses—including the P50, N1, P2, and others, all occurring in a characteristic cascade—is a relatively prominent and clear response peaking around 100 ms after the stimulus onset, located in auditory cortex in the superior temporal lobe. This response has various designations in the literature depending on the technique employed. In EEG, the deflection has negative polarity, and is named with an N-prefix, either N1 or N100. In MEG common names include N1m (for N1 magnetic) and M100. Obviously, very little of the auditory signal can ascend the auditory pathway in time to produce a clear cortical signal 100 ms after its onset; estimates (Gage and Roberts 2000) are that approximately the first 20–50 ms of the signal conditions the brain response. This makes this response very useful for assessing the information available and used at the beginning of a signal, and also aligns well with useful acoustic correlates of phonetic properties that fall within the first 50 ms (e.g. vowel formants, burst spectra, VOT). Non-linguistically, there are M100 latency differences in (sinusoidal) tone perception such that the shortest latency for the M100 is found near 1000 Hz (Roberts and Poeppel 1996); more recently Monahan et al. (2008) show the same pattern of responses for pitch inferred from higher harmonics. Here we will conflate the EEG and MEG results and the various names for this early response.

20.5.2.1.1 *Vowels*

Given the MEG findings for tones below 1000 Hz—low-frequency tones of 100–300 Hz are associated with M100 latencies up to 30 ms longer than higher-frequency tones of 500–3000 Hz—we have a reasonable expectation of tracking properties of the first formant (F1). We should expect longer M100 latencies for high vowels (with F1 distant from 1000 Hz) and the shortest latencies for low vowels (with F1 closest to 1000 Hz). And indeed this is exactly what was found for English

listeners hearing synthesized tokens of /u/ and /a/ (Poeppel et al. 1997). They failed, however, to find any consistent influence of the vowel pitch on the M100 response, perhaps due to the relative complexity of the vowel stimuli as compared to the simpler tone stimuli. Roberts et al. (2004) more closely examined a continuum of synthesized back-vowel tokens ranging from English /u/ to /a/. Rather than finding a smooth 1/f curve (as in the case of matched sinusoidal stimuli) they instead found a staircase effect, which they interpret as evidence of categorical perception of the vowel stimuli. Monahan and Idsardi (2010) demonstrate that the M100 response is not a simple response to only the first formant, but integrates information from F3 (but not from F2). By manipulating F3 while holding F1 and F2 constant they were able to modulate the M100 latency in accord with predictions based on an F1/F3 ratio derived from the previously reported results. This suggests an early response to a derived vowel measure, one at least partially normalized for speaker.

Responses to simple tones have been investigated across a wide range of frequencies, and we see increasing latencies in these responses to tones above 1000 Hz (Roberts and Poeppel 1996). Consequently, we should also expect to be able to find some aspect of the M100 response which tracks F2. However, thus far no such finding has been reported. One possible explanation is that because the amplitude of F1 is substantially greater than that for F2 in most cases, the effect of F2 on the M100 is concomitantly weaker and more difficult to detect, given the limitations of the recording techniques. Nevertheless, the lack of a response latency correlated with F2 dramatically limits the present usefulness of the M100 latency in mapping vowel space perception.

Animal studies (using various methods) have often revealed tonotopic organization within auditory cortex (grey squirrel, Merzenich et al. 1976; ferret, Kelly et al. 1986), and this has been extended to the human auditory cortex as well (Romani et al. 1982; Pantev et al. 1989). Although the resolution of MEG is not sufficient to resolve place-coding of the granularity revealed by the single-unit studies in animals, it is possible that populations of neurons will have different "centers" for different formant frequencies. Obleser et al. (2004), using MEG, and building on earlier related work by Diesch et al. (1996), calculated the Equivalent Current Dipole (ECD) of the source for several distinct German vowels. They found that front vowels tend to map onto a more anterior portion of auditory cortex while back vowels map onto a more posterior region of auditory cortex. Thus, the front/back distinction of vowel categories (correlated with F2) is retained on the anterior/posterior dimension of the auditory cortex. Results for vowel height were not as clear, though the Euclidean distance between the dipole locations for high and low vowels was greater than that for high and mid vowels. The allure of a cortical vowel map is plainly powerful (Poeppel 2008; Scharinger et al. 2011) but as yet premature.

20.5.2.1.2 *Consonants*

Turning to consonants, we might expect to find spectral differences signaling place of articulation differences in consonants reflected in various aspects of the M100 response. Obleser et al. (2003) report differences for dorsal and coronal stops parallel to those reviewed above for back and front vowels. The dorsal consonants were localized more posteriorally in the auditory cortex. Gage et al. (2002) report M100 latency differences for /ba/ (longest), /da/, and /ga/ (shortest); however this was significant only in the right hemisphere.

Differences in the M100 amplitude and latency have been reported for short-lag versus long-lag VOT differences (Phillips et al. 1995; Simosa et al. 1998; Sharma and Dorman 1999). These studies were designed so that the materials crossed a phonemic boundary for the listeners, and clear VOT differences were visible only with "double on" responses. The "double on" responses exhibit essentially two separate M100 responses, one to the burst and one to the vowel onset, and occur only with reasonably long VOTs (>40 ms). This provided support for a "refractory period" explanation of category boundary near this VOT lag—two abrupt events less than 40 ms apart were treated as a single event and those further than 40 ms apart would show direct tracking of the VOT lag in the second response. However, more recently, Frye et al. (2007) report a decrease in amplitude and increase in latency as VOT increases, even for the "single on" responses, suggesting that there are auditory cortex mechanisms to track a wide range of VOT values. Thus, so far, the M100 seems to track relatively low-level aspects of VOT. While it would be desirable to see if the response can be influenced by factors such as speech rate, the necessity of presenting simple materials time-locked to the onset of stimulus presentation makes speech-rate manipulations logistically difficult.

20.5.2.2 *Mismatch responses (MMN/MMNm/MMF)*

Mismatch designs rely on a "surprise" (violation of expectation) response to an uncommon "oddball" or "deviant" in an ongoing series of common "standards." This is a very general and useful experimental paradigm, applicable to a wide range of stimulus types and conditions (Näätänen et al. 2007). The mismatch response (abbreviated in various ways including MMN, MMNm, and MMF) is robust and clear in both EEG and MEG (which we will again conflate here) and remarkably unaffected by listeners' states of attention or even wakefulness. Because of the flexibility of the contextual definitions of "standard" and "deviant" it is possible to construct sophisticated tests of classes of speech sounds with this technique. The method also has clear affinities to classic habituation techniques in infant speech perception (Eimas et al. 1971; Maye, this chapter). This allows for the exploration of categorical perception effects, the dissection of natural classes, and the degree of abstraction in phonological representations. Thus, for laboratory phonology

researchers, the mismatch paradigm is probably the most generally useful of the neurophysiological techniques.

20.5.2.2.1 *Inventories*

Since vowels can reasonably be presented as unitary speech events, it is easy to construct mismatch experiments for vowels. In a series of studies (Näätänen et al. 1997; Winkler et al. 1999), listeners have consistently shown larger mismatch responses to standard-deviant contrasts that map onto vowel prototypes in their native language. Peltola et al. (2005) found no effect of second-language instruction on the mismatch responses—advanced Finnish learners of English behaved like monolingual Finns.

A more sophisticated design is employed in Phillips et al. (2000). That study examined the status of English [ta] and [da] while introducing substantial acoustic variations in the tokens (see Figure 20.5.1). If the tokens were encoded categorically, then the task was a standard mismatch design. If, however, each token was treated as a separate exemplar, then no particular standard dominated the sequence, and there should be no mismatch response. In fact, a clear mismatch response was observed, consistent with a *categorical encoding of the stimuli*. Furthermore, no mismatch response was observed when the VOTs were all shifted upwards so that all tokens were in the [ta] range, indicating the subjects' inability to distinguish standards and deviants in that condition.

A criticism of these studies is that they confound phonetic and phonological inventories—the non-native contrasts are not allophonically present to any substantial extent. Kazanina et al. (2006) address this deficiency by testing a contrast [da]/[ta] between Russian listeners (for whom it is phonemically contrastive) and Korean listeners (for whom [d] is the intervocalic allophone of /t/). Russian speakers showed a clear mismatch response whereas the Koreans did not, a result consistent with the Koreans' failure to register a difference between "standards" and "deviants" (though they must control this difference in production in order to produce the contextually appropriate sounds). This showed that the mismatch negativity response is sensitive to abstract organizations of speech sounds into phonemic categories.

Eulitz and Lahiri (2004) discovered an asymmetrical mismatch effect. They found that for German vowels coronal (= front) deviants presented amongst dorsal (= back) standards produced larger mismatch responses than did dorsal deviants presented amongst coronal standards. This was a notable finding as most mismatch responses are symmetrical when the condition is reversed. They attribute this difference to the different status of coronal in the FUL model of speech perception (Lahiri and Reetz 2002; Lahiri, this volume) in which [coronal] is represented in the auditory input, but not in memory representations, leading to a three-valued matching procedure (match, mismatch, no mismatch). Their

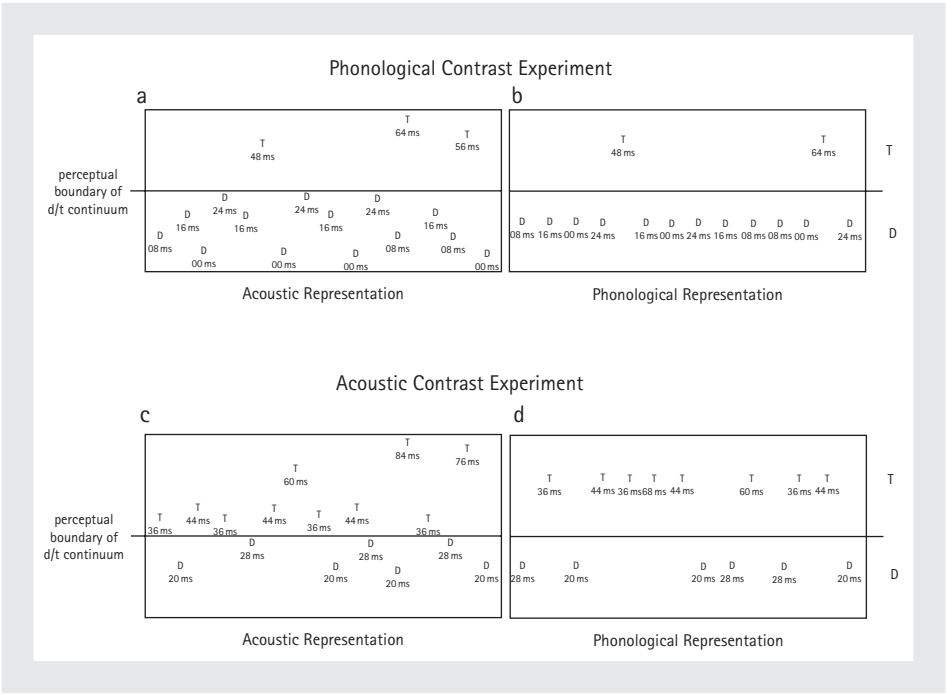

Figure 20.5.1. Design of phonological mismatch experiment, illustrating acoustic and phonological representation of sequence of stimuli: (a,b) phonological contrast experiment; (c,d) acoustic contrast experiment. (Reprinted with permission from Phillips, Pellathy, Marantz, Yellin, Wexler, Poeppel, McGinnis, and Roberts 2000.)

explanation of the results is that coronal standards induce a "blank" standard, which dorsal deviants neither match nor mismatch, producing a reduced response relative to the full mismatch of coronal deviants against the (working) memory representation of the dorsal standard. Ikeda et al. (2002) report a similar asymmetry, with native-language prototype standards producing a larger mismatch with non-prototypical deviants than non-prototypical standards with prototypical deviants.

Dehaene-Lambertz et al. (2000), based on a behavioral study by Dupoux et al. (1999), were able to use a variant of the mismatch design to test whether, in Japanese speakers versus French speakers, native-language syllable structure conditions vowel perception. They showed that the language-typical vowel epenthesis for Japanese speakers has a distinct neural correlate in their mismatch design, providing further evidence that such metrics can be productively employed to probe inventories.

20.5.2.2.2 *Sequences*

Investigations of phonetic and phonological inventories reveal certain aspects of natural language sound systems, but they address neither the dynamic character of speech, nor the contextual realizations of speech sounds due to coarticulation effects and phonotactic restrictions. However, to date relatively few neurophysiological studies have examined speech sound sequences. Flagg et al. (2006) examined the interaction of nasality between vowels and following consonants in English. English vowels are typically nasalized before nasals, yielding the simple (if idealized) pattern of licit [ab, ãm] and illicit *[ãb, am] sequences. They report mismatch-like effects in the early auditory responses to the consonant—the responses are faster for licit than for illicit sequences (though the effect is asymmetric: [ab] is facilitated, but *[ãb], *[am], and [ãm] are all statistically equivalent). Hwang et al. (2010) extend such findings to English final voicing sequences, with voiced stops facilitating the processing of a subsequent voiced fricative, but interfering with the processing of a following voiceless one: [dz] < [ts], [tz] < [ds]. Monahan et al. (2009) also report similar results for velar stops and following vowels: unfronted velar stops facilitate the processing of following back vowels.

20.5.2.3 *Summary*

Evoked responses and mismatch designs in both EEG and MEG have proven useful for the exploration of the neural processing of speech sounds. M100 latency varies according to the spectral properties of the incoming signal, revealing integration of information across different spectral regions (F1 and F3) and modulation by abstract category structure. The localization of the M100 response within auditory cortex may also reveal a cortical vowel map. Mismatch responses have been used to investigate phonetic and phonological inventories and phonotactic constraints on sound sequences. To date, very little work has been done on under-studied languages; there is enormous opportunity in this area for the investigation of typologically unusual patterns.

20.5.3 Pragmatic recommendations

In our opinion, MEG represents the best current compromise of spatial and temporal resolutions, non-invasiveness, and data acquisition in quiet conditions for work with adult subjects. When combined with structural MRI scans MEG is capable of reasonable spatial resolution, certainly enough to distinguish areas associated with the motor control of different articulators (for example the larynx, tongue, and lips) or to confirm that responses are indeed in auditory cortex, and perhaps to begin to investigate possible cortical maps for speech sounds. One practical consideration is the efficiency of data collection, given the necessity of averaging

brain responses over a fairly large number of repetitions of stimulus presentations. For this purpose, evoked responses such as the M100 are more efficient, but are only known for a limited number of attributes (some spectral information and for VOT). Mismatch studies are, in contrast, very flexible, allowing various manipulations of category structures, but are open to some nagging questions of interpretation (what is a "standard" representation exactly?) and are relatively inefficient in terms of data-points obtained per stimulus presentation. However, it is still true that EEG equipment is less costly and more available than MEG machines, and we can expect to continue to see EEG and MEG research on evoked responses relevant to speech for the foreseeable future. In the next decade or so we should expect to see an increasing number of such studies, and we also expect to see a shift in focus to investigation of the online processing of the unfolding sequence of sounds in speech, including a proliferation of studies of phonotactic constraints in various languages.

CHAPTER 21

EXPERIMENTAL DESIGN AND DATA COLLECTION

SOCIALLY STRATIFIED SAMPLING IN LABORATORY-BASED PHONOLOGICAL EXPERIMENTATION

JAMES M. SCOBBIE AND
JANE STUART-SMITH

METHODS FOR STUDYING SPONTANEOUS SPEECH

NATASHA WARNER

METHODS AND EXPERIMENTAL DESIGN FOR STUDYING SOCIOPHONETIC VARIATION

PAUL WARREN AND
JENNIFER HAY

The contributions in the chapter cover major considerations in data collection and analysis especially as they relate to sociophonetics. Scobbie and Stuart-Smith address the use of sociolinguistically defined populations in experimental cross-dialectal research. Warner discusses how to reliably investigate spontaneous speech in a controlled experimental setting. Warren and Hay highlight the importance of studying production and perception in tandem and the need for statistical modeling of such data in the investigation of sociolinguistic variation.

21.1 Socially stratified sampling in laboratory-based phonological experimentation*

James M. Scobbie and Jane Stuart-Smith

21.1.1 Introduction

There has been a division in phonology for several decades between sociolinguistic research, with a theoretical focus on variation and change at the level of the community, and generative research, which has taken a purportedly cognitive perspective. These different theoretical interests are not logically incompatible, of course, and hope springs eternal for positive interaction, since individual speakers encode aspects of variation as part of their own phonological competence, and open-ended and overlapping socially embedded patterns emerge from a collection of individual grammars. Since, at the most fundamental level, it is difficult to disagree with Labov's observation that there could never be "a successful linguistic theory or practice which is not social" (Labov 1972a: xiii), the challenge is rather how to go about constructing such a theory, and on what grounds.

The different theoretical perspectives come complete with their own methodological practices and insights, and professional specialization by linguists tends to hamper crossover and the advancement of understanding in areas of joint interest. Even sociolinguists and laboratory phonologists display a reticence to adopt *any* aspects of methodology from the other camp by those interested in what can be called core theoretical questions (rather than interface issues) due to a misplaced belief that theory and methodology are mutually indissoluble. Yet at heart,

* Scobbie thanks ESRC (Fellowship R000271195) and Stuart-Smith the Leverhulme Trust for support for the reported work on /ai/. They thank Eleanor Lawson for collaboration in the creation of the ECB08 corpus, and acknowledge ESRC (RES 000-22-2032) for funding it, and Claire Timmins for collaboration in collecting the Glasgow data.

sociolinguistics and laboratory phonology are deeply similar empirical quantitative disciplines: Labov's early work in the 1960s in Martha's Vineyard should be seen by laboratory phonologists as part of the history of their own subdiscipline in terms of methodology. After all, it was chosen "as a laboratory for an initial investigation of social patterns in linguistic change" (Labov 1972a: 4) in order "to avoid the inevitable obscurity of texts, the self-consciousness of formal elicitations, and the self-deception of introspection" (Labov 1972a: xix).

Explicit experimentation (i.e. manipulating speaker behavior in a premeditated task-based manner to elicit particular measurable outcomes) is also a fundamental method in sociolinguistics, as indicated by Labov's canonical "fourth floor" piece of experimental fieldwork (Labov 1972a). Despite this, laboratory-based or highly technical experimental studies in sociolinguistics are less usual than the more central variationist methods of interview and spontaneous speech analysis (but see e.g. Thomas 2002a; Campbell-Kibler 2007; see also Warren and Hay, this chapter; Docherty and Mendoza-Denton, this volume). Of course, analysis (e.g. transcription or acoustic analysis) of a "sociolinguistic wordlist" is actually a simple experiment.

This section explores the adoption of one main aspect of sociolinguistic methodology to see what it can offer to the traditional experimental or theoretical phonologist for whom formal elicitation of key low-frequency forms remains the order of the day (though the same lessons apply to, say, analysis of spontaneous corpora (see Warner, this chapter). Specifically, we propose the use of social or dialectal stratification of participants as an integral part of laboratory-based research. After reviewing some basic issues and exemplifying some of them from our own work, we provide an initial checklist of methodological ideas.

21.1.2 The fallacy of the neutral participant

It is well known that subjects or participants in social-science experiments often have to be blinded to the goals and methods of the experimenter, because knowledge of the purpose of a particular experimental task often alters the behavior of the participant. The "Observer's Paradox" (Labov 1972a) has been an important topic for debate and discussion in the context of sociolinguistic methodology, the aim of which has been characterized as studying "how people talk when they are not being systematically observed [despite the fact that. . .] we can only obtain these data by systematic observation" (Labov 1972a: 209). In sociolinguistics there is a strong interest in the most vernacular, least standard, most unmonitored forms that can be observed, since it is argued this is where variation is strongest, thus providing the most evidence for theoretical investigation (for discussion, see Milroy 1987b; Bucholtz 2003; Coupland 2007: 180 ff.). (By definition, non-vernacular varieties comprise standardized and less variable systems, which therefore provide less data

for variation-oriented theoretical investigation.) In the psycholinguistic or phonetic laboratory, avoidance of such a paradox may be part and parcel of experimental design, but it is just one concern amongst a number of possible problems.

Typically, in the laboratory, unwanted participant bias is avoided through recruitment processes and briefings which mask, misdirect from, or do not mention the real purpose of the experiment; and through the use of protocols, tasks, and materials which often distract and conceal from inherent patterns in the design. A more recent and more sophisticated approach is to manipulate not merely experimental tasks and materials, but to treat participant knowledge, both explicit and implicit, as a conditioning factor, in order to see what effect different levels of knowledge have on the task (see Warren and Hay, this chapter). Contrasting implicit linguistic knowledge and behavior against explicit knowledge in turn enables the study of, for example, the salience of sociolinguistic variables, or the relationship between different levels of metalinguistic awareness (e.g. Campbell-Kibler 2007).

Generally speaking, laboratory-based approaches to phonology (Ohala and Jaeger 1986; Pierrehumbert, Beckman, and Ladd 2000/this volume) downplay and even reject traditional non-empirical methods in theorization. So there is no more introspection about contrast, alternation, and identity by the lone phonologist and little emphasis on fieldwork in collaboration with a small group of native speakers. Instead, experimental and quantitative research occupies the methodological center ground, enjoying particular success when these methods have been used to probe areas of grammar in which it is most clear that introspection and even self-recording are unreliable. The study of "low-level" effects provides data crucial to our understanding of pretty much everything in phonological systems more subtle than bare phonemic contrast. Understanding such phenomena requires the study of relatively impromptu, unmonitored, natural speech which illuminates crucial (and often rarely used) word combinations and phone sequences, often with prosodic variants of the same. Yet on the whole these laboratory experiments rely on participants who are highly intelligent, fluent, literate, and who are often colleagues or students (see Docherty and Foulkes 2000). Easy recruitment of research subjects is one reason for this bias, particularly when the research is tedious or uncomfortable. In the case of English, convenience sampling tends to involve speakers of Standard English: and though English may well be the most experimented-on language to date, its standardized forms are merely one narrow aspect. Indeed, restricting research to standard varieties lays the field open to the criticism that the effects of literacy, education, and prescriptive attitudes are filtering through into the data unobserved or unquantified.

So, though things are changing fast, a substantial proportion of earlier work in laboratory phonology is open to one of the criticisms that was leveled at generative phonology: namely that it does not exploit socially and dialectally mediated variation as a methodological tool for narrow phonological theoretical concerns. And,

of course, it cannot have anything significant to contribute to a wider and more inclusive theoretical understanding of the individual's representation and implementation of a grammar which encodes (whether modularly or holistically) their "own" narrow generative system as part of the social system around them (Foulkes and Docherty 2006; Docherty 2007a). We need to reject the idea that we can study, impressionistically or experimentally, some idealized neutral speaker of a variety without understanding better that "a linguistic variety" always includes variation.

21.1.3 Structured heterogeneity in the sampling of speakers and listeners: From cross-linguistic to cross-dialectal designs

There is an assumption in generative research that any single speaker of a language is a complete topic for study, yet for reasons that are unclear, there is often no interest in exploring differences *between* the numerous minimally different systems that exist, despite the opportunities this provides for a more subtle version of cross-linguistic research (Scobbie 2007a). Given that participants are not neutral and the population to be sampled is not homogeneous, the standard "homogeneous" sample will always exhibit interspeaker variation. This can be ignored as noise through pooling and statistical analysis or attributed to a combination of noise, investigatable artifacts (like speech rate) and indexical setting (as in the VOT study of Allen et al. 2003).

A quite different approach is to recruit experimental subjects who are known to vary, perhaps along such traditional sociolinguistic lines as sex, age, social class, or geographical micro-dialect, or to use structures and practices identified through ethnography (e.g. Drager 2009 and Warren and Hay this chapter), or some combination or extension of these (e.g. the rather different approach to VOT of Scobbie 2006). Whether the sample is intentionally bimodal, multimodal, or even continuously varying, the structure of the participant pool can begin to explain some of the noise. Such participant stratification is hardly an innovation: speaker sex is a common factor used in experiments due to the actual or potential physiological effects of speaker sex on speech production (or age, hearing loss, or medication), and geographical dialect differences of a rather coarse grain are sometimes exploited—but what is extraordinary is that social differences are exploited only very rarely, despite the opportunities they provide. It is therefore of particular theoretical and methodological interest when speakers exploit physiologically related differences, e.g. sex, for social ends (Stuart-Smith 2007b).

Social stratification extends the traditional cross-linguistic design (e.g. the VOT study of Cho and Ladefoged 1999) to what are essentially different accents of the same language (which we can call cross-dialectal for convenience). For perception studies, see Warren and Hay (this chapter). For cross-dialect research there is no

need to randomly sample on non-linguistic grounds, as a sociolinguistic study might, where sampling the proportion of different behavior in the community is important. (For different possible methods of sampling in sociolinguistics, see e.g. Milroy and Gordon 2003, and see Trochim and Donnelly 2006 for a more general review.) But for the laboratory phonologist, it may be enough that certain patterns exist, though knowing something about the prevalence of different patterns is indeed preferable. It seems obvious that an investigation of French would not randomly sample Western Europeans, or even people living in France: French speakers would be recruited! It somehow seems more contentious to take this approach in the study of highly vernacular, socially defined varieties of a language. Yet of course a laboratory-based phonological study of, say, spirantization of stops must include, by design, speakers who exhibit that phenomenon. They can then be compared to more standard speakers who are likely to not show it at all. In such diversive sampling methods it is necessary to define each speaker group and to accept that there will be random variation within the groups, but representativeness is an extra condition which is optional, not obligatory. Such groups each have their own accent or dialect, whether they vary along physiological parameters (e.g. age or sex), geographical differences (traditionally the remit of dialectologists), or social ones. A combination of approaches can be taken. Diversity sampling provides a number of specific speakers of different linguistic varieties, and snowball sampling uses them to help recruit new experimental participants from their own networks (e.g. Milroy and Gordon 2003: 32).

Cross-dialectal research avoids an obvious confound which affects cross-linguistic research, which is that different languages by definition have different lexicons, and probably different phonologies, prosodic, morphological, and syntactic systems. Thus looking at purely phonetic/phonological effects across languages (a central goal of laboratory phonology) is actually very difficult, because experimentally, too many factors are changing at once. Research into related accents (which we can for convenience call "cross-dialectal"), on the other hand, is ideal for laboratory phonology because dialects tend to vary in sound system, to greater and lesser extent. Ideally, the same materials can be used. Such variation frequently crosses boundaries in phonetic space that opens up a more complex understanding of contrast and categorization (Scobbie 2006). One major limitation is that the extent of variation is confined to genetically related systems, so it does not reflect the range of cross-linguistic differences that exist.

Crossover research faces the challenge of satisfying the methodological requirements of more than one field, but when it does successfully it can explore fundamental theoretical issues in more than one domain simultaneously, e.g. in both sociolinguistics and phonetics (e.g. Foulkes and Docherty 2006). Our argument is that traditional laboratory-based and indeed consultant-based phonetic and phonological research can also benefit from social stratification and cross-dialectal

designs, even within the narrower remit of research without specific or explicit sociolinguistic theoretical goals.

Recruiting participants by making explicit reference to their social dialect (e.g. broad Glaswegian) rather than their language (e.g. English) naturally exacerbates the Observer's Paradox. However, if the experimental task is kept homogeneous across all participants, then comparisons are still possible if variation is induced by knowledge of direct observation. For many laboratory phonologists, it will not be a problem that some participants are less vernacular than they might be in unmonitored spontaneous speech, or even if they exaggerate vernacular features in a type of hyperspeech. However, further fundamental methodological research in this area is required, not least because it is actually of great theoretical interest, since social dialect effects cut across differences in task and style in production (Wassink et al. 2007) and in perception (but see Warren and Hay, this chapter), both of which are fundamental to phonological theory. Laboratory-based studies into socially varying phonological categorization of phonetic substance or core phonological phenomena such as phonotactics, alternation, and contrast are an area ripe for future study.

21.1.4 Case study A, part 1: Social vs. contrastive functions of phonetic correlates of variants of Scottish English /ai/

In this section we exemplify some of the points above, drawing on our own research into Glasgow English (e.g. Stuart-Smith 1999, 2003; Stuart-Smith et al. 2007). Thirty-two Glaswegian participants were stratified into cells of four, the pool halved by male/female, older/younger, and working-class/middle-class parameters. We included a mini-experiment on the Scottish Vowel Length Rule "SVLR" (McKenna 1988; Scobbie, Hewlett, and Turk 1999; Watt and Ingram 2000) to probe possible phonemic splits, by supplementing a standard sociolinguistic wordlist with appropriate materials (see Stuart-Smith 1999 for details).

The SVLR basically describes short and long allophones, but includes an environment with a "quasi-phonemic contrast" (Scobbie and Stuart-Smith 2008), namely before tautomorphemic (word-final) /d/ vs. suffix /#d/. Our socially stratified sample, even with only a few tokens per speaker, confirmed the finding of the more traditional experimental study of McKenna (1988), with more tokens and contexts, but whose participants were only a few university students (Scobbie, Hewlett, and Turk 1999). Only three vowels are affected: the monophthongs /i/ and /ʉ/ (*need* vs. *kneed* and *brood* vs. *brewed*) and the diphthong /ai/ (*side* vs. *sighed*). The monophthongs were longer in duration by 50 percent or more before /#d/ relative to /d/.

The diphthong /ai/ was more complex phonetically. The allophonic and quasi-phonemic variants of /ai/ (short [ʌ̟ɪ] or [əɪ], and long [ɑ̆ːɐ]) differed in three ways

(Scobbie, Turk, and Hewlett 1999). These were (a) duration, (b) the location of the first (but not second) mora in F1/F2 space, and (c) the relative timing of the transition. Typically, long vowels are about 25 percent greater in duration; have a first mora that is lower and backer; and have a transition between the two moras that is later. To see the effect of the latter two correlates in our experiment, consider Figure 21.1.1, showing pooled data (based on the grand means of each sex-age cell's mean for *side* vs. *sighed* and *tide* vs. *tied*) over the first 200 ms of the vowel (cutting off some of the end of longest /ai/). The first half of the long variant of /ai/ is low and back followed by transitional raising and fronting. Only about a quarter of the short variant precedes the transition, and up to half its duration occurs near the second, offglide target.

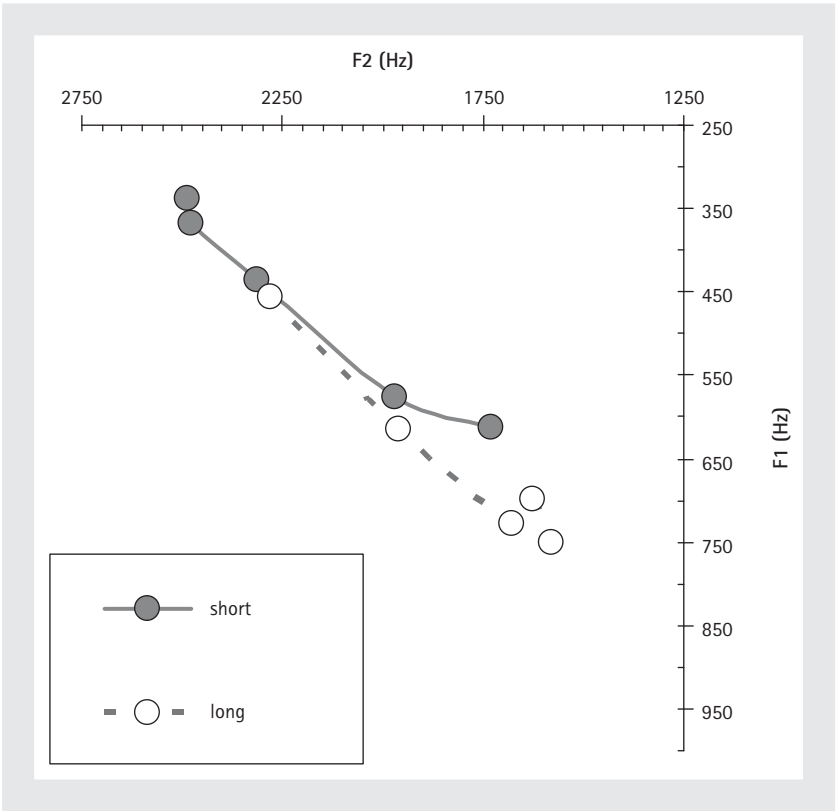

Figure 21.1.1. Quasi-phonemic variants of /ai/ before /d/ (short) and /#d/ (long) in 32 Glaswegian speakers, showing a low back first mora and an offglide trajectory of raising and fronting. The circle tickmarks indicate duration in 50 ms intervals over the first 200 ms of the vowel (approximately quartiles).

We could investigate these possible cues to the distinction through traditional laboratory-based measurement, from either a perception or production point of view, using statistical analysis (e.g. Gordeeva 2008). Instead, our stratified subject pool reveals that one possible cue, the quality of the first mora, varies *socially* (Scobbie, Turk, and Hewlett 1999 and further unpublished work). In *side* vs. *sighed* or *tide* vs. *tied*, all speakers distinguish SVLR length using all the correlates mentioned above, but in addition, middle-class speakers in general seem to have lower/backer targets (for short variants) than working-class speakers (Figure 21.1.2). The first mora quality in F1/F2 space is therefore unlikely to be an important cue to length.

The quality difference in the first mora between long and short variants, is, we think, due to target undershoot in short /ai/ caused by its reduced duration. Given time constraints, the articulatory system tends not to lower fully, failing to reach the target. A lesser duration for the first mora could have repercussions for overall duration and the timing of the transition. Lengthening/shortening affects the start of the vowel. Figure 21.1.2 shows the short vs. long SVLR difference and the class difference in real time over the first 200 ms of /ai/, pooling gender and age.

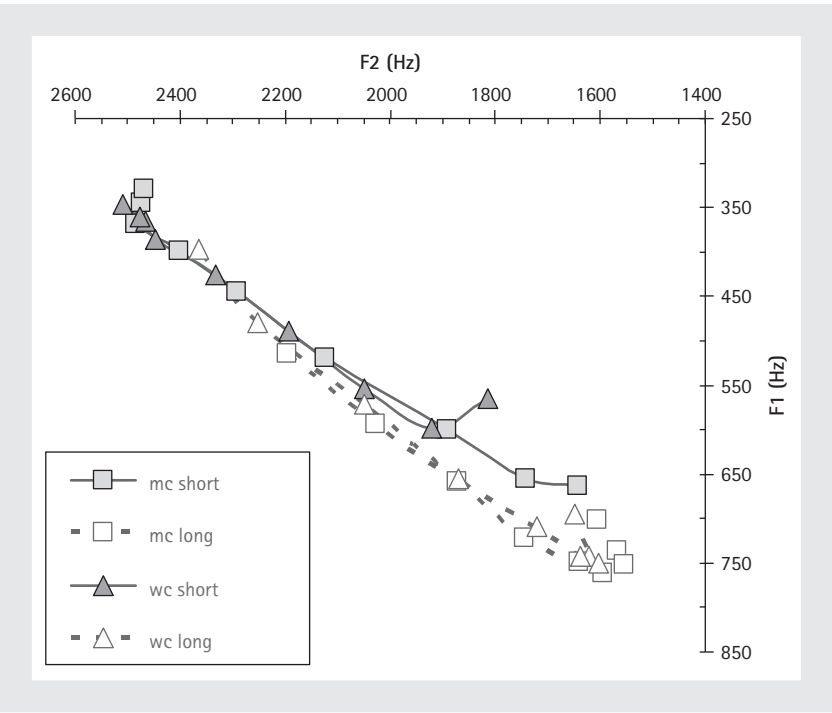

Figure 21.1.2. Quasi-phonemic variants of /ai/ with marks at 25 ms intervals, showing a social difference in the F_1/F_2 location of the first mora of the short variant.

We think that the SVLR short/long difference in /ai/ is less likely to be cued by raw duration: more important is the relative timing of (the start of) the diphthong transition from the first mora to the second: all eight speaker groups have a relatively early transition for the short variant and a relatively late one for the long variant. Also, though normally the transitions are parallel in formant space, heading towards the same formant target, in word-final open syllables, the second mora target can be undershot (and in this context, only the long variant occurs). Let us turn now to word-internal /ai/, which provides evidence to support this analysis.

21.1.5 Case study A, part 2: Cross-speaker minimal pairs

The phonetic nature of any contrast will vary, and so should be examined in a range of different environments. Unfortunately there are no minimal pairs and no near-minimal pairs which could provide experimental materials to investigate quasi-phonemic contrast in /ai/ in any context other than tautomorphemic /d/ vs. suffix /#d/, because elsewhere, long and short variants are allophonically conditioned.

However, our materials included some trochaic words with /ai/ in the initial stressed syllable, e.g. *sidle* and *bible*, and so it is possible to examine near-minimal pairs on a speaker-internal basis (e.g. if someone has a long *libel* vs. a short *bible*), but this introduces coarticulatory confounds. However, we repeatedly found individual differences in the lexical incidence of short vs. long /ai/ among young female speakers for a single word, e.g. *bible* (Scobbie, Stuart-Smith 2008). We realized we could therefore extend the fundamental concept of the minimal pair test, examining structurally matched pairs across speakers. This methodological innovation is a powerful tool for the controlled experimental study of phenomena which cannot be examined within a single speaker's grammar. A word like *bible*, for example, must have either a short /ai/ or a long /ai/, because interspeaker variation is largely bimodal.[1] The cross-speaker minimal pair lets us examine the phonetic properties of *both* /ai/ variants in a novel context (in this case [b__bəl]) which does not exist in the grammar of any individual (because the phonetic difference does not condition a difference in lexical meaning). Free variation, within an otherwise relatively homogenous accent group, or class-based variation is most useful, since both avoid phonetic effects of age or sex.

Applying this technique, we found that vowel duration was not representative of the long/short variant difference (Scobbie, Turk, and Hewlett 1999). Figure 21.1.3a shows that within words, duration of /ai/ is unlikely to be a robust cue to the category. These young females from both social groups have, for comparison,

[1] Whether intraspeaker variation is gradient or categorical is a topic for future research.

normal duration differences in the quasi-phonemic context, where /ai/ before word-final /t/ and /d/ is 198 ms and 204 ms (n.s.) but before /#d/ is 230 ms (where /d/ vs. /#d/ is significant in a paired samples t-test, $t(8) = 1.89$, $p < 0.05$), and where *crisis* (short and/or long) vs. *miser* (182 ms) also has a clear duration difference.

Figure 21.1.3b (for *crisis*) suggests that transition timing (with first mora undershoot) is again a strong correlate of the short/long distinction, while overall duration is far less important (125 ms vs. 139 ms for *crisis* in Figure 21.1.3a), or irrelevant. Basically, short /ai/ has early transitions (within 25 ms of the start of the approx 125 ms vowel) whereas long /ai/ begins its transition after a delay of about 50 ms. Comparable findings from other words and other speaker groups support this conclusion. For example, among the older males, all four working class use a short variant in *crisis*, and all four middle class the long one (Figure 21.1.4).

The lack of an overall durational difference is probably caused by undershoot due to durational compression, and perhaps different speech rates could be used to examine this more closely. However, the cross-speaker method is clear and generally applicable.

21.1.6 Sparsely populated phonotactic cells and low frequency items

Lexical frequency is an important factor in much laboratory-based research, and extending this work on /ai/, for example, means looking at rare words (e.g. *sisal*, *taigon*, *Krug*, *Beeb*, *oblige*) (Scobbie 2005). This turns a simple fact into a problem: people do not have the same lexicon, nor are frequency counts necessarily meaningful for unusual words (see Jurafsky 2003 for more discussion), and elicitation is hard other than through reading. In the sociolinguistics context, researchers have experience of the performance problems associated with reading aloud from text while being observed by pronunciation researchers, the disparity between oral vernacular lexical items and standard written ones (but see Macaulay 1991), and cross-dialectal borrowing.

Reading skills become an issue for even high-frequency items in simple sentences, so appropriate methods may be required (see below). One approach, piloted in Scobbie (2005), is to use carrier sentences which include the meaning of the relevant word and ask the participant for their estimated age of acquisition after reading the sentence twice (with an additional single citation form). This draws attention away from pronunciation and provides useful information on whether the item is known or not. Another alternative we have piloted is to use semantic sets in the elicitation of single citation forms, to focus attention away from phonological relationships. To investigate Scottish /w/ vs. /ʍ/, for example, one set might be "*Scotland*, *Ireland*, *Wales*, *England*" and another "*dolphins*, *turtles*, *whales*, *fish*." For efficiency, sets should probe multiple topics.

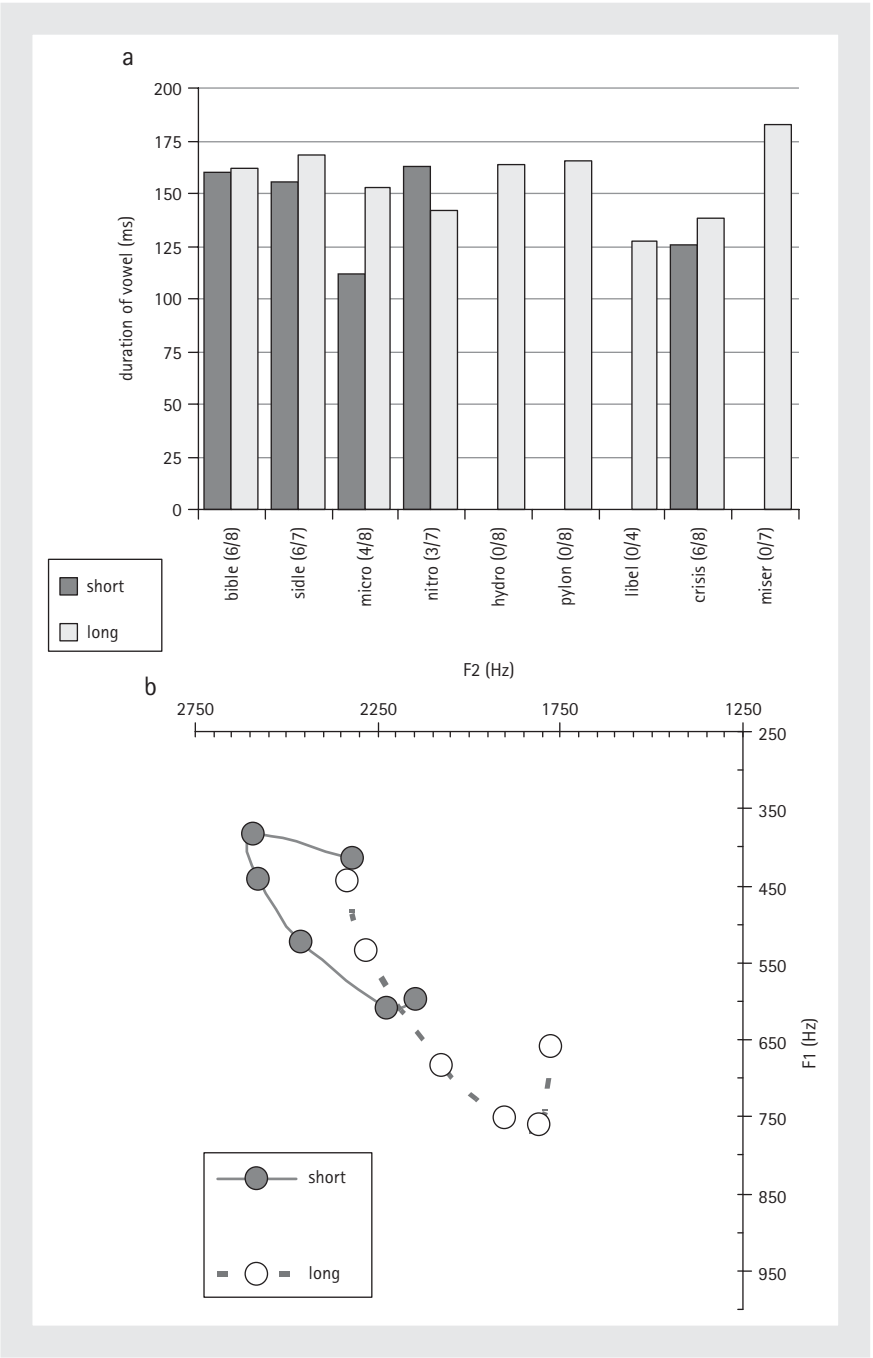

Figure 21.1.3. a. Mean duration of /ai/ in trochees, separated by impressionistic categorization of vowel into long vs. short variants, for eight young female Glaswegians. The number of short tokens and the total number of tokens are indicated in brackets. b. Mean formant transitions of the two variants of /ai/ in *crisis* with marks at 25 ms intervals from the same speakers.

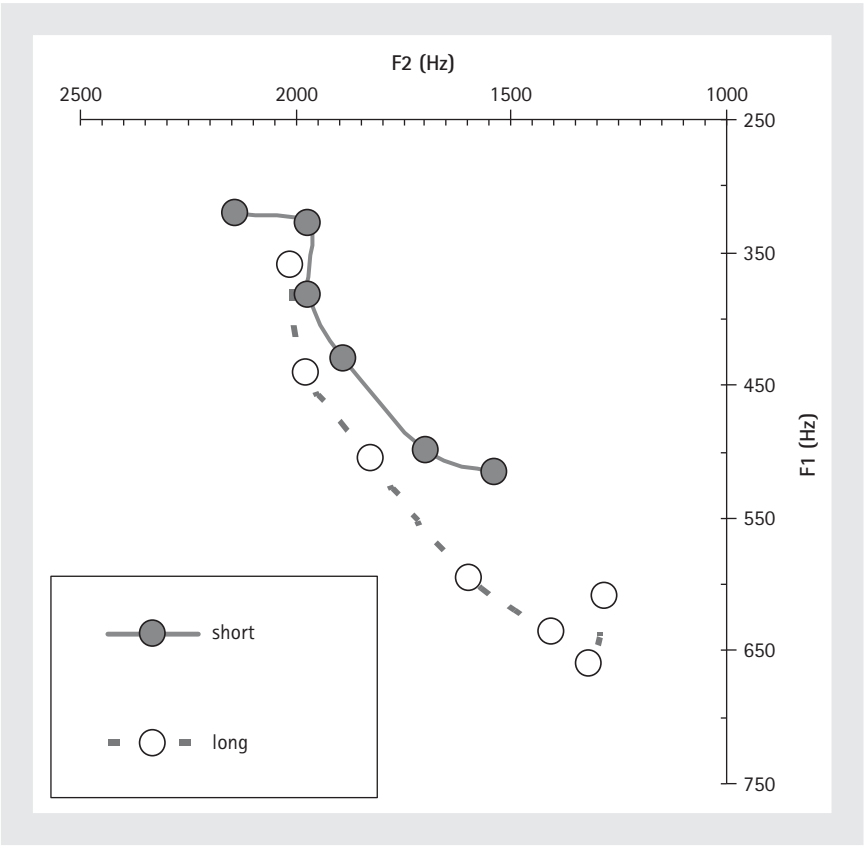

Figure 21.1.4. Short (all WC) and long (all MC) socially meaningful variants of *crisis* /ai/ by eight older male Glasgow speakers.

21.1.7 Case study B: Vernacular articulations in the laboratory

For technical quality and for logistical reasons it is probably preferable to collect vernacular data in the laboratory, and despite the obvious difficulties in relaxing the participants, it is possible to collect good casual/vernacular speech data in this setting (see Warner, this chapter; Anderson et al. 1999). However, such research rarely if ever employs social stratification, and may even fail to record simple social demographic sociolinguistic information. It may also not be explicitly aiming for vernacular speech.

However, using more intrusive phonetic instrumentation would still seem genuinely problematic. Another study that we have conducted has investigated the speech production of a socially structured sound change in progress, namely derhoticization of coda /r/ (Scobbie, Stuart-Smith, and Lawson 2008). This Ultrasound Tongue Imaging (UTI) study used an ultrasound probe mounted on a stabilizing

headset, attached to a processing box, in conjunction with an audio recording, with a team of experimenters/technicians present. Unsurprisingly, typical sociolinguistic data elicitation is naturally biased towards the acoustic analysis of speech, because an anonymous microphone appears to offer the prospect of reducing the Observer's Paradox in a straightforward way. Articulatory phonetics, on the other hand, is *a priori* more intrusive. Yet it is important to merge these different methodologies to help provide data that exemplified a wider set of systems and is more ecologically sound.

We investigated the relative impact of different experimental conditions on fine-grained aspects of vernacular speech performance in young working-class adolescent males from West Lothian, Scotland. First, pairs of participants were recorded chatting to each other and reading a wordlist in their school, with and then without the UTI equipment. No significant differences in consonantal variation were found between the two conditions (Lawson et al. 2008). Then four of the same participants were brought to the laboratory, where audio and ultrasound recordings of read and spontaneous speech were made. Informal impressions are that there are few differences in the level of vernacular obtained between the field and the laboratory recordings. This same impression was maintained for the subsequent socially stratified corpus (ECB08) which was collected from a further fifteen adolescents. It emerged from these experiments that physical context, i.e. a university laboratory, mattered less than the more personal factor of the interlocutor. Having a peer—and friend—present with a participant seemed to be a strong predictor of more natural speech. This finding also highlights another aspect of data collection which is often manipulated in sociolinguistic fieldwork and laboratory phonology alike, namely the use of varied tasks to provoke variation within speakers. From sociolinguistics, however, we can see speakers control a range of repertoires from the more to the less vernacular, a systematic range of linguistic specification. This range can be tapped into by recording more than one speech style by varying formality, for example (though note that recent studies show that reading tasks may not always elicit the most standard forms from younger speakers, Stuart-Smith et al. 2007).

The exciting outcomes of the research include, of course, data on how vernacular Scottish is articulated. Just as important are the insights that emerge because the data comprises part of a socially structured set of linguistic variation. For example, the complex articulatory-acoustic relationships found in /r/ (Mielke et al. forthcoming) now have to be seen in the context of the range of rhoticity observed, from the traditional rhotic Scottish English (which is comparable to North American English) through to the covert rhotic articulations observed in vernacular Scottish English (Scobbie, Stuart-Smith, and Lawson 2008). In other words, social factors affect the much-researched phenomena of intergestural coordination, gestural strength, and the articulatory-acoustics relationship just as much as linguistic ones such as segmental context or prosodic context.

21.1.8 Conclusions and a recommendation checklist

Laboratory phonology has successfully shown that phonological research must not restrict itself to introspective study by highly educated middle-class linguists. Our point is that we must not fall into the trap of undertaking laboratory studies on a similar narrow group of research study participants. Having rejected introspection, empirically minded phonologists (and phoneticians) should be drawn to the lure of vernacular variation, because purportedly neutral experimental participants exemplify only a small proportion of the sound system phenomena within any given language, and their linguistic systems are subject to standardization and other subconscious or conscious prescriptive pressures. Moreover, there are methodological and theoretical advantages in incorporating socially structured pools of participants into experimental design, whether they include speakers of standard varieties or not, simply because this offers a fresh dimension of conditioning factors that are relevant to phonology and phonetics. Socially and geographically structured microvariationism are extremely useful additions, we believe, to the experimental toolkit, particularly suitable for exploring the structuring of systematic fine phonetic detail (e.g. Hawkins and Smith 2001).

There are many advanced and introductory works on sociolinguistic methodology, too numerous to list in full here, which should be consulted for insights into how to address certain core issues, such as defining social class, recruitment, and the tracking of social networks (e.g. Labov 2001; Chambers et al. 2002; Milroy and Gordon 2003; Llamas et al. 2006; Tagliamonte 2006; Meyerhoff 2006). In addition, bilingualism research offers useful guidelines on encouraging the use of a particular language mode (e.g. Grosjean 2008). Instead, and to conclude, we offer some methodological topics to consider, which we think are useful starting points for the standard laboratory phonologist.

- Consider your question from a wider perspective—have sociolinguistic studies been carried out on the variety and/or the feature you are examining?
- Consider stratifying your participants on some basis which is likely to provide structure to interspeaker variation which appears to cross phonological category boundaries.
- Avoid exclusive use of graduate students, but widen your pool to undergraduates, including from other subject areas, such as engineering/physical sciences.
- Brief your participants that natural relaxed speech is the goal and misdirect them by stressing that the recordings are being made for some purpose other than judgment of correctness, such as measuring the "noise" of talking, and that they need not aim to speak properly or carefully.
- Experimental participants can often be encouraged to be more vernacular by being accompanied or even observed by a (quiet) friend from the same background, who is permitted to sit in the lab to exert subliminal pressure to conform to type. Let the friendship pair converse spontaneously and unmonitored before

experimental materials are used to let them become accustomed to the lab. Vernacular jokes or stories may also make a useful ice-breaker. Use a non-native or alternatively a vernacular research assistant or technician.

- Pictures and oral prompts may work better than orthographic prompts, particularly for vernacular lexical items which are not often written down. Try designs in which participants can repeat or respond to audio prompts in vernacular or other accents and not merely to read aloud from standard language.
- If you can control for the generally poor acoustic environment, try fieldwork in schools, shopping centers, or museums for mass observation research where a high number of participants can be recruited in a short time.
- Articulatory instrumentation can be sold to speakers as being tools for understanding the physiology, the size, shape, and functioning of the vocal organs, where speech is just a way of getting the organs to move.
- Finally, and most importantly: collaborate with sociolinguists.

Our final remark is to make a plea for the consideration of varieties of language which are "outside the box." We think in particular of so-called "mixed" systems, arising from multilingualism or multi-dialectalism. Both laboratory phonology and sociolinguistics find such systems challenging, and yet a great deal can be learned from such a natural part of typical language acquisition and function. Thus simpler is not necessarily better. By incorporating what may sometimes be dismissed as extraneous social factors in addition to the typical phonetic, phonological, and psycholinguistic ones, laboratory phonology will provide an evidence base more powerful for those with narrow interests and more representative for those with broad ones.

21.2 Methods for studying spontaneous speech*

Natasha Warner

21.2.1 Introduction and terminology

The overwhelming majority of research in phonetics and laboratory phonology has used *careful* speech, but interest in *spontaneous*, non-careful speech is now surging. This could lead to a very different understanding of how speech and

* The author would like to thank Mirjam Ernestus, Ben Tucker, Anne Cutler, Holger Mitterer, and Rob Podesva for helpful discussion and feedback on the issues in this section. All errors, of course, are the author's own.

communication work. Spontaneous speech often includes sequences with such strong reduction phenomena, that one could never have predicted them and is rather surprised to see them when one examines the spectrogram (e.g. Figure 21.2.1, with multiple deletions, reduction of stops to fricatives, and changes to vowel qualities). Yet these sequences usually sound intelligible and normal, at least to native listeners. But what types of speech are at issue? This section offers a brief discussion of terminology (see also Warner 2011).

One could establish a continuum of carefulness or naturalness in speech data collection. On one end might be vowels or nonsense monosyllables read in isolation (perhaps while wearing an airflow mask). At the other end might be informal conversation among family or friends, perhaps at home with no microphone present. Several terms would fall along this continuum: careful or laboratory speech (near the careful end), any non-read speech (including responses to prompts), connected speech (anything in a longer utterance, read or not), spontaneous speech (nothing read, but including monologues and structured speech such as Map Task dia-

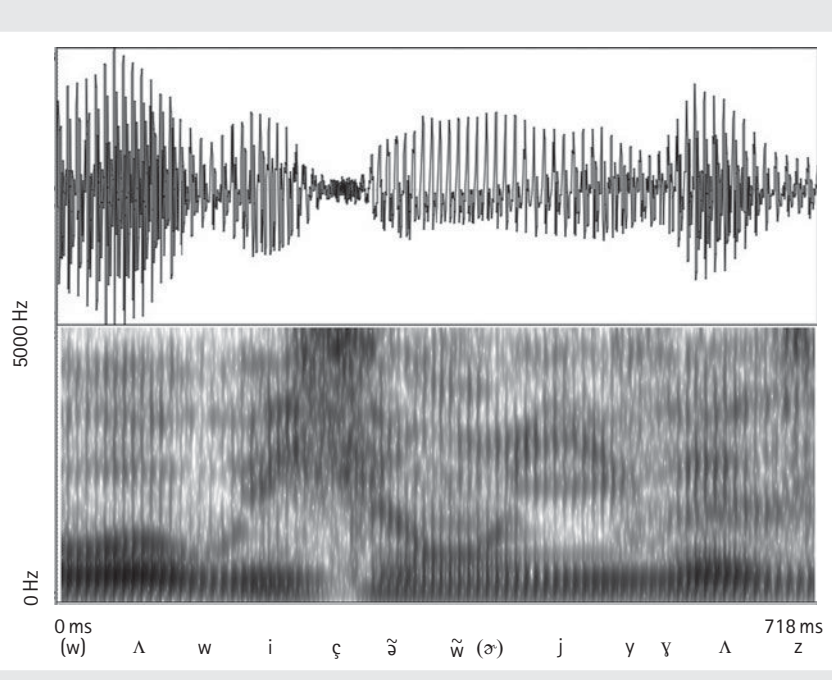

Figure 21.2.1. Waveform and spectrogram from conversational speech, . . .*what weekend were you guys*. . . Symbols in parentheses are segments for which there is little acoustic evidence. Heard in isolation, the portion corresponding to *-end were* does not consist of any identifiable segments, but the entire utterance sounds quite natural and clear.

logues), and conversational speech (even with interviewers). The *clear speech* that Bradlow's group has studied (e.g. Bradlow and Bent 2002; Smiljanić and Bradlow 2009), for addressing hearing-impaired or L2 listeners, is near the careful end of the continuum, more careful than typical read lab speech. The term *natural speech* could be taken as the other end of this continuum, but researchers of differing backgrounds use *natural* very differently. *Natural speech* can mean anything produced by a human vocal tract (not synthesized), or in linguistic anthropology it can set strict requirements on the interactional setting. Therefore, natural speech will be left undefined.

Sociolinguists have put considerable effort into defining various ways in which speakers vary their speech style, some of which overlap with the carefulness dimension delineated here. Schilling-Estes (2002) provides a clear overview of sociolinguistic approaches to speech style, and of shortcomings of simple explanations for why speakers vary style as they do. The carefulness continuum here does not claim any of the particular explanations Schilling-Estes (2002) discusses, but is simply a description of a continuum along which several types of speech fall. Speakers of course vary their speech style in many other ways not covered by *carefulness*, for example in order to show affiliation with a variety of groups, their attitude toward interlocutors' utterances, etc.

The terms *reduced* and *fast* speech are not on the same continuum as carefulness. Careful, read speech and casual conversation can both be fast or slow, and speech rate can be measured acoustically, unlike spontaneity. I take *reduced speech* to refer to any speech exhibiting reduction from the canonical, careful pronunciation, e.g. speech with segments or syllables deleted, with expected stops realized as approximants, with vowels approaching the center of the vowel space, with incomplete tongue closures, etc. That is, reduction is defined by the results (acoustic or articulatory), not by the circumstances under which it is recorded. One finds reductions even in isolated word list reading, but not as often as in conversation. Figure 21.2.2 shows a schematic representation of carefulness, speech rate, and degree of reduction as separate continua.

With the terminology defined, the rest of this section turns first to methods, then to theory. Regarding methods, the major issues for reduced speech are how to obtain recordings of reduced or less-careful speech (21.2.2), how to obtain or create stimuli for perception experiments on such speech (21.2.3), and how to analyze the resulting data (21.2.4). Section 21.2.5 discusses implications of reduced speech for phonetic, phonological, and psycholinguistic theories.

21.2.2 Methods for eliciting less-careful speech

This section addresses recording methodologies (see also Post and Nolan, this volume on related issues). The main purposes are acoustic phonetic analysis, obtaining

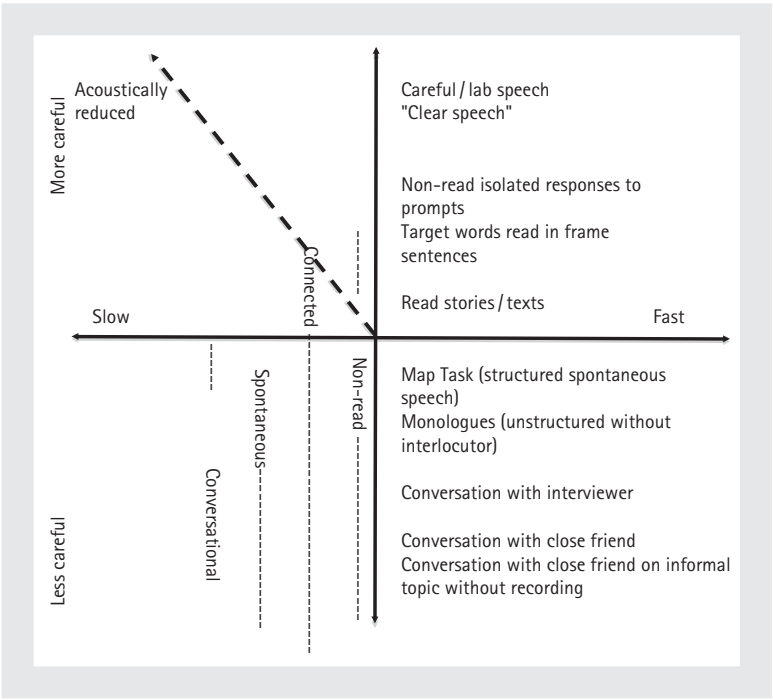

Figure 21.2.2. Schematic representation of carefulness, speech rate, and acoustic reduction as three distinct dimensions along which speech can fall. The carefulness dimension indicates specific examples of speech settings on the right, and ranges covered by the terms non-read, connected, spontaneous, and conversational on the left. The ordering of specific examples on the right is approximate. For example, target words in frame sentences might be more or less careful than non-read responses to prompts in a particular experimental task.

stimuli for perception studies, or development of Automatic Speech Recognition (ASR) systems. None of the methods are optimal, and they are very much under exploration.

One common method to balance control and naturalness is to record subjects conversing with an interviewer/experimenter, as in the Buckeye Corpus (Pitt et al. 2005). Such speech can be rather natural, but still have very good acoustic conditions. The interviewer can attempt to make the interactions casual, and less like a formal interview. The interviewer can keep the subject talking, and can steer the conversation to include topics that are likely to elicit target words, to obtain some matched words across speakers. There are a few disadvantages: the subject does not know the interviewer, making speech more formal than in conversation with friends

or family. The subject's and interviewer's voices are likely to overlap sometimes, although the interviewer can attempt to avoid overlaps. This can be avoided if the interviewer is outside the sound-protected booth and the subject inside, and they hear each others' speech over headphones, but that creates an unfamiliar and less natural setting. (See Scobbie and Stuart-Smith, this chapter, on the overall topic of the effect of presence of experimenter or recording equipment on naturalness of speech.)

The most obvious way to record conversational speech might be to put two subjects who know each other well in a sound booth together, with head-mounted directional microphones that pick up little of the other speaker's voice, and have them converse. I do not know of many studies that have done exactly this (but see the French CID corpus, Bertrand et al. 2008), perhaps because most recording booths are small. However, Ernestus and colleagues use a clever variant on this method (Torreira et al. 2010), by seating two subjects who know each other and a confederate in the booth, so that the experimenter/confederate can help get the conversation going well (making the subjects comfortable), and can steer the conversation toward certain topics. The confederate then leaves the booth on the pretense of switching out a broken microphone, leaving the subjects to converse. This method has some danger of overlapping speech in the recordings, and requires a rather large sound booth. Speakers might also be distracted by a head-mounted microphone on their interlocutor's face, reminding them of the unusual conversational setting. However, Scobbie and Stuart-Smith (this chapter) find that even obvious ultrasound equipment has little negative effect on naturalness if speakers speak casually with an interlocutor who is a peer, so an unobtrusive microphone may not be a problem.

Recording telephone conversations is another method to obtain very spontaneous speech. It avoids several problems of recording two speakers conversing in person: overlapping speech will be recorded separately, and speakers need not sit together in a sound booth. Speakers are also very comfortable with conversing casually on the phone, so speech is very natural. The Switchboard, CALLHOME, and CALLFRIEND corpora all exemplify this approach (e.g. Canavan and Zipperlen 1996, or Switchboard as analyzed by Bell et al. 2009). Conversations can be between acquaintances (e.g. CALLHOME, CALLFRIEND) or between two volunteers introduced for the phone call (Godfrey and Holliman 1997). The naturalness of such recordings is a clear advantage to this method, particularly when the speakers know each other well. However, the recordings retain only telephone speech bandwidth (500–3500 Hz), and speakers call from locations with highly variable, sometimes extremely loud background noise. It may be difficult to collect detailed information about speakers' language and dialect backgrounds. There is, of course, no control whatsoever over what the speakers say. One can take this approach even further and simply attach a recording device to the speaker and leave it recording while they carry on with their daily-life activities, without the researcher present (Mehl

and Pennebaker 2003; Podesva 2006). Podesva uses this method for sociophonetic analysis, while Mehl and Pennebaker use it to obtain social psychology data. Although Podesva's data allowed for detailed phonetic analysis of unusually natural speech, this method has a clear danger of failure to obtain high acoustic quality in recordings.

The favored method in my own lab is to have a speaker sit in a sound booth and talk on the telephone to a close friend or family member, while wearing a head-mounted microphone over the opposite ear from the telephone. The recording only includes one side of the conversation (losing discourse information), but the acoustics are excellent, and the speech is extremely natural and casual. Speakers rapidly become comfortable with the sound booth, and begin animatedly discussing informal topics (e.g. gossiping about one's boyfriend to one's best friend). This retains all advantages of recording telephone speech (except losing the interlocutor's side), but provides high-quality acoustic recordings as well. One might be able to recover discourse information about the interlocutor's utterances (although not phonetic information) from a weak signal the microphone might pick up from the telephone. However, if the microphone picks up enough of the interlocutor's speech to be intelligible, this might require that the interlocutor also be a consented human subject (depending on local regulations), which would present logistical problems.

A method one step less natural is to record spontaneous monologues (e.g. "now please tell us about yourself") over the telephone or in the lab. The Oregon Graduate Institute (OGI) corpora for various languages use this method (Muthusamy et al. 1992). This is easier to set up than conversations: subjects call a toll-free number and hear recorded prompts, so no pairs of subjects need be arranged. This method gives spontaneous but not conversational speech. Some speakers find it difficult to speak with no interlocutor, or to speak naturally to an answering machine, but surprisingly many subjects do quite well at this task (Warner and Arai 2001).

The Map Task (Bard et al. 2001; Shattuck-Hufnagel and Veilleux 2007, among others; and see Warren and Hay, this chapter) elicits relatively spontaneous, conversational speech while maintaining considerable control over target words. In this method, two speakers look at non-identical maps. One speaker directs the other on how to go from one location to another on the map. Because the maps differ, the listener is likely to ask for clarification. Neither speaker is reading a script, although some features on the map might be labeled in order to induce speakers to use specific target words (the labeled items) that contain phonological properties of interest. This method leads to conversational but relatively formal speech.

Moving further toward controlled speech, one can record speakers reading a very large quantity of connected texts. ATR (Kyoto, Japan) in the development of their speech synthesis program might record a speaker reading a newspaper out loud for an hour (Campbell 1992, 1999). This is not spontaneous speech, but when reading

for so long, speakers are likely to speak less carefully. This method gives control over the content and some control over likely intonational patterns. The speech is less variable than conversation, where speakers shift rapidly from enthusiastic speech to slow, tired-sounding utterances.

Recently, it has become possible to obtain large quantities of relatively natural speech over the Internet, even for a variety of languages. Kim (2004) provides just one example of work answering a question about connected speech with such material (see also Loehr and Van Guilder, this volume). Radio and television broadcasts, often available for download, can provide huge publicly available, pre-recorded corpora. One step in using such recordings is to classify the types of speech and speaker, since the material includes both professional newscasters and non-professional speakers (e.g. in interviews). Some speech may be scripted and some spontaneous, and one cannot necessarily tell which. Background noise (recorded in studio vs. on site), background music, dialect, topic, and genre may all vary. Files may be compressed, in a variety of only partially predictable ways, which may make some more detailed acoustic analyses impossible. Language background information is likely unavailable. However, with the number of broadcasts available over the internet rapidly increasing, this provides an exciting opportunity to study relatively natural speech, particularly for languages where large recording experiments might be impossible.

Going a final step toward controlled speech, one can simply manipulate speech rate by instructing speakers to read target sentences quickly, normally, or slowly. If speakers succeed in reading quickly, they are likely to produce some reductions. Research on topics other than reduction uses this method (Ladd et al. 1999; Hirata et al. 2007; and Adank and Janse 2009 provide a few examples), and speakers can vary their speech rates, although this may not be the same as what they do in natural speech. However, speech rate is not the same thing as speech style, spontaneity, or casualness. Overtly asking speakers to vary their speech rate should not be the main method for recording reduction.

21.2.3 Current methods for obtaining stimuli for reduction perception studies

Section 21.2.2 summarized methods for obtaining acoustic recordings, but perception of spontaneous speech may be even more interesting. Perception experiments on reduced speech require stimuli containing reduction, which are even harder to obtain than good spontaneous acoustic recordings. Research on perception and psycholinguistic processing of reduced or spontaneous speech was almost non-existent until a few years ago, with intriguing exceptions such as Mehta and Cutler (1988) and Koopmans-van Beinum (1980). Sociolinguists have long used relatively natural speech in perception experiments (e.g. Labov 1989), but these

studies are often for the purpose of studying perception across dialectal varieties, not perception of reduced speech of one's own variety. (See Warren and Hay, this chapter on the importance of perception studies for sociophonetic topics, as well.) Researchers are now developing an array of methods for obtaining stimuli, varying in naturalness of the source speech.

The most direct method is to extract stimuli from large, relatively natural corpora that have been collected using the most spontaneous and conversational speech methods above. For example, one can record a conversation and extract stimuli from only the parts without overlapping speech, or record one side of a telephone conversation and extract stimuli from that (e.g. Ernestus et al. 2002; Warner, Brenner, Woods, Tucker, and Ernestus 2009; Brouwer et al. forthcoming; cf. also Labov 1989). This has the advantage that stimuli definitely represent what real speakers produce in conversation, and what listeners hear in their daily lives. However, the stimuli are highly variable and uncontrolled. One cannot make a target word list in advance, but must be able to use a wide variety of words, of varied numbers of syllables, spoken with any intonational pattern, in any context, etc. Items will not match across conditions, either. One can record a long conversation and use only portions that meet criteria as stimuli (e.g. only content words in a particular intonational context), but the materials will still vary widely. For example, in one study in my lab, utterances for the target *he's* include *Well, because he's turning 24 and he hasn't accomplished anything in his life*, and *He's like, "I just. . ."* despite their different lengths. In another study, target content words for use in a cross-modal priming lexical decision task included both *kindergarten* and *free*. In psycholinguistic studies not on reduction, all targets often contain the same number of syllables, are controlled for some of the phonemes, and are recorded in isolation or in a consistent frame sentence. For example, Gaskell and Marslen-Wilson (1996), although they go considerably further toward connected speech materials than many studies, use prime words such as *broad*, *cloud*, *crowd*, *bread*, etc., with the most varied items being *concede*, *horrid*, *wicked* etc. Not surprisingly, one may not obtain significant results with the more variable stimuli one takes from open conversation. However, with tasks or questions for which varied stimuli can work (e.g. Ernestus et al. 2002), this method may be optimal.

A related method is to record spontaneous speech, extract usable stimuli, then bring the same speaker back to read those word strings again as careful speech, out of context. One can thus compare listeners' reactions to spontaneous vs. careful speech using the same targets, with words and voice controlled. This requires two recording sessions with each speaker though, and speakers may read casually because the phrases are from their own spontaneous conversations, minimizing style effects. Intonation may differ unpredictably between the spontaneous and read utterances. Mehta and Cutler (1988) use this method successfully with a phoneme-monitoring task, and my own lab has attempted this recording method. However,

we were unable to obtain even a priming effect using cross-modal identity priming with such varied stimuli.

To obtain controlled reduced stimuli, one can record spontaneous or careful speech and resynthesize or splice to manipulate acoustic characteristics one sees in reductions (with PSOLA, LPC, or intensity resynthesis; see Reetz, this volume for a discussion on (re)synthesis). One can also instruct the speaker to say the words over and over, sometimes "in a sloppy way" and sometimes "normally," to obtain reduced and careful tokens of each target. Examples include Mitterer and Ernestus (2006), Niebuhr (2008), and Warner, Brenner, Woods, Tucker, and Ernestus (2009). Sociolinguists have also used these methods to study listeners' perception of sociolinguistically marked variables. Campbell-Kibler (2008), for example, uses splicing between [ɪŋ] and [ən] versions of the English *-ing* suffix, then uses resynthesis to match duration, intensity, and pitch to a target pronunciation. The degree of control with resynthesis is a clear advantage: one can know that only one acoustic aspect of reduction varies at a time (Figure 21.2.3). However, one can never know whether the stimuli are truly representative of spontaneous speech, although one can resynthesize beginning from both a careful and a reduced production, or from a variety of productions (Warner, Fountain, and Tucker 2009), for example, to determine whether the perceptual effect holds despite other cues that may be present. One could also use parametric synthesis from scratch and vary specific acoustic

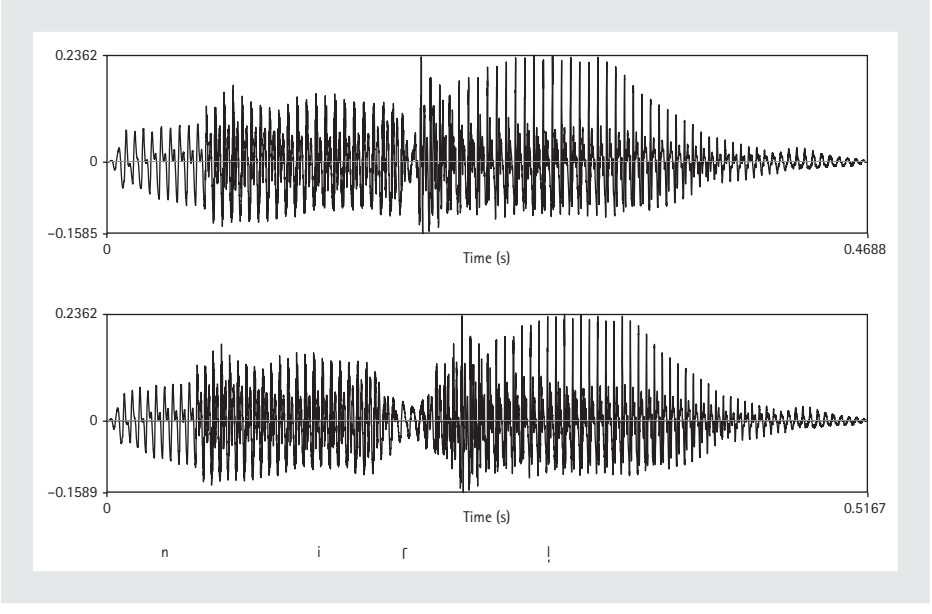

Figure 21.2.3. Waveforms of two steps on a flap duration continuum for the word *needle* (Warner, Fountain, and Tucker 2009), resynthesized to simulate reduction vs. careful speech.

characteristics that mimic what one sees in natural reductions. This provides even more control, but departs from the external validity of spontaneous speech stimuli.

A final method is to record a phonetician intentionally producing reduced and unreduced (careful) forms of target items. One obtains well-matched stimuli in a consistent voice, without synthesis, and the stimuli may represent what naïve speakers do in daily-life speech, but one cannot be absolutely sure of this. Acoustic measurements can partially confirm that stimuli match natural reductions (Tucker 2007).

None of these methods is perfect, but all can contribute—there is no clear way to study perception of reduced speech under controlled circumstances. Therefore, rather than discarding methods as flawed, or even worse avoiding the entire topic of reduction, we should use multiple methods and look for convergent evidence. Since the field of phonetics has been dominated by careful speech, and acoustics is ahead of perception for reduction studies, it is no surprise that the methods are just being explored. A wide variety of methods in flux may signal an innovative research area.

21.2.4 Spontaneous speech analysis methods

After collecting data, one must determine how to analyze it. For acoustic work, researchers are developing some novel methods in order to measure the variable and unexpected segments of reduced speech. With controlled word lists, one knows what segments to expect and can define specific measurement criteria, e.g. offset of voicing for voiceless stops vs. offset of F2 for voiced ones. With reduced speech, though, one can rarely predict what segments will be present, or what manner of articulation they will have. This forces flexibility in measurements.

One relatively common method (Greenberg 1999; Johnson 2004; Shattuck-Hufnagel and Veilleux 2007) is to transcribe a corpus at phonetic and word levels, then compare the segments in the phonetic transcription to the segments given for the same words in a searchable dictionary. One can then tally deletions and substitutions relative to the canonical form as it is transcribed in an electronic dictionary. This method can answer the overall, descriptive question of how much reduction is happening, and it can be applied regardless of what words speakers use and what segments are realized how. However, this method is heavily influenced by transcription conventions, and it assumes that the signal can be transcribed as distinct, categorical segments. In reduced speech, one often hears a segment that one can identify as vocalic, but one cannot say if it is a segment of the language, let alone which one. The heavy coarticulation of reduced speech makes counts of transcription mismatches suspect, even if the segments do seem to be identifiable. Current theories do not all assume that each word has a single, invariant lexical entry, making comparison of the realization to the single form in the electronic lexicon less meaningful, but this is a theoretical issue that goes beyond the methodological

point. A related method is to use automatic speech recognition (ASR) to (help) produce a transcription or locate segment boundaries (Pluymaekers et al. 2006; see also Cole and Hasegawa-Johnson, this volume).

One can avoid transcription problems by using a more global measure such as syllable count (surface perceived syllable count vs. underlying expected count) or overall speech rate (in underlying syllables per second, for example). ASR can also offer innovative global measures of reduction (Nakamura et al. 2007).

Alternatively, one can go to a more detailed, rather than global, method by using traditional phonetic measures such as duration, intensity, etc. to measure particular segments in detail. However, one must define criteria that cover any manner of articulation speakers might produce. In my lab's work, we study reduction of stops and flaps, and define criteria conditionally depending on whether the target is realized as a stop/flap (voiced or voiceless, with or without burst), an approximant (with or without weakening of formants), or is deleted. Riehl (2003) offers a partially automated method of measuring variable flaps.

For perception studies, however one obtains stimuli, one can use them in standard phonetic and psycholinguistic perception tasks (e.g. phonetic identification, discrimination, phoneme/word monitoring, lexical decision, priming). Some methods require filler non-words, which can be challenging to make from spontaneous conversation, but most methods are possible. Thus, for production studies, the methods issues include both how to obtain and how to analyze data, but for perception, the primary issue is how to obtain stimuli, not how to analyze results.

21.2.5 Relationship of the methods to theoretical issues

This section lists a few theories for which spontaneous speech methods may be relevant. See also Ernestus (this volume), Coetzee (this volume), Gafos and Goldstein (this volume), Nguyen (this volume), Warren and Hay (this chapter), Scobbie and Stuart-Smith (this chapter), and Warner (2011). The most obvious theoretical connections are with articulatory topics, such as Articulatory Phonology and task dynamics, because gestures describe reductions conveniently. Reduction also has clear relevance for Lindblom's H&H model (1990) and the idea of competing constraints in OT favoring ease of articulation vs. perceptibility (critiqued by Hale and Reiss 2000). Reduced speech could also impact most phonetics-phonology interface questions, such as what about language is gradient vs. categorical, and what is conditioned vs. random variability. So much varies unexpectedly in spontaneous speech, that it provides an excellent theoretical testing ground regarding what is under the control of an abstract, categorical phonology. For all theories of speech production (phonetic, phonological, and psycholinguistic; see Bell et al. 2009), reduction tests whether a theory generalizes to daily-life speech, since theories are

nearly always developed based on careful pronunciations. However, Warren and Hay (this chapter) point out that not all questions can be answered using daily-life speech, and some theoretical questions should be addressed using controlled, targeted laboratory speech.

A better model of reduced speech could lead to a change in theories of formal phonology: just the idea that there are many possible realizations of a given word, which are not entirely predictable from a single underlying form, is problematic for most theories (see also Hawkins, this volume). For example, how would one derive [wiçə̃] for "weekend" (Figure 21.2.1), or [pʰɛɾĩ] for "apparently" (Johnson 2004), while still being able to derive the more canonical forms of the words and the many other possible pronunciations? Phonological theories usually generate a single surface form, not the tens of distinct pronunciations documented by Greenberg (1999), Bell et al. (2009), and others, e.g. 117 for "that," with the most common, [ðæ], representing only 11 percent of tokens.

Turning to theories of perception, reduced speech clearly impacts theories of spoken word recognition, such as TRACE, SHORTLIST, Merge, etc. Recognition of reduced words in their many surface forms is problematic for the same reason as the multiple forms are problematic for production theories: the multiplicity of forms cannot be systematically derived from the underlying form (Ernestus et al. 2002). If *that* has at least 117 distinct pronunciations, what is stored in the lexicon? This can shade into an exemplar model (e.g. Johnson 1997b, 2006; Pierrehumbert 2001a, 2002). One might also expect reduced speech to be relevant for articulatory theories of speech perception (e.g. the Motor Theory and Direct Realism), although this topic has not been well developed yet. Overall, reduced speech is relevant to testing any theory of speech or word perception, because any cues present in it differ so radically from the kinds of perception stimuli that are typically studied.

A few overall findings from spontaneous speech research methods can be summarized. It is clear that individual words are realized with a wide variety of forms (Greenberg 1999), and that listeners can recognize these forms well in context, but at best poorly out of context (Arai 1999; Ernestus et al. 2002). Listeners recognize unreduced forms more easily than reduced forms (Ernestus et al. 2002; Ranbom and Connine 2007; Tucker 2007), even if the reduced form is more common. One thing we have certainly learned from spontaneous speech is that the real speech we all produce and process every day is far, far more variable than one would ever expect based on more controlled methods. Furthermore, we have learned that however listeners perceive speech and recognize words, they must be able to handle far more variability than most theories address.

Some methods include critical theoretical assumptions, and may not be useful for testing anything if those assumptions are not true. The methods for working on reduced speech make only one minimal assumption about theory: that

variability well below the level of the phoneme is interesting. If reduced speech, fast speech, casual speech reduction, etc. are all relegated to *phonetic implementation* and considered external to the grammar, and the grammar is the topic of study, then by definition reduced speech is of no interest. However, if any aspect of reduction is language-specific rather than caused by universal biological constraints on articulation, then speakers would need to know the language-specific aspects of reduction as part of the grammar (cf. Keating 1985, 1990a and Kingston and Diehl 1994 on language-specific phonetic detail in general, regardless of reduction, and Barry and Andreeva 2001 on cross-linguistic patterns of reduction). Many recent phonological theories extend the realm of interest to include low-level gradient variability (Coetzee, this volume).

Another part of how methods relate to theory is that some may feel it is better science to test theories on controlled data, rather than on spontaneous data. When investigating a new topic, about which little is understood (e.g. intonation in a language for which it has never been studied), one should probably begin with controlled, stable data, such as matched target items in frame sentences. However, when studying a topic with extensive past literature, the field may be ready to move to data that is more representative of daily-life speech. Warner and Arai (2001) argue this for a study of Japanese mora-rhythm using spontaneous speech.

21.2.6 Conclusions

There are problems with all methods of obtaining and analyzing spontaneous speech and stimuli. Researchers are exploring a wide variety of methods. While this may seem chaotic, it is exciting. As large speech corpora have appeared, spontaneous speech research has increased rapidly. For example, LabPhon 10 (Paris, 2006) and ICPhS 2007 (Saarbrücken) both had a proliferation of papers using large speech corpora or investigating speech style. 2008 saw the First Nijmegen Speech Reduction Workshop (program at <http://www.u.arizona.edu/~nwarner/>).

Perception studies on reduction have lagged behind production studies, perhaps because of the methodological challenge of obtaining stimuli, but are now leading to fascinating studies. Moving beyond the core areas of native adult production and perception, there has been only the most tentative exploration into the relationship of spontaneous speech to L1 or L2 acquisition (Bleses 2008; Shockey 2008), cross-linguistic and cross-dialectal language use, or disordered speech (dysarthria, Mattys and Liss 2008). We can expect development into these areas soon. Returning to theory, current phonetic theories (and even more so formal phonological ones) have only begun to develop mechanisms for modeling massive reduction phenomena. We can expect, or work toward, an impact of spontaneous speech on many theories in upcoming years.

21.3 Methods and experimental design for studying sociophonetic variation

Paul Warren and Jennifer Hay

21.3.1 Introduction

As laboratory phonology has emerged as an approach to the study of the sound systems of language, it has had considerable impact on the methods and experimental design used in the study of sociophonetic variation. The influence of laboratory phonology has been not so much a case of bringing sociolinguistics into the laboratory, as an example of how laboratory techniques are taken out into a research area that has a traditional base in field data collection. The collection of data in the field remains of course part of our endeavor, and for many researchers the field is the laboratory. However, the types of analysis of the data afforded by developments in the laboratory, including the statistical laboratory as well as the phonetics laboratory, have had a noticeable impact on the nature of the questions asked and the answers sought. In this section we highlight a number of areas where we believe laboratory phonology approaches have left their indelible mark:

- the nature of the research questions now being asked;
- the combination of controlled laboratory-style recordings with recordings of unprompted utterances;
- the use of a combination of impressionistic and instrumental analyses of data;
- the move from categorical analysis of the dependent speech data to a more continuous analysis;
- the extension of the independent variable set beyond the demographic variables typically covered, i.e. region, age, sex, and class, and the inclusion of a greater range of scalar rather than categorical variables in that set;
- an awareness of the participant as both a speaker and also a listener, and of the possible influences that perception and production may have on one another;
- the impact that the experimenter may have on the nature of the speech data being produced by participants.

21.3.2 Research questions

Phonetic variation carries social meaning (see also Docherty and Mendoza-Denton, this volume). The types of social meaning it can carry, and the exact nature of socially meaningful phonetic variation, are questions that have long interested sociolinguists. The increasing use of laboratory phonology techniques to study such

variation is providing new insight into the very fine details of phonetic implementation that can do social work. In addition, laboratory phonologists are bringing their own questions to this type of data. How is knowledge about sociophonetic detail acquired, stored, and accessed? How does it lead to further generalizations about language? How can our models of speech perception and production account for the full range of sociophonetic data, which illustrate the detailed interconnectedness of linguistic and social meaning in speech? These questions are accompanied by an increase in perceptual studies investigating the degree to which people are aware of and use sociophonetic detail in speech processing.

21.3.3 Types of data

Historically, studies of sociophonetic variation have been based on field recordings. Such studies have produced a wealth of informative data on variation. Consider for example Labov's pioneering work (Labov 1966, 1972a), and the discussion of sociolinguistic fieldwork in Feagin (2002). Of course, individuals produce a wide range of variation, from very careful styles (which typify many laboratory recordings) to the most informal. Sociolinguists have long assumed that it is in the more informal styles that vernacular, or diachronically advanced, features are most often heard (Labov 1972a), and that field recordings are the most likely source for the vernacular. Sociolinguists now realize that vernacular features often appear in highly self-conscious dialect performances (Schilling-Estes 1998). However, most kinds of laboratory speech would probably not fall under the domain of such stylized performances.

A major difficulty with a dependency on spontaneous speech is that it can constrain the range of linguistically interesting phenomena that can be studied, and can make systematic comparison difficult. Take for example the prosodic marking of syntactic ambiguities, such as the ambiguity involving the attachment of the prepositional phrase (PP) as a modifier of *saw* or *cop* in (1).

(1) John saw the cop with the binoculars

Since different words have different durations, and since the weight of a constituent such as the PP in (1) affects prosodic phrasing (Gee and Grosjean 1983), it is important for researchers studying such phenomena to control the lexical content of such utterances. This makes it vanishingly unlikely that suitable utterance pairs will be found in spontaneous speech contexts and produced by the same speakers. Laboratory recordings are therefore a vital component of such research. These recordings have shown that speech properties such as the relative likelihood and duration of a pause and of pre-pausal lengthening immediately prior to the PP are used to distinguish the two interpretations of sentences like (1) (Cooper and Paccia-Cooper 1980).

In addition, some phonological variables of interest are so rare that one could record a speaker for quite a long period of time and not have them produce a relevant environment for the phenomenon. Hay and Sudbury (2005), for example, observe only 198 possible environments for intrusive /r/ in the same amount of speech which was able to reap 13,760 tokens of non-pre-vocalic /r/ variable. There are practical limitations pointing towards the usefulness of laboratory speech for the study of intrusive /r/ (see e.g. Hay and Maclagan 2010).

In most early studies, laboratory speech meant read speech. Yet, since a reader's goals are quite different from those of someone speaking spontaneously (Schafer et al. 2005), there is every reason to believe that the results of studying read speech might not generalize to spontaneous contexts. For example, the analysis of careful laboratory speech indicated that there is a cross-linguistic tendency toward intrinsic vowel pitch, i.e. that f0 correlates with vowel height (e.g. Lehiste 1970). Subsequent analyses showed however that this correlation is suppressed in spontaneous speech (Ladd and Silverman 1984).

In an attempt to overcome this problem while preserving constraints on the content and structure of utterances, methodologies have been developed that include semi-structured game tasks (see also Warner, this chapter, and Post and Nolan, this volume). In one such task (Schafer et al. 2000; Schafer et al. 2005), participants use sentence frames, into which they have to insert object names, and use the resulting utterances to negotiate the move of objects around a gameboard. They use expressions such as that in (2), also involving a PP-attachment ambiguity. (In the game, participants can use a triangle to push a square around the board, or they can move a combined square-with-triangle piece.)

(2) I want to change the position of the square with the triangle

In this task, participants rapidly become familiar with the sentence frames and object names, and produce the commands required for the task with fluency. One of the sociophonetic goals of the project is to compare the use of prosodic features and connected speech processes by speakers of different English dialects. One such connected speech process is *wanna*-contraction, illustrated by the utterances in (3). Research indicates that *wanna*-contraction is much less likely when there is a gap site, or trace of the fronted element *triangle*, between *want* and *to*, as in (3a) than when there is not, as in (3b) (Lakoff 1970; Baker and Brame 1972).

(3) a. Which triangle$_i$ do you want $\_\__i$ to change the position of the square?
b. Which triangle$_i$ do you want to change the position of $\_\__i$ this time?

Using the gameboard task, examples of (3a) and (3b) were collected from groups of Midwestern US English and New Zealand English speakers. The data confirmed that *wanna*-contraction is more likely in (3b) than in (3a), for both varieties, but also that it is over twice as likely in speech from US English speakers than from NZ English speakers (Warren et al. 2003). The use of data from the same gameboard

task for the two dialects gives the researchers confidence that this dialect difference is not due to uncontrolled differences in speech materials.

Related elicitation techniques include the map task, in which one participant needs to guide another around a path on a map, each using maps with overlapping but non-identical landmarks marked (Anderson et al. 1991); and diapix "spot the difference" tasks, where each participant has a slightly different picture and the aim is to figure out the differences between them (Bradlow, Baker, Choi, Kim, and Van Engen 2007).

While there are a range of techniques for eliciting data in the laboratory, there is also considerable work underway conducting ethnographic studies of speech communities, and some of this work is certainly using laboratory phonology analysis techniques and addressing questions central to the laboratory phonology community. Drager (2009, 2010), for example, conducted a year-long ethnographic study in a New Zealand high school. Using the speech data gathered at the school she was able to conduct an acoustically sophisticated analysis of socially and linguistically driven variation in the word *like*. The ethnographic work reveals different social orientations held by groups at the school, and demonstrates how these are revealed through subtle phonetic variation. She also ran speech perception experiments in the school to assess the degree to which individuals were sensitive to the patterns of variation uncovered by the acoustic analysis. For example, girls at the school use different phonetic patterns to distinguish between quotative 'like' (as in e.g. *she was like "oh no"*) and discourse particle 'like' (e.g. *she was like falling over laughing*). In perception tasks, using natural stimuli drawn from recorded conversations, Drager shows that the girls are sensitive to these phonetic cues. She draws conclusions relating to questions of representation, production, and perception—thus addressing questions core to the laboratory phonology enterprise. Hay and Drager (2007) argue that analysis of language use which is both phonetically and socially sophisticated is required to make proper headway into understanding the link between phonetic variation and social meaning. They also argue that such work should be accompanied by perception experiments, so that we might begin to be able to understand the various steps in the production-perception loop.

Indeed, there has been a recent increase in speech perception experiments investigating individuals' use of sociophonetic information. Examples of techniques used include speaker or group identification experiments (Clopper and Pisoni 2007; Drager 2009), in which participants are asked to use phonetic cues to identify a speaker's regional or social background; in some cases including classification and word identification in noise (Clopper and Bradlow 2008). Some work has also investigated changes in listener attitudes in response to manipulations of the acoustic signal (Campbell-Kibler 2006). Work using categorical perception (e.g. Strand 1999), has demonstrated how listener beliefs about the social identity of the speaker can affect perceived phonemic boundaries. And work using forced choice identification tasks (Hay, Warren, and Drager 2006), has shown that social beliefs

about a speaker can affect listener accuracy in distinguishing between vowel sounds undergoing merger.

21.3.4 Types of analysis

A significant area in which sociolinguistic methodologies have been affected by laboratory techniques has been in the types of analyses made of speech data. This is true both of data collected in the laboratory, and of that collected in the field. First, there has been an increasing use of a combination of impressionistic and instrumental analyses. Second, the instrumental analyses have allowed researchers to move away from categorical investigation of speech phenomena to a more continuous analysis.

A good example of the combination of analysis types is provided by the trend in prosodic analysis over the past twenty years to combine auditory prosodic transcription with acoustic measures such as pitch tracks and amplitude envelopes. This is particularly apparent in the ToBI (Tones and Break Indices) tradition (Beckman et al. 2005), a system set up initially for English, and adapted for an increasing range of languages (see papers in Jun 2005). Furthermore, in the prosodic analysis of sentence ambiguity mentioned above, as well as in studies of dialectal variation in prosody, researchers have combined ToBI-style transcription—itself already based on a mix of impressionistic and acoustic data—with further detail from the acoustic analysis (Warren 2005). This approach has been particularly useful in the analysis of intonational systems of varieties of a language for which it is unclear whether the phonological categories of related varieties are relevant, but also in the description of differences in the realization of specific categories.

There is a reasonable history of using acoustic analysis in sociolinguistics, at least with respect to vowels. Labov et al. (1972), for example, measured F1 and F2 of vowels, and this practice has become relatively widespread in the field (see e.g. review of papers in Thomas 2002b). An ongoing problem with respect to conducting statistics on formant values is that they are strongly affected by the length of the vocal tract, and this has led to a large literature on techniques for normalization, none of which is entirely satisfactory. As Thomas (2002b: 174) expresses it "all normalization techniques have drawbacks; choosing one is a matter of deciding which drawbacks are tolerable for the study at hand." This is one area in which recent advances in statistics may make analysis of traditional data-sets much easier. Mixed-effects models include random effects for individuals—allowing each individual in a data-set to vary randomly with respect to the factor being modeled (see also Baayen, this volume; Kingston, this volume). One can then test whether there are overall effects (e.g. linguistic effects or effects of speaker groups) which exist over and above the variation across individual speakers. Modern statistical techniques, then, are likely to mediate the normalization problem, which has plagued the field for a long time.

More recent acoustic work on vowels has also looked past F1 and F2 to include other measures, such as duration (Wassink 2001), diphthongization (Maclagan and Hay 2007), voice quality (Di Paolo and Faber 1990), and measures of formant curvature (Harrington 2006).

While there is a reasonable tradition of acoustic analysis of vowels (at least with respect to F1/F2 space), the tradition within the sociolinguistic literature is to treat consonants as more categorical, and to rely much more heavily on auditory analysis. Only recently has detailed acoustic work been conducted, with attempts to model consonantal variation as more gradient. Examples include the detailed work on /t/ conducted by Docherty and Foulkes (1999, 2005), and the analysis of intrusive /r/ conducted by Hay and Maclagan (2010). Hay and Maclagan show that there is variation across speakers in the degree of constriction involved in intrusive /r/, and that this patterns the same way across speakers as does a categorical analysis of whether the /r/ is present or not. That is, speakers who are more likely to produce /r/ in a word like *clawing*, are also more likely to have a greater constriction (i.e. lower F3) when they do produce it.

In addition to more sophisticated acoustic analysis, there is also a move toward using articulatory techniques such as ultrasound on sociolinguistic data. Lawson, Stuart-Smith, and Scobbie (2008; see also Scobbie and Stuart-Smith, this chapter), for example, show that some Scottish speakers who appear to be non-rhotic actually produce a covert articulation of a word-final /r/.

The study of sociophonetic variation has also seen a number of changes in the independent variable set that is typically investigated. Under the laboratory phonology approach, further variables have been added to the traditional demographic variables of region, age, sex, and class, and the variables have also been looked at in a less categorical manner as new statistical approaches have been adopted. Particularly notable amongst the additional independent variables are item-related variables such as lexical frequency, the phonetic environments in which a speech sound of interest is being uttered, and other similar factors that have long been included in psycholinguistic analyses of speech production and perception, but less so in phonological studies (though of course sociolinguistic variation studies have long been sensitive to the importance of phonetic environments, e.g. Labov 1963).

The inclusion of additional variables has both necessitated and been facilitated by different statistical approaches to data analysis. Studies of speech perception have used traditional tools such as t-tests, chi-square tests, or ANOVAs. The most common statistical analysis technique in the study of language variation and change is VARBRUL (e.g. Sankoff et al. 2005). This is effectively a logistic regression program. As such, it models binary variables. The widespread use of this software, then, has no doubt helped to encourage a categorical view of many phonological variables, as the statistical technique is just not appropriate for testing continuous measures. The VARBRUL program also requires that independent variables are categorical rather than continuous, which has limited investigations of things such as word frequency.

Many researchers have recently explored the impact and interactions of a greater range of predictor variables, using linear and logistic modeling (outside of VARBRUL) and other regression-type analysis. The use of linear modeling enables investigation of more continuous dependent variables, and the move away from VARBRUL enables investigations of continuous independent factors such as age, social class, lexical frequency, various production measurements, etc. (see, e.g. Hay et al. 2006).

Recent statistical developments offer much more appropriate statistical techniques for language data, and especially sociolinguistic data. Many authors have recently advocated a shift toward mixed-effects modeling (Baayen et al. 2008; Jaeger 2008; Quené and van den Bergh 2008; D. E. Johnson 2009; see Baayen, this volume). This enables the analyst to include one or more "random" effects in their model. A random effect for participant, for example, would give each participant their own coefficient, allowing them to vary randomly with respect to one another. The effect of this is that no individual participant can dominate the significance of any reported effect. Drager (2009), for example, uses mixed-effects modeling in her analysis, enabling her to investigate overall social factors affecting her dataset, while taking into account the variation shown by individuals. In her data, this modeling has the added advantage that the random effects assigned to individuals also reveal some interesting and qualitatively interpretable patterns themselves.

D.E. Johnson (2009) has implemented Rbrul—a tool that provides a VARBRUL-like interface to the mixed-effects modeling functions in R.

21.3.5 Links between production, perception, and context

Work on speech perception is acknowledging the role of the participant as both a speaker and a listener. For instance, recent statistical modeling of the perception of sociophonetic differences includes information about individual participants' speech production. A number of recent approaches in sociophonetics, including in particular those oriented towards experience-based models of speech processing such as exemplar theory (Johnson 1997b, 2007; Pierrehumbert 2001, 2003), examine explicitly the relationship between production and perception. A variety of studies show a link between an individual's production and their perception (Drager 2006; Hay et al. 2006; Harrington et al. 2008), indicating that many speech perception experiments would probably benefit from also collecting production data from participants.

Participants' own production patterns can not only influence their perception patterns, but the extent of this influence can differ across different contexts. For example, Hay et al. (2006) found that the extent of the NEAR-SQUARE merger in New Zealanders' own speech influenced their forced-choice identification of words containing one of these vowels. Unsurprisingly, the less the participants merged the

contrast in their own speech, the more accurately they performed on the task. However, we also found a significant interaction between degree of distinction produced by participants and the dialect of the experimenter who met them. Participants who merge the NEAR-SQUARE vowels in their own speech make more errors in an identification task when they have interacted with a US experimenter. In further work (Hay et al. 2010), we also found that performance in an Odd One Out task (e.g. selecting, from the written forms, which of *beer*, *bare*, *bear* sounds different from the others) was significantly affected by an interaction between the dialect in which the instruction set was recorded and the extent of merger in the participants' own speech. For participants who maintained a degree of distinction in their production of NEAR and SQUARE vowels, prior exposure to a British English speaker rather than to a New Zealand English speaker significantly decreased error rates in the perceptual task.

A long-standing result from the sociolinguistic literature is that speakers have a tendency to converge in the phonetic detail of their productions—a phenomenon known as speech accommodation. This tendency is particularly strong if the speaker positively identifies with the addressee. Recent laboratory work has demonstrated that there are actually very automatic effects of "convergence" (Delvaux and Soquet 2007). These don't involve orientation to a particular addressee; they exist in both production and perception, and they can carry over past the time at which an individual is actually interacting with, or exposed to, the speech of the person to whom they are converging. As mentioned above, we discovered more or less by chance that the identity of the experimenter was influencing participants' performance in production and perception tasks involving the NEAR-SQUARE merger in New Zealand English (Hay et al. 2009; Hay et al. 2010). Subsequently, we started a systematic research program to investigate such effects, and have found that the trigger for "convergence" need not be the speech of an actual person, but can even be some external stimulus associated with a particular accent, such as the presence of a stuffed toy (a kiwi, invoking New Zealand, or a koala, invoking Australia; Hay and Drager 2010).

These apparently automatic effects by no means rule out a certain amount of intentionality in style-shifting, but it requires more work to distinguish which observed effects are automatic, and which are more under the control of the speaker/listener. The fact that perception and production seem to operate to some degree in parallel suggests they should be studied together more, and unified models should be explored. The methodological consequences are also non-trivial since "laboratory" culture often involves different participants meeting with different experimenters. As much as possible we need to control environmental effects such as the identity and dress of the experimenter, the nature of the pre-task interaction, the location of the experiment, and tasks conducted prior to the experiment, in order to set the stage in a controlled way to gather more "naturalistic" speech.

21.3.6 Summary

In this contribution to the handbook we hope to have indicated that the impact of laboratory phonology on sociophonetic research has been felt not only in the range of methodologies now used in this field of endeavor, but also in the types of questions that researchers feel confident that they can ask, and in the level of detailed analysis that can be carried out. Rather than simply moving sociophonetic research from the field into the laboratory, what laboratory phonology can do and is doing for sociophonetics is to enrich the research area by adding to the sociophonetician's toolkit. This toolkit now includes more sophisticated statistical approaches, which importantly allow analyses to proceed without losing detail of the data through processes such as averaging over participants or language items while also allowing the introduction of a range of potential explanatory variables in the analysis of data distributions. These approaches are just as useful for field data as they are for laboratory data. The toolkit also includes experimental approaches that attempt to make laboratory data more "real" and to heighten the ecological validity of results obtained under experimental conditions. The development of such techniques is a clear example of the benefit of interaction between fieldwork and laboratory approaches, with advantages to both sides. The toolkit also includes a more expansive mindset, one that acknowledges that there are countless factors that might influence the way we speak or the types of interpretation we give to what we hear, and one that tries to make the connections between language users as producers, comprehenders, and social beings.

CHAPTER 22

STATISTICAL ANALYSES

STATISTICS IN LABORATORY PHONOLOGY

JOHN KINGSTON

MIXED-EFFECTS MODELS

HARALD BAAYEN

CLUSTERING AND CLASSIFICATION METHODS

CYNTHIA G. CLOPPER

The contributions in this chapter review statistical techniques appropriate for speech research. Kingston presents two in-depth case studies, discussing graphical data exploration and analysis using linear regression models. Baayen discusses the principles and applications of mixed-effects models, including consideration of continuous, binary, and count data. Clopper discusses clustering, multidimensional scaling, and factor analysis.

22.1 Statistical methods in laboratory phonology

John Kingston

22.1.1 Introduction

Statistics is one of the methods that constitute laboratory phonology. In this section, I use them to tell a clear story about two exemplary data sets. In one, the dependent variable is continuous, while in the other it is an ordinal categorical variable. Both kinds of dependent variables are commonly produced by phonological experiments; a third common kind is categorical choices between two alternatives, which Baayen (this chapter) deals with in his section. The analyses all begin with graphical exploration of the data, which is followed up by constructing linear regression models. These models are more informative than analyses of variance about how the kinds of independent variables commonly used in phonological experiments influence the dependent variable. Mixed-effects models are developed for the ordinal categorical variable to demonstrate how random effects of participants and items are accommodated.[1]

22.1.2 A linear model of Finnish vowel durations

22.1.2.1 *Graphical explorations*

The data that demonstrate the analysis of a continuous dependent variable were generously provided by Scott Myers, who with Benjamin Hansen used them to develop a phonetic explanation for the cross-linguistically common process of final vowel shortening (Myers and Hansen 2006, 2007). They consist of the total durations of Finnish vowels, as well as the durations of their voiced and voiceless portions. The potential independent variables are the phonological quantity of the vowel (short versus long), the type of syllable it occurred in (open: V, CV, or GV and closed CVN), and whether the syllable containing the vowel was word-final. A fourth independent variable was the duration of the preceding word *sanoin* 'I

[1] All the analyses presented in this section were carried out in R (R Development Core Team, 2010). Besides the base package, the principal packages used in this section are languageR (Baayen 2009), lattice (Sarkar 2010), ordinal (Christensen 2010), and lme4 (Bates and Maechler, 2010). For more comprehensive applications of R to linguistic data, see Baayen (2008) and Johnson (2008); Dalgaard (2002), Maindonald and Braun (2003), Gelman and Hill (2007), and Everitt and Hothorn (2010) are also very useful introductions.

said,' which may index speaking rate. Four speakers each produced twelve tokens in different words of the sixteen kinds of syllables.

Figure 22.1.1 displays the total durations of the short and long vowels in histograms, density plots superimposed on the histograms, and box plots.

These displays show that most long vowels last longer than most short ones, despite some overlap in their respective ranges. Two modes are visible in the distributions of both short and long vowels' durations for speakers 3 and 4, and the distributions for speakers 1 and 2 also have obvious shoulders on their upper tails. This structure raises the suspicion that the vowels' durations may be determined by another factor than their phonological quantity. The breakdown of the density plots in Figure 22.1.2 by whether the syllable containing the vowel is final or non-final confirms this suspicion. Final and non-final distributions overlap for both short and long vowels, but their distributions remain largely distinct from one another within each quantity. The lack of overlap between the notched intervals in the

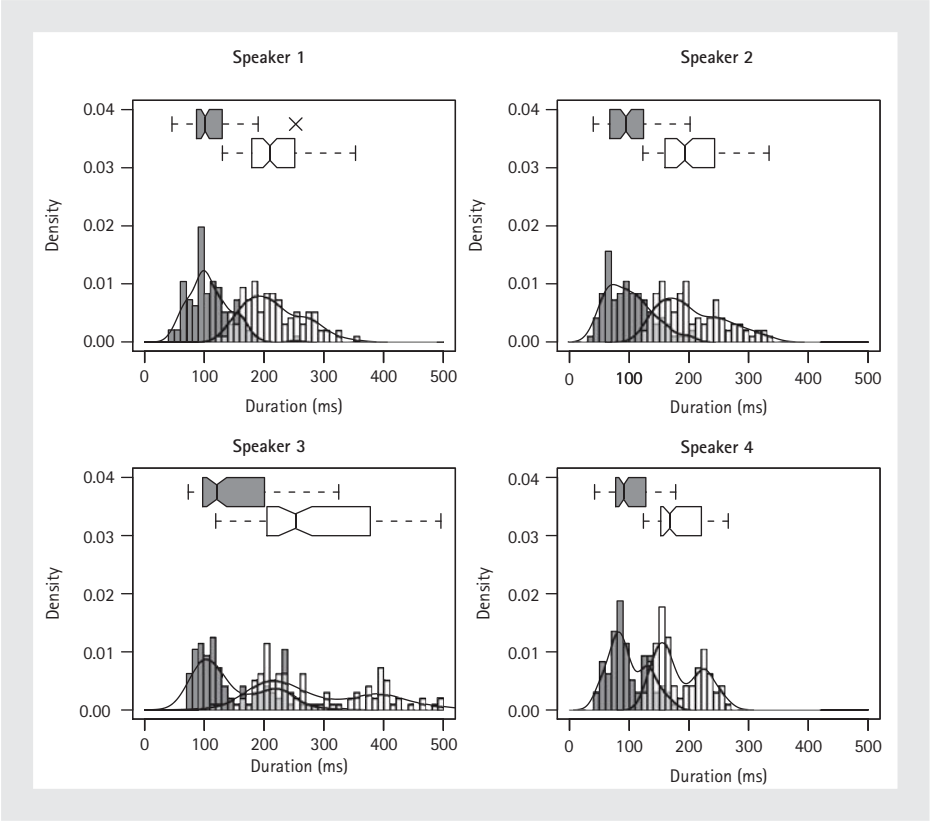

Figure 22.1.1. Histograms, density plots, and box plots of short (dark gray) and long (white) vowel durations produced by four Finnish speakers. The lighter gray bars represent durations common to short and long vowels. The Xs are outliers.

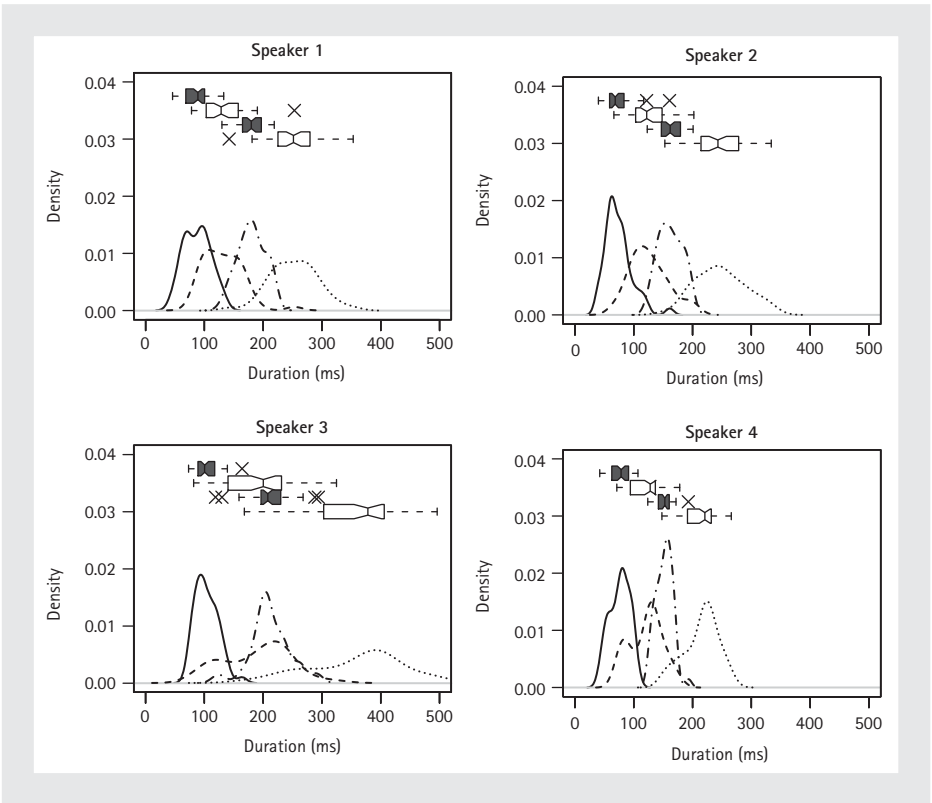

Figure 22.1.2. Density plots for short (solid, dashed) and long (dash-dot, dotted) vowels in final (dashed, dotted) and non-final (solid, dash-dot) position for the four Finnish speakers. Gray box plots represent durations in non-final position, white ones those in final position. The Xs are outliers.

box plots also indicates that the final and non-final distributions differ from one another. The density plots in this figure hint at the presence of yet further structure in the data, perhaps an influence of syllable type.

Evidence of such structure can be seen in Figure 22.1.3, where each panel plots the vowel duration for a particular speaker, vowel quantity, and position against the duration of the word *sanoin*—this figure uses the xyplot() function from the lattice package (Sarkar 2010). The durations of most vowels in open CV, GV, and V ("1–3") are longer than those in closed CVN ("X") syllables in final position for both long and short vowels, except for the short vowels produced by speaker 1 (bottom row), whose vowel durations in open and closed syllables overlap considerably. In non-final position, the distributions of vowel durations in closed and open syllables overlap for all four speakers.

The graphical exploration in Figures 22.1.1–22.1.3 has shown that besides the expected greater duration of phonologically long versus short vowels, vowels in

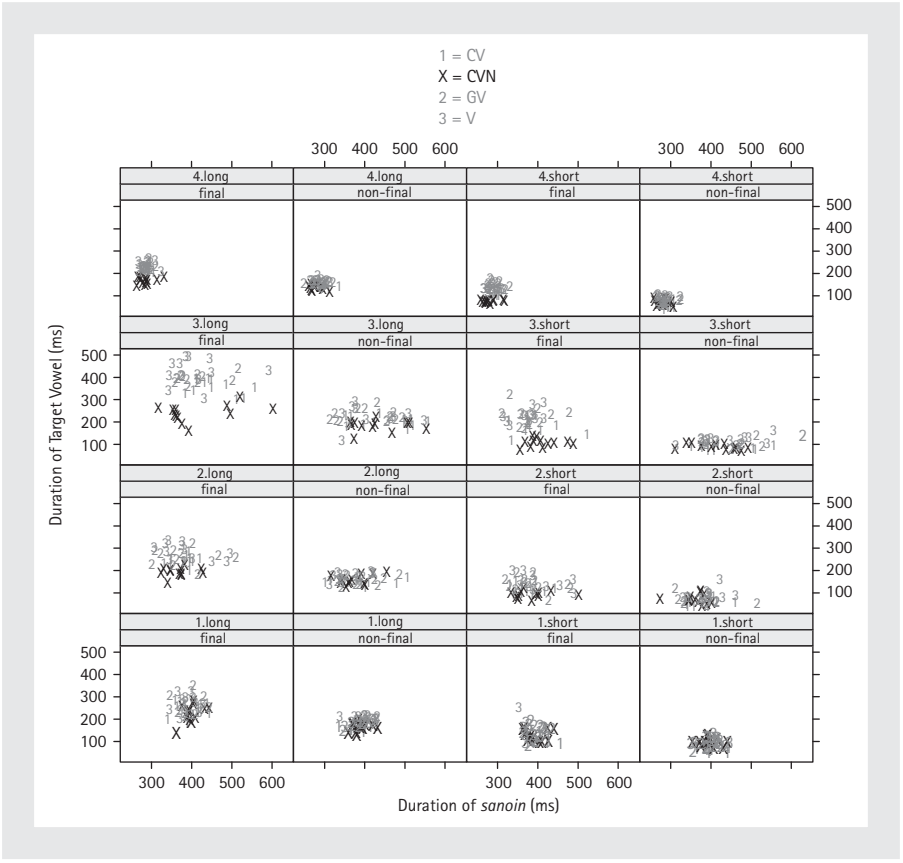

Figure 22.1.3. Vowel durations (vertical axis) by duration of the word *sanoin* in the same utterance by speaker, phonological quantity, position, and syllable type. Black "X" for closed CVN syllables, and gray "1–3" for open CV, GV, and V syllables, respectively.

final position are longer than non-final vowels, and vowels are longer in open than closed syllables when they occur in final position.

22.1.2.2 *Residuals and transforms of the dependent variable*

Unlike the duration of the word *sanoin*, the other independent variables are categorical rather than continuous. This characteristic might prompt submitting these data to an analysis of variance instead of linear regression. The reason for not doing so is that we want to know not only whether any of these independent variables significantly affects vowel durations but also the direction and size of that effect. We may even have hypotheses about the direction of these effects that we would like to test.

To examine deviations of the dependent variable's values from the predicted value, i.e. the residuals, linear regression models were first constructed with just one independent variable at a time. This examination reveals whether the dependent variable needs to be transformed before undergoing further analysis.

The three panels in each row of Figure 22.1.4 display the residuals against the fitted values (left), a QQ-plot of the residuals against the values expected if the residuals were normally distributed (middle), and the Cook's distance values for the residuals (right). Each vertical cluster of residuals in the panels on the left represents a combination of a value for the independent variable and speaker. Ideally, the vertical distances of the residual values from 0 would not vary as a function of the fitted values, but in all three panels we see that they spread out as fitted values increase. This outcome indicates the need to transform the data. The QQ-plots in the middle panels show that the positive residuals have more extreme values than expected,

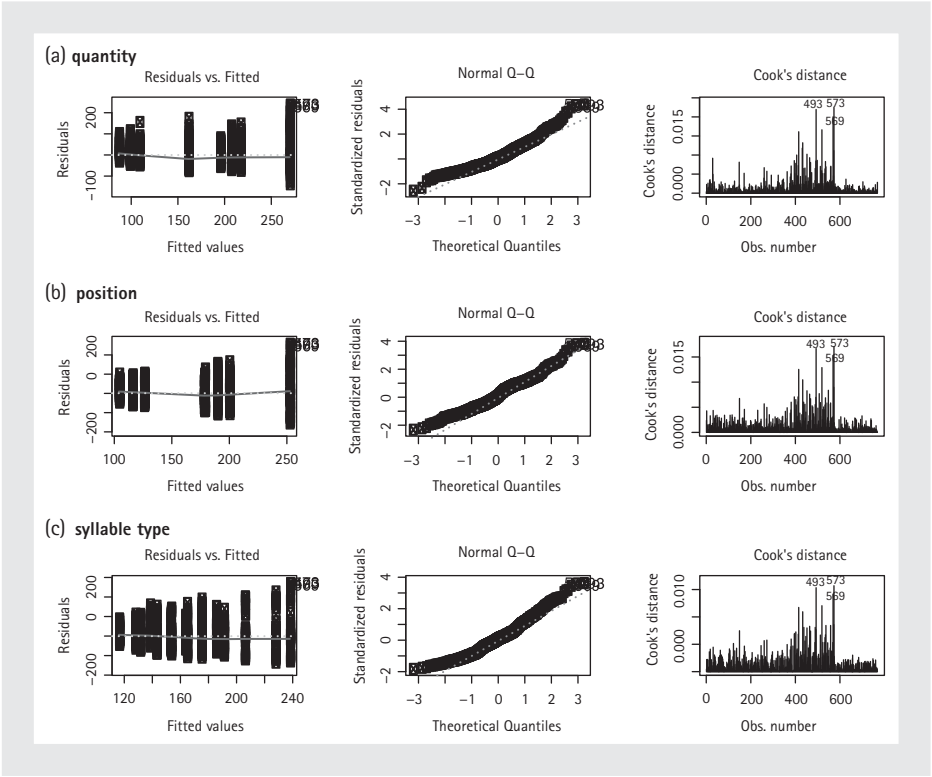

Figure 22.1.4. Residuals by fitted values (left), QQ-plots of standardized residuals by values expected if the residuals were normally distributed (middle), and Cook's distances (right) for the models in which the independent variables are (a) quantity and speaker, (b) position and speaker, and (c) syllable type and speaker.

while the values of the negative residuals are less extreme than expected. This is not surprising given the fanning outward observable in the left-hand panels. Finally, the Cook's distance values in the right-hand panels identify data points that exert "leverage" on the regression line. When data points exert considerable leverage, the analysis should be rerun omitting those data points to determine whether the independent variable's apparent influences depend on just those points. Values of 1 or more indicate influential leverage. Here, the most extreme values are 1–2 orders of magnitude smaller than 1, so none exerts particularly strong leverage, and none should be omitted.

The choice of power transform (1) is determined by the tails of the dependent variable's distribution: a long upper tail (most outliers have high values) indicates an exponent smaller than 1 (λ)—in the limit, the exponent is 0 and the transform is the log transform—but a long lower tail indicates an exponent larger than 1.

$$(1) \quad T(x) = \frac{x^{\lambda} - 1}{\lambda}$$

Figures 22.1.1 and 22.1.2 show that the distributions' upper tails are stretched out, motivating a log transform. The left-hand panels in Figure 22.1.5 show that this transformation successfully eliminates the residual values' fanning outward as the fitted values increase, but the QQ plots in the middle panels show that the residuals' distribution now differs more from normality than before the transformation. The deviation is also different: both negative and positive residuals are less extreme than if they were normally distributed. These deviations are also much greater for the models of position (b) and syllable type (c) than quantity (a). This complementarity between the deviations in the models of position and syllable type, on the one hand, and quantity, on the other, motivates including quantity as a predictor.

22.1.2.3 *Multiple linear regression models, with and without interactions*

We begin with a model that includes all three independent variables as well as the log-transformed duration of the word *sanoin*, a main effects model (Table 22.1.1), and follow up by adding an interaction between position and syllable type (Table 22.1.2). The dependent variable is the log-transformed vowel durations.

The vowel durations do not depend significantly on the duration of *sanoin* ($t = -0.622$, $p = 0.534$; recall Figure 22.1.3), the vowel is significantly shorter when it is phonologically short ($t = -49.826$, $p < 2e-16$) or non-final ($t = -31.735$, $p < 2e-16$), but significantly longer in a GV ($t = 7.513$, $p = 1.64e-13$) or V ($t = 11.322$, $p < 2e-16$) syllable compared to a CV syllable. It is marginally shorter ($t = -1.895$, $p = 0.0585$) in a CVN syllable. To transform the predicted vowel durations back into ms, the model serves as the exponent of the base *e*

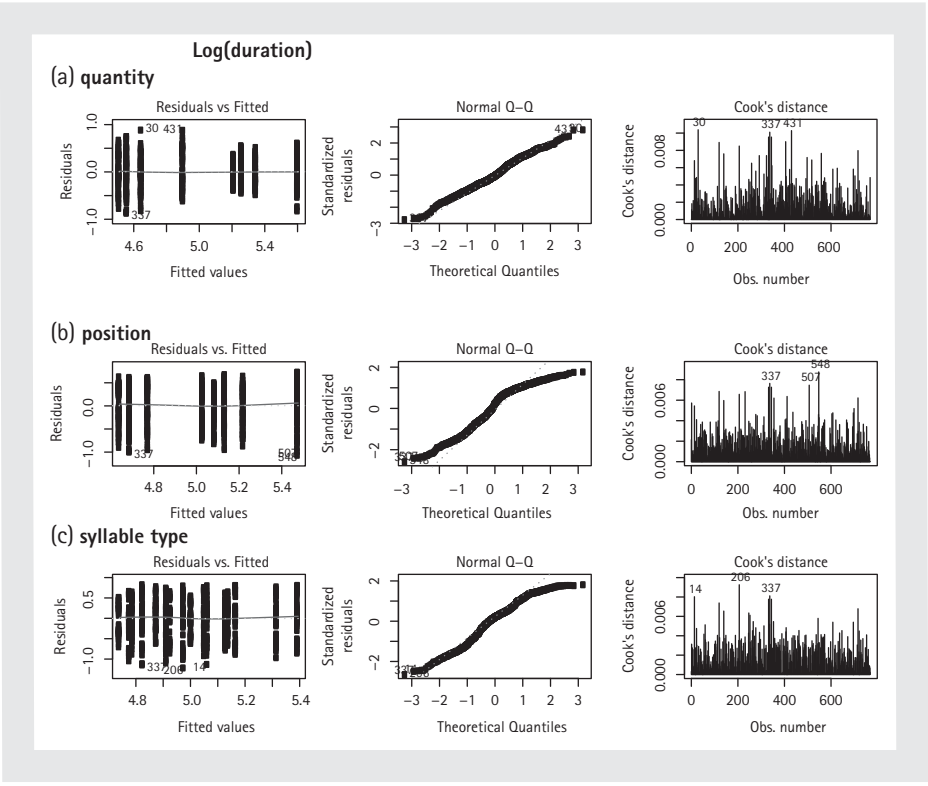

Figure 22.1.5. For log-transformed durations, residuals by fitted values (left), QQ-plots of standardized residuals by values expected if the residuals were normally distributed (middle), and Cook's distances (right) for the models in which the independent variables are (a) quantity and speaker, (b) position and speaker, and (c) syllable type and speaker.

Table 22.1.1. Predictor estimates in a linear regression model of log-transformed Finnish vowel durations in which the log-transformed duration of *sanoin*, quantity, position, syllable type, and speaker are independent variables

Predictor	Estimate	Std. Error	*t* value	Pr(> |*t*|)
(Intercept)	5.74149	0.40949	14.021	< 2e-16
log(sanoin)	−0.04253	0.06835	−0.622	0.5340
short	−0.69988	0.01405	−49.826	< 2e-16
non-final	−0.44621	0.01406	−31.735	< 2e-16
CVN	−0.03764	0.01987	−1.895	0.0585
GV	0.14927	0.01987	7.513	1.64e-13
V	0.22529	0.01990	11.322	< 2e-16
Speaker 2	−0.08865	0.02004	−4.424	1.11e-05
Speaker 3	0.25689	0.02016	12.744	< 2e-16
Speaker 4	−0.14900	0.02924	−5.096	4.37e-07

Table 22.1.2. Predictor estimates in a linear regression mode of Finnish vowel durations in which the log-transformed duration of *sanoin*, quantity, position, syllable type, speaker, and the interaction between position and syllable type are independent variables

Predictor	Estimate	Std. Error	*t* value	Pr(> |*t*|)
(Intercept)	5.62074	0.38072	14.763	< 2e-16
log(sanoin)	−0.01143	0.06352	−0.180	0.857309
short	−0.69974	0.01304	−53.674	< 2e-16
non-final	−0.57721	0.02608	−22.135	< 2e-16
CVN	−0.22225	0.02607	−8.526	< 2e-16
GV	0.13628	0.02608	5.225	2.26e-07
V	0.16261	0.02610	6.230	7.75e-10
Speaker 2	−0.08744	0.01860	−4.702	3.06e-06
Speaker 3	0.25532	0.01871	13.646	< 2e-16
Speaker 4	−0.13924	0.02715	−5.128	3.73e-07
non-final:CVN	0.36968	0.03688	10.024	< 2e-16
non-final:GV	0.02655	0.03687	0.720	0.471789
non-final:V	0.12648	0.03687	3.431	0.000635

(or practically the argument of the exp() function in R). (2) shows the non-zero terms in the model and the predicted duration of a short, non-final vowel in a CVN syllable:

(2) 95.4 ms = exp(5.74149 + −0.69988 + −0.44621 + −0.03764)

This is the predicted duration for Speaker 1; to predict the duration for such a vowel for Speaker 3, one would add 0.25689 to the exponent.

The interactions are represented in Table 22.1.2 as "non-final:CVN" etc., which indicates that they are the increment or decrement in duration predicted when the vowel's position is non-final and its syllable type is CVN etc., as compared to when its position is final and/or CV.

(3) shows the duration of a short, non-final vowel in a CVN syllable predicted by this model:

(3) 89.2 ms = exp(5.62074 + −0.69974 + 0.57721 + −0.22225 + 0.36968)

22.1.2.4 *Predicting novel values*

Models are also used to predict values of the dependent variable for new cases. For this model, one cannot of course use the model to predict what duration a vowel will have if it does not belong to any of the categories defined by the independent variables other than the duration of the word *sanoin*. One can nonetheless still

estimate how precisely a new token's duration is predicted, as the square root of the sum of the squares of the standard errors of the fitted values and of the residual standard errors. This value is 0.67576, which is equivalent to just under 2 ms. The model's predictions are so precise because the *n* is large.

22.1.2.5 *Validating and bootstrapping*

Two more informative means of testing the predictive accuracy of the model assess the extent to which the model overfits the data, i.e. how much adding terms to model to improve its fit to the current data reduces its ability to predict novel data. Both begin by dividing the data into training and testing sets.

"Cross-validation" divides the data into 3–10 equal-sized *folds*, where each fold is a random sample from the original data set. Each fold serves in turn as the test set, and the remaining data as the training set. The training data is used to calculate the model, which is then used to predict the observed test values. Table 22.1.3 shows the results of cross-validation with eight folds. The R^2 value is the proportion of variance accounted for by the model. The mean squared error is the mean of the squared residuals. The intercept and slope are for the line obtained when the observed durations are regressed against the fitted values. They are necessarily 0 and 1 for the original data and the training set, but the slope may be less than 1 for the test set. For slopes less than 1, the intercept's value compensates by shifting away from 0. The values for the training and test sets are the averages across the eight folds. The optimism values are the difference between the training and test-set statistics, and they estimate the extent to which the original model overfits the data. The corrected values are obtained by subtracting the optimism values from the original values—these discounted R^2 and MSE values would be used in a conservative assessment of how well the model fits the data. For this data set and model, the optimism values are all tiny, which indicates only very slight overfitting.

Table 22.1.3. R^2, **mean squared error (MSE), intercept, and slope, showing original, training, and test values, the difference between training and test values (optimism), corrected values, and the number of folds (n) in cross-validation of the final model of the Finnish vowel durations**

	original	training	test	optimism	corrected	*n*
R^2	0.86851	0.86879	0.85933	0.00946	0.85905	8
MSE	0.03207	0.03198	0.03336	−0.00138	0.03345	8
Intercept	0.00000	0.00000	0.01349	−0.01349	0.01349	8
Slope	1.00000	1.00000	0.99734	0.002656	0.99734	8

Table 22.1.4. R^2, MSE, intercept, and slope, showing original, training, and test values, optimism, corrected values, and the number of folds (n) in bootstrap validation of the final model of the Finnish vowel durations

	original	training	test	optimism	corrected	n
R^2	0.86851	0.87067	0.86639	0.004274	0.86424	200
MSE	0.03207	0.03141	0.03258	−0.00118	0.03324	200
Intercept	0.00000	0.00000	0.01564	−0.01564	0.01564	200
Slope	1.00000	1.00000	0.99692	0.00308	0.99692	200

In "bootstrapping," the training set is randomly and repeatedly drawn from the original data with replacement to produce a sample of the same size. This training set consists of roughly 485 unique values for an original data set of 768 values like the one here. The samples' values are used to calculate the model, and that model is used to predict the observed values in the original full data set. These steps are repeated many times, here 200. The results in Table 22.1.4 are interpreted in the same way as those in Table 22.1.3, and like those results they show that the model overfits the data very little.

The extent of overfitting is so small because the original data sample is quite large. When overfitting is greater, these procedures would justify omitting one or more of the independent variables or their interactions from the model.

22.1.2.6 *Fitting the data*

The R^2 value for the main-effects model presented in Table 22.1.1 was 0.847, while that which included the interaction in Table 22.1.2 was 0.869, a difference of just 0.022. This increment looks small, but Figure 22.1.3 motivates including this interaction in the final model. The quantitative modeling of the data was guided by the prior graphical exploration of the influence of the independent variables and possible interactions between them, rather than by a blind desire to improve the quantitative fit to the data. The resulting fit is quite good, more than 0.85 of the variance of the data is accounted for by a model with just thirteen predictors (one of them the intercept), which is the right balance.

Similarly, no automatic procedure like step-wise regression was used to decide which variables should be kept in the model. Besides taking model interpretation away from the analyst, such procedures ignore the effects of the variable selection process in calculating standard errors and t-statistics, they produce overoptimistic estimates of standard errors and p-values, and they bias the absolute values of predictor estimates upwards—positive estimates are farther from 0 and negative estimates closer.

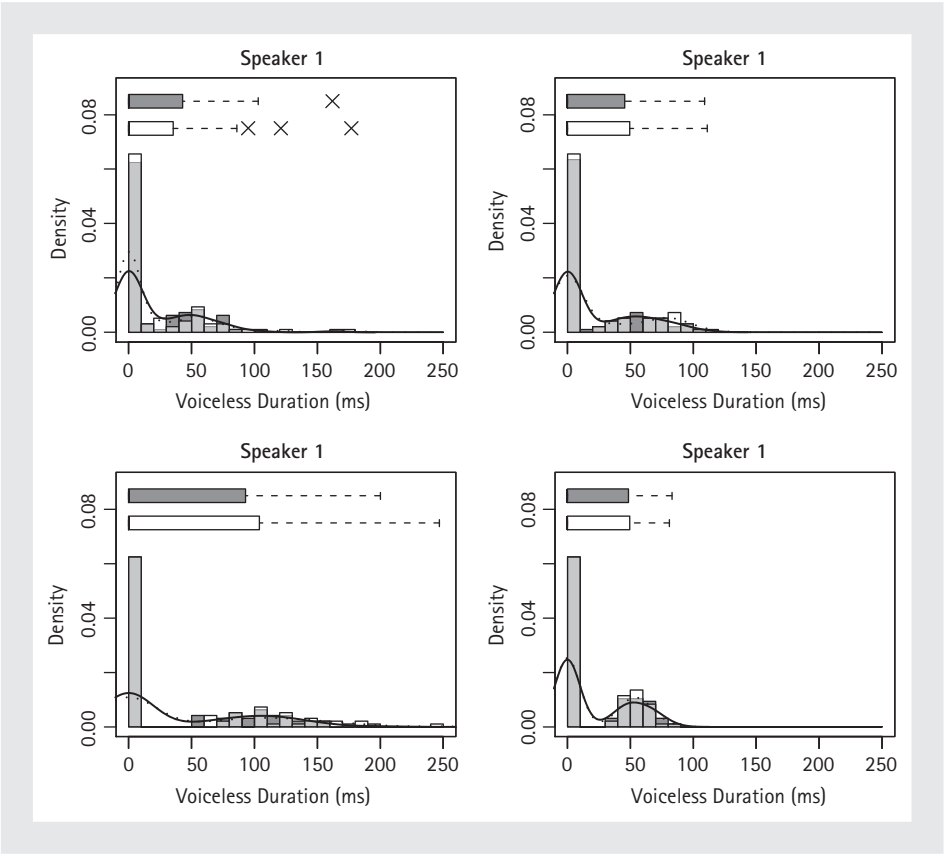

Figure 22.1.6. Histograms, density plots, and box plots of short (dark gray) and long (white) voiceless vowel durations produced by four Finnish speakers. The lighter gray bars represent durations common to short and long vowels. The Xs are outliers.

22.1.2.7 *Voiceless durations*

We now turn briefly to the durations of the voiceless portions of these vowels (Figure 22.1.6). They are of interest because Myers and Hansen propose that final vowels shorten as a result of being partially devoiced (Myers and Hansen 2006, 2007; Baayen 2008). Figure 22.1.6 looks very different from Figure 22.1.1: the short and long distributions are no longer even approximately discrete but instead overlap completely, and now there is a very strong mode at 0, which represents all those vowels which have no voiceless portion. Figure 22.1.7 reveals that no portion of any vowel is voiceless in non-final position or closed syllables. Position and CVN syllables can therefore be left out of modeling the voiceless portions' durations.

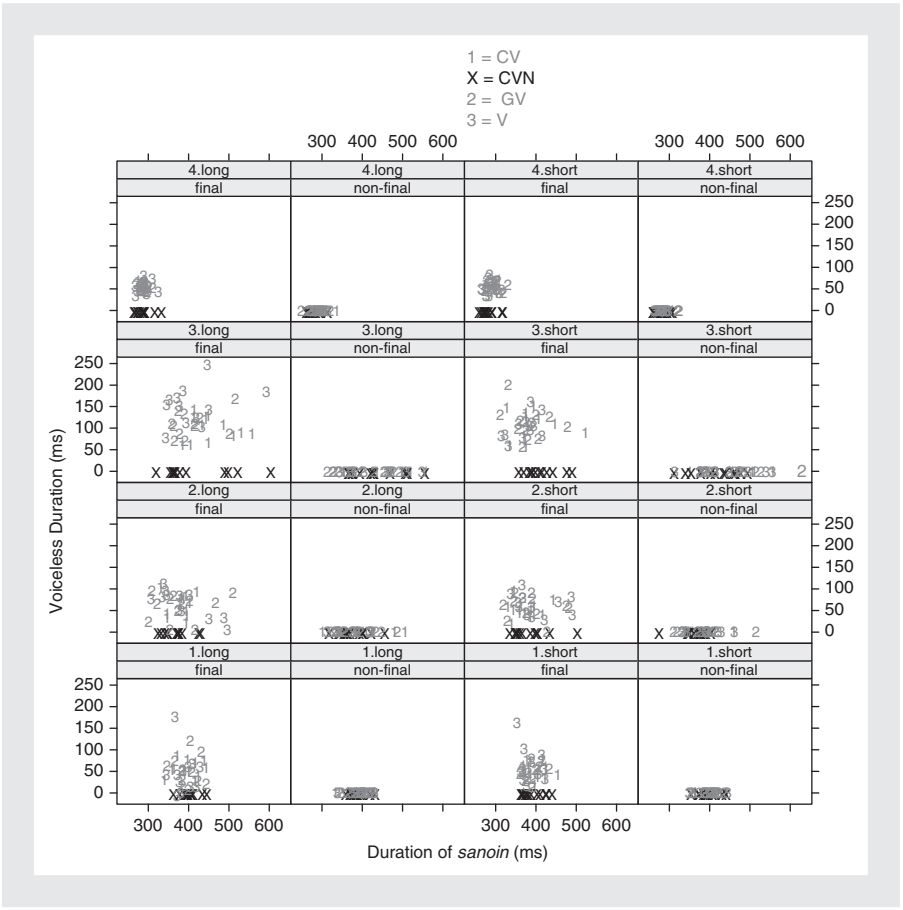

Figure 22.1.7. Durations of voiceless portions of vowels (vertical axis) by duration of the word *sanoin* duration in the same utterance by speaker, phonological quantity, position, and syllable type. Black "X" for closed CVN syllables, and gray "1, 2, 3" for open CV, GV, and V syllables, respectively.

The predictor estimates for the model of the durations of the voiceless portions of the vowels in Table 22.1.5 show that none of the variables have a significant effect on these durations, except that they are significantly longer for speaker 3 than speaker 1. In Table 22.1.6, the log-transformed duration of the voiced portion of the vowel has been added to the model as a predictor. The predictor estimates show that the duration of the voiceless portion shortens significantly as the duration of the voiced portion lengthens ($t = -2.6858$, $p = 0.00767$). A more interesting finding is that the duration of the voiceless portion is now also significantly shorter when the vowel is phonologically short ($t = -2.4136$, $p = 0.01644$). In other words, the voiceless portion's duration varies inversely with the voiced portion's duration but directly with the vowel's phonological quantity.

Table 22.1.5. Predictor estimates in a linear regression model of the durations of the voiceless portions of Finnish vowels in final position in which the log-transformed duration of *sanoin*, quantity, syllable type, and speaker are independent variables

Predictor	Estimate	Std. Error	*t* value	Pr(> |*t*|)
Intercept	6.273839	1.92506	3.2590	0.001256
log(sanoin)	−0.404738	0.32082	−1.2616	0.208150
short	0.008546	0.06327	0.1351	0.892658
GV	−0.084840	0.07708	−1.1007	0.271997
V	−0.028269	0.07742	−0.3651	0.715298
Speaker 2	0.115028	0.08965	1.2831	0.200533
Speaker 3	0.871518	0.08930	9.7598	0.000000
Speaker 4	0.058198	0.13202	0.4408	0.659667

Table 22.1.6. Predictor estimates in a linear regression model of the durations of the voiceless portions of Finnish vowels in final position in which the log-transformed duration of *sanoin*, log-transformed duration of the voiced portion of the vowel, quantity, syllable type, and speaker are independent variables

Predictor	Estimate	Std. Error	*t* value	Pr(> |*t*|)
Intercept	8.64921	2.09935	4.1199	0.00005
log(sanoin)	−0.44829	0.31772	−1.4110	0.15940
log(voiced duration)	−0.41354	0.15397	−2.6858	0.00767
short	−0.38244	0.15845	−2.4136	0.01644
GV	0.01427	0.08470	0.1685	0.86630
V	0.07546	0.08576	0.8799	0.37970
Speaker 2	0.11330	0.08867	1.2778	0.20240
Speaker 3	1.00419	0.10119	9.9234	0.00000
Speaker 4	0.02268	0.13124	0.1728	0.86290

22.1.2.8 *Collinearity and principal components*

There is a problem here, however: the voiced portion's duration and its phonological quantity are probably not independent of one another. The voiced portion is expected to be longer when the vowel is phonologically long. An analysis not shown here confirms this expectation. The covariation of the voiced portion's duration and the vowel's phonological quantity is an example of collinearity. The extent of collinearity here is 8.89, which is greater than negligible (0–6), but less than moderate (around 15) and much less than severe (values greater than 30). Although collinearity is slight enough here that no remedy is required, this example can nonetheless be used to show how to proceed when collinearity is more severe. One

could simply leave out one of the collinear variables, but which one? A more principled approach uses principal components analysis to eliminate the collinearity.

Principal components analysis uses the covariation (collinearity) between the values of individual variables to construct a new composite variable that combines their influences on the data's structure. The first principal component accounts for the largest proportion of the variance in the data's structure, the second for the next largest proportion, and so on. One ends up with as many principal components as the original number of variables, but they are now rank-ordered in terms of how much of the variance they account for. A common rule of thumb is to disregard any principal components that account for less than 0.05 of the variance.

The principal components replace the collinear variables in constructing a new model of the data. Table 22.1.7 lists the statistics of the two principal components extracted when this method was applied to the quantity and voiced durations for the vowels whose voiceless durations were modeled above. The first principal component accounts for just over 0.93 of the variance, and the second for just under 0.07. Table 22.1.8 shows the "loadings" of these two principal components on the original variables, quantity and voiced duration. The signs of these loadings are positive for the loading of PC1 on both quantity and voiced duration, which captures the fact that these two variables co-vary directly: a long vowel has a longer voiced duration. The opposite signs of the loadings for PC2 capture the weaker inverse variation between voiced duration and phonological quantity, namely, that the voiced duration is shorter in long vowels.

Table 22.1.9 presents the linear model of the voiceless durations with PC1 and PC2. The estimate for PC1 is not significant ($t = -1.0454$, $p = 0.29674$), while that

Table 22.1.7. Standard deviations and proportions of variance accounted for by the first two principal components extracted from the covariation between phonological quantity and voiced duration for final vowels in CV, GV, and V syllables

	PC1	PC2
Standard deviation	1.365	0.3702
Proportion of variance	0.931	0.0685

Table 22.1.8. Loadings of the first two principal components on phonological quantity and voiced duration

	PC1	PC2
quantity	0.7071068	−0.7071068
log-transformed voiced duration	0.7071068	0.7071068

Table 22.1.9. Predictor estimates in a linear regression model of the durations of the voiceless portions of Finnish vowels in final position in which the log-transformed duration of *sanoin*, two principal components representing the log-transformed duration of the voiced portion of the vowel and quantity, syllable type, and speaker are independent variables

Predictor	Estimate	Std. Error	*t* value	Pr(> |*t*|)
Intercept	6.44336	1.90150	3.3886	0.00080
log(sanoin)	−0.44829	0.31772	−1.4110	0.15937
PC1	−0.02465	0.02358	−1.0454	0.29674
PC2	−0.29555	0.11335	−2.6073	0.00962
GV	0.01427	0.08470	0.1685	0.86634
V	0.07546	0.08576	0.8799	0.37969
Speaker 2	0.11330	0.08867	1.2778	0.20240
Speaker 3	1.00419	0.10119	9.9234	0.00000
Speaker 4	0.02268	0.13124	0.1728	0.86292

for PC2 is ($t = -2.6073$, $p = 0.00962$). This outcome is not surprising in light of Table 22.1.6, which showed that the voiceless duration was shorter when the vowel was phonologically short and when its voiced duration was longer. This outcome also indicates that we could leave PC1 out of the final model of the voiceless durations. That model accounts for the same proportion of the variance in the voiceless durations as that using quantity and voiced durations as independent variables, 0.294, and does so with one less explanatory predictor.

Besides the slightness of the collinearity in this example, there is another, more general reason to hesitate to apply this method: by collapsing the influences of two or more of the original variables into a single variable, principal components analysis can obscure rather than illuminate the analysis. In some instances, those variables may simply be alternative ways of measuring the same psychologically real linguistic property; then, principal components analysis reveals that underlying reality. But in this instance, the original analysis with quantity and voiced duration as distinct variables provides a more straightforward description and explanation of what influences the duration of the vowel's voiceless portion.

22.1.2.9 *Summary*

This section has presented linear models of continuous dependent variables, the total durations of Finnish vowels and of their voiceless portions. Graphical explorations preceded and guided model construction. The models were then criticized, by examining their residuals, through cross-validation and bootstrapping, and assessing the extent to which the independent variables were collinear. These critiques

led to the log transformation of the dependent variable to bring the residuals into line.

22.1.3 Mixed-effects models of ordinal dependent variables: First try

22.1.3.1 *Random versus fixed effects*

Myers and Hansen presumed that they had drawn a random yet representative sample from the population of Finnish speakers. Presuming that the sample is representative does not entail that one member's data values will not differ from other potential members, but only that the differences will be idiosyncratic rather than systematic. The expected idiosyncrasies were built into the models by representing speakers 2–4 with their own predictor, whose value showed how their vowel durations differed overall from default speaker 1. Interactions were excluded between the speaker predictors and any of the others because I tacitly assumed that the effects of phonological quantity, position, and syllable type would not differ substantially between speakers. That assumption represents a fundamental difference between kinds of effects, random versus fixed effects.

Speaker is a random effect in that the speakers are a random sample from the population of possible speakers. We would not expect to *repeat* the idiosyncrasies of one sample of speakers in another. Fixed effects (also referred to as "conditions" or "treatments") such as phonological quantity etc. are repeatable, in that they can be applied to another sample (see also Baayen, this chapter, for further discussion of how random effects differ from fixed effects).

Models which combine random and fixed effects are called "mixed-effects" or simply "mixed" models. Mixed models of ratings are presented here. Ratings are an example of an ordered or "ordinal" categorical variable. The R packages, lme4 and ordinal, used in carrying out these analyses are described in Bates and Maechler (2010) and Christensen (2010). Before beginning these analyses, contrast coding of categorical variables with more than two values must be discussed (see also Baayen this chapter).

22.1.3.2 *Categorical predictors with more than two values: Contrasts*

A common practice when a categorical predictor has more than two values has been to determine its significance overall by means of an analysis of variance, and then to run post-hoc tests comparing pairs of predictor values. Because there is a substantial danger of getting a spuriously significant result when one runs multiple tests, the α value must be corrected to α/m, where m is the number of post-hoc

Table 22.1.10. Recoding of a four-valued predictor (A–D) into three contrasts, using treatment, Helmert, or polynomial recoding

Predictor	Treatment			Helmert			Polynomial		
	T1	T2	T3	H1	H2	H3	Linear	Quadratic	Cubic
A	0	0	0	−1	−1	−1	−3	1	−1
B	1	0	0	1	−1	−1	−1	−1	3
C	0	1	0	0	2	−1	1	−1	−3
D	0	0	1	0	0	3	3	1	1

comparisons—this is the "Bonferroni" correction. Applying this correction can make it difficult to achieve significance if the number of comparisons is large.

A better solution is to recode the original predictor with a set of contrasts that embody the comparisons one wants to do. All ways of recoding categorical predictors require the contrasts be orthogonal; that is, for k predictor values (AKA "treatments"), there are only $k - 1$ contrasts. Otherwise, treatments excluded from a contrast are assigned a value of 0, treatments that are grouped together are assigned the same sign, and those which are contrasted are assigned opposite signs. In some kinds of contrasts, the values assigned to included treatments also sum to 0. Table 22.1.10 illustrates the recoding of a four-valued predictor (A-D) for treatment, Helmert, and polynomial contrasts.

Treatment recoding is identical to simply comparing each non-default treatment to the default treatment (as was done for syllable type in the Finnish vowel duration analysis). The first contrast in Helmert recoding compares the second treatment (B) with the first (A) and excludes the other treatments (C and D), the second compares the third treatment (C) with the mean of the first and second (A, B) and excludes the fourth (D), and the third compares the fourth treatment (D) with the mean of the first, second, and third (A, B, C). By judiciously ordering the treatments, one can obtain the comparisons one wants with Helmert recoding. Polynomial recoding is only appropriate when the original predictor's values are ordered but cannot be assigned a value along a scale. The contrasts model the predictor's effect as linear, quadratic, cubic functions, etc., where the highest order of the polynomial equals one less than the number of contrasts.

22.1.3.3 *Experiment design and a first look at the results*

The data are dissimilarity ratings obtained in an ERP study carried out by Mara Breen, Lisa Sanders, and me. The participants were presented with two syllables on each trial, the first was the "prime" and the second the "target," and their task was to rate how dissimilar the target was to the prime on a four-point scale, where 1 corresponded to maximally similar, 4 to maximally dissimilar and 2 and 3 to lesser

Table 22.1.11. Voiced and unvoiced prime-target pairs for Legal, Illegal, and Absent primes in Identity, Control, and Test trials

Prime status	Legal		Illegal		Absent	
Trial type	Prime	Target	Prime	Target	Prime	Target
Identity	gw	gw	gl	gl	gw	gw
Control	kw	gw	kl	gl	tw	gw
Test	dw	gw	dl	gl	bw	gw
Identity	kw	kw	kl	kl	tw	tw
Control	gw	kw	gl	kl	gw	tw
Test	tw	kw	tl	kl	pw	tw

values of similarity and dissimilarity. Ratings are an example of an ordinal dependent variable. The values of ordinal variables are ordered like those of continuous variables such as vowel duration, but they are also categorical; that is, values such as 1.5, $\sqrt{2}$, 3.14159, etc. are not possible.

Table 22.1.11 shows that primes in test trials consisted of syllables beginning with consonant clusters that are legal in English, e.g. [dw, tw], absent but perhaps not illegal, e.g. [bw, pw], or illegal [dl, tl] (see Moreton 2002, for justification of this classification). The consonant clusters in the targets were always legal, as were all clusters in primes in identity and control trials. Multiple tokens of each syllable were used to compose 100 distinct trials for each of the eighteen possible prime-target combinations. Responses were collected from eighteen native speakers of English.

Figures 22.1.8 and 22.1.9 display the mean dissimilarity ratings across participants and items for the three kinds of primes and the three trial types, collapsed across voicing. Figure 22.1.8 averages across items, while Figure 22.1.9 averages across participants. Both figures show that responses cluster near 1 on identity trials (top rows) and near 4 on Control and Test trials (middle and bottom rows), except when the prime is Illegal (bottom right), where ratings instead cluster near 2. The noticeably greater spread of values in Figure 22.1.8 than Figure 22.1.9 shows that ratings differed more between participants than items—there are also few if any ratings of 2–3 for items.

22.1.3.4 *Mixed-effects model with Helmert contrasts*

Averaging across participants (Figure 22.1.8) or items (Figure 22.1.9) is the first step in what has until recently been standard practice in psycholinguistics, namely, carrying out a by-participants (= by-subjects) analysis, in which participants are treated as random effects, and then a by-items analysis, in which items are treated as a random effect (Clark 1973; Forster and Dickinson 1976). In this approach, a predictor's effect is treated as significant only if it is significant in both analyses. The development of mixed-effects models has largely superseded the need to carry out

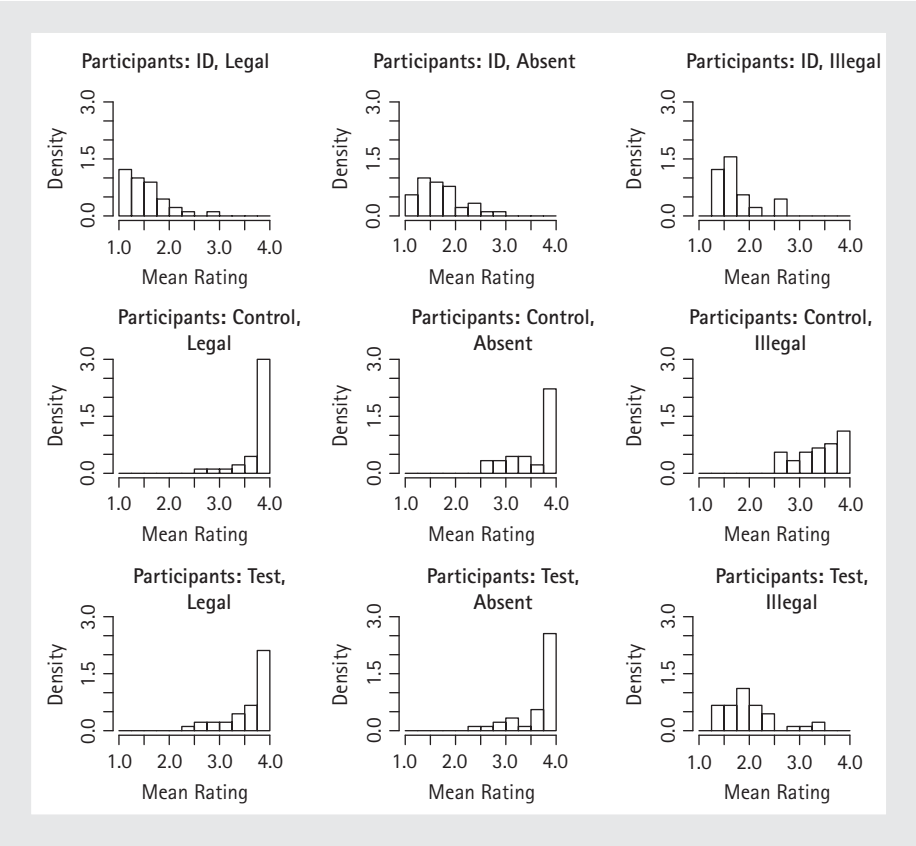

Figure 22.1.8. Mean dissimilarity ratings by participants and by the status of the prime and type of the trial.

separate analyses, because both participants and items can be treated as random effects within a single analysis. However, software to implement such models has not yet been developed for cases like this one where the dependent variable is an ordinal variable, so I carry out separate participant and items analyses here.

Averaging across items or participants replaced the categorical integer values of the original ratings with non-integer values. To turn these means back into integers, they were first multiplied by 4, to preserve to some extent the distinctions between the means, and then rounded to the nearest integer. The combined effect of these two operations is to turn the continous 1–4 scale produced by the averaging into a 4–16 integer scale.

Because the dependent variable is a frequency rather than the measure of some quantity, it is transformed further. The original frequencies or more precisely the ratios of ratings of 4 versus ratings greater than 4, of ratings of 4–5 versus ratings greater than 5, . . . , to ratings of 4–15 versus ratings of 16 are transformed into log

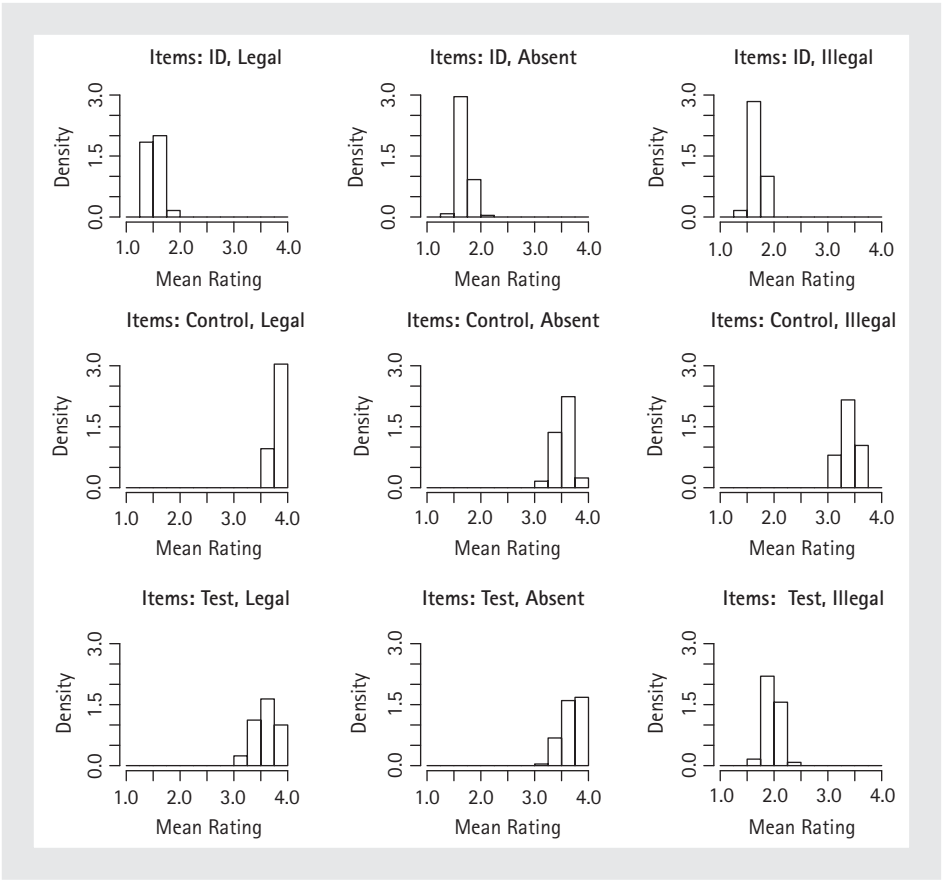

Figure 22.1.9. Mean dissimilarity ratings by items and by the status of the prime and type of the trial.

odds ratios, aka logits (see also Baayen, this chapter). The model for frequency data is therefore a logistic rather than linear regression.

In both analyses, voicing, prime status, and trial type served as independent variables, with prime status and trial type recoded as Helmert contrasts (Table 22.1.12). The signs of the Helmert contrasts for status are opposite to those given in Table 22.1.10 because we expect lower dissimilarity ratings for identity than control or test trials. The models also included all pair-wise interactions between prime status and trial type.

The results of the two analyses are displayed in Tables 22.1.13 and 22.1.14.[2] The variance and standard deviation are much larger for participants, 6.1251 and 2.4749,

[2] The estimates in these tables are logits and can be converted back into odds ratios by using the products of their values with the values of the corresponding Helmert contrasts as exponents of *e*.

Table 22.1.12. Recoding of the three-valued status and type predictors as Helmert contrasts

Status	S1	S2	Type	T1	T2
Legal	1	1	Identity	−1	−1
Absent	−1	1	Control	1	−1
Illegal	0	−2	Test	0	2

Table 22.1.13. Predictor estimates in an ordinal logistic regression by-participants model of dissimilarity ratings in which S1, S2, T1, and T2 are Helmert contrasts representing prime status and trial type

Predictor	Estimate	Std. Error	*z* value	Pr(> |*z*|)
S1	−0.0264	0.1444	−0.1826	0.8550765
S2	1.3181	0.1176	11.2042	< 2.22e-16
T1	6.1420	0.4453	13.7923	< 2.22e-16
T2	1.0297	0.1022	10.0732	< 2.22e-16
Voiced	−0.5804	0.2192	−2.6473	0.0081131
S1:T1	1.1080	0.1862	5.9515	2.6570e-09
S1:T2	−0.2933	0.1037	−2.8284	0.0046778
S2:T1	0.5924	0.0967	6.1233	9.1631e-10
S2:T2	1.0392	0.0958	10.8460	< 2.22e-16

Table 22.1.14. Predictor estimates in an ordinal logistic regression by-items model of dissimilarity ratings in which S1, S2, T1, and T2 are Helmert contrasts representing prime status and trial type

Predictor	Estimate	Std. Error	*z* value	Pr(> |*z*|)
S1	−0.0988	0.0949	−1.0413	0.2977301
S2	11.0210	0.5798	19.0083	< 2.22e-16
T1	50.8508	2.6350	19.2979	< 2.22e-16
T2	6.3377	0.3154	20.0957	< 2.22e-16
Voiced	−1.1321	0.1464	−7.7338	1.0439e-14
S1:T1	1.7209	0.1297	13.2682	< 2.22e-16
S1:T2	−0.2041	0.0663	−3.0788	0.0020783
S2:T1	0.9032	0.0699	12.9237	< 2.22e-16
S2:T2	10.6456	0.5739	18.5485	< 2.22e-16

Table 22.1.15. Values of interaction terms for each combination of prime status and trial type contrasts

Interaction	Leg:ID	Leg:Cntl	Leg:Test	Abs:ID	Abs:Cntl	Abs:Test	Ill:ID	Ill:Cntl	Ill:Test
S1:T1	−1	1	0	1	−1	0	0	0	0
S1:T2	−1	−1	2	1	1	−2	0	0	0
S2:T1	−1	1	0	−1	1	0	2	−2	0
S2:T2	−1	−1	2	−1	−1	2	2	2	−4

than for items, 0.3082 and 0.5552. These outcomes confirm what Figures 22.1.8 and 22.1.9 had already shown: ratings differ more between participants than items.

The results of the two analyses are otherwise strikingly similar: all the predictors significantly influence the dissimilarity ratings, except for S1, the contrast that represents the comparison between legal and absent primes; ratings are predicted to be higher for legal and absent than illegal test primes (S2), on control than identity trials (T1), and on test than control and identity trials (T2); and voiced pairs were judged to be less dissimilar than voiceless ones.

The interactions are also strikingly similar in the two analyses. Table 22.1.15 lists the values which when multiplied by the estimates in Table 22.1.13 or 22.1.14 yield the predicted effects on the odds ratios of the dissimilarity ratings for each combination of prime status and trial type. For example, multiplying postive S2:T2 estimates by the value −4 for the illegal test combination predicts a dramatic drop in the dissimilarity ratings, which can be observed in Figure 22.1.8 and even more dramatically in Figure 22.1.9.

22.1.3.5 *Summary*

In this section, two mixed-effects ordinal logistic regression models of dissimilarity ratings were presented in which two three-valued predictors were recoded as Helmert contrasts to rule out the need for post-hoc tests comparing subsets of predictor values. Each model included a single random effect, either of participants or items, and the dependent variable was the log odds ratios of the dissimilarity ratings for each interval along the ordinal scale averaged over the other random effect.

22.1.4 Mixed-effects models of ordinal dependent variables: Second try

Here, I reanalyze these data by treating the dissimilarity ratings as a series of three binomially distributed variables, 1 versus 2–4, 1–2 versus 3–4, and 1–3 versus 4, in

Table 22.1.16. Random effects for logistic regression models of the partition into a series of two-valued variables: 1 versus 2–4, 1–2 versus 3–4, and 1–3 versus 4

	1:234		12:34		123:4	
Random effects	Variance	Std. Dev.	Variance	Std. Dev.	Variance	Std. Dev.
items	2.622	1.619	1.517	1.232	1.120	1.058
participants	1.848	1.359	1.556	1.247	2.461	1.569

order to include both random effects in the model at once.[3] Three mixed-effects models are constructed, one for each these new variables. The random effects in these models are participants and items, while the fixed effects are the same as those just used, coded once again as Helmert contrasts.

Table 22.1.16 lists the random effects on the intercepts of these three models. The standard deviations differ relatively little between the models, which shows that the range of differences between items and participants is roughly the same for all three. These values are no longer noticeably smaller for items than participants.

Table 22.1.17 lists the estimates of the fixed effects with their standard errors and z-scores; a * follows when the associated p value is less than 0.05. The model predicts that the probability of the lower dissimilarity rating(s) in each partition increases for illegal compared to legal and absent primes (S2 = −2 versus 1), this probability decreases in control compared to identity trials (T1 = 1 versus −1) and in test compared to control and identity trials (T2 = 2 versus −1), and it decreases in test trials compared to identity and control trials for legal and absent primes (S2 : T2 = 2 versus −1), while increasing for test trials compared to identity and control trials for illegal primes (S2 : T2 = −4 versus 2). This analysis is noticeably more conservative than the earlier ones in that only the S2:T2 interaction is significant. Averaging across items or participants hid variation in the random effects that remains exposed in this analysis. This reanalysis thus shows that the traditional approach's insistence on effects being significant in both by-participants (by-subjects) and by-items analyses is not sufficient protection against rejecting the null hypothesis when there is a good chance after all that it's true (Type 1 error).

Ordinal logistic regression was illustrated in this section by treating the dissimilarity ratings as a three-step series of binary and thus binomially distributed dependent variables, and including both participants and items as random

[3] The ordinal logistical regression models just illustrated partition the data similarly.

Table 22.1.17. Fixed effects for a series of logistic regression models in which a four-valued ordinal variable is partitioned into a series of two-valued variables: 1 versus 234, 12 versus 34, and 123 versus 4. $* = p < 0.05$

	1:234				12:34				123:4			
Predictor	Estimate	Std. Error	*z*		Estimate	Std. Error	*z*		Estimate	Std. Error	z	
(Intercept)	−1.984	0.501	−3.963	*	−0.613	0.415	−1.477		0.322	0.447	0.719	
S1	−0.031	0.333	−0.095		−0.015	0.253	−0.058		−0.055	0.218	−0.253	
S2	−0.447	0.192	−2.329	*	−0.558	0.146	−3.818	*	−0.647	0.0126	−5.138	*
T1	−2.773	0.333	−8.326	*	−2.706	0.254	−10.656	*	−2.667	0.219	−12.164	*
T2	−0.491	0.192	−2.559	*	−0.498	0.146	−3.408	*	−0.477	0.126	−3.793	*
Voiced	0.175	0.543	0.323		0.264	0.413	0.638		0.399	0.0356	1.121	
S1:T1	−0.414	0.408	−1.014		−0.395	0.311	−1.272		−0.413	0.268	−1.545	
S1:T2	0.099	0.235	0.420		0.112	0.179	0.627		0.136	0.154	0.881	
S2:T1	−0.218	0.235	−0.927		−0.204	0.179	−1.140		−0.254	0.154	−1.646	
S2:T2	−0.375	0.136	−2.768	*	−0.421	0.103	−4.078	*	−0.459	0.089	−5.165	*

effects. With the random effects unconstrained, fewer interactions turned out significant.

22.1.5 Concluding remarks

In this section, I have tried to illustrate how statistical tools can be used to explore and test hypotheses concerning two kinds of data commonly encountered while studying phonology in the laboratory: continuous variables and ordinal scales. My goal throughout has been to focus on the practicalities of carrying out such explorations and tests. My hope is that these illustrations provide sufficient guidance that you may see how to adapt them to the data you are trying to analyze. To understand the practicalities of these analyses, you will also need to study the sources cited at the beginning of this section. The payoff for doing so is enormous, as the statistical tools used here provide many insights into the data.

22.2 Mixed-effects models

Harald Baayen

22.2.1 Introduction

Consider an experiment in which the duration of the first vowel in a word is studied. It is expected that this duration is determined in part by the number of syllables following in the same word, in part by whether the vowel is in an open syllable (vs. closed syllable), in part by the position of the word in the sentence, by the speech rate, and possibly by the frequency of the word. If our interest is in the generality of vowel shortening, different vowels will be studied, in different words, and produced by different speakers. For this type of experiment, mixed models are an excellent choice.

In this example, the factor `Syllable Type` (with levels *open syllable* and *closed syllable*) is a fixed-effect factor, as its two levels exhaust all possible values that the predictor `Syllable Type` can take. By contrast, the factor `Speaker` is a random-effect factor, as its levels, identifiers for the different speakers, are randomly sampled from a much larger population of speakers. `Word` is another random-effect factor, as the words sampled for the experiment represent only a small proportion of the words known to the speakers (see also Section 22.2.6 for how to define fixed-effect vs. random-effect factors).

Classical analysis of variance and regression analysis run into problems for data sets combining fixed- and random-effect factors, especially when more than one random-effect factor has to be brought into the analysis. Often, researchers aggregate their data to obtain means or proportions for subjects (averaging over items) or for items (averaging over subjects, see also Kingston, this chapter). In psycholinguistics, the work by Clark (1973) and Forster and Dickinson (1976) led to the practice of averaging both over subjects and over items, with an effects accepted as significant only if it reaches significance both 'by subjects' and 'by items'. Mixed-effects models provide the researcher with a more sophisticated tool for analyzing repeated measures data that is both more flexible, more powerful, and more insightful.

22.2.2 Basic concepts

Let X_1 denote the fixed-effect factor `Syllable Type` and let X_2 represent the covariate `Frequency` of occurrence. Suppose that ten vowels are selected, and that the question of interest is whether the duration of the k-th vowel, Y_k, can be predicted from `Syllable Type` (*open* versus *closed syllable*) and `Frequency`. The linear model decomposes the dependent variable into a weighted sum:

$$(1) \quad Y_k = \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \beta_{12} X_{1k} X_{2k} + \epsilon_k, \quad k = 1, 2, \ldots, 10.$$

Fixed-effect factors are coded numerically using dummy coding, such that a factor with n levels contributes $n - 1$ predictors to the model. Of the many ways in which factors can be coded numerically, *treatment coding* is the most straightforward and the most easy to interpret, especially in the case of analysis of covariance. One level of the factor is selected as default or reference level. Although the selection of the reference level can be guided by theoretical considerations, technically, any level can serve as reference level. For the two-level factor `Syllable Type`, treatment coding adds one extra predictor, X_1 in (1), consisting of ones and zeroes. Observations for the reference level, say *closed syllable*, are assigned a zero, and observations for the other, contrasting level (*open syllable*) are assigned a one. As a consequence, the β weight for `Syllable Type` represents the *difference* (or contrast) between the group mean for the vowels in an open syllable and the group mean for the vowels in a closed syllable. This β weight, although technically a slope for a "degenerate" numerical predictor (consisting only of zeroes and ones), is referred to as a contrast coefficient.

The model defined in (1) includes an interaction term for `Syllable Type` by `Frequency`. This interaction allows for the possibility that two different regression lines are required for `Frequency`, one for vowels in closed syllables and a different one for vowels in open syllables. As a consequence, two intercepts and two slopes have to be defined. With treatment coding, the regression line for the reference level (*closed syllable*) is specified by the intercept β_0 and the slope for frequency β_2. The

coefficients of the regression line for *open syllables* is obtained by *adjusting* these slopes and intercepts (by β_1 and β_{12}) respectively (see Table 22.2.1) to make them precise for the data points with the vowels in open syllables. In summary, for a fixed-effect factor, one level is selected as the baseline, and coefficients are invested to adjust slopes and intercepts for the other levels of the factor.

When dealing with a random-effect factor, it does not make sense to select one—arbitrary—level (e.g. a given speaker, or a specific word) as reference level: Such a reference level is unlikely to be representative of the population sampled. Therefore, mixed models dispense with fixing a reference level and contrasts for random-effect factors. Instead, the β coefficients for the intercept, covariates, and fixed-effect factors are taken to represent the population average for each of the populations sampled by the random-effect factors. For any given random-effect factor, adjustments are implemented to allow precise predictions for the individual units sampled, such as the individual speakers in an experiment or corpus. These adjustments (technically referred to as Best Linear Unbiased Predictors or BLUPS) are assumed to follow a normal distribution with mean zero and some unknown standard deviation (to be estimated from the data). Instead of investing $n - 1$ coefficients for a simple main effect for a random-effect factor with n levels (e.g. n speakers), only one parameter is invested, a standard deviation characterizing the spread of the adjustments.

By way of example, consider a data set in which vowels are elicited in m words from n speakers, and that a simple main-effects model is appropriate. A first model,

$$(2)\quad \begin{aligned} &Y_{ij} = [\beta_0 + b_{0i}] + [\beta_1 + b_{1i}]X_{1j} + [\beta_2 + b_{2i}]X_{2j} + \epsilon_{ij}, \\ &i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m, \\ &b_{0i} \sim \mathcal{N}(0, \sigma_1),\ b_{1i} \sim \mathcal{N}(0, \sigma_2),\ b_{2i} \sim \mathcal{N}(0, \sigma_3),\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma), \end{aligned}$$

calibrates the model, for each speaker i, for that speaker's speech rate (through the adjustments b_{0i} to the intercept β_0), as well as for that speaker's sensitivity to the type of syllables (through the adjustments b_{1i} to the contrast coefficient β_1) and for that speaker's specific sensitivity to frequency of occurrence (through the

Table 22.2.1. Treatment coding in analysis of covariance: the contrast coefficients β_1 and β_{12} specify the differences in intercept and slope between the vowels in *open* and *closed* syllables

β_0	the intercept (group mean) for the reference-level *closed syllable*
$\beta_0 + \beta_1$:	the intercept (group mean) for *open syllables*
β_2	the slope for frequency for vowels *in closed syllables*
$\beta_2 + \beta_{12}$:	the slope for frequency for vowels *in open syllables*

adjustments b_{2i} to the slope β_2). Each of the sets of adjustments $b_{.i}$ is assumed to be normally distributed with zero mean. In other words, a random-effect factor (whether speaker, word, text, or syllable) is represented as a source of random variation around the population parameters $\{\beta\}$. This is the sense in which a random-effect factor is "random."

Model (2) is incomplete, in that it does not take into account that the words in which the vowels are embedded are repeated across speakers. To incorporate word as a second random-effect factor, (2) has to be modified as follows,

$$
\begin{aligned}
(3)\quad & Y_{ij} = [\beta_0 + b_{0i} + b_{0j}] + [\beta_1 + b_{1i} + b_{1j}]X_{1j} + [\beta_2 + b_{2i}]X_{2j} + \epsilon_{ij}, \\
& i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, m; \\
& b_{0i} \sim \mathcal{N}(0, \sigma_1), \ b_{1i} \sim \mathcal{N}(0, \sigma_2), \ b_{2i} \sim \mathcal{N}(0, \sigma_3), \\
& b_{0j} \sim \mathcal{N}(0, \sigma_4), \ b_{1j} \sim \mathcal{N}(0, \sigma_5), \ \epsilon \sim \mathcal{N}(0, \sigma),
\end{aligned}
$$

with crossed random effects for speaker and word. Adjustments to the intercept are often referred to as random intercepts. Similarly, adjustments to slopes are known as random slopes. In the case of adjustments to a contrast coefficient, one can speak of random contrasts. In (3), there are by-speaker random intercepts (b_{0i}) as well as by-word random intercepts (b_{0j}). Likewise, there are both by-speaker and by-word random contrasts (b_{1i}, b_{1j}). The model includes random slopes for frequency only for speaker (b_{2i}). It is not possible to include as well by-word random slopes for frequency, as this would lead to an unsolvable confound with frequency itself, which is a word property. In other words, it is only possible to include by-subject random slopes and contrasts for item properties, and by-item random slopes and contrasts for subject properties. For instance, speakers may require adjustments to the slope of the frequency effect, while words may require adjustments to the slope of the effect of aging (see e.g. Baayen and Milin 2010).

Whenever in addition to random intercepts, one or more random slopes (or contrasts) are associated with a given random-effect factor, the possibility arises that the random intercepts and random slopes (or contrasts) are correlated. Assuming multivariate normality, the full specification of the random effects for (3) is therefore given by the matrices

$$
(4)\quad M_{\text{speaker}} = \begin{bmatrix} \sigma_1 & r_{12} & r_{13} \\ r_{21} & \sigma_2 & r_{23} \\ r_{31} & r_{32} & \sigma_3 \end{bmatrix}, \qquad M_{\text{word}} = \begin{bmatrix} \sigma_4 & r_{45} \\ r_{54} & \sigma_5 \end{bmatrix},
$$

where $r_{kl} = r_{lk}$ specifies the correlation of the adjustments k and l estimated for the population of speakers or the population of words. In other words, the adjustments for a given random-effect factor are assumed to be multivariate normal with zero means and unknown standard deviations and correlations.

22.2.3 Advantages of mixed-effects models

Mixed-effects models offer many advantages compared to the classical linear model using dummy coding for random-effect factors (see also Kingston, this chapter). First, a fitted mixed model provides straightforward predictions for unseen levels of random-effect factors. For an unseen speaker and an unseen word, all $b_{\cdot\cdot}$ are set to zero, and predictions based on model (3) for a given position X_1 and frequency X_2 reduce to

(5) $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2.$

For a specific speaker i that contributed observations to the data and an unseen word, more precise predictions can be obtained using the by-subject random-effect adjustments:

(6) $Y_i = [\beta_0 + b_{0i}] + [\beta_1 + b_{1i}] X_1 + [\beta_2 + b_{2i}] X_2.$

Similarly, when the identity of the word is known, even more precise predictions are available by adding in the by-word random intercepts and slopes. For comparison: the classical linear model only provides predictions for the subjects and items sampled in the data, and models with many interactions involving subjects and items may not even be able to estimate all relevant coefficients.

Second, the mixed-effects model allows for fine-grained hypotheses about the random-effects structure of the data. For every data set, it is an empirical question whether all the terms in matrices such as shown in (4) contribute to a significantly better fit of the model to the data. The possibility of including or excluding correlation parameters is not available in the classical linear model, but turns out to be an important tool for understanding, for instance, individual differences between the subjects participating in experiments. In chronometric studies, for instance, one may find that subjects with a large positive adjustment to the intercept reveal a large negative adjustment to the slope of frequency of occurrence. Such a negative correlation suggests that slow responders (with large intercepts) carry the frequency effect (see e.g. Baayen and Milin 2010 for examples).

Third, mixed-effects models are better able to directly model heteroskedasticity. A fundamental assumption of the linear model is that the residual errors have the same variance across all conditions in the data. In many actual data sets, this assumption of homoskedasticity is violated. For instance, the duration of a vowel might be more variable for a sample of non-native speakers than for a sample of native speakers. Given a fixed-effect factor distinguishing between native and non-native speakers, each set of speakers can be assigned its own standard deviation for the by-subject random intercepts, thereby modeling the heteroskedasticity directly (instead of correcting p-values post-hoc for non-sphericity).

Fourth, mixed-effects models can handle autocorrelational structure in data elicited from subjects over time, whether obtained from a stretch of speech or in

an experimental context. Human behavior is consistent over time, and this often gives rise to autocorrelations in language data. For instance, although there are fluctuations in speech rate, the speech rate at time t is likely to be very similar to the speech rate at the immediately preceding timesteps $t-1$, $t-2$, ... If the sequence of responses elicited from a given subject constitutes an autocorrelated time series, then it is essential to bring this autocorrelation into the model. If ignored, the residual errors will enter into autocorrelations, violating the assumption of independence of the residual errors, and giving rise to suboptimal conclusions about significance. The simplest way in which autocorrelations can be brought into a mixed model is by including as a separate predictor the response at the preceding point in time. For detailed discussion of experimental longitudinal effects, the reader is referred to Baayen and Milin (2010).

Fifth, the estimates provided by the mixed-effects model for the adjustments to the population parameters (the BLUPS) are *shrinkage estimates*. A danger inherent in fitting a statistical model to the data is overfitting. By way of example, consider a sample of subjects for which speech rate is recorded. Some subjects will have a faster speech rate than others. The more extreme the speech rate of a given subject is, the less likely it is that in a replication study the speech rate of that subject will be equally extreme (or even more extreme). It is much more likely that in the replication study the speech rate of this subject will have "regressed" or "shrunk" towards the mean. Mixed models anticipate this regression towards the mean and implement estimates for the BLUPS that shrink the adjustments in the direction of the mean. As a consequence, predictions for replication studies with the same subjects or items will be more precise.

Sixth, more than two random-effect factors can be included in the model. Returning to the above example, one possible design is to embed the same vowel in different carrier words. In such a design, vowels are repeated independently of the words, and hence the vowel should be considered as a potential third random-effect factor.

Finally, mixed-effects models tend to be better able to detect effects as significant. Baayen et al. (2008) show, on the basis of simulation studies for several experimental designs, that mixed-effects models offer a slight increase in power without giving rise to inflated Type I error rates, when compared with traditional analyses based on subject and/or item means. More important than the (generally small) increase in power is the much greater flexibility offered by mixed-effects models for bringing into the model specification various sources of variability that are unavailable when working with subject or item means. Even though longitudinal autocorrelational structure is as such often not of specific interest to the researcher, by taking it into account in the statistical model, the data become less noisy, and the effects of actual interest are more likely to reach significance (see e.g. De Vaan et al. 2007, as well as Baayen and Milin 2010).

22.2.4 Generalized linear mixed models (GLMMs)

Thus far, we have considered a dependent variable, duration, that is real-valued, and for which a model assuming normally distributed (Gaussian) errors is reasonable. Two commonly encountered dependent variables require special attention. First, instead of being continuous, the outcome of an experimental observation can be binary: true versus false, correct versus incorrect, success versus failure, present versus absent, etc. This kind of dependent variable is referred to as a binary, or binomial response variable. Second, a response variable can represent how often a phenomenon occurs in a given time window. In this case, we are dealing with count data.

For binary response variables, the traditional approach is to aggregate over trials (by subjects, or by items) to obtain proportions. Subsequently, analysis of variance or multiple regression is applied with these proportions as dependent variable. Three problems arise with this kind of analysis. First, instead of the variance being independent of the mean, the variance changes systematically with the mean, reaching a maximum when the proportion equals 0.5. This violates the assumption of homoskedastic variance that is fundamental to standard regression and analysis of variance. Second, proportions are bounded between 0 and 1, but the linear model assumes the dependent variable can assume any real value. The generalized linear model deals with these problems by taking as dependent variable not the proportion P,

$$(7) \quad P = \frac{\#\text{ successes}}{\#\text{ successes} + \#\text{ failures}},$$

but the log odds ratio (or logit)

$$(8) \quad L = \log \frac{\#\text{ successes}}{\#\text{ failures}}.$$

The log odds ratio ranges from minus infinity to plus infinity, and thus circumvents the problem with the boundedness of proportions. (An alternative to the logit link function that can be attractive for researchers familiar with signal detection theory is the probit link function.) The generalized linear model also implements different options for how the variance changes with the mean. For binary dependent variables, the appropriate variance function is that of a binomial random variable. Given the log odds (or logit) as *link function* and binomial variance, it becomes possible to obtain for each individual observation a good estimate of the probability of a success (or a failure).

A response variable may also represent counts. For example, for a series of interviews of the same length, the number of syllable deletions can be extracted. Just as the normal distribution is often appropriate for measurement data, the Poisson distribution tends to be an approximation for count data. The Poisson distribution has a single parameter, λ, which represents the rate at which a phenomenon occurs.

For one specific syllable, for instance, the rate at which it is deleted might be five times in an interview. For another syllable, the deletion rate might be ten times in an interview. Typical for count data is that the variability in the counts increases with the count itself. The Poisson distribution captures this well, as its single parameter λ represents both the mean and the variance. Thus, a greater mean rate is automatically paired with a greater variance. The generalized linear model for count data takes as dependent variable not the count itself, but its logarithmic transformation. This is the link function for count data. In addition, it uses the Poisson distribution to model how the variance changes with the mean.

The generalized linear model has been extended to incorporate random-effect factors in addition to fixed-effect factors. Crucially, generalized linear mixed-effects models, or GLMMS, do not require any prior aggregation into proportions, as the ambition is to provide estimates of the likelihood of a success (or failure), or the rate at which a phenomenon occurs (in the case of count data), for each individual observational unit.

22.2.5 Significance in mixed-effects models

The significance of covariates and fixed-effects factors can be evaluated in two ways. One option is to test whether slopes or contrasts are significantly different from zero. For non-Gaussian GLMMS, evaluation is based on Z-scores and associated p-values. For Gaussian models, the relevant t-tests run into the problem that there is no good analytical solution for the appropriate degrees of freedom. For large data sets, the upper bound for the degrees of freedom, the number of observations minus the number of fixed-effect parameters, often provides a good approximation. Informally, an absolute t-value exceeding 2 is a robust indicator of significance for $\alpha = 0.05$.

As an alternative to the t-test, a Bayesian method estimating the posterior distribution of the parameters can be used to obtain 95 percent credible intervals for the coefficients, as well as estimates of the probability of values more extreme than those actually observed. For data sets with at least several hundreds of observations, these probabilities are very similar to the probabilities obtained with the t-test based on the upper bound for the degrees of freedom. For smaller samples, the Bayesian probabilities are more precise. Informally, the Bayesian method can be conceptualized as generating a long series of parameter estimates as might be observed in replication studies. For each simulated replication study, a new set of parameters (intercept, slopes, contrasts, standard deviations, correlations) is generated. One can then inspect the distribution of a given parameter, for instance, the contrast coefficient for `Syllable Type`. If the observed contrast has a value that is extreme for the distribution of simulated contrasts, it is more likely to be significant.

A second option for evaluating significance of a predictor is to compare a model with and a model without a given predictor in order to ascertain whether the

parameters invested for this predictor lead to a non-trivial increase in goodness of fit. For mixed-effects models fitted to measurement data, a likelihood ratio test is appropriate. When two models are compared that differ with respect to the presence or absence of a factor or covariate, then both models should be fitted using maximum likelihood. In case the models have exactly the same factors and covariates in their model specification, but differ with respect to their random-effects structure, the two models are best fitted with relativized maximum likelihood.

The test statistic used by the likelihood ratio test is two times the difference between the log likelihood of the model with more parameters and the log likelihood of the model with fewer parameters. This test statistic follows a chi-squared distribution with as degrees of freedom the difference in the number of parameters. For this test to be precise, the models entering into the comparison should be nested, i.e. the full set of parameters of the model with fewer parameters should be a subset of the set of parameters of the model with more parameters. For generalized linear mixed models, an analysis of deviance test is the functional equivalent of the likelihood ratio test.

22.2.6 Working with mixed models

Mixed models are implemented in a range of software packages (e.g. SPSS, SAS, MLwiN, ASReml, S-Plus) and can be programmed within WinBUGS as well. Open-source software for carrying out mixed-effects modeling is available in `R` (the de-facto standard in statistical computing, freely available at <http://www.r-project.org>) using the `lme4` package by Bates and Maechler (2009).

When working with mixed models, several questions may arise. First, there are cases where it is not immediately self-evident whether a factor is to be modeled as fixed or random. Consider an experiment targeting the duration of English front high and mid vowels. Let `Vowel` denote the pertinent factor with as its four levels the four targeted vowels. Is `Vowel` fixed or random? English has fourteen vowels, so we are dealing with a sample of vowels. On the other hand, the population of vowels is quite small. In this example, `Vowel` is best modeled as a fixed-effect factor. The front high and mid vowels do not constitute a random sample from the population of vowels. The focus of the study is on specifically the four high and mid front vowels, with no aims to generalize beyond these four vowels to, e.g. back vowels or diphthongs.

Second, for a classical linear model fitted to a data set, an R-squared (or adjusted R-squared) value is generally reported. This R-squared specifies the proportion of the variance accounted for by the model (see Kingston, this chapter, for an example). For mixed models, an R-squared is often not reported, because it is no longer a good measure for understanding the contribution of the linguistic variables to explaining the variance: Parts, often very substantial parts, of the variance

are explained by the random-effect factors. In chronometric studies, for instance, linguistic predictors sometimes contribute less than 1 percent to the R-squared (Baayen 2008). If required, the R-squared can be calculated by squaring the correlation coefficient for the observed and expected values of the dependent variable in the case of Gaussian and Poisson models, and the index of concordance (Harrell 2001) for binomial models.

22.2.7 Selected studies using mixed models

Mixed-effects models are a relatively recent development in statistics, and do not have a long history of use in language studies. In psycholinguistics, mixed-effects models are rapidly becoming the new standard for data analysis with repeated measures. Quené and van den Bergh (2008), Baayen et al. (2008), and Jaeger (2008), all in a special issue in the *Journal of Memory and Language*, provide non-technical introductions, with Quené and van den Bergh discussing an example from phonetics, Baayen et al. presenting simulations of data sets as encountered in psycholinguistics, and Jaeger focusing on generalized linear mixed-effect models for binary data. Chapters 1 and 4 of Pinheiro and Bates (2000) are also highly recommended for introductory reading. Examples of psycholinguistic studies of auditory comprehension using mixed models are Baayen et al. (2007), Ernestus and Baayen (2007), and Balling and Baayen (2008). For application of mixed models to corpus-based data, see Ernestus et al. (2006), Janda et al. (2010), and Keune et al. (2005).

22.2.8 Concluding remarks

Mixed-effects models provide the researcher with a powerful tool for understanding the structure of quantitative data. Mixed models are robust with respect to unequal numbers of observations in different cells of one's experimental design. This is a useful property not only for the statistical analysis of experimental data, where observations may be lost due to errors, hesitations, or false starts, but also to observational data sets compiled from corpora, for which unbalanced distributions tend to be the norm.

However, mixed-effects models also have their limitations that come with the assumption that the correct model is linear or additive, and that the modeling problem is sparse in the sense that only a few predictors are assumed to be involved. An excellent complementary tool, especially for high-dimensional observational data, is the random forest technique (Strobl et al. 2009). For highly unbalanced data, random forests may yield fits that are as good or better than those provided by mixed-effects models, as observed by Tagliamonte and Baayen (2010) for a sociolinguistic

data set. As each method has its own strengths and weaknesses, statistical analysis often profits from the insights and perspectives offered by different techniques.

22.3 Clustering and classification methods

Cynthia G. Clopper

22.3.1 Introduction

Clustering, multidimensional scaling (MDS), and factor analysis are all data reduction methods that can be used to visualize and interpret the relationships between variables in high-dimensional spaces. Unlike most of the statistical analyses described in this chapter, clustering, MDS, and factor analysis do not involve rejection of a null hypothesis and do not return a p-value or other metric for assessing statistical significance. The researcher is therefore responsible for selecting and interpreting an appropriate model. Clustering analyses produce a tree (dendrogram) visualization of similarity data, allowing for the identification of hierarchical structure and/or subsets (clusters) of data within the larger set. Multidimensional scaling analyses produce a spatial representation of similarity data in one or more dimensions, in which distance in the space corresponds to dissimilarity, and allows for the identification of the primary dimensions of similarity. Factor analyses identify correlations among variables, allowing for the reduction of the data set to a smaller number of hidden, or unobserved, factors.

22.3.2 Research questions and data types

Clustering and MDS are well suited for exploring the similarity structure of a set of items, including identifying subgroups of similar items and the dimensions along which similarity is defined. In the domain of laboratory phonology, similarity may be defined in terms of perception or production, and may be computed over linguistic units, such as segments, words, or phrases, or over indexical units, such as talkers, dialects, or languages. The perceptual data used in clustering and MDS analyses are typically either confusion matrices of identification responses or explicit similarity rating or classification judgments. Items that are highly confusable, rated as highly similar, or classified together are interpreted as perceptually more similar than items that are less confusable, rated as less similar, or classified

separately. Clustering analyses have been used to examine the relationship between phonological features and perceptual confusions among vowels (Warner 2003) and consonants (Zhang et al. 1982), the phonetic similarity of unfamiliar languages (Bradlow et al. 2007), and the effects of native language on the perceptual similarity of linguistic tones (Gandour 1983) and regional dialects (Clopper and Bradlow 2009). The production data used in clustering and MDS analyses are typically distance metrics, such as difference scores, Euclidean distances in a multidimensional space, or Levenshtein distances, calculated from a set of acoustic (e.g. Heeringa et al. 2009) or phonetic (e.g. van de Velde and van Hout 1999; Heeringa et al. 2009) features.

MDS analyses have also been used to examine the effects of phonological structure and linguistic experience on the perceptual similarity of vowels (Fox 1983; Warner 2003), consonants (Goldstein 1977; Iverson and Kuhl 1996; Harnsberger 2001), tones (Gandour 1983; Francis et al. 2008), intonation contours (Grabe et al. 2003), talkers (Kreiman and Papcun 1991), dialects (Clopper and Pisoni 2007; Heeringa et al. 2009), and languages (Stockmal et al. 2000; Bradlow et al. 2007). Clustering and MDS techniques can also be used together to simultaneously explore the subgroupings of items within the larger set and the dimensions of similarity. For example, Warner (2003) used clustering to examine the hierarchical structure of phonological features in perceptual vowel similarity and MDS to determine the primary dimensions of similarity.

Factor analyses are used for data reduction in projects involving large numbers of independent variables that are correlated with one another. In the domain of laboratory phonology, these variables may be acoustic, articulatory, and/or perceptual. Factor analyses have been used to explore the relationships among different acoustic measures of the glottal source spectrum (Kreiman et al. 2007) and vowel variation across talkers (van Nierop et al. 1973), genders (Bachorowski and Owren 1999), and dialects (Clopper and Paolillo 2006), as well as articulatory measures of vowel production (Story 2005), and factors affecting lexical access in production (Bates et al. 2001). Bates et al. (2001) used factor analysis to reduce a set of fifteen intercorrelated variables related to lexical access to a smaller set of four interpretable factors representing the frequency, length, phonetic content, and meaning of the target word. Factor analysis results are often used in further statistical analyses to show the relationship between the underlying factors and other variables of interest. For example, the results of factor analyses on variable productions of consonants and vowels have been used to predict accentedness ratings (van Bezooijen and van Hout 1985) and to identify social categories such as age, ethnicity, gender, and social class (Horvath and Sankoff 1987).

Clustering and MDS analyses require square (N × N) matrices, where N is the number of items in the data set and the value of any given cell is a pairwise distance, similarity, or dissimilarity measure for the pair of items represented by that cell. For the examples of clustering and MDS analyses discussed in this section, the data set

was a square talker similarity matrix obtained from 22 listeners in an unpublished auditory free classification task (e.g. Clopper and Bradlow 2009). The stimulus materials included 20 male talkers (five from each of four American English regional dialects) producing the sentence *She had your dark suit in greasy wash water all year*. A subset of the 20 × 20 talker similarity matrix is shown in Table 22.3.1. The possible values in the cells range from 0 (for pairs of talkers who were not classified together by any of the listeners) to 22 (for pairs of talkers who were classified together by all of the listeners). Thus, larger numbers (e.g. Midland1 and Midland4) indicate greater perceptual similarity than smaller numbers (e.g. Midland5 and North2). The similarity between any talker and himself is 0 and the similarities are symmetric (North1 to North2 equals North2 to North1). Most implementations of clustering and MDS analyses assume that the distance between any item and itself is 0, and that the distance relationships between items are symmetric, although asymmetric similarities are theoretically possible, particularly for perceptual similarity data (e.g. North Korea is more similar to China than China is to North Korea, Tversky and Gati 1982).

Factor analyses require rectangular (N × M) matrices, where N is the number of items in the data set, M is the number of variables, and the number of variables is smaller than the number of items (M < N). For the factor analysis example discussed in this section, the data set was the rectangular matrix shown in Table 22.3.2. The 20 × 6 matrix includes six acoustic measures for each of the 20 talkers in the example free classification task. The measures were selected to reflect phonetic differences between the dialects (see Clopper and Bradlow 2009), including r-lessness in New England (Rhotic = F_3 midpoint – F_3 offset of /ɑ/ in *dark*); intrusive /r/ in the South (No Intrusive R = F_3 midpoint of /ɑ/ in *wash*); pronunciation of *greasy* as [grizi] in the South (Greazy = proportion of voicing of /s/

Table 22.3.1. A 10 × 10 square matrix showing the perceptual similarity of the five Northern (N) talkers and the five Midland (M) talkers in the sample free classification data

	N1	N2	N3	N4	N5	M1	M2	M3	M4	M5
North1	0	3	6	11	7	6	10	9	6	8
North2	3	0	5	4	3	4	8	3	6	2
North3	6	5	0	5	10	8	3	4	8	4
North4	11	4	5	0	8	11	5	9	7	8
North5	7	3	10	8	0	5	5	6	9	8
Midland1	6	4	8	11	5	0	6	11	12	8
Midland2	10	8	3	5	5	6	0	5	6	5
Midland3	9	3	4	9	6	11	5	0	6	10
Midland4	6	6	8	7	9	12	6	6	0	6
Midland5	8	2	4	8	8	8	5	10	6	0

Table 22.3.2. A 20 × 6 rectangular matrix showing the values for each of the six acoustic variables for the twenty talkers in the sample free classification data

Talker	Rhotic (Hz)	No Intrusive R (Hz)	Greazy (%)	Greasy Duration (s)	/u/ Retraction (Hz)	Speaking Rate (s)
NewEngland1	139	2350	0.00	0.367	926	4.46
NewEngland2	324	2095	0.22	0.347	393	4.23
NewEngland3	324	2709	0.00	0.286	347	4.88
NewEngland4	265	2533	0.06	0.398	885	4.45
NewEngland5	266	2589	0.20	0.302	487	4.56
North1	343	2213	0.07	0.352	432	4.30
North2	619	2378	0.00	0.390	487	4.20
North3	244	2190	0.08	0.370	796	3.83
North4	453	2356	0.07	0.358	730	4.72
North5	464	2312	0.00	0.297	299	4.49
Midland1	188	2412	0.00	0.327	841	4.21
Midland2	542	2334	0.00	0.331	398	4.23
Midland3	321	2334	0.00	0.334	520	4.33
Midland4	332	2235	0.00	0.380	465	4.00
Midland5	465	2412	0.00	0.353	332	4.34
South1	576	2423	0.00	0.385	420	4.02
South2	376	2113	0.00	0.363	166	4.84
South3	487	2445	1.00	0.197	487	4.24
South4	465	2190	1.00	0.242	244	4.19
South5	465	2257	1.00	0.249	420	5.08

in *greasy*, Greasy Duration = duration of /s/ in *greasy*); /u/ fronting in the Midland and South (/u/ Retraction = F2 midpoint of /u/ in *suit* normalized to F2 of /i/ in *year*); and speaking rate (Speaking Rate = duration of the sentence). Factor analysis variables must be numeric and continuous, but, like the variables in Table 22.3.2, do not need to share the same scale.

22.3.3 Clustering

Two different approaches to clustering, hierarchical and additive similarity, have been used in laboratory phonology and related fields. Both hierarchical and additive similarity models build trees iteratively by identifying the most similar items in the matrix, grouping them together, and then recalculating the matrix by treating the grouped items as a single unit. In the matrix in Table 22.3.1, the cell with the highest value represents the two talkers with the greatest similarity in the set (Midland1 and Midland4), and those two talkers would be grouped together in the first iteration. For hierarchical clustering analyses, different algorithms have been

developed for recalculating the similarity matrix at each iteration. These different methods define the similarity between clusters and individual items in different ways, and can, therefore, produce different results. For example, the Ward and complete methods are compact methods and tend to produce many small clusters that are later joined together. The single method is a chaining method and tends to add single items to existing clusters. Hierarchical clustering algorithms have been implemented in R and SPSS. Baayen (2008) provides examples of hierarchical clustering and R. Everitt et al. (2001) provide a comprehensive introduction to clustering analyses, with an entire chapter dedicated to hierarchical methods.

Figure 22.3.1 shows hierarchical clustering solutions for the talker similarity matrix obtained from the free classification task. The solution using the Ward method is on the left and the solution using the single method is on the right. Distance between items is represented in these figures by the height on the y-axis at which the items are connected. The height values reflect the model distance between items and do not have inherent units. Higher connections indicate more dissimilar objects, whereas lower connections indicate more similar objects. The distance between NewEngland2 and NewEngland3 in the left panel of Figure 22.3.1 is about 5, whereas the distance between NewEngland2 and North1 is about 60.

In clustering analyses, the researcher interprets the clusters by deciding where in the tree to make the cut between objects within a cluster and objects between clusters. The Ward method solution shown on the left in Figure 22.3.1 could be interpreted as showing three clusters with a break at Height $\approx$ 40 or as showing four clusters with a break at Height $\approx$ 30. The single method solution shown on the right in Figure 22.3.1 clearly shows three clusters with a break at Height $\approx$ 17. If we interpret the Ward method solution as having three clusters, the overall structure of the two solutions is similar with Southern, New England, and mixed Midland and Northern clusters. The structures of the Midland and Northern clusters exhibit the primary difference between the two methods: the Ward method clearly separated

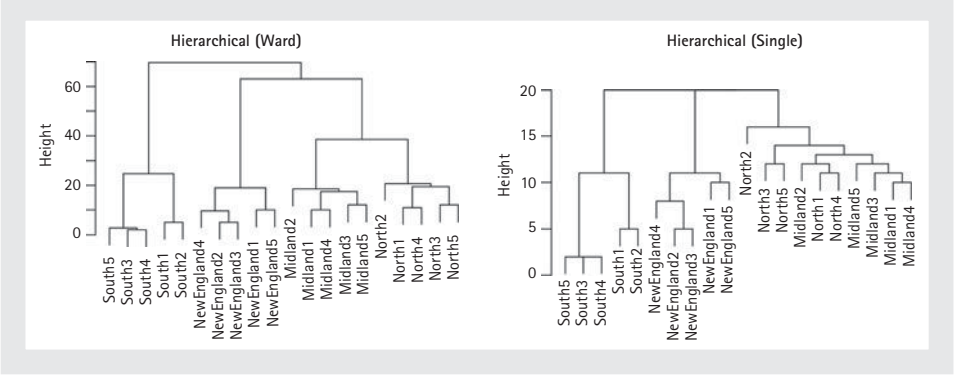

Figure 22.3.1. Hierarchical clustering solutions using the Ward (left) and single (right) clustering methods for the perceptual talker similarity data.

the Midland and Northern talkers into two smaller clusters, whereas the single method produced longer chains of mixed Midland and Northern talkers. The structures of the Southern and New England clusters are virtually identical in the Ward and single method solutions.

Given the strict hierarchical structure of these solutions, for any two clusters, all intracluster distances will be smaller than all intercluster distances. In addition, because the New England cluster is attached to the Northern cluster at about 60 in the Ward model, the distance between any New England talker and any Northern talker is also 60. Thus, for any two clusters, all intercluster distances are equal. These two distance relationships (for any two clusters, all intracluster distances are shorter than all intercluster distances, and all intercluster distances are equal) are intrinsic to hierarchical clustering and therefore hold for all hierarchical clustering solutions, but are intuitively false for many kinds of real data. In the free classification data, some Southern talkers may be more similar to the Midland talkers than others, but hierarchical clustering models cannot capture those differences.

Figure 22.3.2 shows the results of an additive similarity analysis of the free classification data. The additive similarity tree was obtained using Corter's (1982)

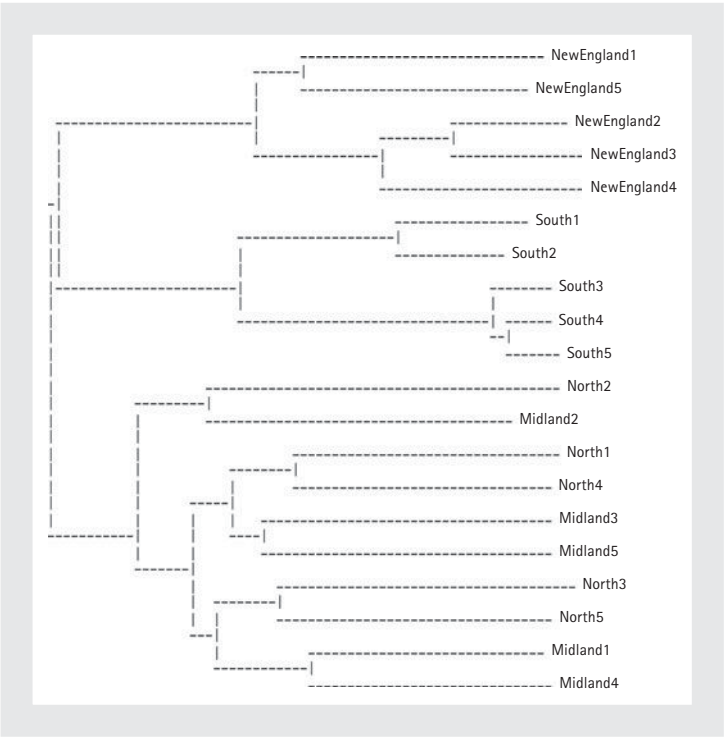

Figure 22.3.2. Additive similarity solution for the perceptual talker similarity data.

ADDTREE program, an implementation of Sattath and Tversky's (1977) Additive Similarity Tree model. Distance between items is represented by the lengths of the horizontal branches connecting the items. Longer branches indicate more dissimilar objects, whereas shorter branches indicate more similar objects. Thus, NewEngland2 and NewEngland3 are the most similar of the New England talkers, because the branches connecting them are shorter than the branches connecting any other pair of New England talkers. As in the hierarchical clustering solutions, the distance between NewEngland2 and North1 is larger than the distance between NewEngland2 and NewEngland3. However, unlike in the hierarchical models, the distance between NewEngland2 and North1 is shorter than the distance between NewEngland3 and North1, because NewEngland2 is closer to the root of the tree than NewEngland3. The overall structure of the additive similarity model is similar to the hierarchical models, with New England, Southern, and mixed Northern and Midland clusters. In the additive similarity model, the Northern and Midland talkers are mixed, similar to the single method hierarchical solution, but the structure is more compact and no chaining is observed. The structure of the New England and Southern clusters is highly similar across the three solutions.

The selection of the clustering model to interpret is based on considerations of the interpretability of the solution as well as the relationship between the data set and the model assumptions. In Figure 22.3.1, the Ward method might be preferred because the separate clusters of Midland and Northern talkers are highly interpretable, and it is less clear how to interpret the chaining of talkers in the single method solution. The additive similarity solution includes a mixed Midland and Northern cluster, but captures the relative similarity of talkers across clusters better than the hierarchical clustering solutions. In the additive similarity solution, South1 and South2 are more similar to the other dialects than South3, South4, and South5, whereas in the hierarchical models, the Southern talkers are all equally similar to the other talkers. Additive similarity clustering is more appropriate for modeling the similarity structure of data that do not exhibit the intracluster and intercluster distance relationships assumed by hierarchical clustering, but can also be used with data where those relationships hold. If hierarchical clustering is used, the linkage method should be chosen based on the interpretability of the solution.

22.3.4 Multidimensional scaling (MDS)

The two most common MDS models in laboratory phonology and related fields are non-metric MDS and individual differences scaling (INDSCAL). MDS analyses are iterative procedures that attempt to maximize the monotonicity of the relationship between the input similarity data and the output distance space. In non-metric and INDSCAL analyses, similarities in the data matrix are rank-ordered (from most to

least similar), and monotonicity is achieved if the rank ordering of the input data is preserved in the rank-ordered output distances (from closest to farthest away). Smaller distances in the MDS solution therefore correspond to greater similarity in the data matrix than larger distances. For the data in Table 22.3.1, Midland1 and Midland4 should be closer in the MDS space than Midland5 and North2. The lack of monotonicity in the output model is reflected in the stress (or badness-of-fit) of the model. Lower stress indicates better model fit than higher stress. Metric MDS analyses are also possible, but they treat the input matrix as ratio data, rather than ordinal data, and are therefore less flexible with respect to the kinds of data that they can be used to model. Kruskal and Wish (1978) provide an excellent introduction to the conceptual and numerical foundations of MDS.

The number of dimensions returned by MDS models is specified by the researcher. However, as a general rule, the number of items in the analysis should be greater than four times the number of dimensions. For an MDS analysis of the 20×20 talker similarity matrix obtained from the free classification task, the maximum number of dimensions is 4. As the number of dimensions increases, the number of parameters in the model also increases, and the fit of the model will improve. The number of dimensions to interpret is selected by the researcher by considering the relative fit and interpretability of models with different numbers of dimensions. The goal is to select a model with a small number of dimensions that has low stress and is interpretable. A scree plot is typically produced to examine the relationship between stress and dimensionality, as shown on the left in Figure 22.3.3 for four independent non-metric MDS analyses of the talker similarity data from the free classification task. The dimension selected for interpretation is usually at the elbow in the scree plot. That is, the selected dimensionality should substantially reduce stress from the next lowest dimension, but not be substantially worse than the next highest dimension. In Figure 22.3.3, the elbow is at two dimensions. The space is interpretable in two dimensions, so the two-dimensional space was selected for interpretation.

The right panel of Figure 22.3.3 shows the two-dimensional space produced by the MDS analysis of the talker similarity data. Non-metric MDS algorithms have been implemented in R and SPSS. The implementation in R is based on Kruskal's (1964) method, whereas the implementation in SPSS is based on Takane et al.'s (1977) ALSCAL model. In non-metric MDS analyses, interpretation of the space and the dimensions of similarity is not restricted to the dimensions returned by the model. The perceptual similarity space in Figure 22.3.3 could be rotated clockwise approximately 30° prior to interpretation, so that one dimension clearly separated the Southern and Northern talkers, and the other dimension clearly separated the Midland and New England talkers. In addition, while most implementations of non-metric MDS analyses center the space at the origin (0, 0), the space can be reflected across either axis and the scale of the space is arbitrary.

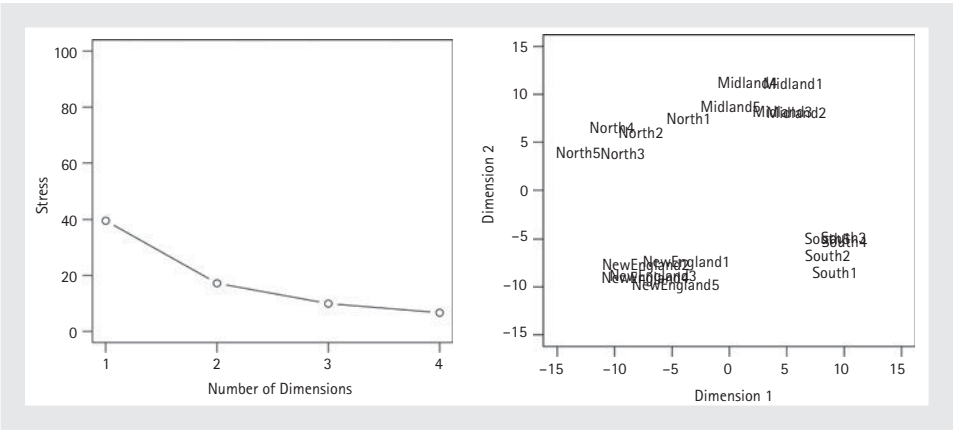

Figure 22.3.3. Scree plot (left) and two-dimensional non-metric MDS solution (right) for the perceptual talker similarity data.

The perceptual similarity space shown on the right in Figure 22.3.3 is also interpretable without rotation. The talkers from the two northern dialects (North and New England) are to the left of Dimension 1, whereas the talkers from the two non-northern dialects (Midland and South) are to the right of Dimension 1. The talkers from the two more stereotyped dialects (New England and South) are to the bottom of Dimension 2, whereas the talkers from the two less stereotyped dialects (North and Midland) are to the top of Dimension 2. As in any statistical analysis, the interpretation of the MDS solution is driven not only by the results themselves, but also by our knowledge and understanding of the data and how they were collected. Thus, the two dimensions of the unrotated similarity space are interpreted as reflecting two important aspects of regional dialect variation in the United States: geography (northern vs. non-northern) and stereotypes (more vs. less). The MDS solution is also consistent with the clustering analyses, and shows separate groups of New England and Southern talkers, but a more mixed group of Northern and Midland talkers.

The interpretation of the dimensions of an MDS solution can be confirmed by regression analyses demonstrating the relationship between the values of the items along a given dimension and some other measure related to the interpretation of that dimension, such as perceptual judgments of voice quality (Kreiman and Papcun 1991) or theoretical vowel features (Fox 1983). The interpretation of the dimensions in Figure 22.3.3 could be supplemented by correlating the values along each dimension for each talker with the acoustic measures shown in Table 22.3.2 to determine which acoustic properties are perceptually salient in the free classification task.

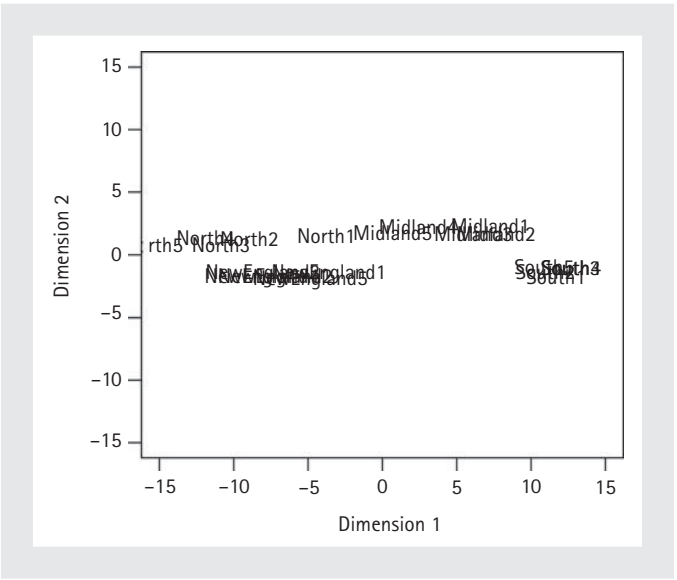

Figure 22.3.4. Hypothetical listener-specific INDSCAL solution with a stretched x-axis and a compressed y-axis.

Given that most MDS models produce solutions that can be rotated, reflected, and rescaled, it is not possible to directly compare two or more MDS solutions. However, INDSCAL analyses can be used to compare solutions across different participants, participant groups, or experimental conditions. For example, INDSCAL has been used to model the effects of native language on the perceptual similarity of tones (Gandour 1983; Francis et al. 2008), and the effect of native dialect on the perceptual similarity of vowels (Fox 1974) and regional dialects (Clopper and Pisoni 2007). The INDSCAL model was develop by Carroll and Chang (1970) and has been implemented in SPSS and Praat. The INDSCAL model accepts a series of square matrices (one per participant, group, or condition) and returns a single similarity space for the set of items, as well as weights for each dimension for each input matrix. The weights reflect the relative strength of each dimension for each participant, group, or condition, and can be visualized as stretching or shrinking the space. If some listeners attended more to geography than stereotypes in the free classification task, an INDSCAL model would return large Dimension 1 weights and smaller Dimension 2 weights for those listeners. Conceptually, the space would be stretched along the x-axis and compressed along the y-axis, as shown in Figure 22.3.4. Unlike non-metric MDS analyses with a single input matrix, INDSCAL solutions cannot be rotated and must be interpreted with respect to the dimensions that the model returns.

22.3.5 Factor analysis

The two most common factor analysis models in laboratory phonology and related fields are factor analysis and principal components analysis. Principal components analysis is a subtype of factor analysis, and the primary difference between them is that principal components analysis uses a single error term to represent all of the variables, whereas factor analysis assigns a different error term to each variable. When the scales of the variables differ (e.g. Hertz vs. seconds in Table 22.3.2), it may be inappropriate to assign the same error term to the distributions of all of the variables. Thus, factor analysis is more appropriate for modeling the structure of data sets that include variables with different scales and/or variances. Factor analysis, including principal components analysis, has been implemented in R and SPSS. Baayen (2008) and Johnson (2008) provide examples of factor analysis and principal components analysis and R. Kim and Mueller (1978a, b) provide a brief, but complete, introduction to factor analysis.

Like MDS solutions, factor analysis solutions can be reflected and rotated. In order to find a unique solution, however, the rotation method must be specified in advance in the analysis, and the resulting space cannot be rotated to improve interpretability. Rotation methods include varimax rotation, which maximizes the variance of the loadings for each factor; quartimax rotation, which maximizes the variance of the loadings for each variable; and oblique rotation, which permits non-orthogonal factors. Varimax rotation is the most commonly used rotation method in laboratory phonology and related fields (e.g. Clopper and Paolillo 2006) because it is more useful for data reduction than quartimax rotation and easier to interpret than oblique rotation. Principal components analysis solutions have a default varimax rotation. Thus, the results of factor analysis with varimax rotation and the results of principal components analysis are typically similar.

The number of factors to interpret is selected by the researcher by considering the eigenvalues and interpretability of the different factors. Typically, factors with eigenvalues greater than 1 are interpreted, although the number of factors to interpret can also be selected by considering the variance accounted for by each factor. The goal is to select a model with a small number of factors that together account for a large proportion of the variance and are interpretable. Eigenvalues can be converted to variance accounted for by dividing each eigenvalue by the total number of input variables. A scree plot can then be produced to examine the relationship between variance accounted for and number of factors. As in the interpretation of MDS solutions, the elbow in the scree plot can be used to select the number of factors to interpret. In the factor analysis with varimax rotation of the six acoustic measures in Table 22.3.2, the analysis returned three factors with eigenvalues greater than 1. However, in the scree plot in Figure 22.3.5, the elbow is at two factors and the third factor was difficult to interpret, so the first two factors were selected for interpretation. The first factor accounts for 38 percent of the variance

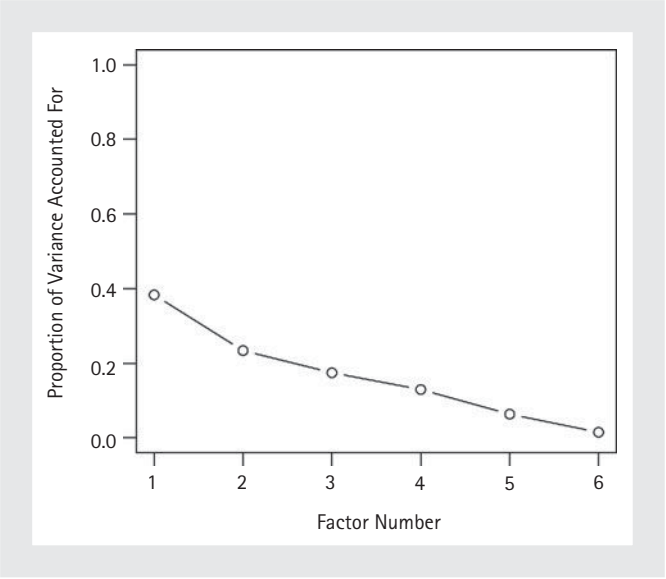

Figure 22.3.5. Scree plot for the factor analysis of the acoustic measures from the free classification stimulus materials.

and the second factor accounts for an additional 23 percent of the variance, for a total of 61 percent of the variance accounted for by the first two factors.

A factor analysis returns factor loadings for each of the variables and, optionally, factor scores for each of the items. The factor loadings of the six variables in the free classification task for the first two factors are shown on the left in Figure 22.3.6. Higher absolute factor loadings indicate greater association between that variable and that factor. For example, Greasy Duration was strongly negatively associated with Factor 1, whereas Greazy was strongly positively associated with Factor 1, suggesting that the variables Greasy Duration and Greazy were strongly negatively correlated. Factor 1 can be interpreted as representing the Southern pronunciation of *greasy* as [grizi]. The variables that were strongly associated with Factor 2 are /u/ Retraction and Rhotic. Rhotic was negatively associated with Factor 2, whereas /u/ Retraction was positively associated with Factor 2, suggesting that backed /u/ productions in *suit* were correlated with r-less productions of *dark* in the sentence analyzed. Factor 2 can be interpreted as representing the New England features of r-lessness and non-fronted /u/s. The other two variables, No Intrusive R and Speaking Rate, were not strongly associated with either factor. Thus, the factor analysis reduced the set of six intercorrelated acoustic variables to two factors that can be interpreted with respect to co-occurring phonetic variation for Southern and New England talkers.

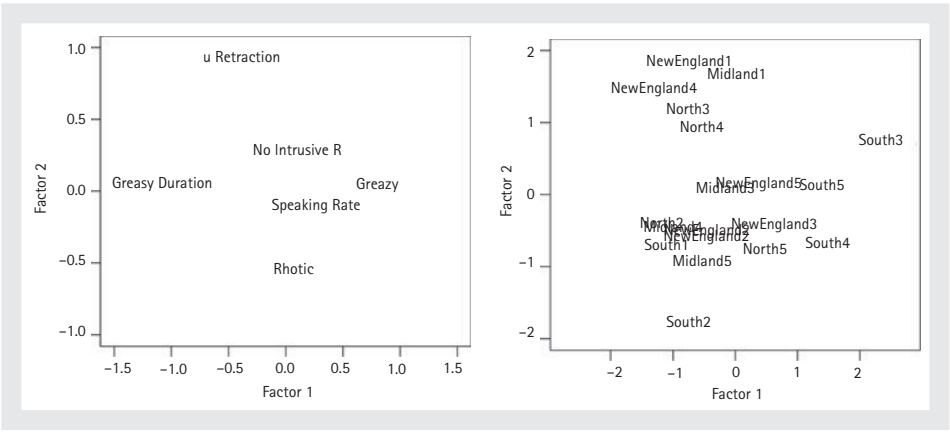

Figure 22.3.6. Factor loadings (left) and factor scores (right) from the two-factor analysis of the acoustic measures from the free classification stimulus materials.

The factor scores of the twenty talkers for the first two factors are shown on the right in Figure 22.3.6. The three talkers who pronounced *greasy* as [grizi] (South3, South4, South5) have high scores on Factor 1. Most of the other talkers, including the other two Southern talkers, have scores below 0 on Factor 1, indicating pronunciation of *greasy* as [grisi]. The talker with the most fronted /u/ (South2) has the lowest score on Factor 2. The talkers with the least constriction of /r/ (i.e. the most r-less productions) in *dark* (NewEngland1, NewEngland4) and with the most retracted /u/s in *suit* (Midland1, North3, North4) have the highest scores on Factor 2. Given that the factor analysis was based on acoustic data and the clustering and MDS analyses were based on perceptual data, the results of the three analyses cannot be directly compared. However, the three Southern talkers who pronounced *greasy* as [grizi] had high Factor 1 scores in the factor analysis and were grouped together in all three clustering analyses and in the MDS analysis, suggesting that [grizi] may be a perceptually salient dialect marker. In general, however, the two-dimensional factor space based on acoustic measures is quite different from the two-dimensional MDS space based on perceptual classification judgments, suggesting that the acoustic measures included in the factor analysis did not fully capture the information available to the listeners in the free classification task.

22.3.6 Summary and future directions

Clustering, MDS, and factor analysis methods have been fruitfully applied to research questions in laboratory phonology and related fields. Clustering and MDS analyses have been used mostly with perception data to explore the perceptual similarity of segmental, suprasegmental, and indexical properties of speech, whereas

factor analysis has been used mostly with production data to explore correlations among acoustic and articulatory measures. However, clustering and MDS analyses could also be used with production data if appropriate similarity metrics could be developed for comparing acoustic or articulatory measures (e.g. Heeringa et al. 2008). In addition, factor analysis could be used with perception data to explore the relationships between different types of tasks and/or responses to the same stimulus materials under different conditions.

This page intentionally left blank

References

Abercrombie, David (1965/1971). *Syllable Quantities and Enclitics in English: Studies in Phonetics and Linguistics*. 3rd edn. Oxford: Oxford University Press.

—— (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press and Chicago, IL: Aldine.

Abramson, Arthur (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech* 23: 19–325.

—— (1979). The coarticulation of tones: An acoustic study of Thai, in T. L. Thongkum, P. Kullavanijaya, V. Panupong, and K. Tingsabadh (eds.), *Studies in Tai and Mon-Khmer Phonetics and Phonology in honour of Eugenie J. A. Henderson*. Bangkok: Indigenous Languages of Thailand Research Project, 127–34.

—— (2004). The plausibility of phonetic explanations of tonogenesis, in G. Fant, H. Fujisaki, J. Cao, and Y. Xu (eds.), *From Traditional Phonology to Modern Speech Processing. Festschrift for Prof. Wu Zongji*. Beijing: Foreign Language Teaching and Research Press, 17–29.

—— and Lisker, Leigh (1985). Relative power of cues: F0 shift versus voice timing, in Victoria Fromkin (ed.), *Linguistic Phonetics: Essays in Honor of Peter Ladefoged*. New York: Academic Press, 25–31.

—— Nye, Patrick, and Luangthongkum, Theraphan (2007). Voice register in Khmu: Experiments in production and perception. *Phonetica* 64: 80–104.

Adank, Patti and Janse, Esther (2009). Perceptual learning of time-compressed and natural fast speech. *Journal of the Acoustical Society of America* 126: 2649–59.

—— Smits, Roel, and van Hout, Roeland (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116(5): 3099–107.

Agwuele, Augustine (2007). Tonal coarticulation in Yoruba: Locus equation analysis. *Journal of the Acoustical Society of America* 122: 3028.

Ahn, Mee-Jin (2000). Phonetic and functional bases of syllable weight for stress assignment. Ph.D. dissertation, University of Illinois, Urbana-Champaign.

Akamatsu, Tsutomu (1997). *Japanese Phonetics: Theory and Practice*. Munich: LINCOM Europa.

Akinlabi, Akin and Liberman, Mark (1995). On the phonetic interpretation of the Yoruba tonal system. *Proceedings of the 13th International Congress of Phonetic Sciences*, August 13–19, 1995. Stockholm, 42–5.

Alam, Farhana (2009). Language and identity among Scottish urban Pakistanis. Paper presented at 1st Sociolinguistics Summer School, University of Edinburgh, June 15, 2009.

Albareda-Castellot, Barbara, Pons, Ferran, and Sebastián-Gallés, Nuria (2011). The acquisition of phonetic categories in bilingual infants: New data from a new paradigm. *Developmental Science* 14: 395–401.

Albright, Adam. (2002a). Islands of reliability for regular morphology: Evidence from Italian. *Language* 78: 684–709.

—— (2002b). The identification of bases in morphological paradigms. Ph.D. dissertation, UCLA.

—— (2009). Modeling analogy as probabilistic grammar, in J. P. Blevins and J. Blevins (eds.), *Analogy in Grammar*. Oxford: Oxford University Press, 185–213.

—— Andrade, Argelia E., and Hayes, Bruce (2001). Segmental environments of Spanish diphthongization, in A. Albright and T. Cho (eds.), *UCLA Working Papers in Linguistics 7: Papers in Phonology 5*: 117–51.

—— and Hayes, Bruce (2003). Rules versus analogy in English past tenses: A computational-experimental study. *Cognition* 90: 119–61.

Alegre, María and Gordon, Peter (1999). Rule-based versus associate processes in derivational morphology. *Brain and Language* 68: 347–54.

Alexander, Joshua M. and Kluender, Keith (2008). Spectral tilt change in stop consonant perception. *Journal of the Acoustical Society of America* 123(1): 386–96.

Allen, George D. (1985). How the young French child avoids the pre-voicing problem for word-initial voiced stops. *Journal of Child Language* 12: 37–46.

Allen, Jont B. (1994). How do humans process and recognize speech? *IEEE Transactions on Speech and Audio Processing* 2(4): 567–77.

Allen, J. Sean and Miller, Joanne L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of the Acoustical Society of America* 106: 2031–9.

—— —— (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception and Psychophysics* 63: 798–810.

—— —— (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America* 115: 3171–83.

—— —— and DeSteno, David (2003). Individual talker differences in Voice-Onset-Time. *Journal of the Acoustical Society of America* 113(1): 544–52.

Allopenna, Paul, Magnuson, James S., and Tanenhaus, Michael K. (1998). Tracking the time course of spoken word recognition using eye-movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38(4): 419–39.

van Alphen, Petra and McQueen, James M. (2006). The effect of voice onset time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance* 32: 178–96.

Altmann, Eduardo G., Pierrehumbert, Janet B., and Motter, Adilson E. (2009). Beyond word frequency: Bursts, lulls, and scaling in the temporal distribution of words. *PLoS One* 4(11), e7678. doi:10.1371/journal.pone.0007678.

—— —— —— (2011). Niche as a determinant of word fate in online groups. *PLoS One* 6(5), e19009 doi:10.1371/journal.pone.0019009.

Altmann, Gerry T. M. (1997). *The Ascent of Babel: An Exploration of Language, Mind, and Understanding*. Oxford: Oxford University Press.

—— (2004). Language-mediated eye movements in the absence of a visual world: The blank screen paradigm. *Cognition* 93: 79–87.

—— and Kamide, Yuki (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73: 247–64.

—— —— (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world, in J. M. Henderson and F. Ferreira (eds.), *The Interface of Language,*

Vision and Action: Eye Movements and the Visual World. New York: Psychology Press, 347–85.

—— —— (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language* 57: 502–18.

Alwan, Abeer, Bangayan, Philbert, Gerratt, Bruce R., Kreiman, Jody, and Long, Christopher (1999). Analysis by synthesis of pathological voices using the Klatt synthesizer, in R. Kent (ed.), *Voice Quality Measurement*. San Francisco: Singular, 307–35.

Ananthakrishnan, Sankaranarayanan and Narayanan, Shrikanth (2008). Automatic prosody labeling using acoustic, lexical, and syntactic evidence. *IEEE Transactions on Speech, Audio and Language Processing* 16(1): 216–28.

Anderson, Anne H., Bader, Miles, Bard, Ellen G., Boyle, Elizabeth, Doherty, Gwyneth, Garrod, Simon, Isard, Stephen, Kowtko, Jacqueline, McAllister, Jan, Miller, Jim, Sotillo, Catherine, Thompson, Henry S., and Weinert, Regina (1991). The HCRC Map Task Corpus. *Language and Speech* 34: 351–66.

Anderson, Jennifer, Morgan, James L., and White, Katherine S. (2003). A statistical basis for speech sound discrimination. *Language and Speech* 46(2–3): 155–82.

Anderson, John M., and Ewen, Colin J. (1987). *Principles of Dependency Phonology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. (1978). Tone features, in V. Fromkin (ed.), *Tone: A Linguistic Survey*. New York: Academic Press, 133–73.

—— (1981). Why phonology isn't "natural." *Linguistic Inquiry* 12: 493–539.

Andruski, Jean E. (2006). Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics* 34: 388–404.

—— Blumstein, Sheila E., and Burton, Martha W. (1994). The effect of subphonetic differences on lexical access. *Cognition* 52: 163–87.

—— and Ratliff, Martha (2000). Phonation types in production of phonological tone: The case of Green Mong. *Journal of the International Phonetic Association* 30: 37–61.

Angermeyer, Philip (2003). Copying contiguous gestures: An articulatory account of Bella Coola reduplication, in E. Kaiser and S. Arunachalam (eds.), *Proceedings of the 26th Annual Penn Linguistics Colloquium*, University of Pennsylvania Working Papers in Linguistics 9.1. Philadelphia: Penn Linguistics Club, 17–30.

Anttila, Arto (1997). Deriving variation from grammar, in F. Hinskens, R. van Hout, and L. Wetzels (eds.), *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins, 35–68.

—— (2002a). Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20: 1–42.

—— (2002b). Variation and phonological theory, in J. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *Handbook of Language Variation and Change*. Malden, MA and Oxford: Blackwell, 206–43.

—— (2007a). Variation and optionality, in Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, 519–36.

—— (2007b). Word stress in Finnish. Handout of a talk given at the Linguistic Society of America Annual Meeting, Anaheim, California, January 7, 2007.

—— (2008a). Gradient phonotactics and the complexity hypothesis. *Natural Language and Linguistic Theory* 26(4): 695–729.

—— (2008b). Word stress in Finnish. MS, Stanford University, Stanford, California.

Anttila, Arto and Andrus, Curtis (2006). T-order generator [computer program], Stanford University. <http://www.stanford.edu/~anttila/research/software.html. [ROA-873]>.

—— and Cho, Young-Mee Yu (1998). Variation and change in Optimality Theory. *Lingua* 104: 31–56. Special issue on Conflicting Constraints.

—— Adams, Matthew, and Speriosu, Michael (2010). The role of prosody in the English dative alternation. *Language and Cognitive Processes* 25(7/8/9): 946–81.

—— Fong, Vivienne, Benus, Stefan, and Nycz, Jennifer (2008). Variation and opacity in Singapore English consonant clusters. *Phonology* 25(2), 181–216. [ROA-981].

Aoyama, Katsura, Flege, James Emil, Guion, Susan G., Akahane-Yamada, Reiko, and Yamada, Tsuneo (2004). Perceived phonetic distance and L2 learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32: 233–50.

—— and Guion, Susan (2007), Prosody in second-language acquisition: Acoustic analyses of duration and F0 change, in O.-S. Bohn and M. Munro (eds.), *Language Experience in Second-language Speech Learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 281–97.

Apfelbaum, Keith and McMurray, Bob (2011). Successes and failures in early word learning: An emergent property of basic learning principles. *Cognitive Science* 35(6): 1105–37.

Arai, Takayuki. (1999). A case study of spontaneous speech in Japanese. *Proceedings of the 14th International Congress of Phonetic Sciences* (ICPhS), San Francisco, 1: 615–18.

Arbib, Michael A. (ed.) (2006). *Action to Language via the Mirror Neuron System*. Cambridge: Cambridge University Press.

Archangeli, Diana (1988). Aspects of underspecification theory. *Phonology* 5: 183–207.

—— and Pulleyblank, Douglas (1994). *Grounded Phonology*. Cambridge, MA: MIT Press.

—— —— (2007). Harmony, in Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, 353–78.

Arvaniti, Amalia (1998). Phrase accents revisited: Comparative evidence from Standard and Cypriot Greek, in *Proceedings of the 5th International Conference on Spoken Language Processing*, 7, 2883–6.

—— (2007a). On the relationship between phonology and phonetics (Or why phonetics is not phonology), in *Proceedings of the 16th International Congress of Phonetic Sciences* (Special Session: Between Meaning and Speech: On the Role of Communicative Functions, Representations and Articulations), 19–24.

—— (2007b). Greek phonetics: The state of the art. *Journal of Greek Linguistics* 8: 97–208.

—— (2007c). On the presence of final lowering in British and American English, in T. Riad and C. Gussenhoven (eds.), *Tones and Tunes*, vol. 2. Berlin: Mouton de Gruyter, 317–47.

—— (2011). Levels versus configuration and the representation of intonation, in M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice (eds.), *The Blackwell Companion to Phonology*. Malden, MA & Oxford: Wiley-Blackwell, 757–80.

—— and Baltazini, Mary (2005). Intonational analysis and prosodic annotation of Greek spoken corpora, in Sun-Ah Jun (ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 84–117.

—— and Garding, Gina (2007). Dialectal variation in the rising accents of American English, in J. Cole and J. H. Hualde (eds.), *Laboratory Phonology* 9. Berlin and New York: Mouton de Gruyter, 547–76.

—— and Godjevac, Svetlana (2003). The origins and scope of final lowering in English and Greek, in M. J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: UAB, 1077–80.

——and Ladd, D. Robert (1995). Tonal alignment and the representation of accentual targets, in *Proceedings of the 13th International Congress of Phonetic Sciences*, 4, 220–3.
—— —— (2009). Greek wh-questions and the phonology of intonation, *Phonology* 26: 43–74.
—— ——and Mennen, Ineke (1998). Stability of tonal alignment: The case of Greek prenuclear accents, *Journal of Phonetics* 26: 3–25.
—— —— —— (2000). What is a starred tone? Evidence from Greek, in M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 119–31.
—— —— —— (2006a). Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication* 48: 667–96.
—— —— —— (2006b). Tonal association and tonal alignment: Evidence from Greek polar questions and contrastive statements. *Language and Speech* 49: 421–50.
Ash, Sharon and Myhill, John (1986). Linguistic correlates of inter-ethnic contact, in D. Sankoff (ed.), *Diversity and Diachrony*. Amsterdam: John Benjamins, 33–44.
Astruc, Lluïsa, Prieto, Pilar, Payne, Elinor, Post, Brechtje, and Vanrell, Maria del Mar (under review). Tonal targets in early child Catalan, Spanish, and English. *Language and Speech*.
Asu, Eva Liina. (2004). The phonetics and phonology of Estonian intonation. Ph.D. dissertation, University of Cambridge.
Atal, B. S., Chang, J. J., Mathews, M. V., and Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformations in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America* 64: 1535–55.
Atkins, Joseph E., Jacobs, Robert A., and Knill, David C. (2003). Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. *Vision Research* 43: 2603–13.
Atterer, Michaela and Ladd, D. Robert (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics* 32: 177–97.
Auger, Julie (2001). Phonological variation and Optimality Theory: Evidence from word-initial vowel epenthesis in Picard. *Language Variation and Change* 13: 253–303.
——and Villeneuve, A.-J. (2008). Ne deletion in Picard and in regional French: Evidence for distinct grammars, in M. Meyerhoff and N. Nagy (eds.), *Social Lives in Language: Sociolinguistics and Multilingual Speech Communities* [Celebrating the work of Gillian Sankoff]. Amsterdam: Benjamins, 223–47.
Avery, Peter and Rice, Keren (1989). Segment structure and coronal underspecification. *Phonology* 6: 179–200.
Aylett, Matthew (2000). Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech. Ph.D. dissertation, University of Edinburgh.
——and Turk, Alice (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1): 31–56.
Baars, Bernard J., Motley, Michael T., and MacKay, Donald G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior* 14: 382–91.
Baayen, R. Harald (2002). *Word Frequency Distributions*. Dordrecht: Klewer Academic Publishers.

Baayen, R. Harald (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

—— (2009). languageR: Data sets and functions with *Analyzing Linguistic Data: A Practical Introduction to Statistics*, <http://CRAN.R-project.org/package=languageR>. R package version 0.955.

—— Davidson, Doug J., and Bates, Doug (2008). Mixed-effects modeling with crossed random effects for subjects and items, *Journal of Memory and Language* 59: 390–412.

—— and Milin, Petar (2010). Analyzing reaction times. *International Journal of Psychological Research* 3: 12–28.

—— Piepenbrock, Richard, and Gulikers, Leon (1995). The CELEX Lexical Database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium.

—— Wurm, Lee H., and Aycock, Joanna (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon* 2: 419–63.

Babel, Molly E. (2009). Phonetic and social selectivity in speech accommodation. Ph.D. dissertation, Department of Linguistics, University of California, Berkeley, CA.

Bachorowski, Jo-Anne and Owren, Michael J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America* 106: 1054–63.

Badin, Pierre (1989). Acoustics of voiceless fricatives: Production theory and data. *STL-QPSR* 30(3): 33–55.

—— Beautemps, Denis, Laboissière, Rafael, and Schwartz, Jean-Luc (1995). Recovery of vocal tract geometry from formants for vowels and fricative consonants using a midsagittal-to-area function conversion model. *Journal of Phonetics* 23: 221–9.

—— Hertegård, Stellan, and Karlsson, Inger (1990). Notes on the Rothenberg mask. *STL-QPSR* 31(1): 1–7.

—— Perrier, Pascal, Boe, Louis-Jean, and Abry, Christian (1991). Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *Journal of the Acoustical Society of America* 87: 1290–300.

—— Shadle, Christine H., Pham Thi Ngoc, Y., Carter, J N., Chiu, Wilson S. C., Scully, Celia, and Stromberg, Karen (1994). Frication and aspiration noise sources: Contribution of experimental data to articulatory synthesis. *Proceedings of ICSLP* 94, vol. 1. Yokohama, 163–6.

Baer, Tom, Gore, John C., Gracco, L. Carol, and Nye, Patrick W. (1991). Analysis of vocal-tract shape and dimensions using magnetic-resonance imaging: Vowels. *Journal of the Acoustical Society of America* 90: 799–828.

—— Löfqvist, Anders, and McGarr, Nancy S. (1983). Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. Haskins Laboratories Status Report on Speech Research SR-73: 283–91.

Bagshaw, Andrew P., Kobayashi, Eliane, and Dubeau, François (2006). Correspondence between EEG-fMRI and EEG dipole localisation of interictal discharges in focal epilepsy. *Neuroimage* 30: 417–25.

Bailey, Guy and Thomas, Erik (1998). Some aspects of African-American vernacular phonology, in S. S. Mufwene, J. R. Rickford, G. Bailey, and J. Baugh (eds.), *African-American English*. London: Routledge, 85–109.

—— Wikle, Tom, Tillery, Jan, and Sand, Lori (1991). The apparent time construct. *Language Variation and Change* 3: 241–64.

Bailey, Todd M. and Hahn, Ulrike (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44: 568–91.

—— —— (2005). Phoneme similarity and confusability. *Journal of Memory and Language* 52: 347–70.

Bailly, G., Laboissière, R., and Schwartz, J. L. (1991). Formant trajectories as audible gestures: An alternative for speech synthesis. *Journal of Phonetics* 19: 9–23.

Baken, Ronald J. (1996). *Clinical Measurement of Speech and Voice*. San Diego: Singular.

Baker, Carl L. and Brame, Michael K. (1972). Global rules: A rejoinder. *Language* 48: 51–75.

Balling, Laura and Baayen, R. Harald (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes* 23: 1159–90.

Baltazani, Mary (2006a). Intonation and pragmatic interpretation of negation in Greek. *Journal of Pragmatics* 38: 1658–76.

—— (2006b). Focusing, prosodic phrasing, and hiatus resolution in Greek, in L. Goldstein, D. Whalen, and C. Best (eds.), *Laboratory Phonology* 8. Berlin: Mouton de Gruyter, 473–94.

Bangayan, P., Alwan, Abeer, and Narayanan, S. (1996). From MRI and acoustic data to articulatory synthesis: A case study of the lateral approximants in American English. *Proceedings of the Fourth International Conference on Spoken Language Processing*, Philadelphia (ICSLP 96) 2, 793–6.

Baran, Jane A., Laufer, Marsha Z., and Daniloff, Ray (1977). Phonological contrastivity in conversation: A comparative study of Voice Onset Time. *Journal of Phonetics* 5: 339–50.

Bard, Ellen G., Anderson, Anne, Sotillo, Catherine, Aylett, Matthew, Doherty-Sneddon, Gwyneth, and Newlands, Alison (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 42: 1–22.

—— Robertson, Dan, and Sorace, Antonella (1996). Magnitude estimation of linguistic acceptability. *Language* 72: 32–68.

—— Sotillo, Catherine, Kelly, M. Louise, and Aylett, Matthew P. (2001). Taking the hit: Leaving some lexical competition to be resolved post-lexically. *Language and Cognitive Processes* 16: 731–7.

Barnes, Jonathan, Shattuck-Hufnagel, Stefanie, Brugos, Alejna, and Veilleux, Nanette (2006). The domain of realization of the L-Phrase Tone in American English, in *Speech Prosody 2006*, <http://aune.lpl.univ-aix.fr/~sprosig/sp2006/contents/papers/PS3-11_0163.pdf>.

—— Veilleux, Nanette, Brugos, Alejna, and Shattuck-Hufnagel, Stefanie (2008). Alternatives to F0 turning points in American English intonation. *Journal of the Acoustical Society of America* 124: 2497.

Barr, Dale J. (2008). Analyzing "visual world" eye-tracking data using multilevel logistic regression. *Journal of Memory and Language: Special issue on emerging data analysis techniques* 59: 457–74.

Barrie, Michael. (2007). Contour tones and contrast in Chinese languages. *Journal of East Asian Linguistics* 16: 337–62.

Barry, Morgan. (1991). Temporal modeling of gestures in articulatory assimilation, in *Proceedings of the 12th International Congress of Phonetic Sciences*. Aix-en-Provence: University of Provence, 14–17.

—— (1992). Palatalisation, assimilation and gestural weakening in connected speech. *Speech Communication* 11: 393–400.

Barry, William and Andreeva, Bistra (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association* 31: 51–66.

—— Klein, Cordula, and Köser, Stephanie (1999). Speech production evidence for ambisyllabicity in German. *Phonus* 4: 87–102 (Institute of Phonetics, University of the Saarland).

Bartels, Christine and Kingston, John (1994). Salient pitch cues in the perception of contrastive focus, in P. Bosch and R. van der Sandt (eds.), *Focus and Natural Language Processing*. IBM Working Papers on Logic and Linguistics 6. Heidelberg, 1–10.

Bates, Doug and Maechler, Martin (2009). lme4: Linear mixed-effects models using S4 classes. <http://CRAN.R-project.org/package=lme4>. R package version 0.999375-32.

—— —— (2010). lme4: Linear mixed-effects models using S4 classes. <http://CRAN.R-project.org/package=lme4>. R package version 0.999375-33.

Bates, Elizabeth, Burani, Cristina, D'Amico, Simona, and Barca, Laura (2001). Word reading and picture naming in Italian. *Memory and Cognition* 29: 986–99.

Baum, Shari and McNutt, James (1990). An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics* 18: 51–63.

Baumann, Stefan, Becker, Johannes, Grice, Martine, and Mücke, Doris (2007). Tonal and articulatory marking of focus in German, in J. Trouvain and W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: Universität des Saarlandes, 1029–32.

—— Grice, Martine, and Steindamm, Susanne (2006). Prosodic marking of focus domains: Categorical or gradient?, in *Speech Prosody 2006*, <http://aune.lpl.univ-aix.fr/~sprosig/sp2006/contents/papers/PS3-09_0065.pdf>.

Baxter, G. J., Blythe, Richard A., Croft, William, and McKane, Alan J. (2009). Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21: 257–96.

Bayley, Robert (1994). Consonant cluster reduction in Tejano English. *Language Variation and Change* 6: 303–26.

—— (2002). The Quantitative Paradigm, in J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell, 117–41.

Beaver, David., Clark, Brady Z., Flemming, Edward, Jaeger, Florian T., and Wolters, Maria (2007). When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language* 83: 245–76.

Becker, Frank and Reinvang, Ivar (2007). Mismatch negativity elicited by tones and speech sounds: Changed topographical distribution in aphasia. *Brain and Language* 100: 69–78.

Becker, Michael (2009). Phonological trends in the lexicon: The role of constraints. Ph.D. dissertation, University of Massachusetts, Amherst.

—— Ketrez, Nihan F., and Nevins, Andrew (2011). The surfeit of the stimulus: Analytical biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1): 84–125.

Beckman, Mary E. and Edwards, Jan (1990). Lengthenings and shortenings and the nature of prosodic constituency, in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 179–200.

—— —— (1992). Intonational categories and the articulatory control of duration, in E. Vatikiotis-Bateson, Y. Tohkura, and Y. Sagisaka (eds.), *Speech Perception, Production, and Linguistic Structure*. Tokyo: OHM Publishing, 356–75.

——— ——— (1994). Articulatory evidence for differentiating stress categories, in P. A. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, 7–33.

——— ——— (2000a). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development* 71: 240–9.

——— ——— (2000b). Lexical frequency effects on young children's imitative productions, in M. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 208–17.

——— ——— (2010). Generalizing over lexicons to predict consonant mastery. *Laboratory Phonology* 1(2): 319–43.

——— ——— and Fletcher, Janet (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics, in G. J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, 68–86.

——— Hirschberg, Julia and Shattuck-Hufnagel, Stefanie (2005). The original ToBI system and the evolution of the ToBI framework, in S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 9–54.

——— de Jong, Kenneth, Jun, Sun-Ah, and Lee, Sook-Hyang (1992). The interaction of coarticulation and prosody in sound change. *Language and Speech* 35: 45–8. Cambridge: Cambridge University Press.

——— and Kingston, John (1990). Introduction, in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 1–16.

——— Munson, Benjamin, and Edwards, Jan (2007). The influence of vocabulary growth on developmental changes in types of phonological knowledge, in J. Cole and J. Hualde (eds.), *Laboratory Phonology 9*. New York: Mouton de Gruyter, 241–64.

——— and Pierrehumbert, Janet B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook* 3: 255–310.

——— ——— (2003). Interpreting "phonetic interpretation" over the lexicon, in J. Local, R. Ogden, and R. Temple (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 13–37.

——— Yoneyama, Kiyoko, and Edwards, Jan (2003). Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan* 7: 18–28.

Beddor, Patrice S. (2007). Nasals and nasalization: The relationship between segmental and coarticulatory timing, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 249–54.

——— Brasher, Anthony, and Narayan, Chandan (2007). Applying perceptual methods to the study of phonetic variation and sound change, in M.-J. Solé, P. Beddor, and M. Ohala (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 127–43.

——— Harnsberger, James D., and Lindemann, Stephanie (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30: 591–627.

——— Krakow, Rena, and Goldstein, Louis (1986). Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology Yearbook* 3: 197–217.

Beddor, Patrice S. and Lindemann, Stephanie (2001). Patterns of perceptual compensation and their phonological consequences, in E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*. San Diego: Academic Press, 55–78.

Bell, Alan (1984). Language style as audience design. *Language in Society* 13: 145–204.

—— Brenier, Jason, Gregory, Michelle, Girand, Cynthia, and Jurafsky, Dan (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60: 92–111.

—— and Hooper, Joan B. (eds.) (1978). *Syllables and Segments.* Amsterdam: North-Holland.

—— Jurafsky, Daniel, Fosler-Lussier, Eric, Girand, Cynthia, Gregory, Michelle, and Gildea, Daniel (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113: 1001–24.

Bell-Berti, Frederica (1980). Velopharyngeal function: A spatio-temporal model, in N. J. Lass (ed.), *Speech and Language: Advances in Basic Research and Practice*, vol. IV. New York: Academic Press, 291–316.

Benkí, José (1998). Evidence for phonological categories from speech perception. Ph.D. dissertation, University of Massachusetts, Amherst.

Bennett, Clinton W. and Ingle, Bobbie H. (1984). Production of /s/ as a function of word frequency, phonetic environment, and phoneme position. *Journal of Communication Disorders* 17: 361–9.

Benus, Stefan (2005). Dynamics and transparency in vowel harmony. Ph.D. dissertation, New York University.

—— and Gafos, Adamantios (2005). Qualitative and quantitative aspects of vowel harmony: A dynamics model, in B. G. Bara, L. Barsalou, and M. Bucciarelli (eds.), *CogSci2005, XXVII Annual Conference of the Cognitive Science Society*, Stresa, Italy, 2005. New Jersey: Lawrence Erlbaum, 226–31.

—— —— (2007). Articulatory characteristics of Hungarian "transparent" vowels. *Journal of Phonetics* 35: 271–300.

—— —— and Goldstein, Louis (2004). Phonetics and phonology of transparent vowels in Hungarian, in P. M. Nowak, C. Yoquelet, and D. Mortensen (eds.), *Proceedings of the 29th Annual Meeting of the Berkeley Linguistic Society*. Berkeley Linguistic Society, 485–97.

—— Smorodinsky, Iris, and Gafos, Adamantios (2004). Gestural coordination and the distribution of English "geminates," in S. Arunachalam and T. Scheffler (eds.), *Proceedings of the 27th Annual Penn Linguistic Colloquium*. University of Pennsylvania Working Papers in Linguistics 10.1. Philadelphia: Penn Linguistics Club, 33–46.

Beranek, Leo (1954). *Acoustics*. New York: McGraw-Hill.

—— (1988). *Acoustical Measurements*, rev. edn. Published for the Acoustical Society of America. New York: American Institute of Physics.

Berent, Iris, Lennertz, Tracy, Smolensky, Paul, and Vaknin-Nusbaum, Vered (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology* 26: 75–108.

van den Berg, Janwillem (1956). Direct and indirect determination of the mean subglottic pressure. *Folia Phoniatrica et Logopaedica* 8: 1–24.

van den Berg, R. (1986). The effect of varying voice and noise parameters on the perception of voicing in Dutch two-obstruent sequences. *Speech Communication* 5: 355–67.

van Bergem, Dick R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress and vowel class. *Speech Communication* 12: 1–23.

Bergen, Benjamin K. (2004). The psychological reality of phonaesthemes. *Language* 80: 290–311.

Berinstein, Ava E. (1979). A cross-linguistic study on the perception and production of stress. *UCLA Working Papers in Phonetics* 47: 1–59.

Berko, Jean (1958). The child's learning of English morphology. Reprinted 2004 in B. Lust and C. Foley (eds), *First Language Acquisition: The Essential Readings*. Oxford: Blackwell, 253–73.

Berkovits, Rochele (1994). Durational effects in final lengthening, gapping, and contrastive stress. *Language and Speech* 37(3): 237–50.

Berlin, Brent, and Kay, Paul (1991). *Basic Color Terms: Their Universality and Evolution.* Berkeley, CA: University of California Press.

Bermúdez-Otero, Ricardo (2006). Phonological change in optimality theory, in K. Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn, vol. 9. Oxford: Elsevier, 497–505.

Bernhardt, Barbara, Gick, Bryan, Bacsfalvi, Penelope, and Adler-Bock, Marcy (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics and Phonetics* 19: 605–16.

——— ——— ——— and Ashdown, Julie (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics and Phonetics* 17(3): 199–216.

Bernstein, Nikolai (1967). *Coordination and Regulation of Movement*. New York: Pergamon Press.

Bernstein-Ratner, Nan (1982). Acoustic study of mothers' speech to language-learning children: An analysis of vowel articulatory characterstics. Doctoral dissertation, Boston University.

——— (1987). The phonology of parent-child speech, in K. E. Nelson and A. van Kleeck (eds.), *Children's Language*, vol. 6. Hillsdale, NJ: Erlbaum, 159–74.

Bertinetto, Pier Marco (1999). Psycholinguistic evidence for syllable geometry: Italian and beyond, in J. Rennison and K. Kühnhammer (eds.), *Phonologica 1996. Syllables!?* The Hague: Holland Academic Graphics, 1–28.

——— (2001). The syllable: Fragments of a puzzle, in C. Schaner-Wolles, J. R. Rennison, and F. Neubarth (eds.), *Naturally! Linguistic Studies in Honour of Wolfgang Ulrich Dressler Presented at the Occasion of his 60th Birthday*. Torino: Rosenberg & Sellier, 35–45.

Bertrand, Roxane, Blache, Philippe, Espesser, Robert, Ferre, Gaëlle, Meunier, Christine, Priego-Valverde, Béatrice, and Rauzy, Stéphane (2008). *Le CID—Corpus of Interactional Data—Annotation et Exploitation Multimodale de Parole Conversationnelle*. Traitement Automatique des Langues, vol. 49, no. 3.

Besle, Julien, Fischer, Catherine, Bidet-Caulet, Aurélie, Lecaignard, Francoise, Bertrand, Olivier, and Giard, Marie-Hélène (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *Journal of Neuroscience* 28: 14301–10.

Best, Catherine T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development, in B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year*. Dordrecht: Kluwer Academic, 289–304.

Best, Catherine T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model, in J. Goodman and H. Nusbaum (eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge, MA: MIT Press, 167–224.

—— (1995). A direct realist view of cross-language speech perception, in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press, 171–204.

—— Halle, Pierre A., Bohn, Ocke-Schwen, and Faber, Alice (2003). Cross-language perception of non-native vowels: Phonological and phonetic effects of listeners' native languages, in M. J. Sole, D. Recsens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions, 2889–92.

—— and Jones, Cathleen (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development* 21: 295.

—— McRoberts, Gerald. W., and Goodell, Elizabeth (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109: 775–94.

—— —— LaFleur, Rosemarie, and Silver-Isenstadt, Jean (1995). Divergent developmental patterns for infants' perception of two nonnative speech contrasts. *Infant Behavior and Development* 18: 339–50.

—— —— and Sithole, Nomathemba M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345–60.

—— and Tyler, Michael (2007). Nonnative and second-language speech perception: Commonalities and complementarities, in O.-S. Bohn and M. Munro (eds.), *Language Experience in Second-Language Speech Learning. In honor of James Emil Flege*. Amsterdam: John Benjamins, 13–34.

van Bezooijen, Renée and van Hout, Roeland (1985). Accentedness ratings and phonological variables as measures of variation in pronunciation. *Language and Speech* 28: 129–42.

Bharucha, Jamshed J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception* 5: 1–30.

Bialystok, Ellen and Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in Click for Articleage differences for second language acquisition, in D. Birdsong (ed.), *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Biber, Douglas (2004). Representativeness in corpus design, in G. Sampson and D. McCarthy (eds.), *Corpus Linguistics: Readings in a Widening Discipline*. London and New York: Continuum International, 174–97.

—— Conrad, Susan, and Reppen, Randi (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge Approaches to Linguistics. Cambridge: Cambridge University Press.

Bickmore, Lee (1995). Tone and stress in Lamba. *Phonology* 12: 307–41.

Bikhchandani, Sushil, Hirshleifer, David, and Welch, Ivo (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives* 12(3): 151–70.

Bird, Steven (1995). *Computational Phonology: A Constraint-Based Approach.* Cambridge: Cambridge University Press.

—— (2001). Linguistic annotation, <http://www.ldc.upenn.edu/annotation>, accessed March 13, 2009.

—— and Klein, Ewan (1990). Phonological events. *Journal of Linguistics* 26: 33–56.

—— and Liberman, Mark (1999). Annotation graphs as a framework for multidimensional linguistic data analysis, in *Proceedings, Towards Standards and Tools for Discourse Tagging Workshop*, Association for Computational Linguistics.

Birdsong, David (1992). Ultimate attainment in second language acquisition. *Language* 68: 706–55.

—— (2007). Nativelike pronunciation among late learners of French as a second language, in O.-S. Bohn and M. Munro (eds.), *Language Experience in Second-language Speech Learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 99–116.

Bishop, Dorothy V. M. and Hayiou-Thomas, Marianna E. (2008). Heritability of specific language impairment depends on diagnostic criteria. *Genes, Brain and Behavior* 7: 365–72.

Blacklock, Oliver. S. (2004). Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods. Ph.D. dissertation, School of Electronics and Computer Science, University of Southampton, UK.

Bladon, Anthony (1986). Phonetics for hearers, in Graham McGregor (ed.), *Language for Hearers*. Oxford: Pergamon Press, 1–24.

Bleses, Dorothe (2008). The struggle of Danish word-learning babies: The role of sound structure in word learning in a cross-linguistic framework. Presentation at the First Nijmegen Speech Reduction Workshop, MPI, Nijmegen, The Netherlands.

Blevins, Juliette (1995). The syllable in phonological theory, in J. A. Goldsmith (ed.), *The Handbook of Phonological Theory*. Cambridge, MA: Blackwell, 206–44.

—— (2003). The independent nature of phonotactic constraints: An alternative to syllable-based approaches, in C. Féry and R. van de Vijver (eds.), *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press, 375–403.

—— (2004). *Evolutionary Phonology: The Emergence of Sound Patterns.* Cambridge: Cambridge University Press.

—— and Garrett, Andrew (1998) The origins of consonant-vowel metathesis. *Language* 74: 508–56.

—— —— (2004). The evolution of metathesis, in B. Hayes, R. Kirchner, and D. Steriade (eds.), *Phonetically based Phonology*. Cambridge: Cambridge University Press, 117–56.

—— and Wedel, Andrew (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26: 143–83.

Blicher, Deborah, Diehl, Randy, and Cohen, Leslie (1990). Effects of syllable duration on the perception of the Mandarin tone 2/tone 3 distinction: Evidence of auditory enhancement, *Journal of Phonetics* 18: 37–49.

Blumstein, Sheila E. (1973). *A Phonological Investigation of Aphasic Speech*. The Hague: Mouton.

—— and Stevens, Kenneth N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America* 66: 1001–17.

Boatman, Dana (2004). Cortical bases of speech perception: Evidence from functional lesion studies. *Cognition* 92: 47–65.

de Boer, Bart (2000). Self-organization in vowel systems. *Journal of Phonetics* 28(4): 441–65.

—— (2001). *The Origins of Vowel Systems*. Oxford: Oxford University Press.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences* 21. University of Amsterdam, 43–58.

—— (1998). *Functional Phonology*. The Hague: Holland Academic Graphics. Doctoral dissertation, University of Amsterdam.

—— (2003). The odds of eternal optimization in Optimality Theory, in D. Eric Holt (ed.), *Optimality Theory and Language Change*, Studies in Natural Language and Linguistic Theory 56. Dordrecht: Kluwer Academic Publishers, 31–65.

—— (2007). Some listener-oriented accounts of h-aspiré in French. *Lingua* 117: 1989–2054.

—— (2008). Emergent ranking of faithfulness explains markedness and licensing by cue. Rutgers Optimality Archive 954, <http://roa.rutgers.edu>.

—— (2009). Cue constraints and their interactions in phonological perception and production, in P. Boersma and S. Hamann (eds.), *Phonology in Perception*. Berlin: Mouton De Gruyter, 55–110.

—— (forthcoming). A programme for bidirectional phonology and phonetics and their acquisition and evolution, in A. Benz and J. Mattausch (eds.), *Bidirectional Optimality Theory*.

—— and Escudero, Paola (2008). Learning to perceive a smaller L2 vowel inventory: An Optimality Theory account, in P. Avery, E. Dresher, and K. Rice (eds.), *Contrast in Phonology: Theory, Perception, Acquisition*. Berlin: Mouton de Gruyter, 271–301.

—— —— and Hayes, Rachel (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013–16.

—— and Hamann, Silke (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25: 217–70.

—— —— (2009a). Loanword adaptation as first-language phonological perception, in A. Calabrese and W. L. Wetzels (eds.), *Loanword Phonology*. Amsterdam: John Benjamins, 11–58.

—— —— (2009b). Introduction: Models of phonology in perception, in P. Boersma and S. Hamann (eds.), *Phonology in Perception*. Berlin: Mouton de Gruyter, 1–24.

—— and Hayes, Bruce (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45–86.

—— and Weenink, David (2009). Praat: Doing Phonetics by Computer (Version 5.1.01) [computer program], <http://www.praat.org/>, accessed February 26, 2009.

Bohn, Ocke-Schwen (1995). Cross language speech production in adults: First language transfer doesn't tell it all, in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Crosslanguage Research*. Baltimore: York Press, 279–304.

—— and Flege, James E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14: 131–58.

—— —— (1997). Perception and production of a new vowel category by adult second language learners, in A. James and J. Leather (eds.), *Second-language Speech: Structure and Process*. Berlin and New York: Mouton de Gruyter, 53–73.

——and Steinlen, Anja K. (2003). Consonantal context affects cross-language perception of vowels, in M. J. Sole, D. Recsens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions, 2289–92.

Bongaerts, Theo (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners, in D. Birdsong (ed.), *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Lawrence Erlbaum Associates, 133–60.

——Mennen, Susan, and Van der Slik, Frans (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language, *Studia Linguistica* 54: 298–308.

Booij, Geert (1995). *The Phonology of Dutch*. Oxford: Clarendon Press.

——(1996). Cliticization as prosodic integration: The case of Dutch. *Linguistic Review* 13: 219–42.

Boothroyd, Arthur and Nittrouer, Susan (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America* 84(1): 101–14.

Bosch, Laura and Sebastián-Gallés, Nuria (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition* 65: 33–69.

————(2003). Simultaneous bilingualism and the perception of a language specific vowel contrast in the first year of life. *Language and Speech* 46: 217–44.

Bosshardt, Hans-Georg, Sappok, C., Knipschild, M., and Hölscher, C. (1997). Spontaneous imitation of fundamental frequency and speech rate by nonstutterers and stutterers. *Journal of Psycholinguistic Research* 26: 425–48.

Bouchhioua, Nadia (2008). The acoustic correlates of stress and accent in Tunisian Arabic: A comparative study with English. Ph.D. dissertation, Université de 7 Novembre, Carthage, Tunisia.

Bourne, Lyle E. and Restle, Frank (1959). Mathematical theory of concept identification. *Psychological Review* 66, 278–96.

Bowen, Caroline (2008). Minimal pairs, listening lists, and more. <http://www.speech-language-therapy.com/wordlists.html>, accessed March 13, 2009.

Bowers, Jeffrey S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review* 116: 220–51.

Boyce, Suzanne E. (1988). The influence of phonological structure on articulatory organization in Turkish and in English: Vowel harmony and coarticulation. Ph.D. dissertation, Yale University, New Haven, CT.

——(1990). Coarticulatory organization for lip rounding in Turkish and English. *Journal of the Acoustical Society of America* 88: 2584–95.

——Krakow, Rena A., Bell-Berti, Frederica, and Gelfer, C. (1990). Converging sources of evidence for dissecting articulatory movements into core gestures. *Journal of Phonetics* 18: 173–88.

de Boysson-Bardies, Benedicte, Hallé, Pierre, Sagart, Laurent, and Durand, Catherine (1989). A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language* 16: 1–17.

de Boysson-Bardies and Vihman, Marilyn M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67: 297–319.

Bradley, Cornelius (1911). Graphic analysis of the tone-accents of the Siamese language. *Journal of the American Oriental Society* 31: 282–9.

Bradley, Travis G. (2002). Gestural timing and derived environment effects in Norwegian clusters, in L. Mikkelsen and C. Potts (eds.), *WCCFL 21 Proceedings*. Somerville, MA: Cascadilla Press, 43–56.

—— (2006). Spanish rhotics and Dominican hypercorrect /s/. *Probus* 18: 1–33.

Bradlow, Ann R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America* 97: 1916–24.

—— Akahane-Yamada, Reiko, Pisoni, David B., and Tohkura, Yoh'ichi (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in speech perception and production. *Perception and Psychophysics* 61: 977–85.

—— and Alexander, Jennifer A. (2007). Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America* 121(4): 2339–49.

—— Baker, Rachel E., Choi, Arim, Kim, Midam, and Van Engen, Kristin J. (2007). The Wildcat Corpus of Native and Foreign-Accented English. *Journal of the Acoustical Society of America* 121(5): 3072.

—— and Bent, Tessa (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America* 112: 272–84.

—— —— (2008). Perceptual adaptation to non-native speech. *Cognition* 106: 707–29.

—— Clopper, Cynthia, and Smiljanic, Rajka (2007). A perceptual similarity space for languages, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany.

—— —— —— and Walter, Mary Ann (2010). A perceptual similarity space for languages. *Speech Communication* 52 (11–12): 930–42.

—— Nygaard, Lynne C., and Pisoni, David B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception and Psychophysics* 61: 206–19.

—— Pisoni, David, Akahane-Yamada, Reiko, and Tohkura, Yoh'ichi (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101: 2299–310.

Braine, Martin D. S. (1992). What sort of innate structure is needed to "bootstrap" into syntax? *Cognition* 45: 77–100.

Brame, Michael K. and Bordelois, Ivonne (1973). Vocalic alternations in Spanish. *Linguistic Inquiry* 4: 111–68.

Bréa-Spahn, María Rosa (2009). Spanish-specific patterns and nonword repetition performance in English-language learners. Ph.D. dissertation, University of South Florida.

Breen, Gavan and Pensalfini, Robert (1999). Arrernte: A language with no syllable onsets. *Linguistic Inquiry* 30: 1–25.

Brent, Michael R. and Siskind, Jeffrey M. (2001). The role of exposure to isolated words in early vocabulary. *Cognition* 81: B33–B44.

Bresch, Erik, Kim, Yoon-Chul, Nayak, Krishna, Byrd, Dani, and Narayanan, Shrikanth (2008). Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging. *IEEE Signal Processing Magazine* 25(3): 123–32.

Bressmann, Tim. (2008). Quantitative assessment of tongue shape and movement using ultrasound imaging, in L. Colantoni and J. Steele (eds.), *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Somerville, MA: Cascadilla Proceedings Project, 101–6.

—— Thind, Parveen, Uy, Catherine, Bollig, Catherine, Gilbert, Ralph, and Irish, Jonathan (2005). Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving, and symmetry: Data from 12 normal speakers and a partial glossectomee. *Clinical Linguistics and Phonetics* 19(6/7): 573–88.

—— Uy, Catherine, and Irish, Jonathan (2005). Analysing normal and partial glossectomee tongues using ultrasound. *Clinical Linguistics and Phonetics* 19: 35–52.

Britain, David and Trudgill, Peter (1999). Migration, new-dialect formation and sociolinguistic refunctionalisation: Reallocation as an outcome of dialect contact. *Transactions of the Philological Society* 97: 245–56.

Broe, Michael (1993). Specification theory: The treatment of redundancy in generative phonology. Ph.D. dissertation, University of Edinburgh.

Brokx, Jan and Nooteboom, Sieb (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics* 10: 23–36.

Bromberger, Sylvain and Halle, Morris (1992). The ontology of phonology, in S. Bromberger (ed.), *On What We Know We Don't Know.* Chicago: University of Chicago Press, CSLI Publications, 209–28.

Broselow, Ellen, Chen, Su-I, and Huffman, Marie (1997). Syllable weight: Convergence of phonology and phonetics. *Phonology* 14: 47–82.

Brouwer, Susanne, Mitterer, Holger, and Huettig, Falk (forthcoming). Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics.*

Browman, Catherine P. and Goldstein, Louis (1986). Towards an articulatory phonology. *Phonology Yearbook* 3: 219–52.

—— —— (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica* 45: 140–55.

—— —— (1989). Articulatory gestures as phonological units. *Phonology* 6: 201–51.

—— —— (1990a). Tiers in articulatory phonology, with some implications for casual speech, in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.* Cambridge: Cambridge University Press, 341–97.

—— —— (1990b). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics* 18: 411–24.

—— —— (1991). Gestural structures: Distinctiveness, phonological processes, and historical change, in I. G. Mattingly and M. Studdert-Kennedy (eds.), *Modularity and the Motor Theory of Speech Perception. Proceedings of a Conference to honor Alvin M. Liberman*. Haskins Laboratories, New Haven, CT: Lawrence Erlbaum Associates, 313–38.

—— —— (1992). Articulatory Phonology: An overview. *Phonetica* 49: 155–80.

—— —— (1995). Gestural syllable position effects in American English, in F. Bell-Berti and L. Raphael (eds.), *Producing Speech: Contemporary Issues for Katherine Safford Harris*. New York: American Insititue of Physics, 19–33.

—— —— (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée* 5, 25–34.

Brown, Gillian, Currie, Karen L., and Kenworthy, Joanne (1980). *Questions of Intonation*. London: Croom Helm.

Brown, Roger (1973). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

Brown-Schmidt, Sarah and Tanenhaus, Michael K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science* 32: 643–84.

Bruce, Gösta (1977). *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

—— (1987). How floating is focal accent?, in K. Gregersen and H. Basbøll (eds.), *Nordic Prosody* IV. Odense: Odense University Press, 41–9.

—— (1990). Alignment and composition of tonal accents, in J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge: Cambridge University Press, 107–15.

Brunelle, Marc. (2005). Register in Eastern Cham: Phonological, phonetic and sociolinguistic approaches. Ph.D. dissertation, Cornell University.

—— (2008). Speaker control in the phonetic implementation of Cham registers. Presentation at the Third Conference on Tone and Intonation in Europe, Lisbon, Portugal.

—— (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37: 79–96.

Bucholtz, Mary (2003). Sociolinguistic nostalgia and the authentication of identity. *Journal of Sociolinguistics* 7(3): 398–416.

Buchwald, Adam, Rapp, Brenda, and Stone, Maureen (2007). Insertion of discrete phonological units: An ultrasound investigation of aphasic speech. *Language and Cognitive Processes* 22(6): 910–48.

Buckley, Eugene (2000). What should phonology explain? Handout from SUNY Buffalo Linguistics Colloquium.

Buder, Eugene and Stoel-Gammon, Carol (1994). Cross-language differences in phonological acquisition: Swedish and American /t/. *Phonetica* 51: 146–58.

Bullock, Daniel, and Grossberg, Steven (1988). Neural dynamics of planned arm movements: emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review* 95: 49–90.

Burdick, Charles K. and Miller, Joanne D. (1975). Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America* 58: 961–70.

Burkard, Robert (2009). The auditory steady-state response: Generation, recording, and clinical applications. *Ear and Hearing* 30: 384–5.

Bûrki, Audrey, Ernestus, Mirjam, and Frauenfelder, Uli (2010). One or two phonological representations for words with two phonological variants? Evidence from French schwa.

Burnham, Denis (2003). Language-specific speech perception and the onset of reading. *Reading and Writing: An Interdisciplinary Journal* 16: 573–609.

—— and Mattock, Karen (2007). The perception of tones and phones, in O.-S. Bohn and M. Munro (eds.), *Language Experience in Second-language Speech Learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 258–80.

Burns, Tracey C., Werker, J. F., and McVie, Karen (2003). Development of phonetic categories in infants raised in bilingual and monolingual environments, in B. Beachley,

A. Brown, and F. Conlin (eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 173–84.

—— Yoshida, Katherine A., Hill, Karen, and Werker, Janet F. (2007). Bilingual and monolingual infant phonetic development. *Applied Psycholinguistics* 28: 455–74.

Burton-Roberts, Noel (2000). Where and what is phonology?, in N. Burton-Roberts, P. Carr, and G. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues.* Oxford: Oxford University Press, 39–66.

—— Carr, Philip, and Docherty, Gerard (eds.) (2000). *Phonological Knowledge: Conceptual and Empirical Issues.* Oxford: Oxford University Press.

Burzio, Luigi (1994). Metrical consistency, in E. Ristad (ed.), *Proceedings of the DIMACS Workshop on Human Language*. Providence, RI: American Mathematical Society.

Busà, Maria Grazia (2003). Vowel nasalization and nasal loss in Italian, in M.-J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, Spain, August 2003, 711–14.

Bush, Robert R. and Mosteller, Frederick (1951). A model for stimulus generalization and discrimination. *Psychological Review* 58(6): 413–23.

Buzsaki, Gyorgy (2006). *Rhythms of the Brain*. Oxford: Oxford University Press.

Bybee, Joan (1985). *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.

—— (2000). The phonology of the lexicon: Evidence from lexical diffusion, in M. Barlow and S. Kemmer (eds.), *Usage-Based Models of Language*. Stanford, CA: CSLI, 65–85.

—— (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

—— (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14: 261–90.

—— (2006). From usage to grammar: The mind's response to repetition. *Language* 82: 711–33.

—— (2007). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

—— (2008) Formal universals as emergent phenomena: The origins of structure preservation, in J. Good (ed.), *Linguistic Universals and Language Change*. Oxford: Oxford University Press, 108–21.

—— and McClelland, James L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Linguistic Review* 22: 381–410.

—— and Pardo, Elly (1981). On lexical and morphological conditioning of alternations: A nonce-probe experiment with Spanish verbs. *Linguistics* 19: 937–68.

—— and Scheibman, Joanne (1999). The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37: 575–96.

Byrd, Dani (1995). C-centers revisited. *Phonetica* 52: 263–82.

—— (1996a). A phase window framework for articulatory timing. *Phonology* 13: 139–69.

—— (1996b). Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24: 209–44.

—— Kaun, Abigail, Narayanan, Shrikanth, and Saltzman, Elliot (2000). Phrasal signatures in articulation, in M. B. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 70–87.

Byrd, Dani, Krivokapic, Jelena, and Lee, Sungbok (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America* 120(3): 1589–99.

—— and Saltzman, Elliot L. (1998). Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics* 26: 173–99.

—— —— (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31(2): 149–80.

—— Tobin, Stephen, Bresch, Erik, and Narayanan, Shrikanth (2009). Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics* 37: 97–110.

Cacoullos, Rena T. and Walker, James A. (2009). The present of the English future: Grammatical variation and collocations in discourse. *Language* 85: 321–54.

Calhoun, Sasha. (2006). Information structure and the prosodic structure of English: A probabilistic relationship. Ph.D. dissertation, University of Edinburgh.

—— (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86(1): 1–42.

Cambier-Langeveld, Tina (1997). The domain of final lengthening in the production of Dutch, in H. de Hoop and J. Coerts (eds.), *Linguistics in the Netherlands*. Amsterdam: John Benjamins, 13–24.

—— (2000). Temporal marking of accents and boundaries. Ph.D. dissertation, University of Amsterdam. LOT Dissertation Series, 32.

—— and Turk, Alice (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics* 27: 171–206.

Campbell, Donald T. and Fiske, Donald W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56: 81–105.

Campbell, Fiona, Gick, Bryan, Wilson, Ian, and Vatikiotis-Bateson, Eric (2010). Spatial and temporal properties of gestures in North American English /r/. *Language and Speech* 53(1): 49–69.

Campbell, Nick (1992). Segmental elasticity and timing in Japanese speech, in Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (eds.), *Speech Perception, Production, and Linguistic Structure*. Tokyo: Ohmsha, 403–18.

—— (1999). Data-driven speech synthesis. *Journal of the Acoustical Society of America* 105: 1029–30.

—— and Beckman, Mary E. (1997). Stress, prominence, and spectral tilt, in A. Botinis, G. Kouroupetroglou, and G. Carayannis (eds.), *Intonation: Theory, Models and Applications* (Proceedings of the ESCA Workshop on Intonation), Athens, Greece, 67–70.

Campbell-Kibler, Kathryn (2006). Listener perceptions of sociolinguistic variables: the case of (ING). Ph.D. dissertation, Stanford University.

—— (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech* 82(1): 32–64.

—— (2008). I'll be the judge of that: Diversity in social perceptions of (ING). *Language in Society* 37: 637–59.

Canavan, Alexandra and Zipperlen, George (1996). CALLHOME Japanese Speech. Philadelphia: Linguistic Data Consortium.

—— Graff, David, and Zipperlen, George (1997). CALLHOME American English Speech. Philadelphia: Linguistic Data Consortium.

Cao, Yang, Zhang, Shuwu, Huang, Taiyi, and Xu, Bo (2004). Tone modeling for continuous Mandarin speech recognition. *International Journal of Speech Technology* 7: 115–28.

Caramazza, Alfonso and Yeni-Komshian, Grace H. (1974). Voice onset time in two French dialects. *Journal of Phonetics* 2: 239–245.

—— —— Zurif, Edgar B., and Carbone, Ettore (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 54: 421–6.

Cardoso, Walcir (2001). Variation patterns in regressive assimilation in Picard. *Language Variation and Change* 13(3): 305–42.

Carney, Arlene E., Widin, Gregory P., and Viemeister, Neal F. (1977). Noncategorical Perception of Stop Consonants Differing in VOT. *Journal of the Acoustical Society of America* 62: 961–70.

Carpenter, Gail and Grossberg, Stephen (1987). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics* 26: 4919–30.

Carré, René (2004). From acoustic tube to speech production. *Speech Communication* 42: 227–40.

—— Bourdeau, Marc, and Tubach, Jean-Pierre (1995). Vowel-vowel production: the distinctive region model (DRM) and vocalic harmony. *Phonetica* 52: 205–14.

—— and Mrayati, Mohamad (1990). Articulatory–acoustic–phonetic relations and modeling, regions and modes, in A. Marchal and W. J. Hardcastle (eds.), *Speech Production and Speech Modelling*. NATO ASI Series. Dordrecht: Kluwer Academic Publishers, 211–40.

Carroll, J. Douglas and Chang, Jih-Jie (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika* 35: 283–319.

Carter, Allyson and Gerken, LouAnn (2004). Do children's omissions leave traces? *Journal of Child Language* 31: 561–86.

Caspers, Joanneke and van Heuven, Vincent (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50: 161–71.

Castelhano, Monica S. and Rayner, Keith (forthcoming). Eye movements during reading, visual search, scene perception: An overview, in K. Rayner, D. Shem, X. Bai, and G. Yan (eds.), *Cognitive and Cultural Influences on Eye Movements*. Tianjin: Tianjin People's Press/Psychology Press.

Cattuto, Ciro, Barrat, Alain, Baldassarri, Andrea, and Schehr, Gregory (2009). Collective dynamics of social annotation. *Proceedings of the National Academy of Sciences, USA* 106: 10511–15.

Cavé, Christian, Guaitella, Isabelle, Bertrand, Roxane, Santi, Serge, Harlay, Françoise, and Espesser, Robert (1996). About the relationship between eyebrow movements and F0 variations. *Proceedings of ICSLP 1996*, Philadelphia, 2175–9.

Cebrian, Juli (2006). Experience and the use of duration in the categorization of L2 vowels. *Journal of Phonetics* 34: 372–87.

Cedergren, Henrietta (1973). Interplay of social and linguistic factors in Panama. Ph.D. dissertation, Cornell University.

—— and Sankoff, David (1974). Variable rules: performance as a statistical reflection of competence. *Language* 50: 333–55.

Cena, Richard M. (1978). *When is a Phonological Generalization Psychologically Real?* Bloomington, IN: Indiana University Linguistics Club.

Cenoz, Jasone, Hufeisen, Britta, and Jessner, Ulrike (2001). Introduction, in J. Cenoz, B. Hufeisen, and U. Jessner (eds.), *Crosslinguistic Influences in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters, 1–7.

Chahal, Dana and Hellmuth, Sam (forthcoming). The intonation of Lebanese and Egyptian Arabic, in S.-A. Jun (ed.), *Prosodic Typology* II. Oxford: Oxford University Press.

Chambers, Jack K. (1995). *Sociolinguistic Theory. Linguistic Variation and its Social Significance*. Oxford: Blackwell.

—— Trudgill, Peter, and Schilling-Estes, Natalie (eds.) (2002). *The Handbook of Language Variation and Change*. Oxford: Blackwell.

Chambers, Kyle, Onishi, Kristine, and Fisher, Cynthia (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87: B69–B77.

Chao, Yuanren (1930). A system of "tone letters." *Le Maître Phonétique* 45: 24–7.

Charles-Luce, Jan and Luce, Paul A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language* 17: 205–15.

Chen, Aoju (2003). Reaction time as an indicator of discrete intonational contrasts in English. *Proceedings of Eurospeech*, 97–100.

—— den Os, Els, and de Ruiter, Jan P. (2007). Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *Linguistic Review* 24(2): 317–44.

Chen, Jenn-Yeu, Chen, Train-Min, and Dell, Gary S. (2002). Word form encoding in Mandarin Chinese as assessed by the implicit priming paradigm. *Journal of Memory and Language* 46: 751–81.

Chen, Ken, Hasegawa-Johnson, Mark, and Cohen, Aaron (2004). An automatic prosody labeling system using ANN-based syntactic-prosodic model and GMM-based acoustic-prosodic model. *Proceedings of the International Conference on Acoustics, Speech, Signal Processing*, 1: 509–12.

—— —— —— Borys, Sarah, Kim, Sung-Suk, Cole, Jennifer, and Choi, Jeung-Yoon (2006). Prosody-dependent speech recognition on Radio News corpus of American English. *IEEE Transactions in Speech and Audio Processing* 14(1): 232–45.

Chen, Marilyn (1995). Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers. *Journal of the Acoustical Society of America* 98: 2443–53.

—— (1997). Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America* 102: 2360–70.

Chen, Matthew. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22: 129–59.

—— (1987). The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4: 109–49.

—— (2000). *Tone Sandhi*. Cambridge: Cambridge University Press.

Chen, Yiya (2003). The phonetics and phonology of contrastive focus in Standard Chinese. Ph.D. dissertation, Stony Brook University.

—— (2006). Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics* 34: 176–201.

—— (2008). The acoustic realization of Shanghai vowels. *Journal of Phonetics* 36: 629–48.

—— (2009). Prosodic marking of topic and focus in Shanghai Chinese. *Chinese Journal of Phonetics* 2: 123–33.

—— (2010). Post-focus suppression: Now you see it, now you don't. *Journal of Phonetics* 38: 517–25.

—— and Braun, Bettina (2006). The prosodic categories of information structure, in *Speech Prosody* 2006. Dresden, Germany.

Chen, Yiya and Gussenhoven, Carlos (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics* 36: 724–46.

—— and Xu, Yi (2006). Production of weak elements in speech: Evidence from F0 patterns of neutral tone in Standard Chinese. *Phonetica* 63: 47–75.

Cheshire, Jennifer, Fox, Sue, Kerswill, Paul, and Torgensen, Eivind (2008). Ethnicity, friendship network and social practices as the motor of dialect change: Linguistic innovation in London, in U. Ammon, J. Darquennes, and S. Wright (eds.), *Sociolinguistica: International Yearbook of European Sociolinguistics*, vol. 22. Max Niemeyer Verlag, 1–23.

Chiba, Tsutomu, and Kajiyama, Masato (1941). *The Vowel: Its Nature and Structure*. Tokyo: Phonetic Society of Japan.

Childers, D. G., Hicks, D. M., Moore, G. P., Eskenazi, L., and Lalwani, A. L. (1990). Electroglottography and vocal fold physiology, *Journal of Speech and Hearing Research* 33: 245–54.

Chistovich, Ludmilla, Sheikin, R. L., and Lublinskaya, V. V. (1979). Centers of gravity and the spectral peaks as the determinants of vowel quality, in B. Lindblom and S. Ohman (eds.), *Frontiers of Speech Communication Research*. London: Academic Press, 143–58.

Chitoran, Ioana and Hualde, José Ignacio (2007). From hiatus to diphthong: The evolution of vowel sequences in Romance. *Phonology* 24: 37–75.

Cho, Taehong (2002). *The Effects of Prosody on Articulation in English*. New York: Routledge.

—— (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32: 141–76.

—— (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English. *Journal of the Acoustical Society of America* 117(6): 3867–78.

—— (2006). Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English, in L. M. Goldstein, D. H. Whalen, and C. T. Best (eds.), *Laboratory Phonology 8*. Berlin: Mouton de Gruyter, 519–48.

—— Jun, Sun-Ah, and Ladefoged, Peter (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30: 193–228.

—— and Ladefoged, Peter (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27: 207–29.

—— and McQueen, James (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33(2): 121–57.

—— —— and Cox, Ethan A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35: 210–43.

Cho, Young-Mee Y. (1990). Syntax and phrasing in Korean. In S. Inkelas and D. Zec (eds.), *The Phonology-Syntax Connection.* Chicago: University of Chicago Press, 47–62.

Choi, Jeung-Yoon, Hasegawa-Johnson, Mark, and Cole, Jennifer (2005). Finding intonational boundaries using acoustic cues related to the voice source. *Journal of the Acoustical Society of America* 118(4): 2579–88.

Choi, John D. (1995). An acoustic-phonetic underspecification account of Marshallese vowel allophony. *Journal of Phonetics* 23: 323–47.

Cholin, Joana and Levelt, Willem J. M. (2009). Effects of syllable preparation and syllable frequency in speech production: Further evidence for syllabic units at a post-lexical level. *Language and Cognitive Processes* 24: 662–84.

—— —— and Schiller, Niels O. (2006). Effects of syllable frequency in speech production. *Cognition* 99: 205–35.

Cholin, Joana, Schiller, Niels O., and Levelt, Willem J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language* 50: 47–61.

Chomsky, Noam. (1964). *The Logical Structure of Linguistic Theory.* The Hague: Mouton.

—— (1977). On Wh-movement, in A. Akmajian, T. Wasow, and P. Culicover (eds.), *Formal Syntax.* Cambridge, MA: MIT Press, 71–133.

—— (1993). A minimalist program for linguistic theory, in K. Hale and S. J. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger.* Cambridge, MA: MIT Press, 53–109.

—— (1995). *The Minimalist Program.* Cambridge, MA: MIT Press.

—— (1998). Minimalist inquiries: The framework. MS, Massachusetts Institute of Technology.

—— and Halle, Morris (1968). *The Sound Pattern of English*. New York: Harper and Row.

—— and Lasnik, Howard (1995). The theory of Principles and Parameters, in N. Chomsky (ed.), *The Minimalist Program.* Cambridge, MA: MIT Press, 13–128.

Christensen, Rune H. B. (2010). Ordinal—Regression models for ordinal data. R package version <http://www.cran.r-project.org/package=ordinal/ 2010.03-04>.

Christoffels, Ingrid K., Firk, Christine, and Schiller, Niels O. (2007). Bilingual language control: An event-related brain potentials study. *Brain Research* 1147: 192–208.

Christophe, Anne, Millotte, Séverine, Bernal, Savita, and Lidz, Jeffrey (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech* 51(1–2): 61–75.

—— Peperkamp, Sharon, Pallier, Christophe, Block, Elisa, and Mehler, Jacques (2004). Phonological phrase boundaries constrain lexical access, I: Adult data. *Journal of Memory and Language* 51: 523–47.

Church, Barbara A. and Schacter, Daniel L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice, intonation, and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 521–33.

Church, Kenneth W. and Gale, William A. (1995). Poisson Mixtures. *Journal of Natural Language Engineering* 1: 163–90.

Cieri, Christopher, Graff, David, Kimball, Owen, Miller, Dave, and Walker, Kevin (2005). Fisher English Training Speech, Part 2 Transcripts. Philadelphia: Linguistic Data Consortium.

—— Miller, David, and Walker, Kevin (2004). The Fisher Corpus: A resource for the next generations of speech-to-text. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, Lisbon, 69–71.

Clark, Eve V. (1987). The principle of contrast: A constraint on language acquisition, in B. MacWhinney (ed.), *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1–34.

Clark, Herbert. (1996). *Using Language*. Cambridge: Cambridge University Press.

—— (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12: 335–59.

Clark, Lynn (2009). Variation, change and the usage-based approach. Doctoral dissertation, Edinburgh University.

Clark, Mary (1990). *The Tonal System of Igbo*. Dordrecht: Foris.

Clayards, Meghan, Tanenhaus, Michael K., Aslin, Richard N., and Jacobs, Robert A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108(3): 804–9.

Clément, Philippe, Hans, Stéphane, Hartl, Dana M., Maeda, Shinji, Vaissière, Jacqueline, and Brasnu, Daniel (2007). Vocal tract area function for vowels using

three-dimensional magnetic resonance imaging: A preliminary study. *Journal of Voice* 21: 522–30.

Clements, G. N. (1976). Vowel harmony in nonlinear generative phonology: An autosegmental model. [Published in 1980 by Indiana University Linguistics Club, Bloomington, IN.]

—— (1977). Neutral vowels in Hungarian vowel harmony: An autosegmental interpretation. *Proceedings of the North Eastern Linguistic Society* 7, 49–64.

—— (1981). The hierarchical representation of tone features. *Harvard Studies in Phonology* 2: 50–115.

—— (1984). Principles of tone assignment in Kikuyu, in G. N. Clements and J. Goldsmith (eds.), *Autosegmental Studies in Bantu Tone*. Dordrecht: Foris Publications, 281–339.

—— (1985). The geometry of phonological features. *Phonology Yearbook* 2: 225–52.

—— (1986). Compensatory lengthening and consonant gemination in Luganda, in L. Wetzels and E. Sezer (eds.), *Studies in Compensatory Lengthening*. Dordrecht: Foris Publications, 37–78.

—— (2001). Representational economy in constraint-based phonology, in T. A. Hall (ed.), *Distinctive Feature Theory*. Phonology and Phonetics Series. Berlin: Mouton, 71–146.

—— and Hume, Elizabeth (1995). The internal organization of speech sounds, in J. Goldsmith (ed.), *The Handbook of Phonological Theory*. London: Blackwell, 245–306.

—— and Ridouane, Rachid (2006). Quantal phonetics and distinctive features: A review, in A. Botinis (ed.), *Proceedings of the ISCA Tutorial and Research Workshop on Experimental Linguistics*, August 28–30, 2006. Athens: University of Athens, 17–24.

Cliff, Emily and Kirchner, Robert (in progress). An exemplar-based account of type-frequency effects in pattern generalization. MS, University of Alberta.

Clifton, Rachel, Freyman, Richard, and Meo, Jennifer (2002). What the precedence effect tells us about room acoustics. *Perception and Psychophysics* 64: 180–8.

Clopper, Cynthia G. and Bradlow, Ann R. (2008). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech* 51(3): 175–98.

—— —— (2009). Free classification of American English dialects by native and non-native listeners. *Journal of Phonetics* 37: 436–51.

—— and Paolillo, John C. (2006). North American English vowels: A factor-analytic perspective. *Literary and Linguistic Computing* 21: 445–62.

—— and Pisoni, David. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics* 35: 421–38.

Coady, Jeffry A., Evans, Julia L., Mainela-Arnold, Elina, and Kluender, Keith R. (2007). Children with specific language impairments perceive speech most categorically when tokens are natural and meaningful. *Journal of Speech, Language, and Hearing Research* 50: 41–57.

—— Kluender, Keith R., and Evans, Julia L. (2005). Categorical perception of speech by children with specific language impairments. *Journal of Speech, Language, and Hearing Research* 48: 944–59.

Coates, Jennifer (1993). *Women, Men, and Language: A Sociolinguistic Account of Gender Differences in Language*. London: Longman.

Coenen, Else, Zwitserlood, Pienie, and Bölte, Jens (2001). Variation and assimilation in German: Consequences of assimilation for word recognition and lexical representation. *Language and Cognitive Processes* 16: 535–64.

Coetzee, Andries W. (2004). What it means to be a loser: Non-optimal candidates in Optimality Theory. Ph.D. dissertation, University of Massachusetts, Amherst.

Coetzee, Andries W. (2006). Variation as accessing non-grammatical candidates. *Phonology* 23: 337–85.
—— (2009a). An integrated grammatical/non-grammatical model of phonological variation, in Y.-S. Kang, J.-Y. Yoon, H. Yoo, S.-W. Tang, Y.-S. Kang, Y. Jang, C. Kim, K.-A. Kim, and H.-K. Kang (eds.), *Current Issues in Linguistic Interfaces*, vol. 2. Seoul: Hankookmunhwasa, 267–94.
—— (2009b). Phonological variation and lexical frequency, in A. Schardl, M. Walkow, and M. Abdurrahman (eds.), *NELS 38*, vol. 1, Amherst: GLSA, 189–202.
—— (2011). Syllables in speech processing: Evidence from perceptual epenthesis, in C. Cairns and E. Raimy (eds.), *Handbook of the Syllable*. Leiden: Brill, 295–328.
—— and Kawahara, Shigeto (forthcoming). Frequency biases in phonological variation. *Natural Language and Linguistic Theory*.
—— and Pater, Joe (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26: 289–337.
—— —— (forthcoming). The place of variation in Phonological Theory, in J. Goldsmith, J. Riggle, and A. Yu (eds.), *Handbook of Phonological Theory*, 2nd edn. Oxford: Blackwell. [ROA-946]
—— —— and Kager, René (2009). Introduction: Phonological models and experimental data. *Phonology* 26: 1–8.
—— and Pretorius, Rigardt (2010). Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38(3): 404–21.
Cohen, Jacob (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1): 37–46.
Cohn, Abigail C. (1990). Phonetic and phonological rules of nasalization. Ph.D. dissertation, UCLA. Distributed as *UCLA Working Papers in Phonetics* 76.
—— (1993a). Nasalisation in English: Phonology or phonetics. *Phonology* 10: 43–81.
—— (1993b). The status of nasalized continuants, in M. Huffman and R. Krakow (eds.), *Nasals, Nasalization, and the Velum*. San Diego: Academic Press, 329–67.
—— (2005). Levels of abstractness in phonology and the lexicon: Evidence from English homophones. Paper presented at the 79th Meeting of the LSA, Oakland, CA, January 2005, and the 13th Manchester Phonology Meeting, May 2005. <http://ling.cornell.edu/docs/CohnhomophonesHO.pdf>.
—— (2006). Is there gradient phonology? in G. Fanselow, C. Féry, M. Schlesewsky, and R. Vogel (eds.), *Gradience in Grammar: Generative Perspectives*. Oxford: Oxford University Press, 25–44.
—— (2010). Laboratory Phonology: Past successes and current questions, challenges, and goals, in C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (eds.), *Papers in Laboratory Phonology 10*. Berlin: Mouton, 3–29.
Cole, Desmond T. (1955). *An Introduction to Tswana Grammar*. Cape Town: Longmans, Green & Co.
Cole, Jennifer (2009). Emergent feature structures: Harmony systems in exemplar models of phonology. *Language Sciences* 31: 144–60.
Cole, Jennifer and Hualde, José Ignacio (eds.) (2007). *Laboratory Phonology 9*. Berlin and New York: Mouton de Gruyter.
—— Kim, Heejin, Choi, Hansook, and Hasegawa-Johnson, Mark (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* 35: 180–209.

—— Linebaugh, Gary, Munson, Cheyenne, and McMurray, Bob (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics* 38(2): 167–84.

Coleman, John S. (1992). York Talk: "Synthesis-by-rule" without segments or rewrite rules, in G. Bailly, C. Benoit, and T. R. Sawallis (eds.), *Talking Machines: Theories, Models, and Designs.* Amsterdam: Elsevier, 211–24.

—— (1994). Polysyllabic words in the York Talk synthesis system, in P. A. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III.* Cambridge: Cambridge University Press, 293–324.

—— (1998). *Phonological Representations—Their Names, Forms, and Powers.* Cambridge: Cambridge University Press.

—— (2002). Phonetic representations in the mental lexicon, in J. Durand and B. Laks (eds.), *Phonetics, Phonology, and Cognition.* Oxford: Oxford University Press, 96–130.

—— and Local, John K. (1992). Monostratal phonology and speech synthesis, in P. Tench (ed.), *Studies in Systemic Phonology.* London: Pinter Publishers, 183–93.

—— and Pierrehumbert, Janet B. (1997). Stochastic phonological grammars and acceptability, in *Computational Phonology. Third Meeting of the ACL Special Interest Group*, Association for Computational Linguistics, 49–56.

Collier, René, Lisker, Leigh, Hirose, Hajime, and Ushijima, Tatsujiro (1979). Voicing in intervocalic stops and fricatives in Dutch. *Journal of Phonetics* 7: 357–73.

Coltheart, Max, Rastle, Kathleen, Perry, Conrad, Langdon, Robyn, and Ziegler, Johannes (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* 108: 204–56.

Connell, Bruce (2000). The perception of lexical tone in Mambila. *Language and Speech* 43: 163–82.

—— (2002). Tone languages and the universality of intrinsic F0: Evidence from Africa. *Journal of Phonetics* 30: 101–29.

—— and Ladd, D. Robert (1990). Aspects of pitch realisation in Yoruba. *Phonology* 7: 1–29.

Connine, Cynthia M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review* 11(6): 1084–9.

—— and Pinnow, Eleni (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *Linguistic Review* 23: 235–45.

—— Ranbom, Larissa J., and Patterson, David J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception and Psychophysics* 70(3): 403–11.

Content, Alain, Meunier, Christine, Kearns, Ruth K., and Frauenfelder, Uli H. (2001). Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes* 16: 609–36.

Cooper, Nicole, Cutler, Anne, and Wales, Roger (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech* 45(3): 207–28.

Cooper, Roger M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* 6: 84–107.

Cooper, William E. and Paccia-Cooper, Jeanne (1980). *Syntax and Speech.* Cambridge, MA: Harvard University Press.

Cooper, William E. and Eady, Stephen (1986). Metrical phonology in speech production. *Journal of Memory and Language* 25: 369–84.

——— and Mueller, Pamela (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77: 2142–56.

——— and Sorensen, John (1981). *Fundamental Frequency in Sentence Production*. Heidelberg: Springer.

Corter, James E. (1982). ADDTREE/P: A PASCAL program for fitting additive trees based on Sattath and Tversky's ADDTREE algorithm. *Behavior Research Methods and Instrumentation* 14: 353–4.

Costa, Albert and Sebastián-Gallés, Núria (1998). Abstract phonological structure in language production: Evidence from Spanish. *Journal of Experimental Psychology: Learning, Memory and Cognition* 24: 886–903.

Côté, Marie-Hélène (2000). Consonant cluster phonotactics: A perceptual approach. Ph.D. dissertation, MIT, Cambridge, MA.

——— and Kharlamov, Viktor (2011). The impact of experimental tasks on syllabification judgments: A case study of Russian, in C. Cairns and E. Raimy (eds.), *Handbook of the Syllable*. Leiden: Brill, 271–94.

Couper-Kuhlen, Elizabeth and Ford, Cecilia E. (eds.) (2004). *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation*. Amsterdam: John Benjamins.

——— Selting, Margret (1996). *Prosody in Conversation: Interactional Studies*. Cambridge: Cambridge University Press.

Coupland, Nikolas (1980). Style-shifting in a Cardiff work setting. *Language in Society* 9(1): 1–12.

——— (2007). *Style: Language Variation and Identity*. Cambridge: Cambridge University Press.

Crane, Riley and Sornette, Didier (2008). Robust dynamic classes revealed by measuring the response function of a social system, in *Proceedings of the National Academy of Sciences* 105, 15649–53.

Cranen, Bert and Boves, Louis (1985). Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production. *Journal of the Acoustical Society of America* 77(4): 1543–51.

——— ——— (1988). On the measurement of glottal flow. *Journal of the Acoustical Society of America* 84(3): 888–900.

Crawford, Clifford J. (2009). Adaptation and transmission in Japanese loanword phonology. Ph.D. dissertation, Cornell University.

Creative Commons (2009). Attribution-Share Alike 3.0, <http://creativecommons.org/licenses/by-sa/3.0/>, accessed April 20, 2009.

Creel, Sarah C., Aslin, Richard N., and Tanenhaus, Michael K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition* 106(2): 633–64.

Cristià, Alejandrina (2009). Individual variation in infant speech processing: Implications for language acquisition theories. Doctoral dissertation, Purdue University.

——— and Seidl, Amanda (2008). Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development* 4: 203–27.

Crocker, Laura and Munson, Benjamin (2006). Speech characteristics of gender-nonconforming boys. Oral presentation given at the Conference on New Ways of Analyzing Variation in Language, Columbus, OH <http://www.tc.umn.edu/~munso005/Crocker&Munson_NWAV2006_PostConference.pdf>, accessed March 2, 2011.

Croot, Karen (2010). The emergent paradigm in Laboratory Phonology: Phonological categories and statistical generalisation in Cutler, Beckman and Edwards, Frisch and Bréa-Spahn, Kapatsinski, and Walter. *Laboratory Phonology* 1: 415–24.

Crosswhite, Katherine (2004). Vowel reduction, in B. Hayes, R. Kirchner, and D. Steriade (eds.), *Phonetically based Phonology*. Cambridge: Cambridge University Press, 191–231.

Crothers, John (1978). Typology and universals of vowel systems, in J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (eds.), *Universals of Human Language, vol. 2, Phonology*. Stanford: Stanford University Press, 93–152.

Crowhurst, Megan J. and Michael, Lev (2005). Iterative footing and prominence-driven stress in Nanti (Kampa). *Language* 81: 47–95.

Cummins, Fred (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics* 31(2): 139–48.

Curtin, Suzanne, Fennell, Christopher. T., and Escudero, Paola (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science* 12: 725–31.

—— Goad, Heather, and Pater, Joe (1998). Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research* 14: 389–405.

Cutillas-Espinosa, Juan Antonio (2004). Meaningful variability: A sociolinguistically-grounded approach to variation in Optimality Theory. *International Journal of English Studies* 4(2): 165–84.

Cutler, Anne (1997). The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes* 12: 839–45.

—— (2008). The abstract representations in speech processing. *Quarterly Journal of Experimental Psychology* 61: 1601–19.

—— and Butterfield, Sally (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language* 31: 218–36.

—— and Carter, David M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer, Speech and Language* 2: 133–42.

—— Eisner, Frank, McQueen, James M., and Norris, Dennis (2010). How abstract phonemic categories are necessary for coping with speaker-related variation, in C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (eds.), *Laboratory Phonology 10*. Berlin: Mouton, 91–111.

—— and Otake, Takashi (2004). Pseudo-homophony in non-native listening. *Journal of the Acoustical Society of America* 115: 2392.

—— Webber, Andrea, Smits, Roel, and Cooper, Nicole (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America* 116: 3668–78.

—— Weber, Andre, and Otake, Takashi (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics* 34: 269–84.

Dahan, Delphine, Drucker, Sarah J., and Scarborough, Rebecca A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* 108(3): 710–18.

—— and Gaskell, M. Gareth (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language* 57: 483–501.

—— Magnuson, James S., and Tanenhaus, Michael K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology* 42(4): 317–67.

Dahan, Delphine, Magnuson, James S., Tanenhaus, Michael K. and Hogan, Ellen M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes* 16: 507–34.

——Tanenhaus, Michael K., and Chambers, Craig G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47: 292–314.

————and Salverda, Anne Pier (2007). How visual information influences phonetically-driven saccades to pictures: Effects of preview and position in display, in R. P. G. van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill (eds.), *Eye Movements: A Window on Mind and Brain*. Oxford: Elsevier.

Dainora, Audra. (2001). An empirically based probabilistic model of intonation in English. Ph.D. dissertation, University of Chicago.

——(2006). Modelling intonation in English, in L. Goldstein, D. H. Whalen, and C. T. Best (eds.), *Laboratory Phonology* 8. Berlin: Mouton de Gruyter, 107–32.

Daland, Robert, Pierrehumbert, Janet B., and Sims, Andrea D. (2007). Much ado about nothing: A social network model of Russian paradigmatic gaps. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 936–43.

Dalby, Jonathan (1984). Phonetic structure of fast speech in American English. Ph.D. dissertation, Indiana University.

Dalgaard, Peter (2002). *Introductory Statistics with R*. New York: Springer.

Dalton, Martha and Ní Chasaide, Ailbhe (2006). Tonal alignment in Irish dialects. *Language and Speech* 43: 441–64.

Daly, John and Hyman, Larry (2007). On the representation of tone in Peñoles Mixtec. *International Journal of American Linguistics* 73: 165–207.

Daly, Nicola and Warren, Paul (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics* 5(1): 85–96.

Damian, Markus F. and Bowers, Jeffrey S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language* 49: 119–32.

Damper, R. I. (1998). The role of the auditory periphery in the categorization of stop consonants. *Proceedings of the Joint Meeting of the International Conference Acoustics and the Acoustical Society of America*: 1973–74.

——and Harnad, S. R. (2000). Neural network models of categorical perception. *Perception and Psychophysics* 62: 843–67.

Darcy, Isabelle (2003). Assimilation phonologique et reconnaissance des mots. Ph.D. dissertation, École des hautes études en sciences sociales, Paris.

Davidson, Lisa (2003). The atoms of phonological representation: Gestures, coordination, and perceptual features in consonant cluster phonotactics. Ph.D. dissertation, Department of Cognitive Science, Johns Hopkins University.

——(2005). Addressing phonological questions with ultrasound. *Clinical Linguistics and Phonetics* 19(6/7): 619–33.

——(2006a). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120(1): 407–15.

——(2006b). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34(1): 104–37.

——(2006c). Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica* 63: 79–112.

——— (2007a). Coarticulation in contrastive Russian stop sequences. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany: University of the Saarland, 417–20.

——— (2007b). The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology* 24: 261–86.

——— and De Decker, Paul. (2005). Stabilization techniques for ultrasound imaging of speech articulations. *Journal of the Acoustical Society of America* 117(4/2): 2544.

——— Jusczyk, Peter, and Smolensky, Paul (2004). The initial and final states: Theoretical implications and experimental explorations of richness of the base, in R. Kager, W. Zonneveld, and J. Pater (eds.), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press, 321–68.

——— Klein, Harriet, and Grigos, Maria (2007). Perceptual, kinematic, and ultrasound measurement of /r/ development in children with phonological delay. Talk presented at Ultrafest IV, New York University, September 28–9, 2007, <http://jerome.linguistics.fas.nyu.edu/presentations/Ultrafest_IV_DavKleGri.pdf>, accessed March 16, 2009.

Davis, Matthew H., Marslen-Wilson, William D., and Gaskell, M. Gareth (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28: 218–44.

De Lacy, Paul (2002a). The formal expression of markedness. Doctoral dissertation, University of Massachusetts, Amherst. [ROA-542].

——— (2002b). The interaction of tone and stress in Optimality Theory. *Phonology* 19: 1–32.

——— (2004). Markedness conflation in Optimality Theory. *Phonology* 21: 145–99.

——— (2007). Quality of data in metrical stress theory. *Cambridge Extra* magazine, Issue 2.

De Vaan, Laura, Schreuder, Robert, and Baayen, R. Harald (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon* 2: 1–23.

De Wachter, Mathias (2007). Example-based continuous speech recognition. Doctoral dissertation, Katholieke Universiteit Leuven.

DeCasper, Anthony J. and Fifer, William P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science* 208: 1174–6.

Dehaene-Lambertz, Ghislaine (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport* 8: 919–24.

——— Dupoux, E., and Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience* 12: 635–47.

——— and Pena, Marecella (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport* 12: 3155–8.

Delattre, Pierre (1946). Stages of Old French phonetic changes observed in Modern Spanish. *Publications of the MLA* 61(1): 7–41.

——— and Freeman, Donald (1968). A dialect study of American r's by x-ray motion picture. *Linguistics: An International Review* 44: 29–68.

——— Liberman, Alvin M., and Cooper, Franklin S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America* 27: 769–73.

Delgutte, Bertrand (1997). Auditory neural processing of speech, in W. J. Hardcastle and J. Laver (eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 507–38.

Dell, Francois and Elmedlaoui, Mohamed (2002). *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht and Boston: Kluwer Academic Publishers.

Dell, Gary S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review* 93(3): 283–321.

—— (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27: 124–42.

—— (2000). Counting, connectionism and lexical representation, in M. Broe and J. Pierrehumbert (eds.), *Papers in laboratory phonology V: Acquisition and the Lexicon.* Cambridge: Cambridge University Press, 334–47.

—— Juliano, Cornell, and Govindjee, Anita (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science* 17: 149–95.

—— and Reich, Peter A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 20: 611–29.

—— Schwartz, Myrna F., Martin, Nadine, Saffran, Eleanor M., and Gagnon, Deborah A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 104: 801–38.

Delvaux, Veronique and Soquet, Alain (2007). The influence of ambient speech on adult speech production through unintentional imitation. *Phonetica* 64: 145–73.

Demolin, Didier (2007). Phonological universals and the control and regulation of speech production, in M.-J. Solé, P. S. Beddor, and M. Ohala (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 75–92.

Démonet, Jean-François, Fiez, Julie A., Paulesu, Eraldo, Petersen, Steven E., and Zatorre, Robert J. (2002). PET studies of phonological processing: A critical reply to Poeppel. *Brain and Language* 55: 352–79.

Demuth, Katherine (1993). Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language* 20: 275–301.

—— (1995a). The acquisition of tonal systems, in J. Archibald (ed.), *The Acquisition of Non-Linear Phonology*. Hillsdale, NJ: Lawrence Erlbaum.

—— (1995b). Markedness and the development of prosodic structure, in J. Beckman (ed.), *Processings of the North Eastern Linguistic Society* 25. Amherst, MA: GLSA, University of MA, 13–25.

—— (2003). The acquisition of Bantu languages, in D. Nurse and G. Phillipson (eds.), *The Bantu Languages*. Surrey, UK: Curzon Press.

—— (2006). Crosslinguistic perspectives on the development of prosodic words. *Language and Speech* 49: 129–35.

—— Culbertson, Jennifer, and Alter, Jennifer (2006). Word-minimality, epenthesis, and coda licensing in the early acquisition of English. *Language and Speech* 49: 137–74.

—— and McCullough, Elizabeth (2009). The prosodic (re)organization of children's early English articles. *Journal of Child Language* 36: 173–200.

—— Shattuck-Hufnagel, Stefanie, Song, Jae Yung, Evans, Karen, Kuhn, Jeremy, and Sinnott-Armstrong, Miranda (2009). Acoustic cues to stop coda voicing contrasts in 1-2-year olds' American English. *Journal of the Acoustical Society of America* 125(4): 2570.

—— and Tremblay, Annie (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language* 35: 99–127.

DePaolis, Rory A. (2006). The influence of production on the perception of speech, in D. Bamman, T. Magnitskaia, and C. Zaller (eds.), *Proceedings of the 30th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 142–53.

Derwing, Bruce (1992). A 'pause-break' task for eliciting syllable boundary judgments from literate and illiterate speakers: preliminary results from five diverse languages. *Language and Speech* 35: 219–35.

—— (2007). What's in CVC-like things? Ways and means to look at phonological units across languages, in M.-J. Solé, P. S. Beddor, and M. Ohala (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 325–38.

—— and Baker, William J. (1980). Rule learning and the English inflections (with special emphasis on the plural), in G. D. Prideaux, B. L. Derwing, and W. J. Baker (eds.), *Experimental linguistics: integration of theories and applications.* Ghent: E. Story-Scientia, 248–72.

Deville, Gérard (1891). Notes sur le developpement du langage II. *Revue de linguistics et de philology comparée* 24: 10–42, 128–43, 242–57, 300–20.

Devonish, Hubert (2007). Nationalism, the State, and Creole language identity. Paper presented at Creoles, Acts of Identity, and Education Workshop. Stanford University, July 15, 2007.

Dewaele, Jean-Marc (1998). Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics* 19: 471–90.

Dewson, James H. (1964). Speech sound discrimination by cats. *Science* 144: 555–6.

Di Paolo, Marianna and Faber, Alice (1990). Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change* 2: 155–204.

Diehl, Randy L. (1991). The role of phonetics within the study of language. *Phonetica* 48: 120–34.

—— (2008). Acoustic and auditory phonetics: The adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society* B 363: 965–78.

—— Walsh, Margaret A., and Kluender, Keith (1991). Auditory discontinuities interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America* 89(6): 2905–9.

Diesch, Eugen, Eulitz, Carsten, Hampson, Scott, and Ross, Bernhard (1996). The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain and Language* 53: 143–68.

Dilley, Laura C. (submitted). The role of F0 alignment in distinguishing categories in American English intonation. *Journal of Phonetics*.

—— and McAuley, J. Devin (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language* 59: 294–311.

—— and Pitt, Mark (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America* 122: 2340–53.

—— Shattuck-Hufnagel, Stefanie, and Ostendorf, Mari (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24: 423–44.

Dimitrova, Snezhina and Turk, Alice (in preparation). Patterns of English phrasal-stress induced lengthening.

D'Imperio, Mariapaola (1995). Timing differences between prenuclear and nuclear pitch accents in Italian. *Journal of the Acoustical Society of America* 98(5): 2894.

—— (1997). Narrow focus and focal accent in the Neapolitan variety of Italian. *Proceedings of ESCA Workshop on Intonation*, Athens, Greece, 87–90.

D'Imperio, Mariapaola (2000). The role of perception in defining tonal targets and their alignment. Ph.D. dissertation, Ohio State University.

—— (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication* 33(4): 339–56.

—— (2002a). Italian intonation: An overview and some questions. *Probus* (Special issue on intonation in Romance languages), 14(1), 37–69.

—— (2002b). Language-specific and universal constraints on tonal alignment: The nature of targets and "anchors", in B. Bel and I. Marlien (eds.), *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, April 11–13, 2002, 101–6.

—— Elordieta, Gorka, Frota, Sónia, Prieto, Pilar, and Vigário, Marina (2005). Intonational phrasing in Romance: the role of syntactic and prosodic structure, in S. Frota, M. Vigário, and M. J. Freitas (eds.), *Prosodies*. Berlin and New York: Mouton de Gruyter, 59–97.

—— Espesser, Robert, Loevenbruck, Hélène, Menezes, Caroline, Nguyen, Noël, and Welby, Pauline (2007). Are tones aligned with articulatory events? Evidence from Italian and French in J. Cole and J. I. Hualde (eds), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 577–608.

—— and Gili Fivela, Barbara (2003). How many levels of phrasing? Evidence from two varieties of Italian, in J. Local, R. Ogden, and R. Temple (eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: CUP, 130–44.

—— —— and Niebuhr, Oliver (2010). Alignment perception of high intonational plateaux in Italian and German. *Proceedings of Speech Prosody 2010*, Chigago, IL.

—— and House, David (1997). Perception of questions and statements in Neapolitan Italian, in G. Kokkinakis, N. Fakotakis, and E. Dermatas (eds.), *Proceedings of Eurospeech'97*, Rhodes, Greece, vol. 1, 251–4.

—— Nguyen, Noël, and Munhall, Kevin G. (2003). An articulatory hypothesis for the alignment of tonal targets in Italian. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, August 3–9, 2003, 253–6.

—— Petrone, Caterina, and Nguyen, Noël (2007). Effects of tonal alignment on lexical identification in Italian, in C. Gussenhoven and T. Riad (eds.), *Tones and Tunes*, vol. 2. Berlin: Mouton de Gruyter, 79–106.

Dinkin, Aaron (2008). The real effect of word frequency on phonetic variation. University of Pennsylvania Working Papers in Linguistics, 14.1, <http://repository.upenn.edu/pwpl/vol14/iss1/8/>.

Dixit, Prakash R. and MacNeilage, Peter F. (1980). Cricothyroid activity and control of voicing in Hindi stops and affricates. *Phonetica* 37: 397–406.

Dmitrieva, Olga and Jongman, Allard (2007). Phonological neutralization by native and non-native speakers: The case of Russian ?nal devoicing. MS, Stanford and KU, <http://www.stanford.edu/~dmitro/Dmitrieva_Jongman.pdf>, accessed March 23, 2010.

Docherty, Gerard (2007a). Speech in its natural habitat: Accounting for social factors in phonetic variability, in J. Cole and J. I. Hualde (eds), *Laboratory Phonology* 9. Berlin: Mouton de Gruyter, 1–35.

—— (2007b). Prosodic factors and sociophonetic variation: Speech rate and glottal variants in Tyneside English. *Proceedings of the 17th International Congress of Phonetic Sciences*, Saarbrücken, Germany: 1517–20.

—— and Foulkes, Paul (1999). Instrumental phonetics and phonological variation: Case studies from Derby and Newcastle, in P. Foulkes and G. J. Docherty (eds.), *Urban Voices: Accent Studies in the British Isles*. London: Arnold, 47–71.

——— ——— (2000). Speaker, speech, and knowledge of sounds, in N. Burton-Roberts, P. Carr, and G. J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford, Oxford University Press, 105–29.

——— ——— (2005). Glottal variants of (t) in the Tyneside variety of English: An acoustic profiling study, in W. Hardcastle and J. M. Beck (eds.), *A Figure of Speech: A Festschrift for John Laver*. London: Lawrence Erlbaum, 173–99.

——— ——— (forthcoming). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*.

——— ——— Milroy, James, Milroy, Lesley, and Walshaw, David (1997). Descriptive adequacy in phonology: A variationist perspective. *Journal of Linguistics* 33: 275–310.

——— ——— Tillotson, Jenny, and Watt, Dominic J. L. (2006). On the scope of phonological learning: Issues arising from socially structured variation, in L. Goldstein, D. H. Whalen, and C. T. Best (eds.), *Laboratory Phonology* 8. Berlin: Mouton de Gruyter, 393–422.

Dogil, Grzegorz (2007). Phonetic dimensions of segmental strength, in *Proceedings of ICPHS XVI*. Saarbrücken, 89–92.

Dohen, Marion, Lœvenbruck, Hélène, Cathiard, Marie-Agnès, and Schwartz, Jean-Luc (2004). Visual perception of contrastive focus in reiterant French speech. *Speech Communication* 44: 155–72.

Dolbey, Andrew E. and Hansson, Gunnar Ó. (1999). The source of naturalness in synchronic phonology, in S. Billings, J. Boyle, and A. Griffith (eds.), *CLS 35*, vol. 1. Chicago: CLS, 59–69.

Dommelen, Wim van (1983). Parameter interaction in the perception of French plosives. *Phonetica* 40: 32–62.

Donegan, Patricia and Stampe, David (1979). The study of natural phonology, in D. A. Dinnsen (ed.), *Current Approaches to Phonological Theory*. Bloomington: Indiana University Press, 126–73.

Dooling, Robert J. and Brown, S. D. (1990). Speech perception by budgerigars (Melopsittacus undulatus): Spoken vowels. *Perception and Psychophysics* 47: 568–74.

——— Best, Carol T., and Brown, S. D. (1995). Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (Melopsittacus undulatus) and zebra finches (Taeniopygia guttata). *Journal of the Acoustical Society of America* 97: 1839–46.

Downing, Laura (1989). The interaction of tone and intonation in Jita yes/no questions. *Studies in the Linguistic Sciences* 19: 91–113.

——— and Pompino-Marschall, Bernd (2004). Prosody and information structure in Chichewa. *ZAS Papers in Linguistics* 37: 167–86.

Drager, Katie (2006). From bad to bed: The relationship between perceived age and vowel perception in New Zealand English. *Te Reo* 48: 55–68.

——— (2008). Sensitivity to grammatical and sociophonetic variability in perception. Oral presentation given at the Eleventh Conference on Laboratory Phonology, July 1, 2008, Wellington, New Zealand.

——— (2009). A sociophonetic ethnography of Selwyn Girls' High. Doctoral dissertation, University of Canterbury, NZ.

——— (2010). Sensitivity to grammatical and sociophonetic variability in perception. *Laboratory Phonology* 1(1): 93–120.

Dresher, B. Elan (2008). The contrastive hierarchy in phonology, in P. Avery, B. E. Dresher, and K. Rice (eds.), *Contrast in Phonology: Perception and Acquisition*. Berlin: Mouton, 11–33.

Dresher, B. Elan (2009). *The Contrastive Hierarchy in Phonology*. Cambridge: Cambridge University Press.

Duanmu, San (1994). Against contour tone units. *Linguistic Inquiry* 25: 555–608.

Dupoux, Emmanuel, Christophe, P., Sebastian-Galles, Núria, and Mehler, Jacques (1997). A distressing deafness in French. *Journal of Memory and Language* 36: 406–21.

——Kakehi, Kazuhiko, Hirose, Yuki, Pallier, Christophe, and Mehler, Jacques (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25(6): 1568–78.

——Pallier, Christophe, Kakehi, Kazuhiko, and Mehler, Jacques (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes* 5: 491–505.

Durand, Jacques and Laks, Bernard (eds.) (1996). *Current Trends in Phonology: Models and Methods.* Salford: University of Salford Publications.

Durand, Marguerite (1955). Du rôle de l'auditeur dans la formation des sons du langage. *Journal de Psychologie Normale et Pathologique* 52: 347–55.

Dyer, Judy M. (2002). "We all speak the same round here": Dialect levelling in a Scottish-English community. *Journal of Sociolinguistics* 6: 99–116.

Dyhr, Niels (1990). The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels. *Phonetica* 47(3–4): 141–54.

Eady, Stephen and Cooper, William (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80: 402–16.

Echols, Catherine and Newport, Elissa (1992). The role of stress and position in determining first words. *Language Acquisition* 2: 189–220.

Eckert, Penelope (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change* 1: 245–67.

——(2000). *Linguistic Variation as Social Practice*. Oxford: Blackwell.

——(2005). Variation, convention, and social meaning. Plenary address delivered at the Linguistic Society of America annual meeting. <http://www.stanford.edu/~eckert/thirdwave.html>, accessed May 30, 2009.

——and McConnell-Ginet, Sally (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology* 21: 461–90.

Eddington, David (1996). Diphthongization in Spanish derivational morphology: An empirical investigation. *Hispanic Linguistics* 8: 1–35.

Edlund, Jens, Beskow, Jonas, Elenius, Kjell, Hellmer, Kahl, Strömbergsson, Sofia, and House, David (2010). Spontal: A Swedish spontaneous dialogue corpus of audio, video and motion capture, in N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias (eds.), *Proceedings of the Seventh Conference on International Language Resources and Evaluation* (LREC'10), Valetta, Malta, 2992–5.

Edmonson, Jerold and Esling, John (2006). The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology* 23: 157–91.

Edwards, Jan (1992). Compensatory speech motor abilities in normal and phonologically disordered children. *Journal of Phonetics* 20: 189–207.

——and Beckman, Mary E. (2008a). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development* 4: 122–56.

——— ——— (2008b). Methodological questions in studying phonological acquisition. *Clinical Linguistics and Phonetics* 22: 939–58.

——— ——— and Fletcher, Janet (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89(1): 369–81.

——— ——— and Munson, Benjamin (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research* 47: 421–36.

——— Fourakis, Marios, Beckman, Mary E., and Fox, Robert A. (1999). Characterizing knowledge deficits in phonological disorders. *Journal of Speech, Language, and Hearing Research* 42: 169–86.

——— Fox, Robert A., and Rogers, Catherine (2002). Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and phonetic inventory size. *Journal of Speech, Language, and Hearing Research* 45: 231–42.

——— Gibbon, Fiona, and Fourakis, Marios (1997). On discrete changes in the acquisition of the alveolar/velar stop consonant constrast. *Language and Speech* 40: 203–10.

Eefting, Wieke (1991). The effect of "information value" and "accentuation" on the duration of Dutch words, syllables, and segments. *Journal of the Acoustical Society of America* 89(1): 412–24.

Eguchi, Satoshi and Hirsch, Ira (1969). Development of speech sounds in children. *Acta Otolaryngologica Supplementum* 257: 1–51.

Eilers, Rebecca E. and Minifie, Fred D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research* 18(1): 158–67.

——— Wilson, Weskey R., and Moore, John M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research* 20: 766–80.

Eimas, Peter D., Miller, Joanne L., and Jusczyk, Peter W. (1987). On infant speech perception and the acquisition of language, in S. Harnad (ed.), *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press, 161–95.

——— Siqueland, Einar R., Jusczyk, Peter, and Vigorito, James (1971). Speech perception in infants. *Science* 171: 303–6.

Eisner, Frank (2006). Lexically-guided perceptual learning in speech processing. Ph.D. dissertation, Nijmegen University.

——— and McQueen, James M. (2006). Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America* 119: 1950–3.

Elbers, Loekie and Wijnen, F. (1992). Effort, production skill, and language learning, in C. A. Ferguson, L. Menn, and C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications.* Timonium, MD: York Press, 337–68.

Elenbaas, Nine (1999). A unified account of binary and ternary stress. Ph.D. dissertation, University of Utrecht. [ROA-397].

——— and Kager, René (1999). Ternary rhythm and the lapse constraint. *Phonology* 16: 273–329.

Ellis, Andrew W. and Lambon Ralph, Matthew A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 1103–23.

Ellis, Lucy and Hardcastle, William J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30: 373–96.

Elman, Jeffrey, Diehl, Randy, and Buchwald, Susan (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America* 62: 971–4.

—— and McClelland, James (1986). Exploiting lawful variability in the speech wave, in J. S. Perkell and D. H. Klatt (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Erlbaum, 360–86.

Elordieta, Gorka and Calleja, Nagore (2005). Microvariation in accentual alignment in Basque Spanish. *Language and Speech* 48: 397–439.

—— Frota, Sónia, Prieto, Pilar, and Vigário, Marina (2003). Effects of constituent weight and syntactic branching on intonational phrasing in Ibero-Romance, in M.-J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions Pty Ltd., 487–90.

—— —— and Vigário, Marina (2005). Subjects, objects and intonational phrasing in Spanish and Portuguese. *Studia Linguistica* (Special issue on Boundaries in Intonational Phonology, ed. M. Horne and M. van Oostendorp) 59: 110–43.

Engel, Andreas K., Fries, Pascal, and Singer, Wolf (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience* 2: 704–16.

Englund, Kjellrun T. (2005). Voice onset time in infant-directed speech over the first six months. *First Language* 25(2): 219–34.

—— and Behne, Dawn (2006). Changes in infant directed speech in the first six months. *Infant and Child Development* 15: 139–60.

Engstrand, Olle (1981). Acoustic constraints of invariant input representation? An experimental study of selected articulatory movements and targets. Reports of the Uppsala University Department of Linguistics 7, Department of Linguistics, Uppsala, Sweden, 67–94.

Engwall, Olov (2003). Combining MRI, EMA and EPG measurements in a three-dimensional tongue model. *Speech Communication* 41: 303–29.

Epstein, Melissa and Stone, Maureen (2005). The tongue stops here: Ultrasound imaging of the palate. *Journal of the Acoustical Society of America* 118(4): 2128–31.

Erickson, Donna (1976). A physiological analysis of the tones of Thai. Ph.D. dissertation, University of Connecticut.

—— (1994). Laryngeal muscle activity in connection with Thai tones. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 27: 135–49.

—— (2002). Articulation of extreme formant patterns for emphasized vowels. *Phonetica* 59: 134–49.

Eriksson, Anders (2007). KatPer: A perception test, replicating a classical experiment on Categorical Perception. <http://www.ling.gu.se/~anders/KatPer/Applet/test.eng.html>, accessed March 13, 2009.

Ernestus, Mirjam (2000). *Voice Assimilation and Segment Reduction in Casual Dutch, a Corpus-Based Study of the Phonology-Phonetics Interface*. Utrecht: LOT.

—— (forthcoming). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*.

—— and Baayen, R. Harald (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79: 5–38.

—— —— (2007). Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch. *Language and Cognitive Processes* 22: 1–24.

——— and Schreuder, Rob (2002). The recognition of reduced word forms. *Brain and Language* 81: 162–73.

——— Lahey, Mybeth, Verhees, Femke, and Baayen, R. Harald (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America* 120: 1040–51.

Ernst, Marc O. and Banks, Martin S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415(6870): 429–33.

Escudero, Paola (2005). Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization. Doctoral dissertation, Utrecht University, Utrecht. *LOT Dissertation Series* 113.

——— (2009). Linguistic perception of similar L2 sounds, in P. Boersma and S. Hamann (eds.), *Phonology in Perception*. Berlin: Mouton de Gruyter, 151–90.

——— and Benders, Titia (2010). Phonetic and phonological approaches to early word recognition: Empirical findings, methodological issues, and theoretical implications, in M. Everaert, T. Lentz, H. de Mulder, Ø. Nilsen, and A. Zondervan (eds.), *The Linguistics Enterprise: From Knowledge of Language to Knowledge in Linguistics.* Amsterdam: John Benjamins, 55–78.

——— ——— and Lipski, Silvia (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics* 37: 452–66.

——— and Boersma, Paul (2002). The subset problem in L2 perceptual development: Multiple category assimilation of Dutch learners of Spanish, in B. Skarabela, S. Fish, and A. H.-J. Doh (eds.), *Proceedings of the 26th Boston University Conference on Language Development.* Somerville, MA: Cascadilla.

——— ——— (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm, in S. Arunachalam, E. Kaiser, and A. Williams (eds.), *Proceedings of the 25th Penn Linguistics Colloquium*. Penn Working Papers in Linguistics 8: 71–85.

——— ——— (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–85.

——— Broersma, Mirjam, and Simon, Ellen (forthcoming). Recognition of auditorily confusable words in native listeners versus L2 and L3 learners. *Language and Cognitive Processes*.

——— Duinmeijer, I., Van den Velde, H., and Adank, P. (under review). Predicting and explaining problems in L2 vowel perception: The case of Spanish learners of Dutch.

——— Hayes-Harb, Rachel, and Mitterer, Holger (2008). Novel L2 words and asymmetric lexical access. *Journal of Phonetics* 36: 345–60.

——— and Simon, Ellen (in preparation). The effect of orthographic cues on L2 word learning: Spanish learners' acquisition of novel words containing Dutch vowel contrasts.

——— and Wanrooij, Karen (2010). The effect of L1 orthography on L2 vowel perception. *Language and Speech* 53(3): 343–65.

Escudier, Pierre, Schwartz, Jean-Luc, and Boulogne, M. (1985). Perception of stationary vowels: internal representation of the formants in the auditory system and two-formant models. *Franco- Swedish Seminar*, Société Française d'Acoustique, Grenoble, 143–74.

Esling, John (1978). The identification of features of voice quality in social groups. *Journal of the International Phonetic Association* 8: 18–23.

Espy-Wilson, Carol Y. (1992). Acoustic measures for linguistic features distinguishing the semi-vowels /w j r l/ in American English. *Journal of the Acoustical Society of America* 92(1): 736–57.

Estow, Sarah, Jamieson, Jeremy P., and Yates, Jennifer R. (2007). Self-monitoring and mimicry of positive and negative social behaviors. *Journal of Research in Personality* 41: 425–33.

Ettema, Sandra L., Kuehn, David P., Perlman, Adrienne L., and Alperin, Noah (2002). Magnetic resonance imaging of the levator veli palatini muscle during speech. *Cleft Palate Journal* 39: 130–44.

Eulitz, Carstens and Lahiri, Aditi (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience* 16: 577–83.

—— —— Obleser, Jonas, and Reetz, Henning (2003). Brain electric activity reflects the underspecification of phonological features in the mental lexicon. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain: 1631–4.

Evans, Bronwen G. and Iverson, Paul (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America* 115: 352–61.

Everitt, Brian S. and Hothorn, Torsten (2010). *Statistical Analyses using R*. Boca Raton, FL: CRC Press.

—— Landau, Sabine, and Leese, Morven (2001). *Cluster Analysis*. New York: Oxford University Press.

Evers, Vincent, Reetz, Henning, and Lahiri, Aditi (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics* 26: 345–70.

Face, Timothy L. (2005). F0 peak height and the perception of sentence type in Castilian Spanish. *Revista Internacional de Lingüística Iberoamericana* 2(6): 49–65.

—— and Prieto, Pilar (2007). Rising accents in Castilian Spanish: A revision of Sp_ToBI. *Journal of Portuguese Linguistics* 5–6: 117–46.

Fagyal, Zsuzsusansa, Swarup, Samarth, Escobar, Anna Marie, Gasser, Les, and Lakkaraju, Kiran (2010). Centers and peripheries: Network roles in language change. *Lingua* 120(8): 2061–79.

Falé, Isabel and Faria, Isabel H. (2005). A glimpse of the time course of intonation processing. *Proceedings of the 9th European Conference on Speech Communication and Technology*. Lisboa, 2377–80.

—— —— (2006). Categorical perception of intonational contrasts in European Portuguese, in R. Hoffmann and H. Mixdorff (eds.), *Proceedings of Speech Prosody*. Dresden: TUDpress Verlag der Wissenschaften GmbH, 69–72.

Fallows, Deborah (1981). Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics* 17: 309–17.

Fanselow, Gisbert, Féry, Caroline, Vogel, Ralph, and Schlesewsky, Matthias (eds.) (2006). *Gradience in Grammar: Generative Perspectives*. Oxford: Oxford University Press.

Fant, Gunnar (1959). Acoustic analysis and synthesis of speech with applications to Swedish. Ericsson Technics Report No. 1.

—— (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

—— (1979). Glottal source and excitation analysis. Speech Trans. Lab. Q. Prog. Stat. Rep. 1. Stockholm: Royal Institute of Technology, 85–107.

—— (1982). Preliminaries to analysis of the human voice source. Speech Trans. Lab. Q. Prog. Stat. Rep. 4. Stockholm: Royal Institute of Technology, 1–27.

——Liljencrants, Johan, and Lin, Qi-guaq (1985). A four-parameter model of glottal flow. Speech Trans. Lab. Q. Prog. Stat. Rep. 4. Stockholm: Royal Institute of Technology, 1–13.

Farnetani, Edda and Busà, M. G. (1994). Italian clusters in continuous speech. *Proceedings of the International Conference on Spoken Language Processing*, vol. 1. Yokohama, 359–62.

——and Recasens, Daniel (1999). Coarticulation models in recent speech production theories, in W. J. Hardcastle and N. Hewlett (eds.), *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press, 31–65.

Feagin, Crawford (2002). Entering the community: Fieldwork, in J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 1. Malden, MA: Blackwell, 20–39.

Felder, Verena, Jönsson-Steiner, Elisabet, Eulitz, Carsten, and Lahiri, Aditi (2009). Asymmetric processing of lexical tonal contrast in Swedish. *Attention, Perception and Psychophysics* 71: 1890–9.

Feldman, Laurie B. (2003). Morphological processing as revealed through the repetition priming task, in J. Bowers and C. Marsolek (eds.), *Rethinking Implicit Memory*. Oxford: Oxford University Press.

Feldman, Naomi H., Griffiths, Thomas L., and Morgan, James L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review* 116(4): 752–82.

Feng, Ching-Mei, Narayana, Shalini, Lancaster, Jack L., Jerabek, Paul A., Arnow, Thomas L., Zhu, Fang, Tan, Lu Hai, Fox, Peter T., and Gao, Jia-Hong (2004). CBF changes during brain activation: fMRI vs. PET. *Neuroimage* 22: 443–6.

——Byers-Heinlein, Krista, and Werker, Janet F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development* 78: 1510–25.

——and Waxman, Sandra R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development* 81(5): 1376–83.

——and Werker, Janet F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech* 46(2–3): 245–64.

————(2004). Infant attention to phonetic detail: Knowledge and familiarity effects. *Proceedings of the 28th Annual Boston University Conference on Language Development*. Boston: Cascadilla Press, 165–76.

Ferguson, Charles, Menn, Lise, and Stoel-Gammon, Carol (eds.). (1992). *Phonological Development*. Timonium, MD: York Press.

Fernald, Anne (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development* 8: 181–95.

——(2000). Speech to infants as hyperspeech: Knowledge-driven process in early word recognitions. *Phonetica* 57: 242–54.

——Pinto, John P., Swingley, David, Weinberg, Amy, and McRoberts, Gerald W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science* 9: 228–31.

——and Simon, Thomas (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology* 20: 104–13.

Ferrand, Ludovic and Grainger, Jonathan G. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology* 45: 353–72.

————(1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society* 31: 119–22.

Ferrand, Ludovic (1994). Effects of orthography are independent of phonology in masked form priming. *Quarterly Journal of Experimental Psychology* 47A: 431–41.

—— Segui, Juan, and Grainger, Jonathan (1996). Masked priming of words and picture naming: The role of syllabic units. *Journal of Memory and Language* 35: 708–23.

—— —— and Humphreys, Glyn W. (1997). The syllable's role in word naming. *Memory and Cognition* 25: 458–70.

Ferreira, Fernanda (1993). Creation of prosody during sentence production. *Psychological Review* 100(2): 233–53.

—— and Tanenhaus, Michael K. (eds.) (2007–8). Special issue(s) on language–vision interactions. *Journal of Memory and Language* 57 and 58.

Féry, Caroline, Fanselow, Gisbert, and Krifka, Manfred (eds.) (2007). The notions of information structure. *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS)* 6. Potsdam: Universitätsverlag Potsdam.

—— and Ishihara, Shin (2009). The phonology of second occurrence focus. *Journal of Linguistics* 45: 285–313.

—— and Kügler, Frank (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36(4): 680–703.

Feyerabend, Paul (1975). *Against Method.* Medawah, NJ: Humanities Press.

Fidelholtz, James L. (1975). Word frequency and vowel reduction in English, in R. E. Grossman, L. J. San, and T. J. Vance (eds.), *Papers from the 11th Regional Meeting Chicago Linguistic Society*, 200–13.

Fifer, William P. and Moon, Christine (2003). Prenatal development, in A. Slater and G. Bremner (eds.), *An Introduction to Developmental Psychology*. Oxford: Blackwell, 95–114.

Fikkert, Paula (1994). *On the Acquisition of Prosodic Structure*. Dordrecht: Holland Institute of Generative Linguistics.

—— (2005). Getting sounds structures in mind. Acquisition bridging linguistics and psychology?, in A. Cutler (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*. Mahwah, NJ: Lawrence Erlbaum, 43–56.

—— and Levelt, Clara C. (2008). How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar, in P. Avery, B. E. Dresher, and K. Rice (eds.), *Contrast in Phonology: Perception and Acquisition*. Berlin: Mouton.

Findlay, John M. (2004). Eye scanning and visual search, in J. M. Henderson and F. Ferreira (eds.), *The Interface of Language, Vision and Action: Eye Movements and the Visual World*. New York: Psychology Press, 135–59.

Fischer-Jørgensen, Eli (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica* 47: 99–140.

Fitch, W. Tecumseh and Giedd, Jay (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America* 106: 1511–22.

Fitts, Paul M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47(6), 381–91.

Fitzpatrick, Jennifer and Wheeldon, Linda R. (2000). Phonology and phonetics in psycholinguistic models of speech perception, in N. Burton-Roberts, P. Carr, and G. J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press, 131–60.

Flack, Kathryn (2005). Lateral acoustics and phonotactics in Australian languages, in K. Flack and S. Kawahara (eds.), *Papers in Experimental Phonetics and Phonology*. UMOP 31. Amherst: GLSA, University of Massachusetts, 37–57.

Flagg, Elissa J., Cardy, Janis E. O., and Roberts, Timothy P. L. (2006). MEG detects neural consequences of anomalous nasalization in vowel-consonant pairs. *Neuroscience Letters* 397: 263–8.

Flanagan, James L. (1972). *Speech Analysis Synthesis and Perception*, 2nd edn. Berlin: Springer-Verlag.

Flege, James E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65.

—— (1991). Age of learning affects the authenticity of voice-onset time (VOT) stop consonants produced in second language, *Journal of the Acoustical Society of America* 89: 395–411.

—— (1992). Speech learning in a second language, in C. Ferguson, L. Menn, and C. Stoel-Gamm (eds.), *Phonological Development: Models, Research, and Implications*. Timonium, MD: York, 565–604.

—— (1995). Second language speech learning: Theory, findings, and problems, in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press, 233–77.

—— (1999). Age of learning and second language speech, in D. Birdsong (ed.), *Second-Language Learning and the Critical Period Hypothesis*. London: Erlbaum, 101–31.

—— (2003). Assessing constraints on second-language segmental production and perception, in A. Meyer and N. Schiller (eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Berlin: Mouton de Gruyter, 319–55.

—— (2006). Language contact in bilingualism: Phonetic system interactions, in J. Cole and A. Hualde (eds.), *Laboratory Phonology* 9. Berlin: Mouton.

—— Birdsong, David, Bialystok, E., Mack, Molly, Sung, H., and Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults, *Journal of Phonetics* 33: 153–75.

—— Bohn, Ocke-Schwen, and Jang, Sunyoung (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437–70.

—— and Eefting, Wieke (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15: 67–83.

—— and Hillenbrand, James (1987). Limits on phonetic accuracy in foreign language speech production, in G. Ioup and S. Weinberger (eds.), *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. Cambridge: Newbury House, 176–201.

—— and MacKay, Ian R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34.

—— —— and Meador, Diane (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106: 2973–87.

—— Munro, Murray, and MacKay, Ian R.A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16: 1–26.

Flege, James E. and Skelton, Laurie (1992). Production of word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America* 92: 128–43.

—— and Wang, Chilin (1989). Native language phonotactic constraints affect how well Chinese subjects perceive the word final English /t/-/d/ contrast. *Journal of Phonetics* 17: 299–315.

——Yeni-Komshian, Grace H., and Liu, Serena (1999). Age constraints on second-language acquisition. *Journal of Memory and Language* 41: 78–104.

Fleiss, Joseph L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5): 378–82.

Flemming, Edward (1995). Auditory representations in phonology. Doctoral dissertation, UCLA.

—— (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7–44.

Fletcher, Janet (2010). The prosody of speech: Timing and rhythm, in W. J. Hardcastle, J. Laver, and F. E. Gibbon (eds.), *The Handbook of Phonetic Sciences*, 2nd edn. Oxford: Blackwell, 521–602.

Fodor, Janet D. (2002). Psycholinguistics cannot escape prosody. *Proceedings of the Speech Prosody 2002 Conference*, Aix-en-Provence, France, 83–8.

Folkins, John W. and Zimmermann, Gerald N. (1981). Jaw-muscle activity during speech with the mandible fixed. *Journal of the Acoustical Society of America* 69: 1441–4.

Fontanari, José F. and Perlovsky, Leonid I. (2004). Solvable null model for the distribution of word frequencies. *Physical Review* E 70(4): 042901.

Foote Michael, Crampton, James S., Beu, Alan G., and Cooper, Roger A. (2008). On the bidirectional relationship between geographic range and taxonomic duration. *Paleobiology* 34: 421–33.

Forrest, Karen, Weismer, Gary, Milenkovic, Paul, and Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America* 84: 115–23.

Forster, Kenneth I. and Dickinson, Rod G. (1976). More on the language-as-fixed effect: Monte-Carlo estimates of error rates for F1, F2, F', and minF. *Journal of Verbal Learning and Verbal Behavior* 15: 135–42.

Fougeron, Cécile (1999). Prosodically conditioned articulatory variations: A review. *UCLA Working Papers in Phonetics* 97: 1–74.

—— and Keating, Patricia (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101: 3728–40.

—— and Steriade, Donca (1997). Does deletion of French schwa lead to neutralization of lexical distinctions? *Proceedings of the 5th European Conference on Speech Communication and Technology* (University of Patras), vol. 2, 943–6.

Foulkes, Paul (2010). Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology* 1: 5–39.

—— and Docherty, Gerard J. (2000). Another chapter in the story of /r/: "labiodental" variants in British English. *Journal of Sociolinguistics* 4: 30–59.

—— —— (2006). The social life of phonetics and phonology. *Journal of Phonetics* 34(4): 409–38.

—— —— and Jones, Mark (2010). Best practices in sociophonetics: Stops, in M. Yaeger-Dror and M. Di Paolo (eds.), *Sociophonetics: A Student's Guide.* New York: Routledge.

——— and Watt, Dominic J. L. (2005). Phonological variation in child-directed speech. *Language* 81: 177–206.

——— Scobbie, James M., and Watt, Dominic (2010). Sociophonetics, in W. Hardcastle and J. Laver (eds.), *Handbook of Phonetic Sciences*, 2nd edn. Oxford: Blackwell, 703–54.

Fournier, Rachel, Gussenhoven, Carlos, Jensen, Ole, and Hagoort, Peter (2010). Lateralization of tonal and intonational pitch processing: An MEG study. *Brain Research* 1328: 79–88.

Fowler, Carol A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8: 113–33.

——— (1984). Segmentation of coarticulated speech in perception. *Perception and Psychophysics* 36: 359–68.

——— (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14: 3–28.

——— (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* 99(3): 1730–41.

——— (2000). Imitation as a basis for phonetic learning after the critical period. Paper presented at the Twenty-fifth Annual Meeting of the Berkeley Linguistics Society, Berkeley, California.

——— (2007). Speech production, in M. G. Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press, 489–502.

——— and Brown, Julie M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception and Psychophysics* 62(1): 21–32.

——— ——— Sabadini, Laura, and Weihing, Jeffrey (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language* 49: 396–413.

——— and Dekle, Dawn J. (1991). Listening with eye and hand: Cross modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 17: 816–28.

——— and Housum, Jonathan (1987). Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26: 489–504.

——— Richardson, Michael, Marsh, Kerry, and Shockley, Kevin (2008). Language use, coordination, and the emergence of cooperative action, in A. Fuchs and V. Jirsa (eds.), *Understanding Complex Systems*. Berlin: Springer, 261–79.

——— and Rosenblum, Lawrence D. (1991). The perception of phonetic gestures, in I. G. Mattingly, A. M. Liberman, and M. Studdert-Kennedy (eds.), *Modularity and the Motor Theory of Speech Perception*. Hillsdale, NJ: Lawrence Erlbaum, 33–60.

——— Rubin, Paul, Remez, Robert, and Turvey, Michael (1980). Implications for speech production of a general theory of action, in B. Butterworth (ed.), *Language Production, Volume 1: Speech and Talk*. London: Academic Press, 373–420.

——— and Smith, Mary R. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance, in J. S. Perkell and D. H. Klatt (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Erlbaum, 123–36.

Fox, Robert A. (1974). An experiment in cross-dialect vowel perception, in M. W. La Galy, R. A. Fox, and A. Bruck (eds.), *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 178–85.

Fox, Robert A. (1983). Perceptual structure of monophthongs and diphthongs in English. *Language and Speech* 26: 21–60.

Francis, Alexander L. and Ciocca, Valter (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *Journal of Acoustical Society of America* 114: 1611–21.

—— —— and Kei Chit Ng, Brenda (2003). On the (non)categorical perception of lexical tones. *Perception and Psychophysics* 65: 1029–44.

—— —— Ma, Lian, and Fenn, Kimberly (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics* 36: 268–94.

—— and Nusbaum, Howard C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance* 28: 349–66.

Frank, Austin F. and Jaeger, T. Florian (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogScio8).* Washington, DC, July, 2008, 939–44.

Frauenfelder, Uli H. and Tyler, Lorraine K. (1987). *Spoken Word Recognition*. Amsterdam: Elsevier.

Friederici, Angela D. and Alter, Kai (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language* 89: 267–76.

—— and Wessels, Jeanine M. I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception and Psychophysics* 54: 287–95.

Friedrich, Claudia K. (2005). Neurophysiological correlates of mismatch in lexical access. *BMC Neuroscience* 6: 64.

—— Eulitz, Carsten, and Lahiri, Aditi (2006). Not every pseudoword disrupts word recognition: An ERP study. *Behavioral and Brain Functions* 2: 1–36. <http://www.behavioralandbrainfunctions.com/content/2/1/36>.

—— Kotz, Sonya A., Friederici, Angela, and Gunter, Thomas C. (2004). ERP correlates of lexical identification in word fragment priming. *Journal of Cognitive Neuroscience* 16: 541–52.

—— Lahiri, Aditi, and Eulitz, Carsten (2008). Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance* 34(6): 1545–59.

Frisch, Stefan A. (1996). Similarity and frequency in phonology. Ph.D. dissertation, Northwestern University.

—— (2000). Temporally organized lexical representations as phonological units, in M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon.* Cambridge: Cambridge University Press, 283–9.

—— Broe, Michael B., and Pierrehumbert, Janet B. (1995). The role of similarity in phonology: Explaining OCP-Place. *Proceedings of the 13th International Conference of the Phonetic Sciences*, Stockholm, Sweden: 544–7.

—— Large, Nathan R., and Pisoni, David B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42: 481–96.

—— —— Zawaydeh, Bushra, and Pisoni, David B. (2001). Emergent phonological generalizations in English and Arabic, in J. L. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 159–80.

—— Pierrehumbert, Janet B., and Broe, Michael (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22: 179–228.
—— and Wright, Richard (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30: 139–62.
—— and Zawaydeh, Bushra A. (2001). The psychological reality of OCP-Place in Arabic. *Language* 77: 91–106.
Fromkin, Victoria A. (1971). The non-anomalous nature of anomalous utterances. *Language* 47: 27–52.
—— (ed.) (1973). *Speech Errors as Linguistic Evidence*. The Hague: Mouton.
—— (1978). Introduction, in V. Fromkin (ed.), *Tone: A Linguistic Survey*. New York: Academic Press, 1–40.
—— (1988). Grammatical aspects of speech errors, in F. J. Newmeyer (ed.), *Linguistics: The Cambridge Survey, Volume II, Linguistic Theory: Extensions and Implications*. Cambridge: Cambridge University Press, 117–38.
Frota, Sónia (2000). *Prosody and Focus in European Portuguese. Phonological Phrasing and Intonation*. New York: Garland Publishing.
—— (2002). Tonal association and target alignment in European Portuguese nuclear falls, in C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology* 7. The Hague: Mouton de Gruyter, 387–418.
—— (forthcoming). The intonational phonology of European Portuguese, in S.-A. Jun (ed.), *Prosodic Typology II*. Oxford: Oxford University Press.
—— D'Imperio, Mariapaola, Elordieta, Gorka, Prieto Pilar, and Vigário, Marina (2007). The phonetics and phonology of intonational phrasing in Romance, in P. Prieto, J. Mascaró, and M.-J. Solé (eds.), *Prosodic and Segmental Issues in (Romance) Phonology*. Amsterdam and Philadelphia: John Benjamins, 131–53.
—— Severino, Cátia, and Vigário, Marina (2009). Syntactic disambiguation: The role of prosody. Paper presented at the Workshop on Prosody and Meaning, Barcelona.
—— and Vigário, Marina (2007). Intonational phrasing in two varieties of European Portuguese, in T. Riad and C. Gussenhoven (eds.), *Tones and Tunes*, vol. 1. Berlin: Mouton de Gruyter, 265–91.
Fry, Daniel B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27(4): 765–8.
—— (1958). Experiments in the perception of stress, *Language and Speech* 1: 126–52.
Frye, Richard E., McGraw Fisher, Janet, Coty, Alexis, Zarella, Melissa, Liederman, Jacqueline, and Halgren, Eric. (2007). Linear coding of voice onset time. *Journal of Cognitive Neuroscience* 19: 1476–87.
Fuchs, Susanne, Brunner, Jana, and Busler, A. (2007). Temporal and spatial aspects concerning the realizations of the voicing contrast in German alveolar and postalveolar fricatives. *Advances in Speech-Language Pathology* 9(1): 1–11.
—— and Koenig, L. L. (2009). Simultaneous measures of electropalatography and intraoral pressure in selected voiceless lingual consonants and consonant sequences of German. *Journal of the Acoustical Society of America* 126(4): 1988–2001.
Fujimura, Osamu (1981). Temporal organization of articulatory movements as a multidimensional phrasal structure. *Phonetica* 38: 66–83.
—— and Lovins, Julie Beth (1977). *Syllables as Concatenative Phonetic Units.* Bloomington, IN: Indiana University Linguistics Club.

Fujimura, Osamu, Macchi, Marian J., and Streeter, Lynn A. (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21: 337–46.

Fujisaki, Hiroya, Wentao Gu, and Ohno, Sumio (2007). Physiological and physical bases of the Command-Response Model for generating fundamental frequency contours in tone languages: Implications for the phonology of tones, in M.-J. Sole, P. Beddor, and M. Ohala (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 228–45.

Gafos, Adamantios (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20: 269–337.

—— (2006). Dynamics in grammar: Comments on Ladd and Ernestus & Baayen, in L. Goldstein, D. Whalen, and C. Best (eds.), *Laboratory Phonology 8*. Berlin and New York: Mouton de Gruyter, 51–79.

—— and Benus, Stefan (2003). On neutral vowels in Hungarian, in M.-J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Universitat Autònoma de Barcelona, 77–80.

—— —— (2006). Dynamics of phonological cognition. *Cognitive Science* 30: 905–43.

—— and Kirov, Christo (2010). A dynamical model of change in phonological representations: The case of lenition, in F. Pellegrino, E. Marsico, I. Chitoran, and C. Coupé (eds.), *Approaches to Phonological Complexity, Phonology & Phonetics Series*. Berlin: Mouton de Gruyter, 225–46.

Gage, Nicole M. and Roberts, Timothy P. L. (2000). Temporal integration: Reflections in the M100 of the auditory evoked field. *Neuroreport* 11: 2723–6.

—— —— and Hickok, Gregory (2002). Hemispheric asymmetries in auditory evoked neuromagnetic fields in response to place of articulation contrasts. *Cognitive Brain Research* 14: 303–6.

Gahl, Susanne (2008). Time and Thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3): 474–96.

—— and Yu, Alan (2006). Introduction to the special issue on exemplar-based models in linguistics. *Linguistic Review* 23(3): 213.

—— and Yu, Alan (eds.) (2006). *Linguistic Review* 23(3). *Special Issue on Exemplar-Based Models in Linguistics*. Berlin: De Gruyter Mouton.

Gallardo del Puerto, Francisco (2007). Is L3 phonological competence affected by the learner's level of bilingualism?, *International Journal of Multilingualism* 4: 1–16.

Gandour, Jackson (1974). On the representation of tone in Siamese, in J. G. Harris and J. R. Chamberlain (eds.), *Studies in Tai Linguistics in honor of William J. Gedney*. Bangkok: Central Institute of English Language, 170–95. (Also published in *UCLA Working Papers in Phonetics* 27: 118–46.)

—— (1978). The perception of tone, in V. Fromkin (ed.), *Tone: A Linguistic Survey*. New York: Academic Press, 41–76.

—— (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics* 9: 20–36.

—— (1983). Tone perception in Far Eastern languages. *Journal of Phonetics* 11:149–75.

—— (2007). Neural substrates underlying the perception of linguistic prosody, in C. Gussenhoven and T. Riad (eds.), *Tones and Tunes, Volume 2: Experimental Studies in Word and Sentence Prosody*. Berlin and New York: Mouton de Gruyter, 3–25.

—— and Harshman, Richard (1978). Cross-language differences in tone perception: A multi-dimensional scaling investigation. *Language and Speech* 21: 1–33.

——Petty, Soranee H., Dardarananda, Rochana, Dechongkit, Sumalee, and Mukongoen, Sunee (1986). The acquisition of the voicing contrast in Thai: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 13: 561–72.

——Ponglorpisit, Suvit, Dechongkit, Sumalee, Khunadorn, Fuangfa, Boongird, Prasert, and Potisuk, Siripong (1993). Anticipatory tonal coarticulation in Thai noun compounds after unilateral brain damage. *Brain and Language* 45(1): 1–20.

——— Potisuk, Siripong, Khunadorn, F., Boongird, P., and Dechongkit, Sumalee (1997). Interaction between tone and intonation in Thai after unilateral brain damage. *Brain and Language* 58: 174–96.

——Potisuk, Siripong, and Dechongkit, Sumalee (1994). Tonal coarticulation in Thai. *Journal of Phonetics* 22: 477–92.

——— Ponglorpisit, Siripong, Dechongkit, Sumalee, Khunadorn, Fuangfa, and Boongird, Prasert (1996). Tonal coarticulation in Thai after unilateral brain damage. *Brain and Language* 52(3): 505–35.

——Wong, Donald, Dzemidzic, Mario, Lowe, Mark, Tong, Yunxia, and Xiaojian, Li (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Human Brain Mapping* 18: 149–57.

——— Hsieh, Li, Weinzapfel, Bret, Van Lancker, Diana, and Hutchins, Gary (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience* 12: 207–22.

Ganong, William F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6(1): 110–25.

Gao, Man (2006). Gestural representation and alignment patterns of Mandarin tones. Presented at the 10th Conference on Laboratory Phonology, Paris, France.

——(2008). Tonal alignment in Mandarin Chinese: An articulatory phonology account. Doctoral dissertation, Yale University.

Gårding, Eva, Kratochvil, Paul, Svantesson, Jan-Olof, and Zhang, Jialu (1986). Tone 4 and tone 3 discrimination in Modern Standard Chinese. *Language and Speech* 29: 281–93.

——Zhang, Jialu, and Svantesson, Jan-Olof (1983). A generative model for tone and intonation in Standard Chinese based on data from one speaker. *Lund Working Papers* 25: 53–65.

Garnica, Olga (1977). Some prosodic and paralinguistic features of speech to young children, in C. Gallaway and B. J. Richards (eds.), *Talking to Children: Language Input and Acquisition*. New York: Cambridge University Press.

Garrod, Simon and Doherty, Gwyneth (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* 53: 181–215.

——and Pickering, Martin J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science* 1: 292–304.

Gaskell, M. Gareth (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics* 31: 447–63.

——and Ellis, Andrew W. (2009). Word learning and lexical development across the lifespan. *Philosophical Transactions of the Royal Society* B 364: 3607–15.

——and Marslen-Wilson, William (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22: 144–58.

Gaskell, M. Gareth and Marslen-Wilson, William (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 24: 380–96.

—— —— (2001). Lexical ambiguity and spoken word recognition: Bridging the gap. *Journal of Memory and Language* 44: 325–49.

—— and Snoeren, Natalie D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance* 34(6): 1632–47.

Gay, Thomas (1977). Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America* 62: 183–91.

—— (1978a). Articulatory units: Segments or syllables?, in A. Bell and J. B. Hooper (eds.), *Syllables and Segments*. Amsterdam: North-Holland Publishing, 121–31.

—— (1978b). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America* 63: 223–30.

Gee, James P. and Grosjean, Francis (1983). Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15: 411–58.

Gelman, Andrew and Hill, Jennifer (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press.

Gerfen, Chip (1999). *Phonology and Phonetics in Coatzospan Mixtec*. Dordrecht: Kluwer.

—— (2001). A critical vue of licensing by cue: Codas and obstruents in Eastern Andalusian Spanish, in L. Lombardi (ed.), *Segmental Phonology in Optimality Theory*. Cambridge: Cambridge University Press, 183–205.

Gerken, LouAnn (1994a). Young children's representation of prosodic structure: Evidence from English-speakers' weak syllable omissions. *Journal of Memory and Language* 33: 19–38.

—— (1994b). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language* 21: 565–84.

—— (1996). Prosodic structure in young children's language production. *Language* 72: 683–712.

—— and Bollt, Alex (2008). Three exemplars allow at least some linguistic generalizations: Implications for generalization mechanisms and constraints. *Language Learning and Development* 4: 228–48.

—— and McIntosh, Bonnie J. (1993). The interplay of function morphemes and prosody in early language. *Developmental Psychology* 29: 448–57.

German, James, Pierrehumbert, Janet, and Kaufmann, Stefan (2006). Evidence for phonological constraints on nuclear accent placement. *Language* 82: 151–68.

Gerrits, Ellen and Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics* 66(3): 363–76.

Ghini, Mirco (1993). f-formation in Italian: a new proposal. *Toronto Working Papers in Linguistics* 12(2): 41–79.

—— (2001a). *Asymmetries in the Phonology of Miogliola*. Berlin: Mouton. [Doctoral dissertation, University of Konstanz 1998].

—— (2001b). Place of articulation first, in T. A. Hall (ed.), *Distinctive Feature Theory. Phonology & Phonetics Series*. Berlin: Mouton, 147–76.

Giannelli, Luciano and Savoia, Leonardo (1979). Indebolimento consonantico in Toscana. *Revista Italiana di Diallettologia* 2: 23–58.

Gibbon, Daffyd, Moore, R., and Winski, Richard (eds.) (1997). *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter.

Gick, Bryan (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association* 32(2): 113–22.

—— (2007). A lingual motor differentiation model for liquid substitutions in children's speech. Abstract for ASA meeting, Salt Lake City, Utah, June 4–8.

—— Bird, Sonya, and Wilson, Ian (2005). Techniques for field application of lingual ultrasound imaging. *Clinical Linguistics and Phonetics* 19(6/7): 503–14.

—— Campbell, Fiona, Oh, Sunyoung, and Tamburri-Watt, Linda (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34: 49–72.

—— Pulleyblank, Douglas, Campbell, Fiona, and Mutaka, Ngessimo (2006). Low vowels and transparency in Kinande vowel harmony. *Phonology* 23: 1–20.

—— and Wilson, Ian (2006). Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets, in L. Goldstein, D. Whalen, and C. Best (eds.), *Laboratory Phonology 8*. New York: Walter de Gruyter, 635–60.

Giezen, Marcel, Escudero, Paola, and Baker, Anne (under review). Rapid learning of minimally different words in children with normal hearing and deaf children with cochlear implants.

Gilbert, Richard J. and Napadow, Vitaly J. (2005). Three-dimensional muscular architecture of the human tongue determined in vivo with diffusion tensor magnetic resonance imaging. *Dysphagia* 20: 1–7.

Giles, Howard (1984). The dynamics of speech accommodation. *International Journal of the Sociology of Language* 46: 1–155.

—— Coupland, Nikolas, and Coupland, Justine (1991a). Accommodation theory: Communication, context, and consequence, in H. Giles, N. Coupland, and J. Coupland (eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press, 1–68.

—— Coupland, Justine, and Coupland, Nikolas (eds.) (1991b). *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.

Gili Fivela, Barbara (2009). From production to perception and back: An analysis of two pitch accents, in S. Fuchs, H. Loevenbruck, D. Pape, and P. Perrier (eds.), *Some Aspects of Speech and the Brain*. Germany: Peter Lang GmbH, 363–405.

—— and D'Imperio, Mariapaola (2008). High peak vs high plateau in the identification of contrastive accents in Italian. Poster presented at Tone and Intonation in Europe (TIE) 3, September 15–17, 2008, Lisbon, Portugal.

—— and Savino, Michelina (2003). Segments, syllables and tonal alignment: A study on two varieties of Italian, in M. J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions, 2933–6.

Gimson, Alfred C. (1966). *An Introduction to the Pronunciation of English* (1st edn). London: Edward Arnold.

Glaser, Wilhelm R. (1992). Picture naming. *Cognition* 42: 61–105.

Gleitman, Lila R., January, David, Nappa, Rebecca, and Trueswell, John C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language* 57: 544–69.

—— and Wanner, Eric (1982). The state of the state of the art, in E. Wanner and L. Gleitman (eds.), *Language Acquisition: The State of the Art*. Cambridge: Cambridge University Press, 3–48.

Goad, Heather and Brannen, Kathleen (2003). Phonetic evidence for phonological structure in syllabification, in J. van de Weijer, V. J. van Heuven, and H. van der Hulst (eds.), *The Phonological Spectrum Vol II: Suprasegmental Structure*. Amsterdam: John Benjamins, 3–30.

Godfrey, John J. and Holliman, Edward (1997). Switchboard-1 Release 2. Philadelphia: Linguistic Data Consortium.

Goffman, Lisa (1999). Prosodic influences on speech production in children with specific language impairments and speech deficits: Kinematic, acoustic, and transcription evidence. *Journal of Speech, Language, and Hearing Research* 42: 1499–517.

—— (2004). Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research* 47: 1088–102.

—— Gerken, LouAnn, and Lucchesi, Julie (2007). Relations between segmental and motor variability in prosodically complex nonword sequences. *Journal of Speech, Language and Hearing Research* 50: 444–58.

Goldinger, Stephen D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22: 1166–83.

—— (1997). Words and voices: Perception and production in an episodic lexicon, in K. Johnson and J. Mullenix (eds.), *Talker Variability in Speech Processing*. San Diego: AP, 33–66.

—— (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2): 251–79.

—— (2000). The role of perceptual episodes in lexical processing, in A. Cutler, J. M. McQueen, and R. Zondervan (eds.), *Proceedings of SWAP (Spoken Word Access Processes)*. Nijmegen: Max Planck Institute for Psycholinguistics, 155–9.

—— (2007). A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 17th International Congress of Phonetic Sciences* , Saarland University, Saarbrücken, 6–10 August, 49–54.

Goldman, Michael D., Smith, H. J., and Ulmer, W. T. (2005). Whole-body plethysmography, in R. Gosselink and H. Stam (eds.), *Lung Function Testing* (European Respiratory Monograph 31). European Respiratory Society, 15–43.

Goldrick, Matthew (2007). Connectionist principles in theories of speech production, in M. G. Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press, 515–30.

—— and Blumstein, Sheila (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes* 21: 649–83.

—— and Larson, Meredith (2008). Phonotactic probability influences speech production. *Cognition* 107: 1155–64.

Goldsmith, John (1976). Autosegmental phonology. Ph.D. dissertation, MIT. [Published, New York: Garland Press, 1979.]

—— (1979). *Autosegmental Phonology*. New York: Garland.

—— (1985). Vowel harmony in Khalka Mongolian, Yaka, Finnish and Hungarian. *Phonology Yearbook* 2: 253–75.

—— (1990). *Autosegmental and Metrical Phonology*. Oxford: Blackwell.

—— (ed.) (1995). *The Handbook of Phonological Theory*. Cambridge, MA: Blackwell.

——— (2002). Probabilistic models of grammar: Phonology as information minimization. *Phonological Studies* 5: 21–46.

Goldstein, Louis (1977). Categorical features in speech perception and production. *UCLA Working Papers in Phonetics* 39: 1–36.

——— (1983). Vowel shifts and articulatory-acoustic relations, in A. Cohen and M. P. R. v. d. Broecke (eds.), *Abstracts of the 10th International Congress of Phonetic Sciences*. Dordrecht: Foris, 267–73.

——— Byrd, Dani, and Saltzman, Elliot (2006). The role of vocal tract gestural action units in understanding the evolution of phonology, in M. Arbib (ed.), *From Action to Language: The Mirror Neuron System*. Cambridge: Cambridge University Press, 215–49.

——— Chitoran, Ioana, and Selkirk, Elisabeth (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber, in J. Trouvain and W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarland University, Saarbrücken, 241–4, <http://www.icphs2007.de>.

——— Pouplier, Marianne, Chen, Larissa, Saltzman, Elliot, and Byrd, Dani (2007). Dynamic action units slip in speech production errors. *Cognition* 103: 386–412.

Goldstein, Michael H. and Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science* 19: 515–22.

——— ——— (2009). From birds to words: Perception of structure in social interactions guides vocal development and language learning, in M. S. Blumberg, J. H. Freeman, and S. R. Robinson (eds.), *The Oxford Handbook of Developmental and Comparative Neuroscience*. Oxford: Oxford University Press.

Goldstone, Robert L. and Medin, Doug L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 29–50.

Goldwater, Sharon and Johnson, Mark (2003). Learning OT constraint rankings using a Maximum Entropy Model, in J. Spenader, A. Eriksson, and Ö. Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Stockholm, Stockholm University Department of Linguistics, 111–20.

Golestani, Narly and Zatorre, Robert J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language* 109: 55–67. doi:10.1016/j.bandl.2008.01.005.

Golinkoff, Roberta M., Hirsh-Pasek, Kathy, Cauley, Kathleen M., and Gordon, Laura (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language* 14: 23–45.

Gómez, Rebecca L. (2002). Variability and detection of invariant structure. *Psychological Science* 13(5): 431–6.

González Ardeo, Juan M. (2001). Engineering students and ESP in the Basque Country: SLA versus TLA, in J. Cenoz, B. Hufeisen, and U. Jessner (eds.), *Looking Beyond Second Language Acquisition: Studies in Tri- and Multilingualism*. Tubingen: Stauffenburg, 75–95.

Gordeeva, Olga B. (2008). The relative importance of laryngeal, supralaryngeal and temporal speech production levels in the implementation of the Scottish Vowel Length Rule. Poster paper presented at the Meeting of the British Association of Academic Phoneticians (BAAP), Sheffield.

Gordon, Elizabeth (1997). Sex, speech, and stereotypes: Why women use prestige forms more than men. *Language in Society* 26: 47–64.

Gordon, Jean K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language* 82: 113–45.

Gordon, Matthew (2004). Syllable weight, in B. Hayes, R. Kirchner, and D. Steriade (eds.), *Phonetically Based Phonology*. Cambridge: Cambridge University Press, 277–312.

—— (2008). Pitch accent timing and scaling in Chickasaw. *Journal of Phonetics* 36: 521–35.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'l' and 'r'. *Neuropsychologia* 9: 317–23.

Gottfried, Terry L. (1984). Perception of temporal and spectral information in French vowels. *Language and Speech* 31: 57–75.

—— (2007). Music and language learning. Effect of musical training on learning L2 speech contrasts, in O.-S. Bohn and M. Munro (eds.), *Language Experience in Second-language Speech Learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins, 222–37.

Goudbeek, Martin, Cutler, Anne, and Smits, Roel (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication* 50: 109–25.

—— Smits, Roel, Swingley, Daniel, and Cutler, Anne (2005). Acquiring auditory and phonetic categories, in H. Cohen, and C. Lefebvre (eds.), *Categorization in Cognitive Science*. Amsterdam: Elsevier, 497–513.

Gout, Ariel, Christophe, Anne, and Morgan, James L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language* 51: 548–67.

Govender, Natasha, Barnard, Etienne, and Davel, Marelie (2007). Pitch modelling for the Nguni languages. *South African Computer Journal* 38: 28–39.

Gow, David W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* 45: 133–59.

—— (2002a). Does assimilation create lexical ambiguity? *Experimental Psychology: Human Performance* 28: 163–79.

—— (2002b). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28(1): 163–79.

—— (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics* 65: 575–90.

—— and Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance* 21: 344–59.

—— and Im, Aaron M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language* 51: 279–96.

—— and McMurray, Bob (2004). From sound to sense and back again: The integration of lexical and speech processes. *The Proceedings of From Sound to Sense: 50+ Years of Discoveries in Speech Communication*. Boston MA.

—— —— (2007). Word recognition and phonology: The case of English coronal place assimilation, in J. S. Cole and J. Hualde (eds.), *Laboratory Phonology 9*. New York: Mouton de Gruyter, 173–200.

Grabe, Esther. (1998). Pitch accent realisation in English and German. *Journal of Phonetics* 26: 129–44.

—— (2001). The IViE Labeling Guide. <http://www.phon.ox.ac.uk/files/apps/IViE//guide.html>, accessed February 3, 2010.

—— (2004). Intonational variation in urban dialects of English spoken in the British Isles, in P. Gilles and J. Peters (eds.), *Regional Variation in Intonation*. Tübingen: Niemeyer, 9–31.

——Kochanski, Greg, and Coleman, John (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech* 50: 281–310.
——and Post, Brechtje (2004). Intonational variation in the British Isles, in G. Sampson and D. McCarthy (eds.), *Corpus Linguistics: Readings in a Widening Discipline*. London and New York: Continuum International, 474–81.
————and Nolan, Francis (2001a). Modelling intonational variation in English. The IViE system, in S. Puppel and G. Demenko (eds.), *Proceedings of Prosody 2000*. Poznan: Adam Mickiewicz University, 51–7.
——————(2001b). The IViE Corpus. Department of Linguistics, University of Cambridge. <http://www.phon.ox.ac.uk/old_IViE/>, accessed May 6, 2009.
——————and Farrar, Kimberly (2000). Pitch accent realization in four varieties of British English. *Journal of Phonetics* 28: 161–85.
——Rosner, Burton S., García-Albea, José E., and Zhou, Xiaolin (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech* 46: 375–401.
——————and Warren, Paul (1995). Stress shift: do speakers do it or do listeners hear it?, in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 95–110.
Graham, Louella W. and House, Arthur S. (1971). Phonological oppositions in children: A perceptual study. *Journal of the Acoustical Society of America* 49: 559–66.
Greenberg, Joseph H. (1950). The patterning of root morphemes in Semitic. *Word* 6: 162–81.
——and Jenkins, James J. (1964). Studies in the psychological correlates of the sound system of American English: I. Measuring linguistic distance from English. II. Distinctive features and psychological space. *Word* 20: 157–77.
Greenberg, Steven (1999). Speaking in shorthand: A syllable-centric perspective for understanding pronunciation variation. *Speech Communication* 29: 159–76.
——and Fosler-Lussier, Eric (2000). The uninvited guest: Information's role in guiding the production of spontaneous speech. *Proceedings of the Crest Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, 129–32.
——Hollenback, Joy, and Ellis, Dan (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard Corpus. *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP 96)* 1. Philadelphia, 24–7.
Gregory, Stanford, Webster, Stephen, and Huang, Gang (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication* 13: 195–217.
Grice, Martine (1995a). *The Intonation of Interrogation in Palermo Italian: Implications for Intonational Theory.* Tübingen: Niemeyer.
——(1995b). Leading tones and downstep in English. *Phonology* 12: 183–233.
——Ladd, D. Robert, and Arvaniti, Amalia (2000). On the place of "phrase accents" in intonational phonology. *Phonology* 17: 143–85.
Grier, J. Brown (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin* 75: 424–9.
Grieser, DiAnne and Kuhl, Patricia K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology* 25: 577–88.
Griffin, Zenzi M. and Bock, J. Kathryn (2000). What the eyes say about speaking. *Psychological Science* 11: 274–9.

Griffiths, Thomas L. and Kalish, Michael L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science* 31: 441–80.

Grimshaw, Jane (1981). Form, function, and the language-acquisition device, in C. L. Baker and J. J. McCarthy (eds.), *The Logical Problem of Language Acquisition*. Cambridge, MA: MIT Press, 165–82.

Grosjean, Francois (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28(4): 267–83.

—— (1996). Gating. *Language and Cognitive Processes* 11(6): 597–604.

—— (2001). The bilingual's language modes, in J. Nicol (ed.), *One Mind, Two Languages: Bilingual Language Processing*. Oxford: Blackwell, 1–22.

—— (2008). *Studying Bilinguals*. Oxford: Oxford University Press.

—— and Collins, Maryann (1979). Breathing, pausing, reading. *Phonetica* 36: 98–114.

Grossberg, Stephen (1976). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics* 23: 121–34.

—— (1980). How does a brain build a cognitive code? *Psychological Review* 87: 1–51.

—— (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science* 11: 23–63.

—— (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics* 31: 423–45.

Grosz, Barbara and Hirschberg, Julia (1992). Some intonational characteristics of discourse structure. *Proceedings of the International Conference on Spoken Language Processing*. Banff, October, 429–32.

Gu, Chong (2002). *Smoothing Spline ANOVA Models*. New York: Springer.

Gu, Wentao and Lee, Tan (2007). Effects of tonal context and focus on Cantonese F0. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarland University, Saarbrücken.

Guenther, Frank H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review* 102: 594–621.

—— (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders* 39: 350–65.

—— and Gjaja, Marin N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America* 100: 1111–21.

—— Husain, Fatima T., Cohen, Michael A., and Shinn-Cunningham, Barbara G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America* 106: 2900–12.

—— Nieto-Castanon, Alfonso, Ghosh, Satrajit S., and Tourville, James A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research* 47: 46–57.

—— and Perkell, Joseph S. (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech, in B. Maassen, R. Kent, H. Peters, P. van Lieshout, and W. Hulstijn (eds.), *Speech Motor Control in Normal and Disordered Speech*. Oxford: Oxford University Press, 29–49.

Guion, Susan G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica* 55: 18–52.

—— (2003). The vowel system of Quichua-Spanish Bilinguals. Age of acquisition effect on the mutual influence of the first and second languages. *Phonetica* 60: 98–128.

—— and Wayland, Ratree (2004). Aerodynmacis of [r] in tonogenesis. Paper presented at the 9th Conference on Laboratory Phonology.

Guo, Taomei and Peng, Danling (2006). Event-related potential evidence for parallel activation of two languages in bilingual speech production. *NeuroReport* 17: 1757–60.

Gussenhoven, Carlos (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

—— (2000a). The boundary tones are coming: On the non-peripheral realization of boundary tones, in M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 132–51.

—— (2000b). The lexical tone contrast of Roermond Dutch in Optimality Theory, in M. Horne (ed.), *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce*. Amsterdam: Kluwer, 129–67.

—— (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

—— (2006). Experimental approaches to establishing discreteness of intonational contrasts, in S. Suhoff, D. Lenertovà, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schliesser (eds.), *Methods in Empirical Prosody Research*. Berlin: Mouton de Gruyter, 321–34.

—— and Jacobs, H. (1998). *Understanding Phonology*. London: Arnold.

—— and Rietveld, Toni (1992). A Target-Interpolation Model for the intonation of Dutch, *Proceedings of the Second International Conference of Speech and Language Processing (ICSLP'92)*. Banff, 1235–8.

—— —— (2000). The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech* 43: 183–203.

—— —— Kerkhoff, Joop, and Terken, Jacques (2003). Transcription of Dutch Intonation: Courseware, <http://todi.let.kun.nl/ToDI/home.htm>, accessed February 3, 2010.

Gut, Ulrike (2007). Learner corpora in second language research and teaching, in J. Trouvain and U. Gut (eds.), *Non-native Prosody: Phonetic Description and Teaching Practice*. Berlin: Mouton de Gruyter, 145–67.

Guy, Gregory R. (1980). Variation in the group and the individual: The case of final stop deletion, in William Labov (ed.), *Locating Language in Time and Space.* New York: Academic Press, 1–36.

—— (1991a). Explanation in variable phonology. *Language Variation and Change* 3: 1–22.

—— (1991b). Contextual conditioning in variable lexical phonology. *Language Variation and Change* 3: 223–39.

—— (1997). Violable is variable: Optimality Theory and linguistic variation. *Language Variation and Change* 9: 333–47.

—— and Boberg, Charles (1997). Inherent variability and the Obligatory Contour Principle. *Language Variation and Change* 9: 149–64.

Haan, Judith (2002). *Speaking of Questions*. Utrecht: LOT dissertation series.

Haas, Mary (1968). Notes on a Chipewyan dialect. *International Journal of American Linguistics* 34(3): 165–75.

Haeri, Niloofar (1996). *The Sociolinguistic Market of Cairo: Gender, Class and Education*. London and New York: Kegan Paul International.

Haggard, Mark, Ambler, Stephen, and Callow, Mo (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America* 47: 613–17.

Hajek, John (1997). *Universals of Sound Change in Nasalization*. Repr. 1999. Oxford: Blackwell.

Haken, H., Kelso, J. A. Scott, and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics* 51: 347–56.

Hale, Mark, Kissock, Madelyn, and Reiss, Charles (2007). Microvariation, variation, and the features of universal grammar. *Lingua* 117: 645–65.

—— and Reiss, Charles (2000). Substance abuse and dysfunctionalism: Current trends in phonology. *Linguistic Inquiry* 31: 157–69.

Hall, Nancy E. (2003). Gestures and segments: Vowel intrusion as overlap. Ph.D. dissertation, University of Massachusetts, Amherst.

Halle, Morris (1959). *The Sound Pattern of Russian*. The Hague: Mouton.

—— (1964). On the bases of phonology, in J. A. Fodor and J. J. Katz (eds.), *The Structure of Language*. Englewood Hills, NJ: Prentice-Hall, 324–33.

—— (1978). Knowledge unlearned and untaught: What speakers know about the sounds of their language, in M. Halle, J. Bresnan, and G. A. Miller (eds.), *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press, 294–303.

—— (1992). Features, in William Bright (ed.), *Oxford International Encyclopedia of Linguistics*. New York: Oxford University Press.

—— 2002. Introduction, in M. Halle (ed.), *From Memory to Speech and Back: Papers on Phonetics and Phonology 1954–2002*. Berlin: Mouton de Gruyter, 1–17.

—— and Marantz, Alec (1993). Distributed morphology and the pieces of inflection, in K. Hale and S. J. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger.* Cambridge, MA: MIT Press, 111–76.

—— and Stevens, Kenneth N. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory* IT-8: 155–9.

—— —— (1971). A note on laryngeal features. *MIT Quarterly Progress Report* 11: 198–213.

—— and Vergnaud, Jean-Roger (1987). *An Essay on Stress.* Cambridge, MA: MIT Press.

Hallé, Pierre A. (1994). Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese. *Language and Speech* 37: 103–24.

—— Best, Catherine T., and Bchrach, A. (2003). Perception of /tl/ and /dl/ clusters: A cross-linguistic perceptual study with French and Israeli listeners, in M. J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions, 2893–6.

—— Chang, Y. C., and Best, Catherine T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics* 31: 395–421.

Halliday, Michael A. K (1967). *Intonation and Grammar in British English*. The Hague and Paris: Mouton.

—— (1970). *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.

Hämäläinen, Matti, Hari, Riita, Ilmoniemi, Risto J., Knuutila, Jukka, and Lounasmaa, Olli V. (1993). Magnetoencephalography: Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics* 65: 413–97.

Hamann, Silke (2003). Norwegian retroflexion: Licensing by cue or prosody?, in A. Dahl, K. Benzen, and P. Svenonius (eds.), *Proceedings of the 19th Scandinavian Conference on Linguistics*. Nordlyd 31, 63–77.

—— Apoussidou, Diana, and Boersma, Paul (forthcoming). Modelling the formation of phonotactic restrictions across the mental lexicon, *Proceedings of the 45th Meeting of the Chicago Linguistic Society*.

—— Boersma, P., and Ćavar, Małgorzata (2010). Language-specific differences in the weighting of perceptual cues for labiodentals. *Proceedings of New Sounds 2010*, Poznań. 167–72.

Hammond, Michael (1999). *The Phonology of English: A Prosodic Optimality-theoretic Approach*. Oxford: Oxford University Press.

—— (2004). Gradience, phonotactics, and the lexicon in English phonology. *International Journal of English Studies* 4: 1–24.

Han, Mieko S. and Kim, Kong-On (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics* 2: 223–32.

Hankamer, Jorge (1989). Morphological parsing and the lexicon in lexical representation and process, in W. Marslen-Wilson (ed.), *Lexical Representation and Process*. Cambridge, MA: MIT Press, 392–408.

Hanna, Joy E. and Tanenhaus, Michael K. (2004). Pragmatic effects on referent resolution in a collaborative task: Evidence from eye movements. *Cognitive Science* 28: 105–15.

Hannula, Deborah E. and Charan, Ranganath (2009). The eyes have it: Hippocampal activity predicts expression of memory in eye movements. *Neuron* 63: 592–9.

Hansen, Jette (2004). Developmental sequences in the acquisition of English L2 syllable codas. *Studies in Second Language Acquisition* 26: 85–124.

Hanson, Helen M. (1995). Glottal characteristics of female speakers. Ph.D. dissertation, Harvard University, Cambridge, MA.

—— (1997a). Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America* 101: 466–81.

—— (1997b). Vowel amplitude variation during sentence production, in *Proceedings of the IEEE ICASSP-97*, Munich, 1627–30.

—— (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *Journal of the Acoustical Society of America* 125(1): 425–41.

—— and Chuang, Erika S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America* 106: 1064–77.

—— Slifka, Janet, Shattuck-Hufnagel, Stefanie, and Kobler, James (2005). Identification of final fall in subglottal pressure contours of speech utterances. *Journal of the Acoustical Society of America* 119: 3393–4.

—— —— —— —— (2007). Tone distribution and its effect on subglottal pressure during speech. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 545–8.

—— and Stevens, Kenneth N. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *Journal of the Acoustical Society of America* 112(3): 1158–82.

Hanson, Kristin and Kiparsky, Paul (1996). A parametric theory of poetic meter. *Language* 72: 287–335.

Hansson, Gunnar Ó. (2001). Theoretical and typological issues in consonant harmony. Ph.D. dissertation, UC Berkeley.

—— (2003). Laryngeal licensing and laryngeal neutralization in Faroese and Icelandic. *Nordic Journal of Linguistics* 26: 45–79.

—— (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass* 2: 859–93.

Hao, Yen-Chen and de Jong, Kenneth (2007). The categorical nature of tones and consonants: Evidence from second language perception and production. *Journal of the Acoustical Society of America* 122: 3018.

Hardcastle, William J. (1972). The use of electropalatography in phonetic research. *Phonetica* 25: 197–215.

—— (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* 4: 247–63.

—— and Hewlett, Nigel (1999). *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press.

Harden, R. Joyce (1975). Comparison of glottal area changes as measured from ultrahigh-speed photographs and photoelectric glottographs. *Journal of Speech and Hearing Research* 18: 728–38.

Hare, Mary (1990). The role of similarity in Hungarian vowel harmony: A connectionist account. *Connection Science* 2: 123–50.

Harnad, Stevan R. (1990). *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.

Harnsberger, James D. (2001). The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. *Journal of Phonetics* 29: 303–27.

Harrell, Frank. (2001). *Regression Modeling Strategies*. Berlin: Springer.

Harrington, Jonathan (2006). An acoustic analysis of "happy-tensing" in the Queen's Christmas broadcasts. *Journal of Phonetics* 34: 439–57.

—— (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts, in J. Cole and J. I. Hualde (eds.), *Laboratory Phonology* 9. Berlin: Mouton de Gruyter, 125–43.

—— Beckman, Mary E., Fletcher, Janet, and Palethorpe, Sallyanne (1998). An electropalatographic, kinematic, and acoustic analysis of supralaryngeal correlates of word and utterance-level prominence contrasts in English. *Proceedings of the 1998 International Conference on Spoken Language Processing*. Australian Speech Science and Technology Association, Inc., 1851–4.

—— Fletcher, Janet, and Beckman, Mary E. (2000). Manner and place conflicts in the articulation of accent in Australian English, in M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 40–51.

—— Kleber, F., and Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America* 123: 2825–35.

—— Palethorpe, Sallyanne, and Watson, Catherine I. (2000). Does the Queen speak the Queen's English? *Nature* 408: 927–8.

—— —— —— (2005). Deepening or lessening the divide between diphthongs? An analysis of the Queen's annual Christmas broadcasts, in W. J. Hardcastle and J. Beck (eds.), *The Gift of Speech (Festschrift for John Laver)*. Hillsdale, NJ: Lawrence Erlbaum, 227–61.

—— —— —— (2007). Age-related changes in fundamental frequency and formants: A longitudinal study of four speakers. *Proceedings of Interspeech 2007*, Antwerp.

Harris, Cyril M. (1953). A study of the building blocks in speech. *Journal of the Acoustical Society of America* 25: 962–9.

—— and Wolpert, Daniel M. (1998). Signal-dependent noise determines motor planning. *Nature* 394: 780–4.

Harris, James W. (1969). *Spanish Phonology*. Cambridge, MA: MIT Press.

—— (1978). Two theories of non-automatic morphophonological alternations. *Language* 54: 41–60.

Harris, John (1994). *English Sound Structure*. Oxford and Cambridge, MA: Blackwell.

—— and Lindsey, Geoff (1995). The elements of phonological representation, in J. Durand and F. Katamba (eds.), *Frontiers of Phonology: Atoms, Structures, Derivations*. Harlow, Essex: Longman, 34–79.

—— —— (2000). Phonology without categorical phonetics, in N. Burton-Roberts, P. Carr, and G. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press, 185–205.

Harris, Katherine S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1: 1–7.

—— (1978). Vowel duration change and its underlying physiological mechanisms, *Language and Speech* 21: 354–61.

Harris, Zellig S. (1951). *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

't Hart, Johan and Cohen, Antonie (1973). Intonation by rule: A perceptual quest. *Journal of Phonetics* 1: 309–27.

—— Collier, René and Cohen, Antonie (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

Hartsuiker, Robert, Corley, Martin, and Martensen, Heike (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars, Motley, and MacKay (1975). *Journal of Memory and Language* 52: 58–70.

Hasegawa-Johnson, Mark and Fleck, Margaret (2007). The ISLEX Project, <http://www.isle.uiuc.edu/dict>, accessed April 20, 2009.

Hattori, Kota and Iverson, Paul (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America* 125(1): 469–79.

Haudricourt, André-Georges (1954). De l'origine des tons en Vietnamien. *Journal Asiatique* 242: 69–82.

Hauser, Marc D., Newport, Elissa L., and Aslin, Richard N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton top tamarins. *Cognition* 78: B53–B64.

Hawkins, Sarah (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31: 373–405.

—— (2004). Puzzles and patterns in 50 years of research on speech perception, in J. Slifka, S. Manuel, J. Perkell, and S. Shattuck-Hufnagel (eds.), *Sound to Sense: 50+ Years of Discoveries in Speech Communication*, <http://www.rle.mit.edu/soundtosense/conference/pages/invited.htm>.

—— (2010a). Phonological features, auditory objects, and illusions. *Journal of Phonetics* 38: 60–89.

—— (2010b). Phonetic variation as communicative system: Perception of the particular and the abstract, in C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (eds.), *Laboratory Phonology 10*. Berlin: Mouton de Gruyter, 479–510.

Hawkins, Sarah and Midgley, Jonathan (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35: 183–99.
—— and Nguyen, Noël (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics* 32: 199–231.
—— and Smith, Rachel (2001). Polysp: a polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics/Rivista di Linguistica* 13: 99–188.
Hay, Jennifer B. (2002). From speech perception to morphology: Affix-ordering revisited. *Language* 78(3): 527–55.
—— (2003). *Causes and Consequences of Word Structure*. New York and London: Routledge.
—— and Baayen, R. Harald (2002). Parsing and productivity, in G. E. Booij and J. van Marle (eds.), *Yearbook of Morphology*. Dordrecht: Kluwer Academic Publishers, 203–55.
—— —— (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9(7): 342–8.
—— and Bresnan, Joan (2006). Spoken syntax: The phonetics of giving a hand in New Zealand English. *The Linguistic Review* 23: 321–49.
—— and Drager, Katie (2007). Sociophonetics. *Annual Review of Anthropology* 36: 89–103.
—— —— (2010). Stuffed toys and speech perception. *Linguistics* 48(4): 865–92.
—— —— and Warren, Paul (2009). Careful who you talk to: An effect of experimenter identity on the production of the NEAR/SQUARE merger in New Zealand English. *Australian Journal of Linguistics* 29(2): 269–85.
—— —— —— (2010). Short-term exposure to one dialect affects processing of another. *Language and Speech* 53(4): 447–71.
—— and Maclagan, Margaret (2010). Social and phonetic conditioners on the frequency and degree of "intrusive /r/" in New Zealand English, in D. Preston and N. Niedzielski (eds.), *A Reader in Sociophonetics*. Trends in Linguistics Studies and Monographs 219. New York: De Gruyter Mouton, 41–69.
—— Nolan, A., and Drager, Katie (2006). From *fush* to *feesh*: Exemplar priming in speech perception. *The Linguistic Review* 23(3): 351–79.
—— Pierrehumbert, Janet, and Beckman, Mary E. (2003). Speech perception, well-formedness, and the statistics of the lexicon, in J. Local, R. Ogden, and R. Temple, (eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 58–74.
—— and Plag, Ingo (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory* 22: 565–96.
—— and Sudbury, Andrea (2005). How rhoticity became /r/-sandhi. *Language* 81(4): 799–823.
—— Warren, Paul, and Drager, Katie (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4): 458–84.
Hayes, Bruce (1984). The phonology of rhythm in English. *Linguistic Inquiry* 15: 33–74.
—— (1986). Assimilation as spreading in Toba Batak. *Linguistic Inquiry* 17: 467–99.
—— (1995). *Metrical Stress Theory.* Chicago: The University of Chicago Press.
—— (2000). Gradient well-formedness in Optimality Theory, in J. Dekkers, F. van der Leeuw, and J. van de Weijer (eds). *Optimality Theory: Phonology, Syntax, and Acquisition*. Oxford: Oxford University Press, 88–120.
—— (2009). *Introductory Phonology*. Malden, MA: Wiley-Blackwell.

——Kirchner, Robert, and Steriade, Donca (2004). *Phonetically Based Phonology*. Cambridge: Cambridge University Press.
——and Lahiri, Aditi (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory* 9: 47–96.
——and Londe, Zsuzsa (2006). Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23: 59–104.
——and McEachern, Margaret (1998). Quatrain form in English folk verse. *Language* 64: 473–507.
——Tesar, Bruce, and Zuraw, Kie (2003). OTSoft 2.1, software package, <http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
——Siptár, Péter, Zuraw, Kie, and Londe, Zsuzsa (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85: 822–63.
——and Wilson, Colin (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379–440.
Hayes, Rachel A. and Slater, Alan (2008). Three-month-olds' detection of alliteration in syllables. *Infant Behavior and Development* 31(1): 153–6.
Hazan, Valerie and Barrett, Sarah (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics* 28: 377–96.
HCRC Map Task Corpus (1993). Philadelphia: Linguistic Data Consortium.
Heeren, Willemijn. F. L. and Schouten, M. E. H. (2008). Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories. *Journal of the Acoustical Society of America* 124: 2291–302.
Heeringa, Wilbert, Johnson, Keith, and Gooskens, Charlotte (2009). Measuring Norwegian dialect distances using acoustic features. *Speech Communication* 51: 167–83.
Heldner, Mattias and Strangert, Eva (2001). Temporal effects of focus in Swedish. *Journal of Phonetics* 29: 329–61.
Hellmuth, Sam (2004). Prosodic weight and phonological phrasing in Cairene Arabic, in N. Adams, A. Cooper, F. Parrill, and T. Wier (eds.), *Proceedings from the 40th Annual Meeting of the Chicago Linguistics Society*, 97–111.
——(2005a). No de-accenting in (or of) phrases: Evidence from Arabic for cross-linguistic and cross-dialectal prosodic variation, in S. Frota, M. Vigário and M. J. Freitas (eds.), *Prosodies*. Berlin and New York: Mouton de Gruyter, 99–112.
——(2005b). Pitch accent alignment in Egyptian Arabic: Exploring the boundaries of cross-linguistic alignment variation. Paper presented at Phonetics and Phonology in Iberia, PaPI 2005, Bellaterra, Spain, June 20–21, 2005.
——(2007). The relationship between prosodic structure and pitch accent distribution: Evidence from Egyptian Arabic. *The Linguistic Review* (Special issue on Prosodic Phrasing, ed. S. Frota and P. Prieto) 24: 291–316.
van der Helm, Peter A. (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin* 126(5): 770–800.
Heman-Ackah, Yolanda and Barr, Arlene (2006). The value of laryngeal electromyography in the evaluation of laryngeal motion abnormalities. *Journal of Voice* 20: 452–60.
Henderson, John M. and Ferreira, Fernanda (2004). Scene perception for psycholinguists, in J. M. Henderson and F. Ferreira (eds.), *The Interface of Language, Vision and Action: Eye Movements and the Visual World*. New York: Psychology Press, 1–58.

Herman, Rebecca (1996). Final lowering in Kipare. *Phonology* 13: 171– 96.

Hermes, Anne, Grice, Martine, Mücke, Doris, and Niemann, Henrik (2008). Articulatory indicators of syllable affiliation in word-initial consonant clusters in Italian, in R. Sock, S. Fuchs, and Y. Laprie (eds.), *Proceedings of the 8th International Seminar on Speech Production*. Strasbourg: INRIA, 433–6, <http://issp2008.loria.fr/proceedings.html>.

Hertegård, Stellan and Gauffin, Jan (1992). Acoustic properties of the Rothenberg mask, Speech Trans. Lab. Q. Prog. Stat. Rep. 2–3, Stockholm: Royal Institute of Technology: 9–18.

—— —— and Lindestad, Per-Åke (1995). A comparison of subglottal and intraoral pressure measurements during phonation. *Journal of Voice* 9: 149–55.

Hertz, Susan R. (1990). The Delta Programming Language: An integrated approach to nonlinear phonology, phonetics, and speech synthesis, in J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 215–57.

—— (1991). Streams, phones and transitions: Toward a new phonological and phonetic model of formant timing. *Journal of Phonetics* 19: 91–109.

Heselwood, Barry and McChrystal, Louise (2000). Gender, accent features and voicing in Panjabi–English bilingual children. *Leeds Working Papers in Linguistics and Phonetics* 8: 45–70.

van Heuven, Vincent (1997). Effects of focus distribution and accentuation on the temporal and melodic organisation of word groups in Dutch, in S. Barbiers, J. Rooryck, and J. van de Weijer (eds.), *Small Words in the Big Picture. Squibs for Hans Bennis*. HIL Occasional Papers 2. Leiden: Holland Institute of Generative Linguistics, 37–42.

—— and Sluijter, Agaath M. C. (1996). Notes on the phonetics of word prosody, in R. Goedemans, H. van der Hulst, and E. Visch (eds.), *Stress Patterns of the World, Part 1: Background*, vol. 2. The Hague: Holland Institute of Generative Linguistics, Leiden/Holland Academic Graphics, 233–69.

Hewitt, Roger (1986). *White Talk, Black Talk*. Cambridge: Cambridge University Press.

Hewlett, Nigel and Waters, Daphne (2004). Gradient change in the acquisition of phonology. *Clinical Linguistics and Phonetics* 18: 523–33.

Hickok, Gregory and Poeppel, David (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92: 67–99.

—— —— (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience* 8: 393–402.

Higgins, John (2009). Minimal pairs for English RP. <http://myweb.tiscali.co.uk/wordscape/wordlist>, accessed March 13, 2009.

Hillenbrand, James M. (1983). Perceptual organization of speech sounds by infants. *Journal of Speech and Hearing Research* 26: 268–82.

—— Clark, Michael J., and Houde, Robert A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America* 108: 3013–22.

—— Getty, Laura A., Clark, Michael J., and Wheeler, Kimberlee (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97(5): 3099–111.

Hillman, Robert E., Holmberg, Eva B., Perkell, Joseph S., Walsh, Michael, and Vaughan, Charles (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research* 32: 373–92.

——— ——— ——— ——— ——— (1990). Phonatory function associated with hyperfunctionally related vocal fold lesions. *Journal of Voice* 4: 52–63.

Hinskens, Frans, van Hout, Roeland, and Wetzels, Leo (1997a). Balancing data and theory in the study of phonological variation and change, in F. Hinskens, R. van Hout, and L. Wetzels (eds.), *Variation, Change and Phonological Theory*. Amsterdam and Philadelphia: John Benjamins, 1–33.

——— ——— ——— (eds.) (1997b). *Variation, Change and Phonological Theory*. Amsterdam and Philadelphia: John Benjamins.

Hirata, Yukari, Whitehurst, Elizabeth, and Cullings, Emily (2007). Training native English speakers to identify Japanese vowel length contrasts with sentences at varied speaking rates. *Journal of the Acoustical Society of America* 121: 3837–45.

Hirose, Hajime (1997). Investigating the physiology of laryngeal structures, in W. J. Hardcastle and J. Laver (eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 116–36.

Hirschberg, Julia (2000). A corpus-based approach to the study of speaking style, in M. Horne (ed.), *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Publishers, 335–50.

——— Gravano, Agustin, Nenkova, Ani, Sneed, Elisa, and Ward, Gregory (2007). Intonational overload: Uses of the H* !H* L- L% contour in read and spontaneous speech, in J. Cole and J. I. Hualde (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 455–82.

——— and Nakatani, Christine H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*, Santa Cruz, California, 286–93.

Hirschfeld, Gerrit, Jansma, Bernadette M., Bölte, Jens, and Zwitserlood, Pienie (2008). Interference and facilitation in overt speech production investigated with event-related potentials. *NeuroReport* 19: 1227–30.

Hirsh-Pasek, Kathy, Kemler Nelson, Deborah, Jusczyk, Peter, Wright, Kimberly, Druss, Benjamin, and Kennedy, Lori (1987). Clauses are perceptual units for prelinguistic infants. *Cognition* 26: 269–86.

Hirst, Daniel, Di Cristo, Albert, and Espesser, Robert (2000). Levels of representation and levels of analysis for intonation, in M. Horne (ed.), *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Publishers, 51–87.

——— and Espesser, Robert (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix* 15: 71–85.

Hixon, Thomas (1966). Turbulent noise sources for speech. *Folia phoniatrica* 18: 168–82.

Hoben, Thomas and Gilmore, Rick O. (2004). Habituation assessment in infancy. *Psychological Methods* 9(1): 70–92.

Hock, Hans H. (1992). Causation in language change, in W. Bright (ed.), *Oxford International Encyclopedia of Linguistics*, vol. 1. London and New York: Oxford University Press, 228–31.

Hockett, Charles F. (1953). Review of Claude L. Shannon and Warren Weaver, *The Mathematical Theory of Communication*. *Language* 29: 69–93.

——— (1954). Two models of grammatical description. *Word* 10: 210–34.

——— (1955). *A Manual of Phonology*. Indiana University Publications in Anthropology and Linguistics 11. Baltimore: Waverly Press and Indiana University.

——— (1961). Linguistic elements and their relations. *Language* 37(1): 29–53.

Hockey, Beth Ann and Fagyal, Zsuzsanna (1999). Phonemic vowel length and pre-boundary lengthening: An experimental investigation on the use of durational cues in Hungarian, in *Proceedings of the XIVth ICPhS*, San Francisco, 313–16.

Hoeksema, Jacob (1985). Formal properties of stress representations, in H. van der Hulst, and N. Smith (eds.), *Advances in nonlinear phonology.* Dordrecht: Foris Publications, 83–99.

Höhle, Barbara, Schmitz, Michaela, Santelmann, Lynn M., and Weissenborn, Jürgen (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities. *Language Learning and Development* 2: 277–300.

Hojen, Anders and Flege, James E. (2006). Early learners' discrimination of second-language (L2) vowels. *Journal of the Acoustical Society of America* 119: 3072–84.

Holes, Clive (1986). The social motivation for phonological convergence in three Arabic dialects. *International Journal of the Sociology of Language* 51: 33–51.

Holmberg, Eva B., Hillman, Robert E., and Perkell, Joseph S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *Journal of the Acoustical Society of America* 84: 511–29.

——— ——— ——— Guiod, Peter, and Goldman, Susan L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research* 38: 1212–23.

Holmes, Janet (1997). Maori and Pakeha English: Some New Zealand social dialect data. *Language in Society* 26: 65–101.

Holmquist, Jonathan (1985). Social correlates of a linguistic variable: A study in a Spanish village. *Language in Society* 14: 191–203.

Holst, Tara and Nolan, Francis (1995). The influence of syntactic structure on [s] to [ʃ] assimilation, in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 315–33.

Holt, Lori L. (2006). The mean matters: Effects of statistically-defined non-speech spectral distributions on speech categorization. *Journal of the Acoustical Society of America* 120: 2801–17.

——— and Lotto, Andrew J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America* 119: 3059–71.

——— ——— (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science* 17: 42–6.

——— ——— (2010). Speech perception as categorization. *Attention, Perception & Psychophysics* 72, 1218–27.

——— ——— and Diehl, Randy L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America* 116: 1763–73.

——— ——— and Kluender, Keith R. (1998). Incorporating principles of general learning in theories of language acquisition, in M. Gruber, C. D. Higgins, K. S. Olson, and T. Wysocki (eds.), *Chicago Linguistic Society, Volume 34: The Panels*. Chicago: Chicago Linguistic Society, 253–68.

——— ——— ——— (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America* 109: 764–74.

Homae, Fumitaka, Watanabe, Hama, Nakano, Tamami, and Taga, Gentaro (2007). Prosodic processing in the developing brain. *Neuroscience Research* 59: 29–39.

Hombert, Jean Marie, Ohala, John, and Ewan, William (1979). Phonetic explanations for the development of tones. *Language* 55: 37–58.

Honda, Kiyoshi (2004). Physiological factors causing tonal characteristics of speech: From global to local prosody, in *Proceedings of Speech Prosody 2004*, Nara, 739–44.

—— and Fujimura, Osamu (1991). Intrinsic vowel F0 and phrase-final F0 lowering: Phonological vs. biological explanations, in J. Gauffin and B. Hammerberg (eds.), *Vocal Fold Physiology*. San Diego: Singular, 149–57.

Honorof, Douglas (1999). Articulatory gestures and Spanish nasal assimilation. Doctoral dissertation, Yale University.

Hoole, Philip (1999). Laryngeal coarticulation. Section A: Coarticulatory investigations of the devoicing gesture, in W. H. Hardcastle and N. Hewlett (eds.), *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press, 105–21.

—— (2006). Experimental studies of laryngeal articulation. Part I: Electromyographic investigation of laryngeal activity in vowel intrinsic pitch and consonant voicing. Unpublished habilitation thesis, Ludwig-Maximilians-Universität, Munich. <http://www.phonetik.unimuenchen.de/~hoole/pdf/habilemg_chap_all.pdf>, accessed September 12, 2009.

—— and Honda, Kiyoshi (2011). Automaticity vs. feature-enhancement in the control of segmental F0, in G. N. Clements and R. Ridouane (eds.), *Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories.* Amsterdam: John Benjamins, 131–71.

—— and Mooshammer, Christine (2002). Articulatory analysis of the German vowel system, in P. Auer, P. Gilles, and H. Spiekermann (eds.), *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer, 129–52.

—— Munhall, K., and Mooshammer, C. (1998). Do airstream mechanisms influence tongue movement paths? *Phonetica* 55(3): 131–46.

Hooper, Joan B. (1976a). *An Introduction to Natural Generative Phonology*. New York: Academic Press.

—— (1976b). Word frequency in lexical diffusion and the source of morphophonological change, in W. M. Christie (ed.), *Current Progress in Historical Linguistics*. Amsterdam: North Holland, 96–105.

Hoover, Jill R. and Storkel, Holly L. (2007). Phonological and lexical cues in word learning by preschool children in a seminar entitled Word Learning in Situ: Interplay between Learners and Learning Environments (Convener: K. K. McGregor). American Speech-Language-Hearing Association Convention, Boston, MA.

Hopper, Paul J. and Traugott, Elizabeth C. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.

Horn, David L., Houston, Derek M., and Miyamoto, Richard T. (2007). Speech discrimination skills in deaf infants before and after cochlear implantation. *Audiological Medicine* 5: 232–41.

Horn, Roger A. and Johnson, Charles R. (1990). Norms for Vectors and Matrices, ch. 5 in *Matrix Analysis.* Cambridge: Cambridge University Press.

Horne, Merle (1990). Empirical evidence for a deletion formulation of the rhythm rule in English. *Linguistics* 28: 959–81.

Horvath, Barbara and Sankoff, David (1987). Delimiting the Sydney speech community. *Language in Society* 16: 179–204.

HOUDE, ROBERT (1968). A study of tongue body motion during selected speech sounds. *Speech Communication Research Laboratory (Santa Barbara), Monograph No. 2* (also available as Ph.D. dissertation, University of Michigan, 1967).

HOUSE, DAVID (1990). *Tonal Perception in Speech*. Lund, Sweden: Lund University Press.

HOUSE, JILL (2006). Constructing a context with intonation. *Journal of Pragmatics* 38: 1542–58.

HOUSTON, DEREK M., HORN, DAVID L., QI, RONG, TING, JONATHAN, and GAO, SUJUAN (2007). Assessing speech discrimination in individual infants. *Infancy* 12: 119–45.

—— and JUSCZYK, PETER (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance* 29: 1143–54.

HOWE, DARIN and PULLEYBLANK, DOUGLAS (2001). Patterns and timing of glottalisation. *Phonology* 18: 45–80.

HOWE, MICHAEL S. and MCGOWAN, RICHARD S. (2005). Aeroacoustics of [s]. *Proceedings of the Royal Society* A, 461: 1005–28.

HRUSCHKA, DANIEL J., CHRISTIANSEN, MORTEN H., BLYTHE, RICHARD, CROFT, WILLIAM, HEGGARTY, PAUL, MUFWENE, SALIKOKO S., PIERREHUMBERT, JANET B., and POPLACK, SHANA (2009). Building social cognitive models of language change. *Trends in Cognitive Sciences* 13: 464–9.

HSIEH, LI, LEONARD, LAURENCE B., and SWANSON, LORI (1999). Some differences between English plural noun inflections and third singular verb inflections in the input: The contributions of frequency, sentence position, and duration. *Journal of Child Language* 26: 531–43.

HUDSON-KAM, CARLA L. and NEWPORT, ELISSA L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology* 59: 30–66.

HUETTIG, FALK and MCQUEEN, JAMES M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language* 57: 460–82.

HUFFMAN, MARIE K. (1991). Time-varying properties of contextually nasalized vowels: Acoustics and perception, in *Proceedings of the 12th International Congress of Phonetic Sciences.* Aix-en-Provence: Université de Provence Aix-Marseille, 130–3.

—— (2007). Laboratory phonology and socio-phonetics: Partners in a conversation whose time has come, in J. Cole and J. I. Hualde (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 113–23.

HULL, DAVID L. (ed.) (1988). *Science as a Process.* Chicago: The University of Chicago Press.

—— (1989). *The Metaphysics of Evolution.* Albany, NY: State University of New York Press.

HULST, HARRY G. VAN DER and RITTER, NANCY A. (1999). Theories of the syllable, in H. G. van der Hulst and N. A. Ritter (eds.), *The Syllable: Views and Facts*. Berlin: Mouton de Gruyter, 13–52.

—— and WEIJER, JEROEN VAN DER (1995). Vowel harmony, in J. Goldsmith (ed.), *Handbook of Phonological Theory*. Cambridge, MA: Blackwell, 495–534.

HUME, ELIZABETH and JOHNSON, KEITH (2001). *The Role of Speech Perception in Phonology*. New York: Academic Press.

HUSSAIN, SARMAD, DURRANI, NADIR, and GUL, SANA (2005). *Pan Localization: Survey of Language Computing in Asia.* Center for Research in Urdu Language Processing, Lahore, Pakistan. <http://www.idrc.ca/uploads/user-S/11446781751Survey.pdf>, accessed April 14, 2009.

Hutters, Birgit (1985). Vocal fold adjustments in aspirated and unaspirated stops in Danish. *Phonetica* 42: 1–24.

Hwang, So-One, Monahan, Philip J., and Idsardi, William J. (2010). Underspecification and asymmetries in voicing perception. *Phonology* 27: 205–24.

Hyman, Larry (1975). *Phonology: Theory and Analysis*. New York, NY: Holt, Rinehart & Winston.

—— (1976). Phonologization, in A. Juilland (ed.), *Linguistic Studies Presented to Joseph H. Greenberg*. Saratoga: Anma Libri, 407–18.

—— (1978). Historical tonology, in V. Fromkin (ed.), *Tone: A Linguistic Survey*. New York: Academic Press, 257–70.

—— (1979). A reanalysis of tonal downstep. *Journal of African Languages and Linguistics* 1: 9–29.

—— (1993). Register tones and tonal geometry, in H. van der Hulst and K. Snider (eds.), *The Phonology of Tone: The Representation of Tonal Register*. Berlin and New York: Mouton de Gruyter, 75–108.

—— (2001). The limits of phonetic determinism in phonology: *NC revisited, in E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*. New York: Academic Press, 141–85.

—— (2007). Universals of tone rules: 30 years later, in C. Gussenhoven and T. Riad (eds.), *Tones and Tunes*, vol. 2. Berlin: Mouton de Gruyter, 1–35.

—— (2008). Tonal and nontonal intonation in Shekgalagari. Presentation at the Third Conference on Tone and Intonation in Europe. Lisbon, Portugal.

Hyslop, Gwendolyn (2009). Kurtop tone: A tonogenetic case study. *Lingua* 119(6): 827–45.

Iacoboni, Marco (2008). The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. *Journal of Physiology Paris* 102: 31–4.

Igarashi, Yosuke (2004). Segmental anchoring of F0 under changes in speech rate: Evidence from Russian, in B. Bel and I. Marlien (eds.), *Proceedings of Speech Prosody 2004*, Nara, Japan, March 23–26, 2004. ISCA, 25–8.

Ijaz, Madiha and Hussain, Sarmad (2007). Corpus-based lexicon development. Paper presented at the Conference on Language and Technology, University of Peshawar, Pakistan.

Ikeda, Kazunari, Hayashi, Akiko, Hashimoto, Souichi, Otomo, Kiyoshi, and Kanno, Atsushi (2002). Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neuroscience Letters* 321: 133–6.

Imbrie, Annika K. K. (2005). Acoustical study of the development of stop consonants in children. Doctoral dissertation, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA.

Indefrey, Peter and Levelt, Willem J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92: 101–44.

Ingram, John C. L. and Park, See-Gyoon (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics* 25: 343–70.

Inkelas, Sharon (1995). The consequences of optimization for underspecification, in E. Buckley and S. Iatridou (eds.), *Proceedings of the Twenty-Fifth Northeastern Linguistics Society*. Amherst: GLSA, 287–302.

Inkelas, Sharon (1998). The theoretical status of morphologically conditioned phonology: A case study of dominance effects, in G. Booij and J. van Marle (eds.), *Yearbook of Morphology 1997*. Dordrecht: Kluwer, 121–55.

—— and Leben, Will (1990). Where phonetics and phonology intersect: The case of Hausa intonation, in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press.

—— Orgun, Orhan, and Zoll, Cheryl (1997). The implications of lexical exceptions for the nature of grammar, in I. Roca, (ed.), *Derivations and Constraints in Phonology*. Oxford: Clarendon Press, 393–418.

Iseli, Markus and Alwan, Abeer (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation, in *Proceedings of the IEEE ICASSP-04* 1, 669–72.

Ishida, Richard (2009). IPA Character Picker, <http://rishida.net/scripts/pickers/ipa>, accessed March 13, 2009.

Ishihara, Shinichiro and Féry, Caroline (2006). The phonology of second occurrence focus. *Journal of Linguistics* 45(2): 285–313.

Ishihara, Takeishi (2003). A phonological effect on tonal alignment in Tokyo Japanese, in M. J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, vol. 1. Barcelona: Causal Productions, 615–18.

Ishikawa, Keiichi (2002). Syllabification of intervocalic consonants by English and Japanese speakers. *Language and Speech* 45: 355–85.

Iskarous, Khalil (2005a). Detecting the edge of the tongue: A tutorial. *Clinical Linguistics and Phonetics* 19(6/7): 555–65.

—— (2005b). Patterns of tongue movement. *Journal of Phonetics* 33: 363–81.

—— Goldstein, Louis M., Whalen, Douglas H., Tiede, Mark K., and Rubin, Philip E. (2003). CASY: The Haskins Configurable Articulatory Synthesizer. *Proceedings of the 15th International Congress of Phonetic Sciences* 1, 185–8.

Isshiki, Nobuhiko (1964). Regulatory mechanism of voice intensity variation. *Journal of Speech, Language and Hearing Research* 7: 17–29.

Itô, Junko and Mester, Armin (1995). Japanese phonology, in J. A. Goldsmith (ed.), *The Handbook of Phonological Theory*. Oxford: Blackwell, 817–38.

—— —— (1999). The phonological lexicon, in N. Tsujimura (ed.), *The Handbook of Japanese Linguistics*. Oxford: Blackwell, 62–100.

Ito, Kiwako and Speer, Shari R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58: 541–73.

Iverson, Paul, Ekanayake, Dulika, Hamann, Silke, Sennema, Anke, and Evans, Bronwen G. (2008). Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers. *Journal of Experimental Psychology: Human Perception and Performance* 34: 1305–16.

—— and Evans, Bronwen G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America* 122: 2842–54.

—— Hazan, Valerie, and Bannister, Kerry (2005). Phonetic training with acoustic cue manipulation: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America* 118(5): 3267–78.

—— and Kuhl, Patricia K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America* 97: 553–62.

——— ——— (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America* 99: 1130–40.

——— ——— Akahane-Yamada, Reiko, Diesch, E., Tohkura, Yoh'ichi, Kettermann, Andreas, and Siebert, Claudia (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: B47–B57.

——— Pinet, Melanie, and Evans, Bronwen G. (2011). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, doi:10.1017/S0142716411000300.

——— Smith, Charlotte A., and Evans, Bronwen G. (2006). Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration. *Journal of the Acoustical Society of America* 120: 3998–4006.

Jablonski, David (2005). Mass extinctions and macroevolution. *Paleobiology* 31: 192–210.

Jackendoff, Ray. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

——— (1997). *The architecture of the language facility.* Cambridge, MA: MIT Press.

Jackson, Philip J. B. and Shadle, Christine H. (2000). Frication noise modulated by voicing, as revealed by pitch-scaled decomposition. *Journal of the Acoustical Society of America* 108(4): 1421–34.

——— ——— (2001). Decomposing speech signals into their simultaneous voiced and unvoiced components. *IEEE Transactions on Speech and Audio Processing* 9(7): 713–26.

Jacobs, Robert A. (2002). What determines visual cue reliability? *Trends in Cognitive Science* 6(8): 345–50.

Jacquemot, Charlotte, Pallier, Christophe, LeBihan, Denis, Dehaene, Stanislas, and Dupoux, Emmanuel (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience* 23: 9541–6.

Jaeger, Florian (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language* 59: 434–46.

Jaeger, Jeri J. (1984). Assessing the psychological status of the vowel shift rule. *Journal of Psycholinguistic Research* 13: 13–36.

Jäger, Gerhard (2007). Maximum entropy models and stochastic Optimality Theory, in A. Zaenen, J. Simpson, T. Holloway King, J. Grimshaw, J. Maling, and C. Manning (eds.), *Architectures, Rules, and Preferences: Variations on Themes by Joan W. Bresnan*. Stanford: CSLI Publications, 467–79.

——— and Rosenbach, Anette (2006). The winner takes it all – almost: Cumulativity in grammatical variation. *Linguistics* 44: 937–71.

Jakobson, Roman. C. (1941). *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton de Gruyter.

——— Fant, Gunnar M., and Halle, Morris (1963/1952). *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates*. Cambridge, MA: MIT Press.

Jamieson, Donald G. and Morosan, David E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology* 43: 88–96.

Janda, Laura A., Nesset, Tore, and Baayen, R. Harald (2010). Capturing correlational structure in Russian paradigms: A case study in logistic mixed-effects modeling. *Corpus Linguistics and Linguistic Theory* 6: 29–48.

Janda, Richard D. (2001). Beyond "pathways" and "unidirectionality": On the discontinuity of language transmission and the counterability of grammaticalization. *Language Sciences* 23: 265–340.

Janda, Richard D. (2003). "Phonologization" as the start of dephoneticization—Or, on sound change and its aftermath: Of extension, generalization, lexicalization, and morphologization, in B. D. Joseph and R. D. Janda (eds.), *The Handbook of Historical Linguistics*. Malden, MA: Blackwell, 401–22.

—— and Joseph, Brian D. (2001). Reconsidering the canons of sound-change: Towards a big bang theory, in *Historical Linguistics 2001. Selected Papers from the 15th International Conference on Historical Linguistics*, Melbourne, August 13–17, 2001, 205–19.

Jannedy, Stefanie and Martins, M. (2008). The sociophonetics of Turkish ethnic youth in Berlin. Paper presented at First Arizona Anthropology and Linguistics Conference, May 12, 2008.

Janssen, Dirk P., Roelofs, Ardi, and Levelt, Willem J. M. (2002). Inflectional frames in language production. *Language and Cognitive Processes* 17: 209–36.

Javkin, Hector R., Antonanzas-Barroso, Norma, and Maddieson, Ian (1987). Digital inverse filtering for linguistic research. *Journal of Speech and Hearing Research* 30: 122–9.

Jesus, Luis M.T. and Shadle, Christine H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics* 30: 437–64.

—— —— (2003). Devoicing measures of European Portuguese fricatives, in N. J. Mamede et al. (ed.), *PROPOR 2003, LNAI 2721*. Berlin/Heidelberg: Springer-Verlag, 1–8.

Jilka, Matthias (2007). Different manifestations and perceptions of foreign accent in intonation, in J. Trouvain and U. Gut (eds.), *Non-native Prosody: Phonetic Description and Teaching Practice*. Berlin: Mouton de Gruyter, 76–96.

Jin, Shunde (1996). An acoustic study of sentence stress in Mandarin Chinese. Ph.D. dissertation, Ohio State University, Columbus.

Johns-Lewis, Catherine (ed.) (1986). *Intonation in Discourse*. London: Croom Helm.

Johnson, Carolyn E. and Wilson, Ian L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *The International Journal of Bilingualism* 6: 271–89.

Johnson, Daniel E. (2009). Getting off the GoldVarb Standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1): 359–83.

Johnson, Keith (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America* 88: 642–54.

—— (1997a). *Acoustic and Auditory Phonetics*, 2nd edn. Cambridge, MA: Blackwell Publishers.

—— (1997b). Speech perception without speaker normalization, in K. Johnson and J. Mullennix (eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press, 9–32.

—— (1997c). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics* 50: 101–13.

—— (2004). Massive reduction in conversational American English, in K. Yoneyama and K. Maekawa (eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*. Tokyo: National International Institute for Japanese Language, 29–54.

—— (2005). Decisions and mechanisms in exemplar-based phonology, in *UC Berkeley Phonology Lab Annual Report*, 289–311.

—— (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34: 485–99.

——(2007). Decisions and mechanisms in exemplar-based phonology, in M. J. Sole, P. Beddor, and M. Ohala (eds.), *Experimental Approaches to Phonology. In honor of John Ohala*. Oxford : Oxford University Press, 25–40.
——(2008). *Quantitative Methods in Linguistics*. Malden, MA: Blackwell.
—— Flemming, Edward, and Wright, Richard (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language* 69: 505–28.
—— Ladefoged, Peter, and Lindau, Mona (1993). Individual differences in vowel production. *Journal of the Acoustical Society of America* 94: 701–14.
—— and Martin, Jack (2001). Acoustic vowel reduction in Creek: Effects of distinctive length and position in the word. *Phonetica* 58: 81–102.
—— Strand, Elizabeth, and D'Imperio, Mariapaola (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics* 27: 359–84.
Johnson, Susan (2006). If you're "heppy" and you know it, front your /æ/. Oral presentation given at NWAV 36, Columbus, OH.
Johnstone, Barbara and Bean, Judith M. (1997). Self-expression and linguistic variation. *Language in Society* 26: 221–46.
Jones, Amanda (2002). A lexicon-independent phonological well-formedness effect: Listeners' sensitivity to inappropriate aspiration in initial /st/ clusters. *UCLA Working Papers in Phonetics* 100: 33–72.
Jones, Daniel (1950). *The Phoneme: Its Nature and Use.* Cambridge: Heffer.
de Jong, Kenneth J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97(1): 491–504.
——(2003). Temporal constraints and characterising syllable structuring, in J. Local, R. Ogden, and R. Temple (eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 253–68.
——(2007). Temporal structure and the nature of syllable-level timing patterns, in J. Cole and J. I. Hualde (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 657–68.
Jongman, Allard, Wayland, Ratree, and Wong, Serena (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustic Society of America* 108(3), 1252–63.
Ju, Min and Luce, Paul A. (2006). Representational specificity of within-category phonetic variation in the long-term mental lexicon. *Journal of Experimental Psychology: Human Perception and Performance* 32(1): 120–38.
Jun, Sun-Ah (1994a). The status of lenis stop voicing rule in Korean, in Y.-K. Kim-Renaud (ed.), *Theoretical Issues in Korean Linguistics*. Stanford: CSLI, 101–14.
——(1994b). The domains of laryngeal feature lenition effects in Chonnam Korean, *Ohio State University Working Papers in Linguistics*, 43: 15–20.
——(1995). Asymmetrical prosodic effects on the laryngeal gesture in Korean, in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 235–53.
Jun, Sun-Ah (1996). *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure*. New York: Garland Publishing.
——(2003). The effect of phrase length and speech rate on prosodic phrasing, in M. J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: UAB, 483–6.
——(2005a). *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Jun, Sun-Ah (2005b). Korean intonational phonology and prosodic transcription, in S.-A. Jun (ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 201–29.

—— (2007). The intermediate phrase in Korean: Evidence from sentence processing, in T. Riad and C. Gussenhoven (eds.), *Tones and Tunes*, vol. 2. Berlin: Mouton de Gruyter, 143–67.

Jurafsky, Dan (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production, in R. Bod, J. Hay, and S. Jannedy (eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press, 39–95.

—— Bell, Alan, and Girand, Cynthia (2002). The role of the lemma in form variation, in C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology 7*. Berlin and New York: Mouton de Gruyter, 1–34.

—— —— Gregory, Michelle, and Raymond, William D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production, in J. L. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 229–54.

Jusczyk, Peter W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics* 21: 3–28.

—— (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

—— and Aslin, Richard N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* 29(1): 1–23.

—— Cutler, Anne, and Redanz, Nancy J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development* 64: 675–87.

—— Friederici, Angela D., Wessels, Jeanine, Svenkerud, Vigdis, and Jusczyk, Ann M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* 32: 402–20.

—— Goodman, Mara B., and Baumann, Angela (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language* 40: 62–82.

—— Hohne, Elizabeth A., and Bauman, Angela (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics* 61: 1465–76.

—— Houston, Derek, and Newsome, Mary (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39: 159–207.

—— Luce, Paul, and Charles-Luce, Jan (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33: 630–45.

—— Smolensky, Paul, and Allocco, Theresa (2002). How English-learning infants respond to markedness and faithfulness constraints. *Language Acquisition* 10: 31–73.

Kabak, Baris and Idsardi, William J. (2007). Perceptual distorsions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech* 50: 23–52.

—— and Revithiadou, Anthi (2006). The phonology of clitic groups: Prosodic recursivity revisited. Paper given at the 13th International Conference on Turkish Linguistics, Uppsala.

Kaburagi, Tokihiko and Honda, Masaaki (1996). A model of articulator trajectory formation based on the motor tasks of vocal-tract shapes. *Journal of the Acoustical Society of America* 99: 3154–70.

Kagan, Jerome and Lewis, Michael (1965). Studies of attention in the human infant. *Merrill Palmer Quarterly* 11: 95–127.

Kager, René (1996). On affix allomorphy and syllable counting, in U. Kleinhenz (ed.), *Interfaces in Phonology*. Berlin: Akademie Verlag, 155–71.

—— (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

—— (2008). Lexical irregularity and the typology of contrast, in K. Hanson and S. Inkelas (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*. Cambridge, MA: MIT Press, 397–432.

Kahn, Daniel (1976). Syllable-based generalizations in English phonology. Ph.D. dissertation, MIT, Cambridge, MA. [Published, New York: Garland Press, 1980.]

Kainada, Evia. (2009). The phonetic and phonological nature of prosodic boundaries: Evidence from Modern Greek. Doctoral dissertation, University of Edinburgh.

Kaiser, Eden, Munson, Benjamin, Li, Fangfang, Holliday, Jeffrey J., Beckman, Mary E., Edwards, Jan, and Schellinger, Sarah K. (2009). Why do adults vary in how categorically they rate the accuracy of children's speech? *Journal of the Acoustical Society of America* 125: 2753, <http://www.ling.ohio-state.edu/~edwards/ASA09_Kaiser_etal_poster.pdf>, accessed June 14, 2009.

Kalikow, D. N., Stevens, Kenneth N., and Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America* 61: 1337–51.

Kallayanamit, Saovapak (2004). The phonetics and phonology of Thai intonation: Contours, registers, and boundary tones. Ph.D. dissertation, Georgetown University.

Kamide, Y., Altmann, G. T. M., and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49: 133–56.

Kanerva, Jonni. (1989). Focus and phrasing in Chichewa phonology. Ph.D. dissertation, University of Stanford, Palo Alto.

Kang, Kyoung-H. and Guion, Susan G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant system of Korean-English bilinguals. *Journal of the Acoustical Society of America* 119: 1672–83.

Kang, Yoonjung (2000). The phonetics and phonology of coronal markedness and unmarkedness. Ph.D. dissertation, MIT, Cambridge, MA.

Kapatsinski, Vsevolod (2009). Testing theories of linguistic constituency with configural learning: The case of the English syllable. *Language* 85(2): 248–77.

Kaplan, Aaron F. (2006). Vowel length and coda cluster interactions in Misantla Totonac, in A. Eilam, T. Scheffler, and J. Tauberer (eds.), *Proceedings of of the 29th Annual Penn Linguistics Colloquium*. Penn Working Papers in Linguistics 12(1): 161–74.

Karlsson, Fred (1982). *Suomen kielen äänne- ja muotorakenne* [*The Phonological and Morphological Structure of Finnish*]. Helsinki: Werner Söderström Osakeyhtiö.

Karttunen, Lauri (2006). The insufficiency of paper-and-pencil linguistics: The case of Finnish prosody, ROA-818, < http://roa.rutgers.edu/files/818-0406/818-KARTTUNEN-0-0.PDF>, accessed May 22, 2011.

Karvonen, Dan (2005). Word prosody in Finnish. Ph.D. dissertation, University of California, Santa Cruz.

Kawahara, Hideki, Masuda-Katsuse, Ikuyo, and de Cheveigne, Alain (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication* 27: 187–207.

Kawahara, Shigeto (2011). Experimental approaches in generative phonology, in M. van Oostendorp, C. Ewen, E. Hume, and K. Rice (eds.), *The Blackwell Companion to Phonology*. Malden, MA: Blackwell, 2283–303.

Kawasaki, Haruko (1982). An acoustical basis for universal constraints on sound sequences. Ph.D. dissertation, University of California, Berkeley.

Kazanina, Nina, Phillips, Colin, and Idsardi, William (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences* 103: 11381–6.

Keating, Patricia A. (1984). Phonetic and phonological representation of consonant voicing. *Language* 60: 286–319.

—— (1985). Universal phonetics and the organization of grammars, in V. Fromkin (ed.), *Phonetic Linguistics*. New York: Academic Press, 115–32.

—— (1987). A survey of phonological features. *UCLA Working Papers in Phonetics* 66: 124–42.

—— (1988). Underspecification in phonetics. *Phonology* 5: 275–92.

—— (1990a). Phonetic representations in a generative grammar. *Journal of Phonetics* 18: 321–34.

—— (1990b). The window model of coarticulation: Articulatory evidence, in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 451–70.

—— (1996). The phonetics-phonology interface. *Studia Grammatica* 41: 262–78.

—— (2006). Phonetic encoding of prosodic structure, in J. Harrington and M. Tabain (eds.), *Speech Production: Models, Phonetic Processes, and Techniques*, Macquarie Monographs in Cognitive Science. New York and Hove: Psychology Press, 167–86.

—— Cho, Taehong, Fougeron, Cécile, and Hsu, Chai-Shune (2003). Domain-initial articulatory strengthening in four languages, in J. Local, R. Ogden, and R.Temple (eds.). *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 145–63.

—— and Esposito, Christina (2007). Linguistic voice quality. *University of California Working Papers in Phonetics* 105: 85–91.

Kedrova, Galina, Anisimov, Nikolay, Zaharaov, Leonid, and Pirogov, Yurij (2008). Magnetic resonance investigation of palatalized stop consonants and spirants in Russian. *Journal of the Acoustical Society of America* 123(5): 3325.

Keller, Frank (2000). Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Ph.D. dissertation, University of Edinburgh. [ROA-677]

—— (2005). Linear Optimality Theory as a model of gradience in grammar, in G. Fanselow, R. V. Féry, and M. Schlesewsky, (eds.), *Gradience in Grammar: Generative Perspectives*. Oxford: Oxford University Press, 270–87.

—— and Alexopoulou, T. (2001). Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition* 79(3): 301–72.

—— and Asudeh, Ash (2002). Probabilistic learning algorithms and Optimality Theory. *Linguistic Inquiry* 33(2): 225–44.

Kelly, Jack B., Judge, Peter W., and Phillips, Dennis P. (1986). Representation of the cochlea in primary auditory cortex of the ferret (Mustela putorius). *Hearing Research* 24: 111–15.

Kelso, J. A. Scott, Saltzman, Elliot L., and Tuller, Betty (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics* 14: 29–59.

Kemler Nelson, Deborah G., Hirsh-Pasek, Kathy, Jusczyk, Peter W., and Cassidy, Kimberly W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language* 16: 55–68.

—— Jusczyk, Peter W., Mandel, Denise R., Myers, James, Turk, Alice, and Gerken, LouAnn (1995). The headturn preference procedure for testing auditory perception. *Infant Behavior and Development* 18: 111–16.

Kemps, Rachèl, Ernestus, Mirjam, Schreuder, Rob, and Baayen, R. Harald (2005). Prosodic cues for morphological complexity. *Memory and Cognition* 33: 430–46.

Kenstowicz, Michael (1996). Quality-sensitive stress. *Rivista di linguistica* 9(1), 157–87. [ROA-33].

—— (1997). Uniform exponence: Extension and exemplification, in V. Miglio and B. Morén (eds.), *University of Maryland Working Papers in Linguistics 5: Selected Papers from the Hopkins Optimality Workshop 1997*, 139–54.

—— and Kisseberth, Charles (1977). *Topics in Phonological Theory*. New York: Academic Press.

—— —— (1979). *Generative Phonology: Description and Theory*. New York: Academic Press.

Kent, Raymond D. and Forner, Linda L. (1980). Speech segment duration in sentence recitations by children and adults. *Journal of Phonetics* 8: 157–68.

Kernan, Keith T. and Blount, B. G. (1966). The acquisition of Spanish grammar by Mexican children. *Anthropological Linguistics* 8(9): 1–14.

Kerswill, Paul (1985). A sociophonetic study of connected speech processes in Cambridge English: An outline and some results. *Cambridge Papers in Phonetics and Experimental Linguistics* 4: 25–49.

—— (1994). *Dialects Converging: Rural Speech in Urban Norway*. Oxford: Oxford University Press.

—— (2002). Koineization and accommodation, in J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell, 669–702.

—— Torgersen, Eivind, and Fox, Sue (2008). Reversing "drift": Innovation and diffusion in the London diphthong system. *Language Variation and Change* 20: 451–91.

Kessinger, Rachel H. and Blumstein, Sheila E. (1998). Effects of speaking rate on voice onset time and vowel production: Some implications for perception studies. *Journal of Phonetics* 26: 117–28.

Kessler, Brett and Treiman, Rebecca (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37: 295–311.

Keune, Karen, Ernestus, Mirjam, Van Hout, Roeland, and Baayen, R. Harald (2005). Social, geographical, and register variation in Dutch: From written 'mogelijk' to spoken 'mok'. *Corpus Linguistics and Linguistic Theory* 1: 183–223.

Keyser, Samuel J. and Kiparsky, Paul (1984). Syllable structure in Finnish phonology, in M. Aronoff and R. T. Oehrle (eds.), *Language Sound Structure. Studies in Phonology Presented to Morris Halle by His Teacher and Students*, Cambridge, MA: MIT Press, 7–31.

—— and Stevens, Kenneth N. (1994). Feature geometry and the vocal tract. *Phonology* 11: 207–36.

—— —— (2006). Enhancement and overlap in the speech chain. *Language* 82: 33–63.

Khan, Arfaan (2006). A sociolinguistic study of Birmingham English: Language variation and change in a multi-ethnic British community. Ph.D. dissertation, Lancaster University.

Khan, Sameer ud Dowla (2008). Intonational phonology and focus prosody in Bengali. Ph.D. dissertation, UCLA.

Khattab, Ghada (2000). VOT production in English and Arabic bilingual and monolingual children. *Leeds Working Papers in Linguistics* 8: 95–122.

—— (2007). Variation in vowel production by English-Arabic bilinguals, in J. I. Hualde and J. Cole (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 383–410.

Khouw, Edward and Ciocca, Valter (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics* 35: 104–17.

Kiesling, Scott F. (1998). Variation and men's identity in a fraternity. *Journal of Sociolinguistics* 2(1): 69–100.

Kim, Heejin (2006). Speech rhythm in American English: A corpus study. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

—— and Cole, Jennifer (2005). The stress foot as a unit of planned timing: Evidence from shortening in the prosodic phrase. *Proceedings of Interspeech 2005*, Lisbon, Portugal, 2365–8.

—— Hasegawa-Johnson, Mark, Perlman, Adriene, Gunderson, Jon, Huang, Thomas, Watkin, Kenneth, and Frame, Simone (2008). Dysarthric speech database for universal access research. *Proceedings of the International Conference on Spoken Language Processing (Interspeech '08)*. Brisbane, Australia, September 2008, 1741–4.

Kim, Hyunsoon (2004). Stroboscopic-Cine MRI data on Korean coronal plosives and affricates: Implications for their place of articulation as alveolar. *Phonetica* 61: 234–51.

—— Honda, Kiyoshi, and Maeda, Shinji (2005). Stroboscopic-cine MRI study of the phasing between the tongue and the larynx in the Korean three-way phonation contrast. *Journal of Phonetics* 33: 1–26.

Kim, Jae-On and Mueller, Charles W. (1978a). *Factor Analysis: Statistical Methods and Practical Issues*. Beverly Hills, CA: Sage.

—— —— (1978b). *Introduction to Factor Analysis: What It Is and How to Do It*. Beverly Hills, CA: Sage.

Kim, Mi-Ryoung, Beddor, Patrice S., and Horrocks, Julie (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics* 30: 77–100.

Kim, Sahyang (2004). The role of prosodic phrasing in Korean word segmentation. Doctoral dissertation, Department of Linguistics, UCLA.

Kingston, John (1990). Articulatory binding, in J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 406–34.

—— (1992). The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech* 35: 99–113.

—— (2005). The phonetics of Athabaskan tonogenesis, in S. Hargus and K. Rice (eds.), *Athabaskan Prosody*. Amsterdam: John Benjamins, 137–84.

—— (2007). Segmental influences on F0: Controlled or automatic? in C. Gussenhoven and T. Riad (eds.), *Tones and Tunes*, vol. 2. Berlin: Mouton de Gruyter, 171–210.

—— and Beckman, Mary E. (eds.) (1990). *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press.

—— and Diehl, Randy (1994). Phonetic knowledge. *Language* 70: 419–54.

Kiparsky, Paul (1975). What are phonological theories about?, in D. Cohen and J. Wirth (eds.), *Testing Linguistic Hypotheses*. New York: Wiley, 47–78.

—— (1981). Vowel harmony. MS, Stanford University, Stanford, CA.

——(1982). Lexical morphology and phonology, in *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*. Linguistic Society of Korea. Seoul, Korea: Hanshin Publishing.
——(1985). Some consequences of Lexical Phonology. *Phonology Yearbook* 2: 85–138.
——(1993). An OT perspective on phonological variation. Handout from Rutgers Optimality Workshop 1993, also presented at NWAVE 1994, Stanford University. Available at <http://www.stanford.edu/~kiparsky/Papers/nwave94.pdf>.
——(1995). The phonological basis of sound change, in Goldsmith, J. (ed.), *The Handbook of Phonological Theory*. Cambridge, MA: Blackwell, 640–70.
——(2000). Opacity and cyclicity, *The Linguistic Review* 17: 351–67.
——(2003). Finnish noun inflection, in S. Manninen and D. Nelson (eds.), *Generative Approaches to Finnic and Saami Linguistics*. Stanford, CA: CSLI Publications, 109–61.
Kirchhoff, Katrin and Schimmel, Steven (2005). Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition. *Journal of the Acoustical Society of America* 117(4): 2238–46.
Kirchner, Robert (1999). Preliminary thoughts on phonologization within an exemplar-based speech-processing system, in M. Gordon (ed.), *UCLA Working Papers in Linguistics* (Papers in Phonology 2), 1, 205–31.
——(2004). Consonant lenition, in B. Hayes, R. Kirchner, and D. Steriade, *Phonetically Based Phonology*. Cambridge: Cambridge University Press, ch. 10.
——Moore, Roger K., and Chen, Tsung-Ying (2010). Computing phonological generalization over real speech exemplars. *Journal of Phonetics* 38(4): 540–7.
——and Varelas, Eleni (2002). A cue-based approach to the phonotactics of Upper Necaxa Totonac. MS, University of Alberta.
Kirk, Cecilia and Demuth, Katherine (2005). Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language* 32(4): 709–34.
————(2006). Accounting for variability in 2-year-olds' production of coda consonants. *Language Learning and Development* 2: 97–118.
Kisilevsky, Barbara S., Hains, Sylvia M. J., Lee, Kang, Xie, Xing, Huang, Hefeng, Ye, Hai-Hui, Zhang, Ke, and Wang, Zengping (2003). Effects of experience on fetal voice recognition. *Psychological Science* 14: 220–4.
Kisseberth, Charles and Abasheikh, Mohammad I. (1974). Vowel length in Chi Mwi:ni: A case study of the role of grammar in phonology, in A. Bruck, R. A. Fox, and M. W. LaGaly (eds.), *Papers from the Parasession on Natural Phonology*. Chicago: Chicago Linguistic Society, 193–200.
Klatt, Dennis H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics* 3: 129–40.
——(1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59: 1208–21.
——(1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics* 7: 279–312.
Klatt, Dennis H. (1980). Software for a Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America* 67: 971–95.
——(1989). Review of selected models of speech perception, in W. D. Marslen-Wilson (ed.), *Lexical Representation and Process*. Cambridge, MA: MIT Press, 169–226.
——and Klatt, Laura C. (1990). Analysis, synthesis and perception of voice quality variations among male and female talkers. *Journal of the Acoustical Society of America* 87: 820–56.

Kluender, Keith R., Coady, Jeffry A., and Kiefte, Michael (2003). Sensitivity to change in perception of speech. *Speech Communication* 41(1): 59–69.

—— Diehl, Randy L., and Killeen, Peter R. (1987). Japanese quail can learn phonetic categories. *Science* 237: 1195–7.

—— and Lotto, Andrew J. (1994). Effects of first formant onset frequency on [-voice] judgments result from general auditory processes not specific to humans. *Journal of the Acoustical Society of America* 95: 1044–52.

—— —— and Holt, Lori L. (2005). Contributions of nonhuman animal models to understanding human speech perception, in S. Greenberg and W. Ainsworth (eds.), *Listening to Speech: An Auditory Perspective*. New York: Oxford University Press.

—— —— and Bloedel, Suzi B. (1998). Role of experience for language-specific functional mappings for vowel sounds. *Journal of the Acoustical Society of America* 104: 3568–82.

Kochanski, Greg, Grabe, Esther, Coleman, Jonathan, and Rosner, B. (2005). Loudness predicts prominence; fundamental frequency lends little. *Journal of the Acoustical Society of America* 118(2): 1038–54.

—— Shih, C., and Jing, H. (2003). Quantitative measurement of prosodic strength in Mandarin. *Speech Communication* 41: 625–45.

Kochetov, Alexei (1999). A cue-based analysis of the distribution of palatalized stops in Russian, in O. Fujimura, B. D. Joseph, and B. Palek (eds.), *Proceedings of LP '98: Item Order in Language and Speech*, vol. 1. Prague: Karolinum Press, 247–70.

—— (2004). Perception of place and secondary articulation contrasts in different syllable positions: Language-particular and language-independent asymmetries. *Language and Speech* 47: 351–82.

—— (2006a). Syllable position effects and gestural organization: Evidence from Russian, in L. Goldstein, D. Whalen, and C. Best (eds.), *Laboratory Phonology 8*. Berlin: Mouton de Gruyter, 565–88.

—— (2006b). Testing licensing by cue: A case of Russian palatalized coronals. *Phonetica* 63: 113–48.

—— (2006c). The role of social factors in the dynamics of sound change: A case study of a Russian dialect. *Language Variation & Change* 18: 99–119.

—— (2008). Perception of gestural overlap and self-organizing phonological contrasts, in P. Avery, E. Dresher, and K. Rice (eds.), *Contrast in Phonology: Perception and Acquisition*. Berlin: Mouton de Gruyter, 173–96.

Koenig, Laura L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research* 43: 1211–28.

—— Lucero, Jorge C., and Perlman, Elizabeth (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *Journal of the Acoustical Society of America* 124(5): 3158–70.

Koester, Dirk and Schiller, Niels. O. (2008). Morphological priming in overt language production: Electrophysiological evidence from Dutch. *NeuroImage* 42: 1622–30.

Kohler, Klaus J. (1983). Prosodic boundary signals in German. *Phonetica* 40: 89–134.

—— (1985). F0 in the perception of lenis and fortis plosives. *Journal of the Acoustical Society of America* 78: 21–32.

—— (1987). Categorical pitch perception, in U. Viks (ed.), *Proceedings of the 11th International Congress of Phonetic Sciences*, vol. 5, Tallinn, Estonia, August 1–7, 1987, 331–3.

—— (1990a). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations, in W. J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 21–33.

——(1990b). Macro and Micro F0 in the synthesis of intonation, in J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 115–38.

——(2006). Paradigms in experimental prosodic analysis: From measurements to function, in S. Sudhoff et al. (eds.), *Methods in Empirical Prosody Research*. Berlin and New York: Mouton de Gruyter, 123–52.

Kohonen, Teuvo (1984). *Self-organization and Associative Memory*. Berlin: Springer.

Koike, Yasuo and Hirano, Minoru (1973). Glottal-area time function and subglottal-pressure variation. *Journal of the Acoustical Society of America* 54: 1618–27.

Komarova, Natalia L. and Nowak, Martin (2001). The evolutionary dynamics of the lexical matrix. *Bulletin of Mathematical Biology* 63: 451–84.

Kondaurova, Maria and Francis, Alexander L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America* 124(6): 3959–71.

——(forthcoming). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of Phonetics*.

Kong, Eun Jong (2009). The development of phonation-type contrasts in plosives: Cross-linguistic perspectives. Ph.D. dissertation, Department of Linguistics, Ohio State University, Columbus.

Konno, Kimio and Mead, Jere (1967). Measurement of the separate volume changes of rib cage and abdomen during breathing. *Journal of Applied Physiology* 22: 407–22.

Koopmans-van Beinum, Florien J. (1980). Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions. Amsterdam: Academische Pers B.V.

Koreman, Jacques (2006). Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America* 119: 582–96.

Kornai, András (1991). Formal phonology. Ph.D. dissertation, Stanford University.

Kornfeld, Judith (1971). What initial clusters tell us about a child's speech code. *MIT RLE Quarterly Progress Report* 101: 218–21.

Kotsinas, Ulla-Britt (1998). Language contact in Rinkeby: An immigrant suburb, in J. K. Androutsopoulos and A. Scholz (eds.), *Jugendsprache—langue des jeunes—youth language*. Frankfurt am Main: Peter Lang, 125–48.

Kovelman, Ioulia, Shalinsky, Mark H., Berens, Melody S., and Petitto, Laura-Ann (2008). Shining new light on the brain's "bilingual signature": A functional near infrared spectroscopy investigation of semantic processing. *NeuroImage* 39: 1457–71.

Krakow, Rena A. (1989). The articulatory organization of syllables: A kinematic analysis of labial and velic gestures. Ph.D. dissertation, Yale University.

Krakow, Rena A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics* 27: 2–54.

——Beddor, Patrice S., Goldstein, Louis M., and Fowler, Carol A. (1988). Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America* 83(3): 1146–58.

Kraljic, Tanya, Brennan, Susan E., and Samuel, Arthur G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition* 107(1), 51–81.

——Samuel, Arthur G., and Brennan, Susan E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science* 19(4): 332–8.

Krane, Michael H. (2005). Aeroacoustic production of low-frequency unvoiced speech sounds. *Journal of the Acoustical Society of America* 118(1): 410–27.

Krauss, Robert M. and Pardo, Jennifer S. (2006). Speech perception and social behavior: Bridging social psychology and speech science, in P. A. M. van Lange (ed.), *Bridging Social Psychology: Benefits of Transdisciplinary Approaches*. Mahwah, NJ: Lawrence Erlbaum, 273–8.

Kreiman, Jody, Gerratt, Bruce R., and Antoñanzas-Barroso, Norma (2007). Measures of the glottal source spectrum. *Journal of Speech, Language, and Hearing Research* 50: 595–610.

—— and Papcun, George (1991). Comparing discrimination and recognition of unfamiliar voices. *Speech Communication* 10: 265–75.

Kripke, Saul (1972). Naming and necessity, in D. Davidson and G. Harman (eds.), *Semantics and Natural Language.* Dordrecht: Reidel, 253–355.

Kröger, Bernd J. (1993). A gestural production model and its implications for reduction in German. *Phonetica* 50: 213–33.

Kroll, Judith F., Gerfen, Chip, and Dussias, Paola E. (2008). Laboratory designs and paradigms: Words, sounds, and sentences, in L. Wei and M. Moyer (eds.), *The Blackwell Guide to Research Methods in Bilingualism*. Cambridge, MA: Blackwell Publishers, 108–31.

—— and Sunderman, Gretchen (2003). Cognitive processes in second language acquisition: The development of lexical and conceptual representations, in C. Doughty and M. Long (eds.), *Handbook of Second Language Acquisition*. Cambridge, MA: Blackwell Publishers, 104–29.

Kruijff-Korbayová, Ivana and Steedman, Mark (2003). Discourse and information structure. *Journal of Logic, Language, and Information* 12(3): 249–59.

Krull, Diana (1997). Prepausal lengthening in Estonian: Evidence from conversational speech, in I. Lehiste and J. Ross (eds.), *Estonian Prosody: Papers from a Symposium*. Tallinn: Institute of Estonian Language, 136–48.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29: 115–29.

—— and Wish, M. (1978). *Multidimensional Scaling*. Newbury Park, CA: Sage.

Kubozono, Haruo (1992). Modeling syntactic effects on downstep in Japanese, in G. J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: Cambridge University Press, 368–88.

Kuhl, Patricia K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America* 66: 1668–79.

—— (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development* 6(3): 263–85.

—— (1991). Human adults and human infants show a "perceptual magnetic effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50: 93–107.

—— (1992). Speech prototypes: Studies on the nature, function, ontogeny and phylogeny of the "centers" of speech categories, in Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (eds.), *Speech Perception, Production and Linguistic Structure*. Tokyo: Ohmsha, 239–64.

—— (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory, in B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. Dordrecht: Kluwer Academic Publishers, 259–74.

——— (2000a). A new view of language acquisition. *Proceedings of the National Academy of Science* 97: 11850–7.

——— (2000b). Language, mind, and brain: Experience alters perception, in M. S. Gazzaniga (ed.), *The New Cognitive Neurosciences* (2nd edn). Cambridge, MA: MIT Press, 99–115.

——— Andruski, Jean E., Chistovich, Inna A., Chistovich, Ludmilla A., Kozhevnikova, Elena V., Ryskina, Viktoria L., Stolyarova, Elvira I., Sundberg Ulla, and Lacerda, Francisco (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277: 684–6.

——— Conboy, Barbara T., Coffey-Corina, Sharon, Padden, Denise, Rivera-Gaxiola, Maritza, and Nelson, Tobey (2008). Native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society* B 363: 979–1000.

——— ——— Padden, Denise M., Nelson, Tobey, and Pruitt, Jessica (2005). Early speech perception and later language development: Implications for the critical period. *Language Learning and Development* 1: 237–64.

——— and Iverson, Paul (1995). Linguistic experience and the perceptual magnet effect, in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore: York Press, 121–54.

——— and Miller, James D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* 63: 905–17.

——— and Miller, Joanne D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science* 190: 69–72.

——— ——— and Padden, Denise M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature for macaques. *Journal of the Acoustical Society of America* 71: 1003–10.

——— Stevens, Erika, Hayashi, Akiko, Deguchi, Toshisada, Kiritani, Shigeru, and Iverson, Paul (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* 9: F13–F21.

——— Tsao, Feng-Ming, and Liu, Huei-Mei (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences* 100: 9096–101.

——— Williams, Karen A., Lacerda, Francisco, Stevens, Kenneth N., and Lindblom, Björn (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255: 606–8.

Kuhn, Thomas (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kühnert, Barbara, Hoole, Philip, and Mooshammer, Christine (2006). Gestural overlap and C-center in selected French consonant clusters, in *Proceedings of the 7th International Seminar on Speech Production*, 327–34.

——— and Nolan, Francis (1999). The origin of coarticulation, in W. J. Hardcastle and N. Hewlett (eds.), *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press, 7–30.

Kunze, Luverne H. (1964). Evaluation of methods of estimating sub-glottal air pressure. *Journal of Speech and Hearing Research* 7: 151–64.

Kuperman, Victor, Ernestus, Mirjam, and Baayen, R. Harald (2008). Frequency distributions of uniphones, diphones and triphones in spontaneous speech. *Journal of the Acoustical Society of America* 124: 3897–908.

Kutas, Marta and Van Petten, Cyma K. (1994). Psycholinguistics electrified: Event-related brain potential investigations, in M. A. Gernsbacher (ed.), *Handbook of Psycholinguistics*. San Diego: Academic Press, 83–143.

—— —— and Kluender, Robert (2006). Psycholinguistics electrified II: 1994–2005, in M. A. Gernsbacher and M. J. Traxler (eds.), *Handbook of Psycholinguistics*, 2nd edn. New York: Elsevier, 659–724.

Kuzla, Claudia, Cho, Taehong, and Ernestus, Mirjam (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics* 35: 301–20.

Labov, William (1963). The social motivation of a sound change. *Word* 19: 273–309.

—— (1966). *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.

—— (1969). Contraction, deletion, and inherent variability of the English copula. *Language* 45: 715–62.

—— (1972a). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

—— (1972b). Some principles of linguistic methodology. *Language in Society* 1: 97–120.

—— (1973). Where do grammars stop? In R. W. Shuy (ed.), *Sociolinguistics: Current Trends and Prospects*. 23rd annual round table (Monograph Series on Languages and Linguistics, 25), 43–88.

—— (1981). Resolving the Neogrammarian hypothesis. *Language* 57: 267–308.

—— (1989a). The limitations of context: Evidence from misunderstandings in Chicago, in *Papers from the 25th Annual Regional Meeting of the Chicago Linguistic Society, Part 2: Parasession on Language in Context*. Chicago: Chicago Linguistic Society, 171–200.

—— (1989b). The exact description of the speech community: Short-a in Philadelphia, in R. Fasold and D. Schiffrin (eds.), *Language Change and Variation*. Amsterdam: Benjamins, 1–57.

—— (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2: 205–54.

—— (1994). *Principles of Linguistic Change, vol. 1: Internal Factors*. Oxford: Blackwell.

—— (1997). Resyllabification, in F. Hinskens, R. van Hout, and L. Wetzels (eds.), *Variation, Change and Phonological Theory*. Amsterdam/Philadelphia: John Benjamins, 145–79.

—— (2001). *Principles of Linguistic Change: Social Factors*. Oxford: Blackwell.

—— (2002). Driving Forces in Linguistic Change. Paper presented at the 2002 International Conference on Korean Linguistics, August 2, 2002. Seoul National University. Available at <http://www.ling.upenn.edu/~wlabov/Papers/DFLC.htm>.

—— (2004). Quantitative analysis of linguistic variation, in U. Ammon, N. Dittmar, K. J. Mattheier, and P. Trudgill (eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society*, vol. 1, 2nd edn. Berlin: Mouton de Gruyter, 6–21.

—— (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics* 34: 500–15.

—— Ash, Sharon, and Boberg, Charles (2006). *Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Mouton de Gruyter.

—— and Harris, W. (1986). De facto segregation of black and white vernaculars, in D. Sankoff (ed.), *Diversity and Diachrony*. Philadelphia: John Benjamins, 1–25.

—— Yaeger, Malcah, and Steiner, Richard (1972). *A Quantitative Study of Sound Change in Progress*. Philadelphia: US Regional Survey.

Lacerda, Francisco (1998). An exemplar-based account of emergent phonetic categories. *Journal of the Acoustical Society of America* 103: 2980.

Lachs, Lorin, McMichael, Kip, and Pisoni, David (2003). Speech perception and implicit memory: Evidence for detailed episodic encoding, in J. S. Bowers and C. J. Marsolek (eds.), *Rethinking Implicit Memory*. Oxford: Oxford University Press. 215–35.

Ladd, D. Robert (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

—— (1983). Phonological features of intonational peaks. *Language* 59: 721–59.

—— (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

—— (2006). Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Italian Journal of Linguistics* 18(1): 19–38.

—— (2008). *Intonational Phonology (2nd edition)*. Cambridge: Cambridge University Press.

—— Faulkner, Dan, Faulkner, Hanneke, and Schepman, Astrid (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106: 1543–54.

—— Mennen, Inneke, and Schepman, Astrid (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107(5): 2685–95.

—— and Morton, Rachel (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics* 25: 313–42.

—— and Schepman, Astrid (2003). Sagging transitions between high-pitch accents in English: Experimental evidence. *Journal of Phonetics* 31: 81–112.

—— and Scobbie, James M. (2003). External sandhi as gestural overlap? Counter-evidence from Sardinian, in J. Local, R. Ogden, and R. Temple (eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 164–82.

—— and Silverman, Kim (1984). Vowel intrinsic pitch in connected speech. *Phonetica* 41: 31–40.

—— Verhoeven, Jo, and Jacobs, Karen (1994). Influence of adjacent pitch accents on each other's perceived prominence: Two contradictory effects. *Journal of Phonetics* 22: 87–99.

Ladefoged, Peter (1962). Sub-glottal activity during speech, in *Proceedings of the 4th International Congress of Phonetic Sciences*, Helsinki, 1961. The Hague: Mouton & Co., 73–91.

—— and Maddieson, Ian (1996). *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

Laeufer, Christiane (1992). Patterns of voicing conditioned vowel duration in French and English. *Journal of Phonetics* 20: 411–40.

Lahiri, Aditi and Blumstein, Sheila E. (1984). A re-evaluation of the feature "coronal." *Journal of Phonetics* 12: 133–45.

—— and Fikkert, Paula (1999). Trisyllabic shortening in English: Past and present. *English Language and Linguistics* 3: 229–67.

—— Gewirth, L., and Blumstein, Sheila E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *Journal of the Acoustical Society of America* 76: 391–404.

—— and Kraehenmann, Astrid (2004). On maintaining and extending contrasts: Notker's Anlautgesetz. *Transactions of the Philological Society* 102: 1–55.

—— and Marslen-Wilson, William D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38: 245–94.

Lahiri, Aditi and Marslen-Wilson, William D. (1992). Lexical processing and phonological representation, in G. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, 229–54.

—— and Plank, Frans (2009). What linguistics universals can be true of, in S. Scalise, E. Magni, and A. Bisetto (eds.), *Universals of Language Today*. Berlin: Springer, 31–58.

—— and Reetz, Henning (2002). Underspecified recognition, in C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology 7*. Berlin: Mouton, 637–75.

—— —— (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38: 44–59.

—— Wetterlin, Allison, and Jönsson-Steiner, Elisabet (2005). Lexical specification of tone in North Germanic. *Nordic Journal of Linguistics* 28: 61–96.

Lakoff, George (1970). Global rules. *Language* 46; 627–39.

Lambrecht, Knud (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

Lane, Lisa (2000). Trajectories of linguistic variation: Emergence of a dialect. *Language Variation and Change* 12: 267–94.

Laniran, Yetunde O. (1992). Intonation in tone languages: The phonetic implementation of tones in Yorùbá. Ph.D. dissertation, Cornell University.

—— and Clements, G. N. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics* 31(2): 203–50.

Lass, Roger (1984). Vowel system universals and typology: Prologue to theory. *Phonology Yearbook* 1: 75–111.

Lattner, Sonja, Maess, Burkhard, Wang, Yunhua, Schauer, Michael, Alter, Kai, and Friederici, Angela D. (2003). Dissociation of human and computer voices in the brain: Evidence for a preattentive gestalt-like perception. *Human Brain Mapping* 20(1): 13–21.

Laudan, Larry (1983). *Science and Values.* Berkeley and Los Angeles: University of California Press.

—— (1996). *Beyond Positivism and Relativism: Theory, Method, and Evidence.* Boulder, CO: Westview Press.

Lave, Jean and Wegner, Etienne (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge and New York: Cambridge University Press.

Lavoie, Lisa and Cohn, Abigail C. (1999). Sesquisyllables of English: The structure of vowel-liquid syllables, in J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*. Berkeley: Linguistics Department, University of California, 109–12.

Lawson, Eleanor, Stuart-Smith, Jane, and Scobbie, James M. (2008). Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of derhoticization in Scottish English. *University of Pennsylvania Working Papers in Linguistics* 14(2).

Lawson, Robert (2009). Sociolinguistic constructions of identity among urban adolescents in Glasgow. Ph.D. dissertation, University of Glasgow.

Leach, Laura and Samuel, Arthur G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology* 55: 306–53.

Leather, Jonathan (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics* 11: 373–82.

Leben, William (1973). Suprasegmental phonology. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

——— (1978). The representation of tone, in V. Fromkin (ed.), *Tone: A Linguistic Survey*. New York: Academic Press, 177–220.

Lecanuet, Jean-Pierre, Granier-Deferre, Carolyn, and Busnel, Marie-Claire (1991). Prenatal familiarization, in G. Piéraut-Le Bonniec and M. Dolitsky (eds.), *From Basic Language to Discourse Bases*. Amsterdam: John Benjamin, 31–44.

——— ——— Jacquet, A. Y., Capponi, I., and Ledru, L. (1993). Prenatal discrimination of a male and female voice uttering the same sentence. *Early Development and Parenting* 2: 217–28.

Lee, Chao-Yang (2001). Lexical tone in spoken word recognition: A view from Mandarin Chinese. Ph.D. dissertation, Brown University.

Lee, Sungbok, Potamianos, Alexandros, and Naryanan, Shrikanth (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105: 1455–68.

Lee, Yongeun and Goldrick, Matthew (2008). The emergence of sub-syllabic representations. *Journal of Memory and Language* 59: 155–68.

Leech, Robert, Holt, Lori L., Devlin, Joseph T., and Dick, Frederick (2009). Expertise with nonspeech sounds recruits speech-sensitive cortical regions. *Journal of Neuroscience* 29: 5234–89.

Legendre, Géraldine, Miyata, Yoshiro, and Smolensky, Paul (1990). Harmonic Grammar—A formal multilevel connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum, 388–95.

——— Sorace, Antonella, and Smolensky, Paul (2006). The Optimality Theory–Harmonic Grammar connection, in P. Smolensky and G. Legendre (eds.), *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, vol. 2. Cambridge, MA: MIT Press, 339–402.

Lehiste, Ilse (1960). An acoustic-phonetic study of internal open juncture. *Phonetica* 5 (Suppl.), 1–54.

——— (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

——— (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51(6): 2018–24.

——— and Peterson, Gordon E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America* 33: 419–25.

Lehrer, Adrienne (2007). Blendalicious, in J. Munat (ed.), *Lexical Creativity, Texts and Contexts*. Amsterdam: John Benjamins, 115–36.

Lehtola, Heidi, Tamminen, Henna, Peltola, Maija S., and Aaltonen, Olli (2007). Vowel identification in balanced bilinguals. *Proceedings of the 16th International Congress of Phonetic Sciences*, 793–6.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.

Lenzo, Kevin (2009). The CMU pronouncing dictionary, <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>, accessed March 13, 2009.

Leskovec, Jure, Backstrom, Lars, and Kleinberg, Jon (2009). Meme-tracking and the dynamics of the news cycle, in J. F. Elder IV, F. Fogelman-Soulié, P. Flach, and M. Zaki, (eds.), *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery.

Levelt, Clara C. (1995). Segmental structure of early words: Articulatory frames or phonological constraints, in *The Proceedings of the Twenty-seventh Annual Child Language Research Forum*. Stanford: CSLI, 19–27.

Levelt, Clara C., Schiller, Niels O., and Levelt, Willem J. (2000). The acquisition of syllable types. *Language Acquisition* 8: 237–64.

Levelt, Willem. J. M. (1989). *Speaking. From Intention to Articulation*. Cambridge, MA: MIT Press.

—— Praamstra, Peter, Meyer, Antje S., Helenius, Païvi, and Salmelin, Riitta (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience* 10: 553–67.

—— Roelofs, Ardo, and Meyer, Antje S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1–75.

—— and Schiller, Niels O. (1998). Is the syllable frame stored? *Behavioral and Brain Sciences* 21: 520.

—— and Wheeldon, Linda R. (1994). Do speakers have access to a mental syllabary? *Cognition* 50: 239–69.

Levitt, Andrea and Utman, Jennifer A. (1992). From babbling towards the sound systems of English and French: A longitudinal two-case study. *Journal of Child Language* 19: 19–49.

Levitt, Harry (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America* 49: 467–77.

Levy, Erika S. and Strange, Winifred (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics* 36: 141–57.

Levy, Roger and Jaeger, T. Florian (2007). Speakers optimize information density through syntactic reduction. *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*. Vancouver, Canada, December 4–7, 2006.

Li, Charles N. and Thompson, Sandra A. (1978). Tone acquisition, in V. Fromkin (ed.), *Tone: A Linguistic Survey*. New York: Academic Press, 271–84.

Li, Fangfang (2005). The production and perception of dental vs. retroflex sibilants in the Songyuan dialect of Northeastern Mandarin Chinese. Poster presented in NWAV 35, November 2005, New York.

—— Edwards, Jan, and Beckman, Mary E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics* 37: 111–24.

—— Kong, Eun Jong, Beckman, Mary E., and Edwards, Jan (2008). Adult acoustics and developmental patterns for gender-marked phonetic variants in Mandarin fricatives and Japanese stops. Poster presented at the 11th Conference of Laboratory Phonology, Wellington, New Zealand, June 30, 2008.

—— Mays, Chanelle, Skorniakova, Oksana, and Beckman, Mary E. (2009). Gendered production of sibilants in the Songyuan dialect of Mandarin Chinese. Poster presented at the Annual Meeting of the Linguistic Society of America, San Francisco, January 8–11, 2009.

—— Munson, Benjamin, Edwards, Jan, Yoneyama, Kiyoko, and Hall, Kathleen C. (2011). Language specificity in the perception of voiceless sibilant fricatives in English and Japanese: Implications for cross-language differences in speech-sound development. *Journal of the Acoustical Society of America* 129: 999–1011.

Li, M., Kambhamettu, Chandra, and Stone, Maureen (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics* 19(6/7): 545–54. EdgeTrak available at <http://speech.maryland.edu/software.html>.

Li, Weijun and Yang, Yufang (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience* 158(4): 1416–25.

Li, Xiaoqing, Yang, Yufang, and Hagoort, Peter (2008). Pitch accent and lexical tone processing in Chinese discourse comprehension: An ERP study. *Brain Research* 1222: 192–200.

Liberman, Alvin M. (1957). Some results of research on speech perception: A critical review. *Psychological Review* 72: 275–309.

—— Cooper, Frank S., Shankweiler, David P., and Studdert-Kennedy, Michael (1967). Perception of the speech code. *Psychological Review* 74: 431–61.

—— Harris, Katherine S., Hoffman, H. S., and Griffith, Belver C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54(5): 358–68.

—— and Mattingly, Ignatius G. (1985). The motor theory of speech perception revised. *Cognition* 21: 1–36.

—— (1989). A specialization for speech perception. *Science* 245: 489–94.

—— and Whalen, Douglas H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences* 4: 187–96.

Liberman, Mark Y. (1978). The intonational system of English. Ph.D. dissertation, MIT, Cambridge, MA.

—— and Pierrehumbert, Janet B. (1984). Intonational invariance under changes in pitch range and length, in M. Aronoff and R. T. Öhrle (eds.), *Language Sound Structure*. Cambridge, MA: MIT Press, 157–233.

—— and Prince, Alan (1977). On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.

—— and Streeter, Lynn A. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America* 63: 231–3.

Lickley, Robin J., Schepman, Astrid, and Ladd, D. Robert (2005). Alignment of "phrase accent" lows in Dutch falling-rising questions: Theoretical and methodological implications. *Language and Speech* 48: 157–83.

Lickliter, Robert and Honeycutt, Hunter (2003). Developmental dynamics: Toward a biologically plausible evolutionary psychology. *Psychological Bulletin* 192(6): 819–35.

Lieberman, Erez, Michel, Jean-Baptiste, Jackson, Joe, Tang, Tina, and Nowak, Martin A. (2007). Quantifying the evolutionary dynamics of language. *Nature* 449: 713.

Lieberman, Philip (1967). *Intonation, Perception, and Language*. Cambridge, MA: MIT Press.

Lieven, Elena V. M. (1994). Cross-linguistic and cross-cultural aspects of language addressed to children, in C. Gallaway and B. J. Richards (eds.), *Input and Interaction in Language Acquisition*. Cambridge: Cambridge University Press, 56–73.

Liljencrants, Johan and Lindblom, Björn (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48: 839–62.

Limpert, Eckert, Stahel, Werner A., and Abbt, Marcus (2001). Log-normal distributions across the sciences: Keys and clues. *Bioscience* 51: 341–52.

Lin, Hwei-B. and Repp, Bruno (1989). Cues to the perception of Taiwanese tones. *Language and Speech* 32: 25–44.

Lin, Ying and Mielke, Jeff (2008). Discovering place and manner features: What can be learned from acoustic and articulatory data?, in J. Tauberer, A. Eilam, and L. MacKenzie (eds.), *Penn Working Papers in Linguistics* 14.1: 241–54.

Lindblom, Björn (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35: 1773–81.

Lindblom, Björn (1968). Temporal organization of syllable production. Speech Transmission Laboratory, Quarterly Progress Status Report No. 2–3, 1–5.

—— (1986). Phonetic universal in vowel systems, in J. J. Ohala and J. J. Jaeger (eds.), *Experimental Phonology*. Orlando, FL: Academic Press, 13–43.

—— (1990). Explaining phonetic variation: A sketch of the H and H theory, in A. Marchal and W. Hardcastle (eds.), *Speech Production and Speech Modelling*, NATO ASI Series. Dordrecht: Kluwer Academic Publishers, 403–40.

—— (1992). Phonological units as adaptive emergents of lexical development, in C. A. Ferguson, L. Menn, and C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press, 131–63.

—— (2003). Patterns of phonetic contrast: Towards a unified explanatory framework, in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 39–42.

—— and Engstrand, Olle (1989). In what sense is speech quantal? *Journal of Phonetics* 17: 107–22.

—— Guion, Susan, Hura, Susan, Moon, Seung-Jae, and Willerman, Raquel (1995). Is sound change adaptive? *Rivista di Linguistica* 7: 5–36.

—— Lubker, James, and Gay, Thomas (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics* 7: 147–61.

—— MacNeilage, Peter, and Studdert-Kennedy, Michael (1984). Self-organizing processes and the explanation of phonological universals, in B. Butterworth, B. Comrie, and O. Dahl (eds.), *Explanations for Language Universals*. Berlin: Mouton, 181–203.

—— and Maddieson, Ian (1988). Phonetic universals in consonant systems, in L. M. Hyman and C. N. Li (eds.), *Language, Speech, and Mind: Studies in Honor of Victoria A. Fromkin*. London: Routledge, 62–80.

—— and Studdert-Kennedy, Michael (1967). On the rôle of formant transitions in vowel recognition. *Journal of the Acoustical Socierty of America* 42: 830–43.

—— and Sundberg, Johan (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America* 50: 1166–79.

Lindfield, Kimberly C., Wingfield, Arthur, and Goodglass, Harold (1999). The role of prosody in the mental lexicon. *Brain and Language* 68(1–2): 312–17.

Ling, Feng and Li, Baoyu (2008). A pilot study on the perception space of lexical tones. *Proceedings of the 8th Phonetics Conference of China and the International Symposium on Phonetic Frontiers.* Beijing, China, April 18–20.

Linguistic Data Consortium (2004). Meeting Room Careful Transcription Guidelines, technical report version 1.2, January 16, 2004.

—— (2009), Rapid Transcription Guidelines, <http://www.ldc.upenn.edu/Transcription/quick-trans/index.html>, accessed April 20, 2009.

Lisker, Leigh (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29: 3–11.

—— and Abramson, Arthur S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384–422.

—— (1970). The voicing dimension: Some experiments in comparative phonetics, in *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague, 1967. Prague: Academia.

Liu, Chang and Kewley-Port, Diane (2004). STRAIGHT: a new speech synthesizer for vowel formant discrimination. Acoustical Research Letters Online (ARLO), 5, 31–6.

Liu, Fang and Xu, Yi (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62: 70–87.

Liu, Huei-Mei, Kuhl, Patricia K., and Tsao, Feng-Ming (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science* 6: F1–F10.

Liu, Ran and Holt, Lori L. (2011). Neural changes associated with nonspeech category learning parallel those of speech category acquisition, *Journal of Cognitive Neuroscience* 23: 683–98.

Liu, Siyun and Samuel, Arthur (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech* 47: 109–38.

Liu, Yang, Shriberg, Elizabeth, Stolcke, Andread, Hillard, Dustin, Ostendorf, Mari, and Harper, Mary (2006). Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Transactions on Audio, Speech, and Language Processing* 14(5): 1526–40.

Lively, Scott E., Logan, John S., and Pisoni, David B. (1993). Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94(3): 1242–55.

Livescu, Karen, Bezman, Ari, Borges, Nash, Yung, Lisa, Cetin, Ozgun, Frankel, Joe, King, Simon, Magimai-Doss, Mathew, Chi, Xuemin, and Lavoie, Lisa (2007). Manual transcription of conversational speech at the articulatory feature level. *Acoustics, Speech and Signal Processing* 4: 953–6.

Llamas, Carmen, Mullany, Louise, and Stockwell, Peter (eds.) (2006). *The Routledge Companion to Sociolinguistics*. London: Routledge.

Local, John K. (2003). Variable domains and variable relevance: Interpreting phonetic exponents. *Journal of Phonetics* 31: 321–39.

—— (2007). Phonetic detail and the organisation of talk-in-interaction, in W. J. Barry and J. Trouvain (eds.), *16th International Congress of Phonetic Sciences*. Saarbrücken, 1–10, Paper ID 1785, <http://www.icphs2007.de/>.

Locke, John L. and Pearson, D. M. (1992). Vocal learning and the emergence of phonological capacity, in C. A. Ferguson, L. Menn, and C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications.* Timonium, MD: York Press, 91–129.

Loevenbruck, Hélène, Collins, Michael J., Beckman, Mary E., Krishnamurthy, Ashok K., and Ahalt, Stanley C. (1999). Temporal coordination of articulatory gestures in consonant clusters and sequences of consonants, in O. Fujimura, B. D. Joseph, and B. Palek (eds.), *Proceedings of the 1998 Linguistics and Phonetics Conference.* Prague: Karolinum Press, 547–73.

Löfqvist, Anders (2005). Lip kinematics in long and short stop and fricative consonants. *Journal of the Acoustical Society of America* 117: 858–78.

—— Baer, Thomas, McGarr, Nancy S., and Story, Robin S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America* 85: 131421.

—— and Gracco, Vincent. (1999). Interarticulator programming in VCV sequences: Lip and tongue movements. *Journal of the Acoustical Society of America* 105: 1864–76.

—— —— (2002). Control of oral closure in lingual stop consonant production. *Journal of the Acoustical Society of America* 111(6): 2811–27.

Logan, John S., Lively, Susan E., and Pisoni, David (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89: 874–86.

LOGIOS LEXICON TOOL (2009). <http://www.speech.cs.cmu.edu/tools/lextool.html>, accessed March 13, 2009.

LOMBARDI, LINDA (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17: 267–302.

LOTTO, ANDREW J. (2000). Language acquisition as complex category formation. *Phonetica* 57: 189–96.

—— HICKOK, GREGORY S., and HOLT, LORI L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Science* 13: 110–14.

—— KLUENDER, KEITH R., and HOLT, LORI L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). *Journal of the Acoustical Society of America* 102(2): 1134–40.

—— —— —— (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America* 103: 3648–55.

—— —— —— (2000). Effects of language experience on organization of vowel sounds, in M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon.* Cambridge: Cambridge University Press, 218–26.

—— SATO, MOMOKO, and DIEHL, RANDY L. (2004). Mapping the task for the second-language learner: The case of Japanese acquisition of /r/ and /l/, in J. Slifka, S. Manuel, and M. Matthies (eds.), *From Sound to Sense: 50+ Years of Discoveries in Speech Communication*.

LU, QIMING, KORNISS, G., and SZYMANSKI, BOLESLAW K. (2009). The naming game in social networks: Community formation and consensus engineering. *Journal of Economic Interaction and Coordination* 4: 221–35.

LUCE, PAUL A. (1986). Neighborhoods of words in the mental lexicon. Research on Speech Perception, Tech. Rep. No. 6, Bloomington, Indiana.

—— GOLDINGER, STEPHEN D., AUER, EDWARD T., and VITEVITCH, MICHAEL S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics* 62(3): 615–25.

—— and LARGE, NATHAN R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes* 16: 565–81.

—— and MCLENNAN, CONOR T. (2005). Spoken word recognition: The challenge of variation, in D. B. Pisoni and R. E. Remez (eds.), *The Handbook of Speech Perception*. Oxford: Blackwell, 591–609.

—— and PISONI, DAVID B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19: 1–36.

LUCK, STEVEN J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

LUCY, JOHN A. (1996). The linguistics of 'color,' in C. Hardin and L. Maffi (eds.), *Color Categories in Thought and Language.* Cambridge: Cambridge University Press, 320–46.

LÜTKENHÖNER, B. and POEPPEL, D. (2011). From tones to speech: Magnetoencephalographic studies, in J. A. Winer and C. E. Schreiner (eds.), *The Auditory Cortex*. Berlin: Springer, 597–615.

MACAULAY, RONALD (1991). *Locating Dialect in Discourse: The Language of Men and Bonnie Lassies in Ayr.* Oxford: Oxford University Press.

MACEACHERN, MARGARET R. (1999). *Laryngeal Cooccurrence Restrictions*. New York: Garland.

Mack, Molly (1990). Phonetic transfer in a French-English bilingual child, in P. Nelde (ed.), *Languages, Attitudes and Language Conflict*. Bonn: Dummler, 107–24.

MacKay, Donald G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* 8: 323–50.

MacKay, Ian R. A., Flege, James E., and Imai, S. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics* 27: 157–83.

—— Meador, Diane, and Flege, James E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica* 58: 103–25.

Macken, Marlys A. and Barton, David (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 7: 41–74.

Maclagan, Margaret and Hay, Jennifer (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language Variation and Change* 19: 1–25.

Macleod, Andrea N. and Stoel-Gammon, Carol (2005). Are bilinguals different? What VOT tells us about simultaneous bilinguals. *Journal of Multilingual Communication Disorders* 3: 118–27.

Macmillan, Neil A., Braida, L. D., and Goldberg, R. F. (1987). Central and peripheral effects in the perception of speech and non-speech sounds, in M. E. H. Schouten (ed.), *The Psychophysics of Speech Perception.* Dordrecht: Martinus Nihhoff, 28–45.

—— and Creelman, C. Douglas (1991). *Detection Theory: A User's Guide*. New York: Cambridge University Press.

MacNeilage, Peter F. (1980). The control of speech production, in G. Yeni-Komshian, J. Kaveanagh, and C. A. Ferguson (eds.), *Child Phonology Vol. 1: Production*. New York: Academic Press, 9–21.

—— and Davis, Barbara L. (1990). Motor explanations of babbling and early speech patterns, in M. Jeannerod (ed.), *Attention and Performance XIII: Motor Representation and Control*. Hillsdale, NJ: Lawrence Erlbaum, 567–82.

—— —— (2000). On the origin of internal structure of word forms. *Science* 288: 527–31.

—— —— Kinney, Ashlynn, and Matyear, Christine L. (2000). The motor core of speech: A comparison of serial organization patterns in infants and language. *Child Development* 71(1): 153–63.

MacWhinney, Brian (2000). *The CHILDES Project: Tools for Analyzing Talk. Vol. 2: The Database*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Maddieson, Ian (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

—— (1985). Phonetic cues to syllabification, in V. A. Fromkin (ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Orlando: Academic Press, 203–21.

—— and Ladefoged, Peter (1993). Phonetics of partially nasal consonants, in M. Huffman and R. Krakow (eds.), *Nasals, Nasalization, and the Velum (Phonetics and Phonology 5).* San Diego: Academic Press, 251–301.

Mády, Katalin and Beer, Ambros (2007). Articulatory parameters in consonant production after tumour surgery: A real-time MRI investigation. *Archives of Acoustics* 32: 135–45.

Maeda, Shinji (1991). On articulatory and acoustic variabilities. *Journal of Phonetics* 19: 321–31.

Maess, Burkhard, Friederici, Angela D., Damian, Markus, Meyer, Antje S., and Levelt, Willem J. M. (2002). Semantic category interference in overt picture naming:

Sharpening current density localization by PCA. *Journal of Cognitive Neuroscience* 14: 455–63.

Magen, Harriet S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics* 25: 187–205.

Magnuson, James S., McMurray, Bob, Tanenhaus, Michael K., and Aslin, Richard N. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science* 27(2): 285–98.

—— and Nusbaum, Howard C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance* 33(2): 391–409.

Maindonald, John H. and Braun, John (2003). *Data Analysis and Graphics Using R—An Example-based Approach*. Cambridge: Cambridge University Press.

Mair, Sheila J. and Shadle, Christine H. (1996). The voiced/voiceless distinction in fricatives: EPG, acoustic and aerodynamic data. *Proceedings of the Institute of Acoustics* 18(9): 163–70.

Makashay, Matthew J. and Johnson, Keith (1998). Surveying auditory space using vowel formant data. *Proceedings of the Joint Meeting of the International Conference Acoustics and the Acoustical Society of America* (Acoustical Society of America), 2037–8.

Malhotra, Shveta and Lomber, Stephen G. (2007). Sound localization during homotopic and heterotopic bilateral cooling deactivation of primary and nonprimary auditory cortical areas in the cat. *Journal of Neurophysiology* 97: 26–43.

Malsheen, Bathsheba J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children, in G. H. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (eds.), *Child Phonology*, vol. 2. San Diego, CA: Academic Press, 173–84.

Man, Vicky C. H. (2002). Focus effects on Cantonese tones: An acoustic study. *Proceedings of the 1st International Conference on Speech Prosody*. Aix-en-Provence, France, 467–70.

Mandel, Denise R., Jusczyk, Peter W., and Pisoni, David B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science* 6: 315–18.

Mann, Virginia A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics* 28: 407–12.

—— and Repp, Bruno H. (1980). Influence of vocalic context on perception of the [š] vs. [s] distinction. *Perception and Psychophysics* 28: 213–28.

Manning, Christopher and Schütze, Hinrich (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Manuel, Sharon, Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of Acoustical Society of America* 88: 1286–98.

—— (1999). Cross-language studies: Relating language-particular coarticulation patterns to other language-particular facts, in W. J. Hardcastle and N. Hewlett (eds.), *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press, 179–98.

Marchal, Alain (1988). Coproduction: Evidence from EPG data. *Speech Communication* 7: 287–95.

Marchman, Virginia and Bates, Elizabeth (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language* 21: 339–66.

Marean, G. Cameron, Werner, Lynne A., and Kuhl, Patricia K. (1992). Vowel categorization by very young infants. *Developmental Psychology* 28: 396–405.

Marian, Viorica and Spivey, Michael (2003a). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism* 6: 97–115.

——— (2003b). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics* 24: 173–93.

Marin, Stefania and Pouplier, Marianne (2008). Organization of complex onsets and codas in American English: Evidence for a competitive coupling model, in R. Sock, S. Fuchs, and Y. Laprie (eds.), *Proceedings of the 8th International Seminar on Speech Production*. Strasbourg: INRIA, 437–40, <http://issp2008.loria.fr/proceedings.html>.

Markel, John D. and Gray, Augustine H. (1976). *Linear Prediction of Speech*. Berlin: Springer.

Markram, Henry (2006a). The Blue Brain Project. *Nature Reviews Neuroscience* 7: 153–60.

——— (2006b). Dendritic object theory: A theory of the neural code where 3D electrical objects are formed across dendrites by neural microcircuits. Swiss Society for Neuroscience: Abstract H11.

Marslen-Wilson, William D. (1987). Functional parallelism in spoken word recognition. *Cognition* 25(1–2): 71–102.

——— Nix, Andy, and Gaskell, Gareth (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes* 10: 285–308.

——— Tyler, Lorraine K., Waksler, Rachelle, and Older, Lianne (1994). Morphology and meaning in the English mental lexicon. *Psychological Review* 101: 3–33.

——— and Warren, Paul (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review* 101: 653–75.

Martin, James G. and Bunnell, H. Timothy (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance* 8: 473–88.

Martins, Paula, Carbone, Inês, Pinto, Alda, Silva, Augusto, and Teixeira, António (2008). European Portuguese MRI-based speech production studies. *Speech Communication* 50: 925–52.

Mártony, Janos (1965). Studies of the voice source. Speech Trans. Lab. Q. Prog. Stat. Rep. 1. Stockholm: Royal Institute of Technology, 4–9.

Masakowski, Yvonne and Fifer, William P. (1994). The effects of maternal speech on foetal behavior. International Conference on Infant Studies, Paris.

Mascaró, Joan (1996). External allomorphy as emergence of the unmarked, in J. Durand and B. Laks (eds.), *Current Trends in Phonology: Models and Methods*. European Studies Research Institute, University of Salford, 473–83.

Massaro, Dominic (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press.

Matin, Ethel, Shao, K. C., and Boff, Kenneth R. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception and Psychophysics* 53: 372–80.

Matthews, Peter H. (1997). *The Concise Oxford Dictionary of Linguistics*. Oxford: Oxford University Press.

Matthies, Melanie L., Svirsky, Mario A., Lane, Harlan L., and Perkell, Joseph P. (1994). A preliminary study of the effects of cochlear implants on the production of sibilants. *Journal of the Acoustical Society of America* 96(3): 1367–73.

Mattock, Karen, Polka, Linda, Rvachew, Susan, and Krehm, Madelaine (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science* 13: 229–43.

MATTYS, SVEN L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance* 30: 397–408.

—— and JUSCZYK, PETER W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78: 91–121.

—— —— and LISS, JULIE M. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception and Psychophysics* 70: 1235–42.

—— WHITE, LAURENCE, and MELHORN, JAMES F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General* 134: 477–500.

MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS ONLINE EXPERIMENTS (2009). <http://www.mpi.nl/cgi-bin/exp/default.pl>, accessed March 13, 2009.

MAX, LUDO and CARUSO, ANTHONY J. (1997). Acoustic measures of temporal intervals across speaking rates: Variability of syllable- and phrase-level relative timing. *Journal of Speech, Language, and Hearing Research* 40: 1097–110.

MAY, JANET (1976). Vocal tract normalization for /s/ and /š/. Haskins Laboratories: Status Report on Speech Research SR-48, 67–73.

MAYE, JESSICA (2000). Learning speech sound categories on the basis of distributional information. Doctoral dissertation, University of Arizona.

—— and GERKEN, LOUANN (2000). Learning phoneme categories without minimal pairs, in S. Howel, S. Fisch, and T. Keith-Lucas (eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 522–33.

—— WEISS, DANIEL J., and ASLIN, RICHARD N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science* 11(1): 122–34.

—— WERKER, JANET F., and GERKEN, LOUANN (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82: B101–B111.

MAYO, CATHERINE, SCOBBIE, JAMES, HEWLETT, NIGEL, and WATERS, D. (2003). The influence of phonemic awareness development on acoustic cue weighting in children's speech perception. *Journal of Speech, Language, and Hearing Research* 46: 1184–96.

MAYO, LYNN H., FLORENTINE, MARY, and BUUS, SØREN (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research* 40: 686–93.

MAYR, R. and ESCUDERO, PAOLA (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism, Language and Cognition* 13(3): 279–97.

MCALLISTER, R. (1978). Temporal asymmetry in labial coarticulation. Working Papers, Institute of Linguistics, Stockholm University 35, 1–29.

MCCAFFERTY, KEVIN (1998) Shared accents, divided speech community? Change in Northern Ireland English. *Language Variation and Change* 10: 97–121.

MCCANDLISS, BRUCE D., FIEZ, JULIE A., PROTOPAPAS, ATHANASSIOS, CONWAY, MARY, and MCCLELLAND, JAY L. (2002). Success and failure in teaching the r-l contrast to Japanese adults: predictions of a hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience* 2: 89–108.

MCCARTHY, JOHN J. (1985). *Formal Problems in Semitic Phonology and Morphology.* New York: Garland Press.

—— (1988). Feature geometry and dependency: A review. *Phonetica* 45: 84–108.

——(1994) The phonetics and phonology of Semitic pharyngeals, in P. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, 191–233.
——(2005). Taking a free ride in morphononemic learning. *Catalan Journal of Linguistics* 4: 19–55. ROA 683.
——(2008). *Doing Optimality Theory*. Malden, MA: Blackwell Publishing.
——and Taub, Alison (1992). Review of Paradis and Prunet (1991). *Phonology* 9: 363–70.
McCawley, James D. (1967). Sapir's phonologic representation. *International Journal of American Linguistics* 33(2): 106–11.
——(1968). *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.
——(1978). What is a tone language?, in V. Fromkin (ed.), *Tone: A Linguistic Survey*. New York: Academic Press, 113–32.
McClelland, James L. and Elman, Jeffrey L. (1986). The TRACE model of speech perception. *Cognitive Psychology* 18: 1–86.
——Fiez, Julie A., and McCandliss, Bruce D. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology and Behavior* 77: 657–62.
——Mirman, Daniel, and Holt, Lori L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences* 10(8): 363–9.
——and Rumelhart, David E. (1986). A distributed model of human learning and memory, in J. L. McClelland, D. E. Rumelhart, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*. Cambridge, MA: MIT Press, 170–215.
McCrary, Kristie M. (2004). Reassessing the role of the syllable in Italian phonology. Ph.D. dissertation, UCLA.
McDonough, Joyce (2003). *The Navajo Sound System*. Dordrecht: Kluwer Academic Publishers.
——and Wood, V. (2008). The stop contrasts of the Athabaskan languages. *Journal of Phonetics* 36: 427–49.
McEnery, Tony and Wilson, Andrew (2001). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
McGowan, Richard S. and Nittrouer, Susan (1988). Differences in fricative production between children and adults: Evidence from an acoustic analysis of /S/ and /s/. *Journal of the Acoustical Society of America* 83: 229–32.
McGurk, Harry and MacDonald, John (1976). Hearing lips and seeing voices. *Nature* 264: 746–8.
McKenna, Gordon (1988). Vowel duration in the Standard English of Scotland. Unpublished M.Litt. thesis, University of Edinburgh.
McLennan, Conor T. (2007). Challenges facing a complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, 67–70.
——and Luce, Paul A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition* 31: 306–21.
————and Charles-Luce, Jan (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29: 539–53.
McMahon, April (1994). *Understanding Language Change*. Cambridge University Press: Cambridge.

McLennan, Conor T., Foulkes, Paul, and Tolfree, Laura (1994). Gestural representation and lexical phonology. *Phonology* 11: 277–316.

McMurray, Bob and Aslin, Richard N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy* 6: 203–29.

—— —— (2005). Infants are sensitive to within-category variation in speech perception. *Cognition* 95: B15–B26.

—— —— Tanenhaus, Michael K., Spivey, Michael J., and Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance* 34: 1609–31.

—— —— and Toscano, Joseph (2009). Statistical learning of phonetic categories: Computational insights and limitations. *Developmental Science* 12(3): 369–78.

—— Cole, Jennifer S., and Munson, Cheyenne (2011). Features as an emergent product of perceptual parsing: Evidence from vowel-to-vowel coarticulation. Invited chapter for R. Ridouane and G. N. Clement (eds.), *Where Do Features Come From? The Nature and Sources of Phonological Primitives*. Elsevier: North-Holland Linguistic Series.

—— Horst, Jessica, Toscano, Joseph, and Samuelson, Larissa (2009). Towards an integration of connectionist learning and dynamical systems processing: Case studies in speech and lexical development, in J. Spencer, M. Thomas, and J. McClelland (eds.), *Toward a Unified Theory of Development: Connectionism and Dynamic Systems Theory Reconsidered*. London: Oxford University Press.

—— and Jongman, Allard (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review* 118(2): 219–46.

—— Kovack-Lesh, Kristine, Goodwin, Dresden, and McEchron, William D. (in preparation). Separating phonetic categories in infant directed speech: Intentional enhancement or hyperarticulation?

—— and Spivey, Michael J. (2000). The categorical perception of consonants: the interaction of learning and processing. *Proceedings of the Chicago Linguistics Society* 34(2): 205–20.

—— Tanenhaus, Michael K., and Aslin, Richard N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition* 86(2): B33–B42.

—— —— —— (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language* 60(1): 65–91.

McQueen, James M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language* 39: 21–46.

—— (2005). Speech perception, in K. Lamberts and R. Goldstone (eds.), *The Handbook of Cognition*. London: Sage Publications, 255–75.

—— and Cutler, Anne (1997). Cognitive processes in speech perception, in W. J. Hardcastle and J. Laver (eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 566–85.

—— —— (1998). Morphology in word recognition, in A. Spencer and A. M. Zwicky (eds.), *The Handbook of Morphology*. Oxford: Blackwell, 406–27.

—— —— and Norris, Dennis (2006). Phonological abstraction in the mental lexicon. *Cognitive Science* 30: 1113–26.

—— and Viebahn, Malte C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology* 60(5): 661–71.

Mead, Jere (1969). Volume displacement body plethysmograph for respiratory measurements in human subjects. *Journal of Applied Physiology* 15: 736–40.

Mehl, Matthias R. and Pennebaker, James W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology* 84: 857–70.

Mehler, Jacques, Dommergues, Jean-Yves, Frauenfelder, Uli, and Segui, Juan (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20: 298–305.

—— Jusczyk, Peter, Lambertz, Ghislaine, Halsted, Nilofar, Bertoncini, Josiane, and Amiel-Tison, Claudine (1988). A precursor of language acquisition in young infants. *Cognition* 29: 143–78.

Mehta, Gita, and Cutler, Anne (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech* 31: 135–56.

Meijer, Paul J. A. (1996). Suprasegmental structures in phonological encoding: The CV structure. *Journal of Memory and Language* 35: 840–53.

Ménard, Lucie, Aubin, Jérôme, Brisebois, Amélie, and Thibeault, Mélanie (2007). A study of the development of speech motor control using ultrasound recordings. Talk presented at Ultrafest IV, New York University, September 28–29, 2007.

—— and Demuth, Katherine (in preparation). Articulatory gestures in targeting word-final consonants.

—— Schwartz, Jean-Luc, and Boë, Louise-Jean (2004). The role of the vocal tract morphology in speech development: Perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research* 47: 1059–80.

Mendoza-Denton, Norma (2002). Language and Identity, in J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell, 475–99.

—— (2007). Sociolinguistic extensions of Exemplar Theory, in J. Cole and J. Hualde (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 443–54.

—— (2008). *Homegirls: Language and Cultural Practice among Latina Youth Gangs*. London: Blackwell.

Mendoza-Denton, Norma, Hay, Jennifer, and Jannedy, Stefanie (2003). Probabilistic sociolinguistics: Beyond variable rules, in R. Bod, J. Hay, and S. Jannedy (eds.), *Probability Theory in Linguistics*. Cambridge, MA: MIT Press, 97–138.

Mertens, Piet (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model, in *Proceedings of Speech Prosody 2004*, Nara (Japan), 23–26 March.

Merzenich, Michael M., Kaas, Jon H., and Roth, G. Linn (1976). Auditory cortex in the grey squirrel: Tonotopic organization and architectonic fields. *Journal of Comparative Neurology* 166: 387–401.

Mester, Armin (1994). The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12: 1–61.

Mesthrie, Rajend (1992). *English in Language Shift: The History, Structure, and Sociolinguistics of South African Indian English*. Cambridge: Cambridge University Press.

Metsala, Jamie and Walley, Amanda (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability, in J. L. Metsala and L. C. Ehri (eds.), *Word Recognition in Beginning Literacy*. Mahwah, NJ: Erlbaum, 89–120.

Metz, Dale E., Whitehead, Robert L., and Peterson, Donald H. (1980). An optical-illumination system for high-speed laryngeal cinematography. *Journal of the Acoustical Society of America* 67: 719–21.

Meyer, Antje S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language* 29: 524–45.

—— (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language* 30: 69–89.

—— (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition* 42: 181–211.

—— and Schriefers, Herbert (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition* 17: 1146–60.

—— Sleiderink Astrid M., and Levelt, William J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition* 66(2): B25–B33.

Meyer, Martin, Steinhauer, Karsten, Alter, Kai, Friederici, Angela D., and von Cramon, D. Yves (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language* 89: 277–89.

Meyerhoff, Miriam (2006). *Introducing Sociolinguistics*. London: Routledge.

—— and Nagy, Naomi (eds.) (2008). *Social Lives in Language: The Sociolinguistics of Multilingual Speech Communities. Celebrating the Work of Gillian Sankoff*. Amsterdam and Philadelphia: John Benjamins.

Mielke, Jeff (2005). Modeling distinctive feature emergence, in J. Alderete, C.-H. Han, and A. Kochetov (eds.), *Proceedings of the 24th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 281–9.

—— (2007). Multiple mechanisms of change and influence: Comments on Harrington, Gussenhoven, Gow and McMurray, and Munson, in J. Cole and J. I. Hualde (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 229–40.

—— (2008). *The Emergence of Distinctive Features*. Oxford: Oxford University Press.

—— (2009). Accepting unlawful variation and unnatural classes: A model of phonological generalization, in R. van de Vijver, C. Féry, and F. Kügler, *Variation and Gradience in Phonetics and Phonology*. Berlin: Mouton de Gruyter, 17–42.

—— (forthcoming). Phonologization and the typology of feature behaviour, in A. Yu, (ed.), *Phonologization*. Oxford: Oxford University Press.

—— Baker, Adam, and Archangeli, Diana (2006). Forever young: Inaudible /r/ allophony resists conventionalization. Talk presented at the Linguistic Society of America 80, Albuquerque, NM. <http://dingo.sbs.arizona.edu/~apilab/presentations/LSA2006rtalk.pdf>, accessed March 15, 2009.

—— —— —— (2010). Variability and homogeneity in American English: /r/ allophony and /s/ retraction, in C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (eds.), *Laboratory Phonology 10*. Berlin: Mouton de Gruyter, 699–729.

—— —— —— and Racy, Sumayya (2005). Palatron: a technique for aligning ultrasound images of the tongue and palate. *Coyote Papers* 14: 97–108.

Miller, Amanda (2008). Click cavity formation and dissolution in IsiXhosa: Viewing clicks with high-speed ultrasound, in R. Sock, S. Fuchs, and Y. Laprie (eds.), *Proceedings of the 8th International Seminar on Speech Production*. Strasbourg, 137–40.

——Brugman, Johanna, Sands, Bonnie, Namaseb, Levi, Exter, Mats, and Collins, Chris (2009). Differences in airstream and posterior place of articulation among N|uu clicks. *Journal of the International Phonetic Association* 39, 129–61.

——and Finch, Kenneth (2011). Corrected high-speed anchored ultrasound with software alignment. *Journal of Speech, Language, and Hearing Research* 54, 471–86.

——Namaseb, Levi, and Iskarous, Khalil (2007). Tongue body constriction differences in click types, in J. Cole and J. I. Hualde (eds.), *Laboratory Phonology 9*. New York: Mouton de Gruyter, 643–56.

Miller, George A. and Nicely, Patricia E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27: 338–52.

Miller, Joanne L. (1997). Internal structure of phonetic categories. *Language and Cognitive Processes* 12: 865–9.

——and Dexter, Emily R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 14: 369–78.

——and Eimas, Peter D. (1996). Internal structure of voicing categories in early infancy. *Perception & Psychophysics* 58(8): 1157–67.

——and Volaitis, Lydia E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics* 46(6): 505–12.

Millotte, Séverine, Wales, Roger, and Christophe, Anne (2007). Phrasal prosody disambiguates syntax. *Language and Cognitive Processes* 22(6): 898–909.

Milroy, J. and Milroy, Lesley (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21: 339–84.

————(1993). Mechanisms of change in urban dialects: The role of class, social network and gender. *International Journal of Applied Linguistics* 3(1): 57–78.

Milroy, Lesley (1987a). *Language and Social Networks*, 2nd edn. Oxford: Blackwell.

——(1987b). *Observing and Analysing Natural Language*. Oxford: Blackwell.

——and Gordon, Matthew (2003). *Sociolinguistics: Method and Interpretation*. Oxford: Blackwell.

Mirman, Daniel, Dixon, James A., and Magnuson, James S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language* 59: 475–94.

——Holt, Lori L., and McClelland, Jay L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America* 116: 1198–207.

——McClelland, Jay L., and Holt, Lori L. (2006). Interactive activation and Hebbian learning produce lexically guided tuning of speech perception. *Psychonomic Bulletin and Review* 13: 958–65.

Mitchell, Thomas (1997). *Machine Learning*. New York: McGraw-Hill.

Mitchener, W. Garrett (2003). Bifurcation analysis of the fully symmetric language dynamical equation. *Journal of Mathematical Biology* 46(3): 265–85.

——and Nowak, Martin A. (2004). Chaos and Language. *Proceedings of the Royal Society B: Biological Sciences*, vol. 271: 701–4.

Mitterer, Holger and Blomert, Leo (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics* 65: 956–69.

Mitterer, Holger, Csépe, Valéria, and Blomert, Leo (2006). The role of perceptual integration in the perception of assimilation word forms. *Quarterly Journal of Experimental Psychology* 59: 1395–424.

——— Honbolygo, Ferenc, and Blomert, Leo (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science* 30: 451–79.

——— and Ernestus, Mirjam (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34: 73–103.

——— ——— (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109: 168–73.

——— and McQueen, James M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance* 35: 244–63.

——— Yoneyama, Kumiko, and Ernestus, Mirjam (2008). How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction in speech comprehension. *Journal of Memory and Language* 59: 133–52.

Mitzenmacher, Michael (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics* 1(2): 226–51.

Mixdorff, Hansjörg, Luksaneeyanawin, Sudaporn, Fujisaki, Hiroya, and Charnvivit, Patavee (2002). Perception of tone and vowel quality in Thai. Paper presented at the 7th International Conference on Spoken Language Processing, Denver, Colorado, September 2002.

Miyawaki, Kanika, Strange, Winifred, Verbrugge, Robert, Liberman, Alvin L., Jenkins, James J., and Fujimura, Osamu (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics* 18: 331–40.

Mohammad, M., Sinan, T., and Mohammad, B. (2006). Articulatory models of Arabic vowels computed from magnetic resonance images. *Kuwait Journal of Science and Engineering* 33: 69–79.

Mohanan, K. P. (1991). On the bases of radical underspecification. *Natural Language and Linguistic Theory* 9: 285–325.

——— and Mohanan, T. (1986). Lexical phonology of the consonant system of Malayalam. *Linguistic Inquiry* 15: 575–602.

Molfese, Dennis L., Key, Fonaryova A. P., Maguire, Mandy J., Dove, Guy O., and Molfese, Victoria J. (2005). Event-related evoked potentials (ERPs) in speech perception, in D. B. Pisoni and R. E. Remez (eds.), *The Handbook of Speech Perception*. Malden, MA: Blackwell, 99–120.

Monahan, Philip J., Hwang, So-One, and Idsardi, William J. (2009) (under revision). Predicting speech: Neural correlates of voicing mismatch using MEG. *Brain Research*.

——— and Idsardi, William J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalization. *Language and Cognitive Processes* 25: 808–39.

——— de Souza, Kevin, and Idsardi, William J. (2008). Neuromagnetic evidence for the auditory restoration of fundamental pitch. *PLoS One* 3: e2900.

Moon, Christine, Cooper, Robin P., and Fifer, William P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development* 16(4): 495–500.

Moore, Corinne A. (1992). The correspondence of vocal-tract resonance with volumes obtained from magnetic-resonance images. *Journal of Speech and Hearing Research* 35: 1009–23.

—— and Jongman, Allard (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America* 102: 1864–77.

Moore, Roger K. (2007). Spoken language processing: Piecing together the puzzle. *Speech Communication* 49: 418–35.

—— and Maier, Viktoria (2006). Preserving fine phonetic detail using episodic memory: Automatic speech recognition with MINERVA2. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken.

Mooshammer, Christine and Geng, Christian (2008). Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association* 38: 117–36.

—— Hoole, Philip, and Kühnert, Barbara (1995). On loops. *Journal of Phonetics* 23: 3–21.

Moran, Michael J. (1993). Final consonant deletion in African American children speaking black English. *Language, Speech, and Hearing Services in Schools* 24: 161–6.

Morén, Bruce and Zsiga, Elizabeth (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language and Linguistic Theory* 24: 113–78.

Moreton, Elliott (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84: 55–71.

—— (2008). Analytic bias and phonological typology. *Phonology* 25: 83–127.

—— Feng, Gary, and Smith, Jennifer L. (2008). Syllabification, sonority, and perception: New evidence from a language game, in R. L. Edwards, P. J. Midtlyng, C. L. Sprague, and K. G. Stensrud (eds.), *Proceedings of the Chicago Linguistic Society (CLS 41)*, vol. 1 (main session), 341–55.

—— and Thomas, Erik R. (2007). Origins of Canadian raising in voiceless-coda effects: A case study in phonologization, in J. Cole and J. I. Hualde (eds.), *Papers in Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 37–64.

Morgan, Jane L. and Wheeldon, Linda R. (2003). Syllable monitoring in internally and externally generated English words. *Journal of Psycholinguistic Research* 32: 269–96.

Morris, Richard E. (1998). Stylistic variation in Spanish phonology. Ph.D. dissertation, Ohio State University.

Morrison, Geoff S. (2006). L1 & L2 production and perception of English and Spanish vowels: A statistical modelling approach. Doctoral dissertation, University of Alberta, Edmonton, Alberta, Canada.

Morse, Phillip A. and Snowdon, Charles T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception and Psychophysics* 17: 9–16.

Moskowitz, Breyne A. (1973). On the status of vowel shift in English, in T. E. Moore (ed.), Cognitive *Development and the Acquisition of Language*. New York: Academic Press, 223–60.

Moulines, Eric and Charpentier, Francis (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9: 453–67.

Mowrey, Richard A. and MacKay, Ian R. A. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America* 88: 1299–312.

—— and Pagliuca, William (1995). The reductive character of articulatory evolution. *Rivista di Linguistica* 7: 37–124.

Mrayati, Mohamad, Carré, René, and Guérin, B. (1988). Distinctive region and modes: A new theory of speech production. *Speech Communication* 7: 257–86.

Mücke, D., Grice, Martine, Becker, Johannes, and Hermes, Anne (2009). Sources of variation in tonal alignment: Evidence from acoustic and kinematic data. *Journal of Phonetics* 37(3): 321–38.

——————— Baumann, Stefan (2006). Articulatory and acoustic correlates of prenuclear and nuclear accents, in R. Hoffmann and H. Mixdorff (eds.), *Proceedings of Speech Prosody 2006*. Dresden: TUDpress Verlag der Wissenschaften GmbH, 297–300.

——— and Hermes, Anne (2007). Phrase boundaries and peak alignment: An acoustic and articulatory study, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 997–1000.

——— Nam, Hosung, Prieto, Pilar, and Goldstein, Louis (2009). Coupling of tone and constriction gestures in Catalan and German. Poster presented at PaPI 09 [Phonetics and Phonology in Iberia], Las Palmas de Gran Canaria, June 17–19, 2009.

Mullenix, John W., Pisoni, David B., and Martin, Christopher S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America* 85: 365–78.

Müller, Johannes (1851). *Manuel de Physiologie*. (Trans. from German by A.-J.-L. Jourdan.) Paris: Chez J.-B. Baillière.

Munat, Judith. (ed.) (2007). *Lexical Creativity, Texts and Contexts*. Amsterdam: John Benjamins.

Munhall, Kevin, Fowler, Carol, Hawkins, Sarah, and Saltzman, Elliot (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics* 20: 225–39.

——— Kawato, Mitsuo, and Vatikiotis-Bateson, Eric (2000). Coarticulation and physical models of speech production, in M. Broe and J. Pierrehumbert (eds), *Papers in Laboratory Phonology V: Acquisition and the Lexicon.* Cambridge: Cambridge University Press, 9–28.

——— and Löfqvist, Anders (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics* 20: 111–26.

Munro, Murray J. and Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38: 289–306.

Munson, Benjamin (2001). Phonological pattern frequency and speech production in children and adults. *Journal of Speech, Language, and Hearing Research* 44: 778–92.

——— (2004). Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean. *Journal of Speech, Language, and Hearing Research* 47: 58–69.

——— (2009). Gender biases in fricative perception, revisited. Oral presentation given at the 2009 meeting of the Linguistic Society of America, San Francisco, CA. <http://www.tc.umn.edu/~munso005/LSA2009_Munson.pdf>, accessed on June 13, 2009.

——— (2010). Levels of phonological abstraction and knowledge of socially motivated speech-sound variation: A review, a proposal, and a commentary on the papers by Clopper, Pierrehumbert, and Tamati, Drager, Foulkes, Mack, and Smith, Hall, and Munson. *Laboratory Phonology* 1: 157–78.

——— Baylis, Adriane L., Krause, Miriam O., and Yim, Dongsun (2010). Representation and access in phonological impairment, in C. Fougeron, M. D'Imperio, N. Vallee, and B. Kühnert (eds.), *Laboratory Phonology 10*. New York: Mouton de Gruyter, 381–404.

——— Edwards, Jan, and Beckman, Mary E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research* 48: 61–78.

——— Schellinger, Sarah K., Beckman, Mary E., and Meyer, Marie K. (2009). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. Accepted, pending revisions, in *Clinical Linguistics and Phonetics*.

—— Kaiser, Eden, and Urberg Carlson, Kari (2008). Assessment of children's speech production 3: Fidelity of responses under different levels of task delay. Poster presented at the 2008 ASHA Convention, Chicago, 20–2. <http://www.tc.umn.edu/~munso005/MunsonKaiserUrberg-Carlson_Final.pdf>, accessed June 4, 2009.

—— Kurtz, Beth A., and Windsor, Jan (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research* 48: 1033–47.

—— Li, Fangfang, Yoneyama, Kiyoko, Hall, Kathleen C., Beckman, Mary E., Edwards, Jan, and Sunawatari, Yuki (2008). Sibilant fricatives in English and Japanese: Different in production or in perception? Oral presentation given at the annual meeting of the Linguistic Society of America, Chicago, IL.

Muthusamy, Yeshwant K., Cole, Ronald A., and Oshika, Beatrice T. (1992). The OGI multi-language telephone speech corpus. *Proceedings of the International Conference on Spoken LanguageProcessing,* Banff, Alberta, Canada, October 1992, 895–8.

Myers, Emily B. and Blumstein, Sheila E. (2008). The neural bases of the lexical effect: An fMRI investigation. *Cerebral Cortex* 18: 278–88.

Myers, James and Guy, Gregory (1997). Frequency effects in Variable Lexical Phonology. University of Pennsylvania Working Papers in Linguistics 4(1): 215–27.

Myers, Scott (1987). Vowel shortening in English. *Natural Language and Linguistic Theory* 5: 485–518.

—— (1996). Boundary tones and the phonetic implementation of tone in Chichewa. Studies in African Linguistics 25: 29–60.

Myers, Scott (2000). Boundary disputes: The distinction between phonetics and phonological sound patterns, in N. Burton-Roberts, P. Carr, and G. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues.* Oxford: Oxford University Press, 245–72.

—— and Hansen, Benjamin (2006). The origin of vowel-length neutralization in vocoid sequences. *Phonology* 22: 317–44.

——— —— (2007). The origin of vowel length neutralization in final position: Evidence from Finnish speakers. *Natural Language and Linguistic Theory* 25: 157–93.

Näätänen, Risto, Gaillard, Anthony W. K., Mäntysalo, Sirkka (1978). Early selective attention effects on voluntary and involuntary attention. *Acta Psychologica* 42: 313–29.

—— Lehtokoski, Anne, Lennes, Mietta, Cheour, Marie, Huotilainen, Minna, Iivonen, Antti, Vainio, Martti, Alku, Paavo, Ilmoniemi, Risto J., Luuk, Aavo, Allik, Jüri, Sinkkonen, Janne, and Alho, Kimmo (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385: 432–4.

—— Paavilainen, Petri, Rinne, Teemu, and Kimmo Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology* 118: 2544–90.

Nagy, Naomi and Reynolds, William (1997). Optimality Theory and variable word-final deletion in Faetar, *Language Variation and Change* 9: 37–55.

Nakai, Satsuki, Kunnari, Sari, Turk, Alice, Suomi, Kari, and Ylitalo, Riikka (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37: 29–45.

Nakamura, Masanobu, Iwano, Koji, and Furui, Sadaoki (2007). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language* 22: 171–84.

Nakatani, Christine H., Hirschberg, Julia, and Grosz, Barbara J. (1995). Discourse structure in spoken language: Studies on speech corpora. *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*.

Nakatani, L. H. and Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America* 62(3): 714–19.

Nam, Hosung. (2007). Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure, in J. Cole and J. I. Hualde (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 483–506.

—— Goldstein, Louis, and Saltzman, Elliot (2009). Self-organization of syllable structure: A coupled oscillator model, in F. Pellegrino, E. Marisco, and I. Chitoran (eds.), *Approaches to Phonological Complexity*. Berlin and New York: Mouton de Gruyter, 299–328.

Narayanan, Shirikanth, Nayak, Krishna, Lee, Sungbok, Sethy, Abhinav, and Byrd, Dani (2004). An approach to real-time magnetic resonance imaging for speech production. *Journal of the Acoustical Society of America* 115: 1771–6.

Natale, Michael (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32: 790–804.

Nazzi, Thierry, Bertoncini, Josiane, and Mehler, Jacques (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24: 756–66.

—— Iakimova, Galina, Bertoncini, Josiane, Fredonie, Severine, and Alcantara, Carmela (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language* 54: 283–99.

—— Jusczyk, Peter W., and Johnson, Elizabeth K. (2000). Language discrimination by English-learning 5-month-olds: Effect of rhythm and familiarity. *Journal of Memory and Language* 43: 1–19.

Nearey, Terrance M. and Hogan, John. T. (1986). Phonological contrast in experimental phonetics: Relating distributions of measurements in production data to perceptual categorization curves, in J. Ohala and J. J. Jaeger (eds.), *Experimental Phonology*. New York: Academic Press, 141–61.

Needham, Amy and Baillargeon, Rene (1998). Effects of prior experience in 4.5-month-old infants' object segregation. *Infant Behavior and Development* 21: 1–24.

Nespor, Marina and Vogel, Irene (1986). *Prosodic Phonology*. Dordrecht: Foris.

—— —— (1989). On clashes and lapses. *Phonology* 6: 69–116.

Nettle, Daniel (1999). Using social impact theory to simulate language change. *Lingua* 108: 95–117.

Newman, Rochelle S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *Journal of the Acoustical Society of America* 113(5): 2850–60.

—— (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology* 41(2), 352–62.

—— Ratner, Nan B., Jusczyk, Ann M., Jusczyk, Peter W., and Dow, Kathy A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology* 42(4), 643–55.

——Sawusch, James, and Luce, Paul (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance* 23(3): 873–89.

Nguyen, Hanh thi and Macken, Marlys (2008). Factors affecting the production of Vietnamese tone: A study of American learners. *Studies in Second Language Acquisition* 30: 49–77.

Nguyen, Noël, Wauquier, Sophie, and Tuller, Betty (2009). The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories, in F. Pellegrino, E. Marsico, I. Chitoran, and C. Coupé (eds.), *Approaches to Phonological Complexity*. Berlin: Mouton de Gruyter, 193–217.

Ní Chasaide, Ailbhe and Gobl, Christer (1993). Contextual variation of the vowel voice source as a function of adjacent consonants. *Language and Speech* 36: 303–30.

Niebuhr, Oliver (2003). Perceptual study of timing variables in F0 peaks. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 1225–8.

——(2008). Identification of highly reduced words by differential segmental lengthening. Presentation at the First Nijmegen Speech Reduction Workshop, MPI, Nijmegen, The Netherlands.

Niedzielski, Nancy A. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18: 62–85.

Nielsen, Kuniko Y. (2007). Implicit phonetic imitation is constrained by phonemic contrast. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany, 1961–4.

van Nierop, D. J. P. J., Pols, L. C. W., and Plomp, R. (1973). Frequency analysis of Dutch vowels from 25 female speakers. *Acustica* 29: 110–18.

Nissenbaum, John, Hillman, Robert E., Kobler, James B., Curtin, Hugh D., Halle, Morris, and Kirsch, John E. (2002). High speed MRI of laryngeal gestures during speech production. Presentation at the Acoustical Society of America, Pittsburgh, PA, June.

Nitisaroj, Rattima (2006). Effects of stress and speaking rate on Thai tones. Ph.D. dissertation, Georgetown University.

Nittrouer, Susan (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics* 20: 351–82.

——(1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America* 97: 520–30.

——(1996). The discriminability and perceptual weighting of some acoustic cues to speech perception by three-year-olds. *Journal of Speech and Hearing Research* 39: 278–97.

——(2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *Journal of the Acoustical Society of America* 115: 1777–90.

——Estee, Sandy, Lowenstein, Joanna H., and Smith, Jennifer (2005). The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops. *Journal of the Acoustical Society of America* 117: 351–64.

——and Miller, Marnie E. (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America* 101: 2253–66.

——and Studdert-Kennedy, Michael (1987). The role of coarticulatory effects on the perception of fricatives by children and adults. *Journal of Speech and Hearing Research* 30: 319–29.

Nittrouer, Susan, Studdert-Kennedy, Michael and McGowan, Richard S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research* 32: 120–32.

Nolan, Francis (1992). The descriptive role of segments: Evidence from assimilation, in G. J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: Cambridge University Press, 261–80.

—— Holst, T., and Kühnert, Barbara (1996). Modelling [s] to [ʃ] assimilation in English. *Journal of Phonetics* 24, 113–37.

Nolte, John (2009). *The Human Brain: An Introduction to its Functional Anatomy*, 6th edn. Philadelphia, PA: Mosby.

Nooteboom, Sieb G. (1969). The tongue slips into pattern, in A. G. Sciarone, A. J. von Essen, and A. A. van Raad (eds.), *Nomen: Leyden Studies in Linguistics and Phonetics*. The Hague: Mouton, 114–32.

—— (1972). Production and perception of vowel duration: A study of durational properties in Dutch. Ph.D. dissertation, University of Utrecht.

—— and Quené, Hugo (2008). Self-monitoring versus feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language* 58: 837–61.

Norris, Dennis and McQueen, James M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review* 115: 357–95.

—— —— and Cutler, Anne (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23: 299–370.

—— —— —— (2003). Perceptual learning in speech. *Cognitive Psychology* 47: 204–38.

Nosofsky, Robert M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14: 700–8.

Nowak, Martin (2000). The basic reproductive ratio of a word, the maximum size of a lexicon. *Journal of Theoretical Biology* 204(2): 179–89.

Nusbaum, Howard, Pisoni, David, and Davis, C. K. (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress*, Report No. 10, 357–76.

Nygaard, Lynne C. (2005). Perceptual integration of linguistic and nonlinguistic properties of speech, in D. B. Pisoni and R. E. Remez (eds.), *The Handbook of Speech Perception*. Oxford: Blackwell, 390–414.

O'Connor, J. D. and Arnold, Gordon F. (1961/1973). *Intonation of Colloquial English*. London: Longman.

Oakes, Lisa M., Coppage, Deborah J., and Dingel, Angela (1997). By land or by sea: The role of perceptual similarity in infants' categorization of animals. *Developmental Psychology* 33(3): 396–407.

Obleser, Jonas, Lahiri, Aditi, and Eulitz, Carsten (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *Neuroimage* 20: 1839–47.

—— —— —— (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience* 16: 31–9.

Ochs, Elinor (1986). From feelings to grammar: A Samoan case study, in B. Schieffelin and E. Ochs (eds.), *Language Socialization across Cultures*. Cambridge: Cambridge University Press, 251–71.

Odden, David. (1989). Predictable tone systems in Bantu, in H. Van der Hulst and N. Smith (eds.), *Autosegmental Studies on Pitch Accent Systems*. Dordrecht: Foris, 225–251.
—— (1992). Simplicity of underlying representation as motivation for underspecification. *Ohio State University Working Papers in Linguistics* 41: 83–100.
—— (1995). Tone: African languages, in J. Goldsmith (ed.), *The Handbook of Phonology*. Oxford: Blackwell, 444–75.
—— (2002). The verbal tone system of Zina Kotoko, in B. K. Schmidt, D. Odden, and A. Holmberg (eds.), *Aspects of Zina Kotoko Grammar*. München: Lincom Europa.
—— (2005). *Introducing Phonology*. Cambridge: Cambridge University Press.
Ogden, Richard (2006). Phonetics and social action in agreements and disagreements. *Journal of Pragmatics* 38: 1752–75.
—— Hawkins, Sarah, House, Jill, Huckvale, Mark, Local, John K., Carter, Paul, Dankovicová, Jana, and Heid, Sebastian (2000). ProSynth: An integrated prosodic approach to device-independent, natural-sounding speech synthesis. *Computer Speech and Language* 14: 177–210.
—— and Local, John K. (1994). Disentangling autosegments from prosodies: A note on the misrepresentation of a research tradition in phonology. *Journal of Linguistics* 30: 477–98.
—— and Routarinne, Sara (2005). The communicative functions of final rises in Finnish intonation. *Phonetica* 62: 160–75.
Oglesbee, Eric and de Jong, Kenneth J. (2007). Searching for best exemplars in multidimensional stimulus spaces. *Journal of the Acoustical Society of America* 122: EL101–EL106.
Ohala, John J. (1974). Experimental historical phonology, in J. Anderson and C. Jones (eds.), *Historical Linguistics II: Theory and description in Phonology (Proceedings of the First International Conference on Historical Linguistics, Edinburgh 2nd-7th September 1973)*, North-Holland Linguistics Series, vol 12b. Amsterdam: North-Holland, 353–87.
—— (1977). The physiology of stress, in L. M. Hyman (ed.), *Studies in Stress and Accent*, Southern California Occasional Papers in Linguistics. Los Angeles: University of Southern California, 145–68.
—— (1978). Production of tone, in Victoria Fromkin (ed.), *Tone: A Linguistic Survey*. New York: Academic Press, 5–39.
—— (1981a). The listener as a source of sound change, in C. S. Masek, R. A. Hendrick, and M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society 178–203.
—— (1981b). Speech timing as a tool in phonology. *Phonetica* 38: 204–12.
—— (1983). The origin of sound patterns in vocal tract constraints, in P. F. MacNeilage (ed.), *The Production of Speech*. New York: Springer, 189–216.
—— (1990a). Respiratory activity in speech, in W. J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modeling*. Dordrecht: Kluwer Academic Press, 23–53.
—— (1990b). The phonetics and phonology of aspects of assimilation, in J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 258–75.
—— (1990c). There is no interface between phonetics and phonology. A personal view. *Journal of Phonetics* 18: 153–71.
—— (1992a). What's cognitive, what's not, in sound change, in G. Kellermann and M. D. Morrissey (eds.), *Diachrony within Synchrony: Language History and Cognition*, Duisburger Arbeiten zur Sprach und Kulturwissenschaft 14. Frankfurt am Main: Peter Lang Verlag, 309–55.

Ohala, John J. (1992b). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Proceedings of the Chicago Linguistic Society (CLS 26), vol. 2: Parasession on the Syllable in Phonetics and Phonology*, 319–38.

—— (1993a). Coarticulation and phonology. *Language and Speech* 36: 155–70.

—— (1993b). The phonetics of sound change, in C. Jones (ed.), *Historical Linguistics: Problems and Perspectives*. London: Longman, 237–78.

—— (1995). The perceptual basis of some sound patterns, in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 87–92.

—— (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America* 99(3): 1718–25.

—— (1997). The relation between phonetics and phonology, in W. J. Hardcastle and J. Laver, *Handbook of Phonetic Sciences*. Oxford: Blackwell, 674–94.

—— (2008). The emergent syllable, in B. L. Davis and K. Zajdó (eds.), *The Syllable in Speech Production*. New York: Lawrence Erlbaum, 179–86.

—— and Amador, M. (1981). Spontaneous nasalization. *Journal of the Acoustical Society of America* 69: S54–S55 (abstract).

—— and Busà, Maria Grazia (1995). Nasal loss before voiceless fricatives: A perceptually-based sound change. *Rivista di Linguistica* 7: 125–44.

—— and Eukel, Brian W. (1987). Explaining the intrinsic pitch of vowels, in R. Channon and L. Shockey (eds.), *In Honor of Ilse Lehiste*. Dordrecht: Foris, 207–15.

—— and Ewan, William. G. (1972). Speed of pitch change. *Journal of the Acoustical Society of America* 53: 345.

—— and Jaeger, Jeri J. (eds.) (1986). *Experimental Phonology*. Orlando, FL: Academic Press.

—— and Kawasaki, Haruko (1984). Prosodic phonology and phonetics. *Phonology Yearbook* 1: 113–27.

—— and Ohala, Manjari (1995). Speech perception and lexical representation: The role of vowel nasalization in Hindi and English, in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 41–60.

Öhman, Sven (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151–68.

—— (1967). A numerical model of coarticulation. *Journal of the Acoustical Society of America* 41(2): 310–20.

Okobi, Anthony O. (2006). Acoustic correlates of word stress in American English. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA. <http://hdl.handle.net/1721.1/36350>.

Oller, D. Kimbrough (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54: 1235–47.

—— (1980). The emergence of the sounds of speech in infancy, in G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (eds.), *Child Phonology I: Production*. New York: Academic Press.

Onishi, Kristine H., Chambers, Kyle E., and Fisher, Cynthia (2002). Learning phonotactic constraints from brief auditory experience. *Cognition* 83: B13–B23.

van Oostendorp, Marc (1997). Style registers in conflict resolution, in F. Hinskens, R. van Hout, and L. Wetzels (eds.), *Variation, Change and Phonological Theory*. Amsterdam and Philadelphia: John Benjamins, 207–29.

Oppenheim, Gary M. and Dell, Gary S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* 106: 528–37.

Ostendorf, Mari, Price, Patti, and Shattuck-Hufnagel, Stefanie (1995). The Boston University radio news corpus. Boston University Technical Report ECS-95-001.

——— ——— ——— (1996). Boston University Radio Speech Corpus, Linguistic Data Consortium, Philadelphia.

Ota, Mitsuhiko (1999). Phonological theory and the acquisition of prosodic structure: Evidence from child Japanese. Unpublished doctoral dissertation, Georgetown University, Washington DC.

——— (2006). Input frequency and word truncation in child Japanese: Structural and lexical effects. *Language and Speech* 49: 261–95.

——— Hartsuiker, Robert J., and Haywood, Sarah L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition* 111: 263–9.

Oudeyer, Pierre-Yves (2005a). The self-organization of speech sounds. *Journal of Theoretical Biology* 233: 435–49.

Oudeyer, Pierre-Yves (2005b). How phonological structures can be culturally selected for learnability. *Adaptive Behavior* 13: 269–80.

——— (2006). *Self-Organization in the Evolution of Speech*. Oxford: Oxford University Press.

Padgett, Jay (1995). *Stricture in Feature Geometry*. Stanford: CLSI.

——— (2003). The emergence of contrastive palatalization in Russian, in Eric Holt (ed.), *Optimality Theory and Language Change*. Dordrecht: Kluwer Academic Press, 307–35.

——— and Tabain, M. (2005). Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica* 62: 14–54.

Pagel, Mark, Atkinson, Quentin D., and Meade, Andrew (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449: 717–21.

Pallier, Christophe, Bosch, Laura, and Sebastián-Gallés, Núria (1997). A limit on behavioral plasticity in speech perception. *Cognition* 64: B9–B17.

——— Colomé, Angels, and Sebastián-Gallés, Núria (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science* 12: 445–9.

Palmeri, Thomas J., Goldinger, Stephen D., and Pisoni, David B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition* 19: 309–28.

Pan, Ho-Hsien (1994), The acquisition of Taiwanese (Amoy) initial stops. Ph.D. dissertation, Ohio State University.

——— (2007a). Focus and Taiwanese unchecked tones, in C. Lee, M. Gordon, and D. Büring (eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*. Dordrecht: Springer, 195–213.

——— (2007b). Initial strengthening of lexical tones in Taiwanese, in T. Riad and C. Gussenhoven (eds.), *Tones and Tunes*, vol. 2. Berlin: Mouton de Gruyter, 271–91.

Panagos, John M., Quine, Mary E., and Klich, Richard J. (1979). Syntactic and phonological influences on children's production. *Journal of Speech and Hearing Research* 22: 841–8.

Pantev, Christo, Hoke, M., Lütkenhöner, Bernd, and Lehnertz, Klaus (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science* 246: 486–8.

Paolillo, John C. (2000). A probabilistic model for Optimality Theory. *Indiana Working Papers in Linguistics* 2.1.

—— (2002). *Analyzing Linguistic Variation: Statistical Models and Methods*. Stanford: CLSI.

Pardo, Jennifer S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119: 2382–93.

—— and Fowler, Carol A. (1997). Perceiving the causes of coarticulatory acoustic variation: Consonant voicing and vowel pitch. *Perception and Psychophysics* 59(7): 1141–52.

—— and Remez, Robert E. (2006). The perception of speech, in M. J. Traxler and M. A. Gernsbacher (eds.), *The Handbook of Psycholinguistics*, 2nd edn. Cambridge, MA: Elsevier, 201–48.

Park, Haeil and Iverson, Gregory (2009). Neural correlates of phonological representation and markedness. MS, University of Wisconsin, Milwaukee and University of Maryland Center for the Advanced Study of Language.

Park, Hanyong and de Jong, Kenneth J. (2008). Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects on second language identification skills. *Journal of Phonetics* 36: 704–23.

Parush, Avraham and Ostry, David (1993). Lower pharyngeal wall coarticulation in VCV syllables. *Journal of the Acoustical Society of America* 94(2): 715–22.

Pascual-Leone, Alvaro, Davey, Nick, Rothwell, John, Wassermann, Eric M., and Puri, Besant K. (2002). *Handbook of Transcranial Magnetic Stimulation*. London: Hodder Arnold.

Patel, Aniruddh D. and Balaban, Evan (2001). Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nature Neuroscience* 4: 839–44.

—— —— (2004). Human auditory cortical dynamics during perception of long acoustic sequences: Phase tracking of carrier frequency by the auditory steady-state response. *Cerebral Cortex* 14: 35–46.

Pater, Joe (2000). Nonuniformity in English stress: The role of ranked and lexically specific constraints. *Phonology* 17: 237–74.

—— (2003). The perceptual acquisition of Thai phonology by English speakers: Task and stimulus effects. *Second Language Research* 19: 209–23.

—— (2004). Bridging the gap between receptive and productive development with minimally violable constraints, in R. Kager, J. Pater, and W. Zonneveld (eds.), *Constraints in Phonological Acquisition*. Cambridge University Press, Cambridge, 219–44.

—— (2006). The locus of exceptionality: Morpheme-specific phonology as constraint indexation, in L. Bateman and A. Werle (eds.), *UMOP: Papers in Optimality Theory III*. Amherst, MA: GLSA, 1–36.

—— (2009a). Morpheme-specific phonology: Constraint indexation and inconsistency resolution, in S. Parker (ed.), *Phonological Argumentation: Essays on Evidence and Motivation*. London: Equinox Publishing, 123–54.

—— (2009b). Weighted constraints in generative linguistics. *Cognitive Science* 33: 999–1035.

—— Stager, Christine L., and Werker, Janet F. (2004). The lexical acquisition of phonological contrasts. *Language* 80(3): 361–79.

Patrick, Peter (2002). The speech community, in J. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *Handbook of Language Variation and Change*. Oxford: Blackwell, 573–97.

Patterson, Michelle L. and Werker, Janet F. (2002). Infants' ability to match dynamic information in the face and voice. *Journal of Experimental Child Psychology* 81: 93–115.

Payne, Elinor, Post, Brechtje, Prieto, Pilar, Vanrell, Maria del Mar, and Astruc, Lluïsa (forthcoming). Measuring child rhythm. *Language and Speech*.

Pell, Marc. D. (2005). Nonverbal emotion priming: evidence from the "facial affect decision task." *Journal of Nonverbal Behavior* 29(1): 45–73.

Peltola, Maija. S., Kuntola, Minna, Tamminen, Henna, Hamalainen, Heikki, Aaltonen, Olli (2005). Early exposure to non-native language alters preattentive vowel discrimination. *Neuroscience Letters* 388: 121–5.

—— Tamminen, Henna, Lehtola, Heidi, and Aaltonen, Olli (2007). Balanced bilinguals have one intertwined phonological system, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 1865–8.

Peng, Shu-Hui (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25: 371–400.

Peperkamp, Sharon (1997). *Prosodic Words*. HIL Dissertations 34. The Hague: Holland Academic Graphics.

—— and Dupoux, Emmanuel (2007). Learning the mapping from surface to underlying representations in an artificial language, in J. Cole and J. I. Hualde (eds.), *Change in Phonology (LabPhon 9)*. Berlin and New York: Mouton de Gruyter, 315–28.

—— Skoruppa, Katrin, and Dupoux, Emmanuel (2006). The role of phonetic naturalness in phonological rule acquisition, in D. Bamman, T. Magnitskaia, and C. Zaller (ed.), *Proceedings of the 30th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 464–75.

Perkell, Joseph S. (1986). Coarticulation strategies: Preliminary implications of a detailed analysis of lower lip protrusion movements. *Speech Communication* 5: 47–68.

—— and Cohen, Marc H. (1989). An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/ and /u/. *Journal of Phonetics* 17: 123–33.

—— Matthies, Melanie L., Lane, Harlan, Guenther, Frank H., Wilhelms-Tricarico, Reiner, Wozniak, Jane, and Guiod, Peter (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication* 22: 227–50.

—— —— Svirsky, Mario A., and Jordan, Michael I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot "motor equivalence" study. *Journal of the Acoustical Society of America* 93(5): 2948–61.

Perrachione, Tyler K., Lee, Jiyeon, Ha, Louisa Y. Y., and Wong, Patrick C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America* 130: 461–72.

Perret, Cyril (2007). La syllabe comme unité de traitement en production verbale orale et écrite. Doctoral dissertation, Université Blaise Pascal, Clermont-Ferrand.

Perrier Pascal, Ostry, David J., and Laboissière, Rafael (1996). The equilibrium point hypothesis and its application to speech motor control. *Journal of Speech, Language, and Hearing Research* 39: 365–78.

—— Payan, Yohan, Zandipour, Majid, and Perkell, Joseph (2003). Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America* 114(3): 1582–99.

Peters, Ann M. and Menn, Lise (1993). False starts and filler syllables: Ways to learn grammatical morphemes. *Language* 69: 742–77.

Peterson, Gordon E. and Barney, Harold L. (1952). Control methods used in the study of vowels. *Journal of the Acoustical Society of America* 24(2): 175–84.

Petrone, Caterina (2008). From targets to tunes: Nuclear and prenuclear contribution in the identification of intonation contours in Italian. Ph.D. dissertation, Laboratoire de Parole et Langage, Université de Provence.

——and Ladd, D. Robert (2007). Sentence-domain effects on tonal alignment in Italian?, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, August 6–10, 2007, 1253–6.

Phillips, Betty S. (1984). Word frequency and the actuation of sound change. *Language* 60: 320–42.

——(1994). Southern English glide deletion revisited. *American Speech* 69: 115–27.

——(2001). Lexical diffusion, lexical frequency, and lexical analysis, in J. L. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 123–36.

——(2006). *Word Frequency and Lexical Diffusion*. Basingstoke: Palgrave Macmillan.

Phillips, Colin (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science* 25: 711–31.

——Marantz, Alec, McGinnis, Martha, Pesetsky, David, Wexler, Kenneth, Yellin, Elron, Poeppel, David, Roberts, Timothy, and Rowley, Howard (1995). Brain mechanisms of speech perception: A preliminary report. *MIT Working Papers in Linguistics* 26: 125–63.

——Pellathy, Thomas, and Marantz, Alec (2000). Phonological feature representations in auditory cortex. MS, University of Delaware and MIT.

————Yellin, Elron, Wexler, Kenneth, Poeppel, David, McGinnis, Martha, and Roberts, Timothy (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience* 12: 1038–55.

Phonetic Alphabets (2009). Retrieved March 13, 2009 from Wikipedia, <http://en.wikipedia.org/wiki/Category:Phonetic_alphabets>.

Picanço, Gessiane (2005). Mundurukú: Phonetics, phonology, synchrony, diachrony. Ph.D. dissertation, University of British Columbia.

Pierrehumbert, Janet B. (1980). The phonology and phonetics of English intonation. Ph.D. dissertation, Massachusetts Instititue of Technology, Cambridge, MA.

——(1990a). On the value of reductionism and formal explicitness in phonological models: Comments on Ohala's paper, in J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 276–9.

——(1990b). Phonological and phonetic representation. *Journal of Phonetics* 18: 375–94.

——(1993). Dissimilarity in the Arabic verbal roots. *Proceedings of the 23rd Meeting of the Northeastern Linguistic Society, Graduate Student Association*, U. Mass. Amherst. 367–81.

——(1994a). Syllable structure and word structure: A study of triconsonantal clusters in English, in P. A. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, 168–88.

——(1994b). Knowledge of variation, in K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki, and E. Zeinfeld (eds.), *Papers from the 30th Meeting of the Chicago Linguistics Society. Vol. 2. Papers from the Parasession on Variation*, 232–56.

——(2001a). Exemplar dynamics: Word frequency, lenition and contrast, in J. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 137–58.

——(2001b). Stochastic phonology. *Glot International* 5/6: 195–207.

——(2001c). Why phonological constraints are so coarse-grained, in J. McQueen and A. Cutler (eds.), *Language and Cognitive Processes* 16: 691–8.
——(2002). Word-specific phonetics, in C. Gussenhoven and N. Warner (eds.), *Papers in Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 101–40.
——(2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46(2–3): 115–54.
——(2003b). Probabilistic phonology: Discrimination and robustness, in R. Bod, J. Hay, and S. Jannedy (eds.), *Probability Theory in Linguistics*. Cambridge, MA: MIT Press, 177–228.
——(2006a). The next toolkit. *Journal of Phonetics* 34(6): 516–30.
——(2006b). The statistical basis of an unnatural alternation, in L. M. Goldstein, D. H. Whalen, and C. T. Best (eds.), *Laboratory Phonology 8*. Berlin: Mouton de Gruyter, 81–106.
——and Beckman, Mary E. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.
————and Ladd, D. Robert (2000). Conceptual foundations of phonology as a laboratory science, in N. Burton-Roberts, P. Carr, and G. J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press, 273–303. [Reprinted, this volume.]
——Bent, Tessa, Munson, Benjamin, Bradlow, Anne, and Bailey, Michael (2004). The influence of sexual orientation on vowel production. *Journal of the Acoustical Society of America* 116(4): 1905–8.
——and Clopper, Cynthia (2010). What is LabPhon? And where is it going? in C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (eds.), *Laboratory Phonology 10*. Berlin: Mouton, 113–32.
——and Hirschberg, Julia (1990). The meaning of intonational contours in the interpretation of discourse, in P. Cohen, J. Morgan, and M. Pollack (eds.), *Intentions in Communication*. Cambridge, MA: MIT Press, 271–311.
——and Nair, Rami (1995). Word games and syllable structure. *Language and Speech* 38: 77–114.
——and Steele, Shirley (1989). Categories of tonal alignment in English. *Phonetica* 46: 181–96.
——and Talkin, David (1992). Lenition of /h/ and glottal stop, in G. J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, 90–117.
Pike, Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.
——(1948). *Tone Languages*. Ann Arbor: University of Michigan Press.
Pinheiro, Jose C. and Bates, Douglas M. (2000). *Mixed-effects Models in S and S-PLUS*. New York: Springer.
Pinker, Steven (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
——(1989). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.
——and Prince, Alan (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28: 73–193.
Piske, Thorsten, MacKay, Ian R. A., and Flege, James E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

Pisoni, David B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America* 61: 1352–61.

—— (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication* 13: 109–25.

—— Aslin, Richard N., Perey, A. J., and Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* 8: 297–314.

—— and Lively, Scott E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning, in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press, 433–59.

—— —— and Logan, John S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception, in J. C. Goodman and H. C. Nusbaum (eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge, MA: MIT Press, 121–66.

Pitrelli, John F., Beckman, Mary E., and Hirschberg, Julia (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework, in *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, 123–6.

Pitt, Mark A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language* 61(1): 19–36.

—— Dilley, Laura, Johnson, Keith, Kiesling, Scott, Raymond, William, Hume, Elizabeth, and Fosler-Lussier, Eric (2007). Buckeye Corpus of Conversational Speech (2nd release), <http://www.buckeyecorpus.osu.edu>. Columbus, OH: Department of Psychology, Ohio State University (Distributor).

—— Johnson, Keith, Hume, Elizabeth, Kiesling, Scott, and Raymond, William (2005). The Buckeye Corpus of Conversational Speech: Labeling conventions and a test of transcriber reliability. *Speech Communication* 45: 90–5.

Plaut, David C. and Kello, Christopher T. (1999). The emergence of phonology from the interplay of speech comprehension and production: a distributed connectionist approach, in B. MacWhinney (ed.), *The Emergence of Language*. Mahwah, NJ: Erlbaum, 381–415.

Plug, Leendert (2005). From words to actions: The phonetics of *Eigenlijk* in two communicative contexts. *Phonetica* 62: 131–45.

Pluymaekers, Mark, Ernestus, Mirjam, and Baayen, R. Harald (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62: 146–59.

—— —— —— and Booij, Geert (2006). The role of morphology in fine phonetic detail: The case of Dutch *-igheid*, in C. Fougeron et al. (eds.), *Laboratory Phonology 10*. Berlin: Mouton, 53–4.

—— —— —— —— (2010). Morphological effects on fine phonetic detail: The case of Dutch -igheid*, in C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (eds.), *Laboratory Phonology 10*. Berlin: Walter de Gruyter, 511–31.

Podesva, Robert J. (2006). Phonetic detail in sociolinguistic variation: Its linguistic significance and role in the construction of social meaning. Ph.D. dissertation, Stanford University.

—— (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11: 478–504.

——(2008). Three sources of stylistic meaning. *Texas Linguistic Forum. Proceedings of the Symposium about Language and Society, Austin* 15, 51: 1–10.

Poeppel, David (1996). A critical review of PET studies of phonological processing. *Brain and Language* 55: 317–51.

Poeppel, David (2008). The cartographic imperative: Confusing localization and explanation in human brain mapping, in H. Bredekamp, M. Bruhn, and G. Werner (eds.), *Bildwelten des Wissens 6.1: Ikonographie des Gehirns*. Berlin: Akademie Verlag Berlin.

——Phillips, Colin, Yellin, Elron, Rowley, Howard, Roberts, Timothy, and Marantz, Alec (1997). Processing of vowels in supratemporal auditory cortex. *Neuroscience Letters* 221: 145–8.

Polivanov, Evgenij. D. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4: 79–96. [English translation: The subjective nature of the perceptions of language sounds, in E. D. Polivanov (1974), *Selected Works: Articles on General Linguistics*. The Hague: Mouton, 223–37.]

Polka, Linda (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America* 89: 2961–77.

——(1992). Characterizing the influence of native experience on adult speech perception. *Perception and Psychophysics* 52: 37–52.

——(1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America* 97(2): 1286–96.

——Colantonio, Connie, and Sundara, Megha (2001). A cross-language comparison of /d/~/D/ discrimination: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America* 109: 2190–201.

——and Werker, Janet F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421–35.

Poplack, Shana (2001). *African American English in the diaspora*. Malden, MA: Blackwell.

——and Tagliamonte, Sali (1991). African American English in the diaspora: Evidence from Old-Line Nova Scotians. *Language Variation and Change* 3: 301–39.

————(1999). The grammaticization of *going to* in (African American) English. *Language Variation and Change* 11: 315–42.

Port, Robert (2007a). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology* 25: 143–70.

——(2007b). The problem of speech patterns in time, in M. G. Gaskell (ed.), *The Oxford Handbook of Psycholinguistics.* Oxford: Oxford University Press, 503–14.

Portal for Psychological Experiments on Language (2009). <http://www.surf.to/experiments>, accessed March 13, 2009.

Poser, William J. (1984). The phonetics and phonology of tone and intonation in Japanese. Ph.D. dissertation, MIT, Cambridge, MA.

Post, Brechtje (1999). Restructured phonological phrases in French: Evidence from clash resolution. *Linguistics* 37: 41–63.

——(2000). *Tonal and Phrasal Structures in French Intonation*. The Hague: Holland Academic Graphics.

——(2011). The multi-faceted relation between phrasing and intonation in French, in C. Lleo and C. Gabriel (eds.), *Hamburger Studies in Multilingualism 10: Intonational Phrasing at the Interfaces: Cross-Linguistic and Bilingual Studies in Romance and Germanic*. Amsterdam: John Benjamins, 44–74.

Post, Brechtje, D'Imperio, Mariapaola, and Gussenhoven, Carlos (2007). Fine phonetic detail and intonational meaning. *Proceedings of the International Congress of Phonetic Sciences 16*, Saarbrücken, 191–6.

—— Payne, Elinor, Prieto, Pilar, Vanrell, Maria del Mar, and Astruc, Lluïsa (forthcoming). A multisystemic model of rhythm development.

Potisuk, Siripong, Gandour, Jackson, and Harper, Mary (1997). Contextural variations in trisyllablic sequences of Thai tones. *Phonetica* 54: 22–42.

Pouplier, Marianne (2003a). Units of phonological encoding: Empirical evidence. Ph.D. dissertation, Yale University.

—— (2003b). The dynamics of error, in M.-J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 2245–8.

—— (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech* 50: 311–41.

—— (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics* 36: 114–40.

—— and Goldstein, Louis (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics* 33: 47–75.

—— and Hardcastle, William (2005). A re-evaluation of the nature of speech errors in normal and disordered speakers. *Phonetica* 62: 227–43.

PreBergen, Benjamin K. (2004). The psychological reality of phonaesthemes. *Language* 80: 290–311.

Prieto, Pilar (1998). The scaling of the L tone line in Spanish downstepping contours. *Journal of Phonetics* 26: 261–82.

—— (2005). Syntactic and eurhythmic constraints on phrasing decisions in Catalan. *Studia Linguistica* (Special issue on Boundaries in Intonational Phonology, ed. M. Horne and M. van Oostendorp) 59: 194–222.

—— D'Imperio, Mariapaola, and Gili Fivela, Barbara (2005). Pitch accent alignment in Romance: Primary and secondary associations with metrical structure. *Language and Speech* (special issue on Variation in Intonation, ed. P. Warren) 48(4): 359–96.

—— Mücke, Doris, Becker, Johannes, and Grice, Martine (2007). Coordination patterns between pitch movements and oral gestures in Catalan, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, August 6–10, 2007, 989–92.

—— van Santen, Jan, and Hirschberg, Julia (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics* 23: 429–51.

—— and Torreira, Francisco (2007). The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35(4): 473–500.

—— Torres-Tamarit, Francesc J., and Vanrell, Maria del Mar (2008). The role of tonal scaling in distinguishing intonational categories in Catalan. Paper presented at the Third TIE Conference on Tone and Intonation, Lisbon, September 15–17, 2008.

Prince, Alan S. (1990). Quantitative consequences of rhythmic organization, in M. Ziolkowski, M. Noske, and K. Deaton (eds.), *Papers from the Chicago Linguistic Society* 26(2): 355–98.

—— (2002a). Entailed ranking arguments [ROA-500].

—— (2002b). Arguing optimality [ROA-562].

——(2006). Implication and impossibility in grammatical systems: What it is and how to find it [ROA-880].

——(2007). The pursuit of theory, in P. de Lacy (ed.), *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, 33–60.

——and Smolensky, Paul (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar.* Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, April 1993. [Published: Oxford: Blackwell, 2004].

Pulgram, Ernst (1970). *Syllable, Word, Nexus, Cursus*. Berlin and New York: Mouton.

Pulleyblank, D. (1994). Underlying mora structure. *Linguistic Inquiry* 25(2): 344–53.

Pullum, Geoffrey (1991). *The Great Eskimo Vocabulary Hoax.* Chicago: University of Chicago Press.

Pulvermüller, Friedemann (1999). Words in the brain's language. *Behavioral and Brain Sciences* 22: 253–336.

——(2002). *Neuroscience of Language: On Brain Circuits of Words and Serial Order*. Cambridge: Cambridge University Press.

——and Shtyrov, Yury (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology* 79: 49–71.

————Illmoniemi, Risto J., and Marslen-Wilson, William (2006). Tracking speech comprehension in space and time. *Neuroimage* 31: 1297–305.

Putnam, Hilary (1973). Meaning and reference. *Journal of Philosophy* 7: 699–711.

——(1978). *Meaning and the Moral Sciences.* London: Routledge, and Kegan Paul.

Pycha, Anne, Nowak, Pawel, Shin, Eurie, and Shosted, Ryan (2003). Phonological rule-learning and its implications for a theory of vowel harmony, in G. Garding and M. Tsujimura (eds.), *Proceedings of WCCFL 22*. Somerville, MA: Cascadilla Press, 423–35.

Quené, Hugo (1992). Integration of acoustic-phonetic cues in word segmentation, in M. E. H. Schouten (ed.), *The Auditory Processing of Speech: From Sounds to Words*. Berlin: Mouton de Gruyter, 349–55.

——and van den Bergh, Huub (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language* 59(4): 413–25.

Quine, W. V. O. (1954/1966). The scope and language of science. Reprinted in *The Ways of Paradox and Other Essays.* New York: Random House, 215–32.

——(1960). *Word and Object.* Cambridge, MA: MIT Press.

——(1961). *From a Logical Point of View.* 2nd edn. Cambridge, MA: Harvard University Press.

Quinn, Paul C., Eimas, Peter D., and Rosenkrantz, Stacey L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception* 22(4): 463–75.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, <http://www.R-project.org>.

Rabiner, Lawrence R. (1989). A tutorial on hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE* 77: 257–86.

Rammage, Linda, Morrison, Murray D., Nichol, Hamish, and Pullan, Bruce (2001). *Management of the Voice and Its Disorders*, 2nd edn. Australia: Cengage Learning.

Rampton, Ben (1995). *Crossing: Language and Ethnicity among Adolescents*. London: Longman.

Ramus, Franck, Hauser, Marc D., Miller, Cory, Morris, Dylan, and Mehler, Jacques (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288: 349–51.

—— Peperkamp, Sharon, Christophe, Anne, Jacquemot, Charlotte, Kouider, Sid, and Dupoux, Emmanuel (2010). A psycholinguistic perspective on the acquisition of phonology, in C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (eds.), *Laboratory Phonology 10*. Berlin: Mouton De Gruyter, 311–40.

Ranbom, Larissa and Connine, Cynthia (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language* 57: 273–98.

Raposo, Ana, Moss, Helen E., Stamatakis, Emmanuel A., and Tyler, Lorraine K. (2009). Modulation of motor and premotor cortices by actions, action words, and action sentences. *Neuropsychologia* 47: 388–96.

Rapp, Brenda and Goldrick, Matthew (2000). Discreteness and interactivity in spoken word production. *Psychological Review* 107(3): 460–99.

—— —— (2006). Speaking words: Contributions of cognitive neuropsychological research. *Cognitive Neuropsychology* 23: 39–73.

Rathcke, Tamara and Harrington, Jonathan (2007). The phonetics and phonology of high and low tones in two falling f0-contours in standard German. Interspeech 2007, Antwerp.

Rauber, Andreia S., Escudero, Paola, Bion, Ricardo, and Baptista, Barbara O. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. *Proceedings of Interspeech*, 2913–16.

Rauschecker, Josef P. and Tian, Biao (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences* 97: 11800–6.

Raymond, William D., Dautricourt, Robin, and Hume, Elizabeth (2006). Word-medial /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 18: 55–97.

Reali, Florencia and Griffiths, Thomas L. (2009). The evolution of linguistic frequency distribution: Relating regularization to inductive biases through iterated learning. *Cognition* 111: 317–28.

Recasens, Daniel (2002). An EMA study of VCV coarticulatory direction. *Journal of the Acoustical Society of America* 111(6): 2828–41.

—— (2007). Patterns of CVC coarticulatory direction according to the DAC model, in P. Prieto, J. Mascaró, and M.J. Solé (eds.), *Segmental and Prosodic Issues in Romance Phonology*. Amsterdam: John Benjamins, 25–40.

—— and Espinosa, Alina (2006). Dispersion and variability of Catalan vowels. *Speech Communication* 48: 645–66.

—— —— (2009). Dispersion and variability in Catalan five and six peripheral vowel systems. *Speech Communication* 51: 240–58.

Redford, Melissa A. and Diehl, Randy L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America* 106: 1555–65.

—— and Miikkulainen, Risto (2007). Rate effects on structure in a source-filter model of phonological development. *Language* 83: 737–69.

Reetz, Henning (2000). Automatic speech recognition with features. Habilitationsschrift, Universität des Saarlandes, Saarbrücken.

—— (2003). Underspecified phonological features for lexical access, in W. J. Barry and J. Koreman, with K. Kirchhoff (eds.), *Phonus* 5. Saarbrücken: Institute of Phonetics, Saarland University, 161–73.

Reinholt Petersen, Niels (1978). Intrinsic fundamental frequency of Danish vowels. *Journal of Phonetics* 6: 177–89.

Remez, Robert (2005). Perceptual organization of speech, in David B. Pisoni and R. Remez (eds.), *Handbook of Speech Perception*. Oxford: Blackwell, 28–50.

Remijsen, Bert and Gilley, Leoma (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics* 36(2): 318–44.

—— and van Heuven, Vincent (1999). Categorical pitch dimensions in Dutch: Diagnostic test, in J. J. Ohala, Y. Hasagawa, M. Ohala, D. Granville, and A. C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco: University of California, 1865–8.

Repp, Bruno H. (1981a). Auditory and phonetic trading relations between acoustic cues in speech perception: Preliminary results. Haskins Laboratory Status Report on Speech Research, SR-67/68, 165–89.

—— (1981b). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society* 18: 12–14.

—— (1984). Categorical perception: Issues, methods, findings, in N. J. Lass (ed.), *Speech and Language, Advances in Basic Research and Practice*. Orlando: Academic Press, 243–335.

—— Liberman, Alvin M., Eccardt, Thomas, and Pesetsky, David (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance* 4: 621–37.

—— and Lin, Hwei-Bing (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics* 18: 481–95.

Restle, Frank (1955). A theory of discrimination learning. *Psychological Review* 62(1): 11–19.

Reynolds, William T. (1994). Variation and phonological theory. Ph.D. dissertation, Department of Linguistics, University of Pennsylvania.

Rialland, Annie (1994). The phonology and phonetics of extrasyllabicity in French, in P. A. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, 136–59.

—— (2001). Anticipatory raising in downstep realization: Evidence for preplanning in tone production, in S. Kaji (ed.), *Proceedings of the Symposium on Cross-linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics*, vol. 3. Tokyo, Japan, 301–22.

—— and Somé, Penu A. (2000). Dagara downstep: How speakers get started, in V. Carstens and F. Parkinson (eds.), *Trends in African Linguistics*. Trenton, NJ: Africa World Press, 251–63.

Richards, Douglas S., Frentzen, Barbara, Gerhardt, Kenneth J., McCann, Mary E., and Abrams, Robert M. (1992). Sound levels in the human uterus. *Obstetrics and Gynecology* 80(2): 186–90.

Rickford, John R. (1986). The need for new approaches to social class analysis in sociolinguistics. *Language and Communication* 6(3): 215–21.

—— (1999). *African American Vernacular English: Features, Evolution, Educational Implications*. Malden, MA: Blackwell.

Rickford, John R. and McNair-Knox, Fay (1994). Addressee- and topic-influenced style shift, in D. Biber and E. Finegan (eds.), *Sociolinguistic Perspectives on Register*. Oxford: Oxford University Press, 235–76.

Ridouane, Rachid (2003). Geminate vs. singleton stops in Berber: An acoustic, fiberscopic and photoglottographic study. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 1743–6.

Riehl, Anastasia (2003). American English flapping: Perceptual and acoustic evidence against paradigm uniformity with phonetic features. *Working Papers of the Cornell Phonetics Laboratory* 15: 271–337.

Rietveld, Toni and Gussenhoven, Carlos (1995). Aligning pitch targets in speech synthesis: Effects of syllable structure. *Journal of Phonetics* 23: 375–85.

Riney, Timothy, Takagi, Naoyuki, Ota, Kaori, and Uchida, Yoko (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics* 35: 439–43.

Ringen, Catherine and Heinämäki, Orvokki (1999). Variation in Finnish vowel harmony: An OT account. *Natural Language and Linguistic Theory* 17: 303–37.

Riordan, Carol J. (1977). Control of vocal tract length in speech. *Journal of the Acoustical Society of America* 62: 998–1002.

Rizzolatti, Giacomo and Arbib, Michael A. (1998) Language within our grasp. *Trends in Neurosciences* 21(5): 188–94.

Roark, Brian and Demuth, Katherine (2000). Prosodic constraints and the learner's environment: A corpus study, in C. S. Howell, S. A. Fish, and T. Keith-Lucas (eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 597–608.

Roberts, Julie (1997). Acquisition of variable rules: A study of (-t, d) deletion in preschool children. *Journal of Child Language* 24: 351–72.

Roberts, Timothy P. L., Flagg, Elissa J., and Gage, Nicole M. (2004). Vowel categorization induces departure of M100 latency from acoustic prediction. *Neuroreport* 15: 1679–82.

—— and Poeppel, David (1996). Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport* 7: 1138–40.

Roelofs, Ardi and Meyer, Antje S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition* 24: 922–39.

Roengpitya, Rungpat (2007). The variations, quantification, and generalizations of Standard Thai tones, in M.-J. Solé, P. Beddor, and M. Ohala (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 270–301.

Rogers, Catherine L., Dalby, Jonathan, and Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech* 47, 139–54.

—— Lister, Jennifer, Febo, Dashielle M., Besing, Joan M., and Abrams, Harvey B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics* 27: 465–85.

Romani, Christina and Calabrese, Anna (1998). Syllabic constraints in the phonological errors of an aphasic patient. *Brain and Language* 64: 83–121.

Romani, Gian Luca, Williamson, Samuel J., and Kaufman, Lloyd (1982). Tonotopic organization of the human auditory cortex. *Science* 216: 1339–40.

Rooth, Mats (1992). A theory of focus interpretation. *Natural Language Semantics* 1: 75–116.

—— (1996). Focus, in S. Lappin (ed.), *Handbook of Contemporary Semantic Theory*. London: Blackwell.

Rose, Mary A. (2006). Language, place, and identity in later life. Ph.D. dissertation, Stanford University.

Rose, Sharon and King, Lisa (2007). Speech error elicitation and co-occurrence restrictions in two Ethiopian Semitic languages. *Language and Speech* 50: 451–504.

—— and Walker, R. (2004). A typology of consonant agreement as correspondence. *Language* 80: 475–531.

Rose, Yvan and Waquier-Gravelines, Sophie (2007). French speech acquisition, in S. McLeod (ed.), *The International Guide to Speech Acquisition*. Florence, KY: Thomson Delmar Learning.

Rosenthal, Orna, Fusi, Stefano, and Hochstein, Shaul (2001). Forming classes by stimulus frequency: Behavior and theory. *Proceedings of the National Academy of Sciences* 98(7): 4265–70.

Rost, Gwyneth and McMurray, Bob (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science* 12(2): 339–49.

—— —— (2010). Finding the signal by adding noise: The role of non-contrastive phonetic variability in early word learning. *Infancy* 15(6): 608–35.

Rothenberg, Martin (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *Journal of the Acoustical Society of America* 53: 1632–45.

Rousseau, Pascal and Sankoff, David (1978). Advances in variable rule methodology, in D. Sankoff (ed.), *Linguistic Variation: Models and Methods*. New York: Academic Press, 57–69.

Roux, Justus C. (1995). On the perception and production of tone in Xhosa. *South African Journal of African Languages* 15: 196–204.

Roy, Alice C., Craighero, Laila, Fabbri-Destro, Maddalena, and Fadiga, Luciano (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *Journal of Physiology Paris* 102: 101–5.

Roy, Deb (2005a). Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence* 167: 170–205.

—— (2005b). Grounding words in perception and action: Computational insights. *Trends in Cognitive Sciences* 9: 389–96.

—— (2008). A mechanistic model of three facets of meaning, in M. de Vega, A. Glenburg, and A. Graesser (eds.), *Symbols and Embodiment: Debates on Meaning and Cognition*. Oxford: Oxford University Press.

Rubach, Jerzy (1996). Shortening and ambisyllabicity in English. *Phonology* 13: 197–237.

—— (1999). The syllable in phonological analysis. *Rivista di Linguistica* 11: 273–314.

—— and Booij, Geert. E. (2001). Allomorphy in Optimality Theory: Polish iotation. *Language* 77: 26–60.

Rubin, Philip, Baer, Thomas, and Mermelstein, Paul (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America* 70: 321–8.

Rumelhart, David and Zipser, David (1986). Feature discovery by competitive learning, in D. Rumelhart and J. L. McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press, 151–93.

Russell, Kevin (2008). Sandhi in Plains Cree. *Journal of Phonetics* 36: 450–64.

Rvachew, Susan and Andrews, Ellen (2002). The influence of syllable position on children's production of consonants. *Clinical Linguistics and Phonetics* 16: 183–98.

—— and Grawburg, Meghann (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research* 49: 74–87.

Rvachew, Susan and Jamieson, Donald (1989). Perception of voiceless fricatives by children with a functional articulation disorder. *Journal of Speech and Hearing Disorders* 54: 193–208.

—— Mattock, Karen, Polka, Linda, and Menard, Lucie (2006). Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *Journal of the Acoustical Society of America* 120: 2250–9.

Saffran, Jenny R., Aslin, Richard N., and Newport, Elissa L. (1996). Statistical learning by 8-month-old infants. *Science* 274: 1926–8.

—— and Thiessen, Erik D. (2003). Pattern induction by infant language learners. *Developmental Psychology* 39: 484–94.

Sagey, Elizabeth (1986). The representation of features and relations in non-linear phonology. Ph.D. dissertation, MIT. [Published, New York: Garland Press, 1991.]

—— (1988). On the ill-formedness of crossing association lines. *Linguistic Inquiry* 19: 109–18.

Saltzman, Elliot (1979). Levels of sensorimotor representation. *Journal of Mathematical Psychology* 20: 91–163.

—— (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Series* 15: 129–44.

—— (1995). Dynamics and coordinate systems in skilled sensorimotor activity, in R. F. Port and T. van Gelder (eds.), *Mind as Motion*. Cambridge, MA: MIT Press, 149–73.

—— and Byrd, Dani (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science* 19: 499–526.

—— Löfqvist, A., and Mitra, S. (2000). "Clocks" and "glue"—Global timing and intergestural cohesion, in M. B. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V*: *Acquisition and the Lexicon*. Cambridge: Cambridge University Press. 88–101.

—— and Munhall, Kevin G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1: 333–82.

—— Nam, Hosung, Krivokapic, Jelena, and Goldstein, Louis (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation, in *Proceedings of Speech Prosody 2008*, Campinas, Brazil.

Salverda, Anne Pier, Dahan, Delphine, and McQueen, James M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90: 51–89.

—— —— Tanenhaus, Michael K., Crosswhite, Katherine, Masharov, Mikhail, and McDonough, Joyce (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition* 105: 466–76.

Sambur, Marvin R., Rosenberg, Aaron E., Rabiner, Lawrence R., and McGonegal, Carol A. (1978). On reducing the buzz in LPC synthesis. *Journal of the Acoustical Society of America* 63: 918–24.

Samuel, Arthur G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology* 110(4): 474–94.

—— (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General* 125: 28–51.

Sancier, Michele and Fowler, Carol A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25: 421–36.

Sandler, Wendy (2006). Phonology, phonetics and the nondominant hand, in L. Goldstein, D. H. Whalen, and C. T. Best (eds.), *Laboratory Phonology 8*. Berlin: Mouton de Gruyter, 185–212.

—— and Lillo-Martin, Diane (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.

Sangster, Catherine (2002). Inter- and intra-speaker variation in Liverpool English: A sociophonetic study. D.Phil dissertation, University of Oxford.

Sankoff, David (1970). On the rate of replacement of word-meaning relationships. *Language* 47: 564–9.

Sankoff, David (1988). Variable rules, in U. Ammon, N. Dittmar, and K. J. Mattheier (eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society*, vol. 2. Berlin: Walter de Gruyter, 984–97.

—— and Kruskal, Joseph (1983). *Time Warps, String Edits, and Macromolecules*. Stanford, CA: CSLI Publications.

—— and Labov, William (1979). On the uses of variable rules. *Language in Society* 8(2): 189–222.

—— Tagliamonte, Sali, and Smith, Eric (2005). Goldvarb X: A variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto.

Sankoff, Gillian and Blondeau, Hélèn (2007). Language change across the lifespan: /r/ in Montreal French. *Language* 83: 560–88.

van Santen, Jan P. H. (1992). Contextual effects on vowel durations. *Speech Communication* 11: 513–46.

—— and Hirschberg, Julia (1994). Segmental effects on timing and height of pitch contours. *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, vol. 2, 719–22.

—— and Shih, Chilin (2000). Suprasegmental and segmental timing models in Mandarin Chinese and American English. *Journal of the Acoustical Society of America* 107(2): 1012–26.

Sapir, Edward (1921). *Language*. New York: Harcourt, Brace, & World.

—— (1925). Sound patterns in language. *Language* 1: 37–51.

Sapir, Shimon (1989). The intrinsic pitch of vowels: Theoretical, physiological, and clinical considerations. *Journal of Voice* 3: 44–51.

Sarkar, Deepayan (2010). lattice: Lattice graphics. <http://CRAN.R-project.org/package=lattice>. R package version 0.18-3.

Sasisekaran, Jayanthi, Smith, Anne, Sadagopan, Neeraja, and Weber-Fox, Christine (forthcoming). Nonword repetition in children and adults: Effects on movement coordination. *Developmental Science*.

Sato, Wataru and Yoshikawa, Sakiko (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition* 104: 1–18.

Sattath, Shmuel and Tversky, Amos (1977). Additive similarity trees. *Psychometrika* 42: 319–45.

Saur, Dorothee, Kreher, Björn W., Schnell, Susanne, Kümmerer, Dorothee, Kellmeyer, Philipp, Vry, Magnus-Sebastian, Umarova, Roza, Musso, Mariacristina, Glauche, Volkmar, Abel, Stefanie, Huber, Walter, Rijntjes, Michel, Hennig, Jürgen, and Weiller, Cornelius (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences* 105: 18035–40.

Saussure, Ferdinand de (1916). *Cours de Linguistique Générale*, ed. C. Bally and A. Sechehaye in collaboration with A. Riedlinger. Paris: Payot & Cie. [2nd edn, 1922]

Savela, Janne (2009). Role of selected spectral attributes in the perception of synthetic vowels. Ph.D. dissertation, University of Turku.

Savino, Michelina and Grice, Martine (2011). The perception of negative bias in Bari Italian questions, in S. Frota, P. Prieto, and G. Elordieta (eds.), *Prosodic Categories: Production, Perception and Comprehension*. Berlin: Springer Verlag, 187–206.

Sawashima, Masayuki, Hirose, Hajime, Yoshioka, Hirohide, and Kiritani, Shigeru (1982). Interaction between articulatory movements and vocal pitch control in Japanese word accent. *Phonetica* 39(45): 188–98.

Scarborough, Rebecca (2010). Lexical and contextual predictability: Confluent effects on the production of vowels, in C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (eds.), *Laboratory Phonology 10*. Berlin: de Gruyter.

Schafer, Amy J., Speer, Shari R., and Warren, Paul (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task, in J. C. Trueswell and M. K. Tanenhaus (eds.), *Approaches to Studying World-situated Language Use*. Cambridge, MA: MIT Press, 209–25.

——— ——— ——— and White, S. David (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research* 29: 169–82.

Scharenborg, Odette (2007). Reaching over the gap: A review of efforts to link human and automatic speech recognition research. *Speech Communication* 49: 336–47.

Scharinger, Mathias (2007a). The representation of vocalic features in vowel alternations. Phonological, morphological and computational aspects. Konstanz: Konstanz Online Publication System, <http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-24341)>.

——— (2007b). Inter-word identity, <http://www.inter-word.net/identity>, accessed April 4, 2007.

——— (2008). Minimal representations of alternating vowels. *Lingua*, forthcoming (doi:10.1016/j.lingua.2007.12.009).

——— Idsardi, William J., and Poe, Samantha (2011). A comprehensive three-dimensional cortical map of vowel space. *Journal of Cognitive Neuroscience*, forthcoming (doi:10.1162/jocn_a_00056).

——— Reetz, Henning, and Lahiri, Aditi (2009). Levels of regularity in inflected word form processing. *The Mental Lexicon* 4(1): 77–114.

Scheffers, Michael T. M. and Simpson, Adrian P. (1995). LACS: Label-Assisted Copy Synthesis. *Proceedings of the 13th International Congress of Phonetic Sciences* 2, 346–9.

Schellinger, Sarah K., Edwards, Jan, Munson, Benjamin, and Beckman, Mary E. (2008). The role of listener expectations on judgments of children's /s/ productions. Poster presented at the Symposium on Research in Child Language Disorders, June 5–7, University of Wisconsin, Madison. <http://www.ling.ohio-state.edu/~edwards/SRCLD_schellinger_final.pdf>, accessed June 17, 2009.

Scherer, Klaus R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40: 227–56.

Schiefer, Liselotte (1986). F0 in the production and perception of breathy stops: Evidence from Hindi. *Phonetica* 43: 43–69.

Schiller, Niels O. (1998). The effect of visually masked primes on the naming latencies of words and pictures. *Journal of Memory and Language* 39: 484–507.

——— (2000). Single word production in English: The role of subsyllabic units during speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26: 512–28.

——— (2005). Verbal self-monitoring, in A. Cutler (ed.), *Twenty-first Century Psycholinguistics: Four Cornerstones*. Mahwah, NJ: Lawrence Erlbaum, 245–61.

——— (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research* 1112: 201–12.

——(2008). Syllables in psycholinguistic theory: Now you see them, now you don't, in B. L. Davis and K. Zajdó (eds.), *The Syllable in Speech Production: Perspectives on the Frame/Content Theory*. New York, NY and Hove: Taylor & Francis, 155–76.

Schiller, Niels O., Bles, Mart, and Jansma, Bernadette M. (2003). Tracking the time course of phonological encoding in speech production: An event-related brain potential study. *Cognitive Brain Research* 17: 819–31.

——and Costa, Albert (2006). Activation of segments, not syllables, during phonological encoding in speech production. *The Mental Lexicon* 1: 231–50.

————and Colomé, Angels (2002). Phonological encoding of single words: In search of the lost syllable, in C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 35–59.

——Fikkert, Paula, and Levelt, Claartje C. (2004). Stress priming in picture naming: An SOA study. *Brain and Language* 90: 231–40.

——Jansma, Bernadette M., Peters, Judith, and Levelt, Willem J. M. (2006). Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes* 21: 112–40.

Schilling-Estes, Natalie (1998). Investigating "self-concious" speech: The performance register in Ocracoke English. *Language in Society* 27: 53–83.

——(2002). Investigating stylistic variation, in J. K. Chambers et al. (eds.), *The Handbook of Language Variation and Change*. Malden, MA: Blackwell, 375–401.

Schmidt, Anna Marie (1996). Cross-language identification of consonants. Part 1: Korean perception of English. *Journal of the Acoustical Society of America* 99: 3201–11.

Schmidt, Richard A., Zelaznik, Howard, Hawkins, Brian, Frank, James S., Quinn, John T., Jr. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review* 86(5): 415–51.

Schmidt, Thomas, Duncan, Susan, Ehmer, Oliver, Hoyt, Jeffrey, Kipp, Michael, Loehr, Dan, Magnusson, Magnus, Rose, Travis, and Sloetjes, Han (2008). An exchange format for multimodal annotations, in *Proceedings of the Workshop on Multimodal Corpora: From Models of Natural Interaction to Systems and Applications, Sixth International Conference on Language Resources and Evaluation*.

Schmitt, Bernadette M., Bles, Mart, Schiller, Niels O., and Münte, Thomas F. (2002). Overt naming in a picture-word interference task analyzed with event-related potentials (abstract). *Proceedings of the 9th Annual Meeting of the Cognitive Neuroscience Society*. Durham, NC: Duke University, 80.

——Münte, Thomas F., and Kutas, Marta (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology* 37: 473–84.

Schneider, Katrin and Linftert, Britta (2003). Categorical perception of boundary tones in German, in D. Recasens, M.-J. Solé, and J. Romero (eds.), *Proceedings of 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions Pty Ltd, 631–4.

Schouten, M. E. and Van Hessen, A. J. (1992). Modeling phoneme perception: Categorical perception. *Journal of the Acoustical Society of America* 92: 1841–55.

Schriefers, Herbert, Meyer, Antje S., and Levelt, Willem J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language* 29: 86–102.

Schuhmann, Teresa, Schiller, Niels O., Goebel, Rainer, and Sack, Alex (2009). The temporal characteristics of functional activation in Broca's area during overt picture naming. *Cortex* 45: 1111–16.

Schütze, Carson (2005). Thinking about what we are asking speakers to do, in S. Kepser and M. Reis (eds.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*. Berlin: Mouton de Gruyter, 457–84.

Schwartz, Jean-Luc, Boe, Louis-Jean, Valle, Natalie, and Abry, Christian (1997a). Major trends in vowel system inventories. *Journal of Phonetics* 25: 233–53.

——— ——— ——— ——— (1997b). The dispersion-focalization theory of vowel systems. *Journal of Phonetics* 25: 255–86.

——— and Escudier, P. (1989). A strong evidence for the existence of a large scale integrated spectral representation in vowel perception. *Speech Communication* 8: 235–59.

Schwartz, Richard and Leonard, Larry (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* 9: 319–36.

Schwarzlose, Rebecca and Bradlow, Ann R. (2001). What happens to segment durations at the end of a word? *Journal of the Acoustical Society of America* 109: 2292.

Scobbie, James M. (1991). Attribute-value phonology. Ph.D. dissertation, University of Edinburgh. [Published, New York:Garland Publishing, 1999.]

——— (2005). The "end" of phonology: The theoretical significance of interface phenomena. Oral paper at the 1st International Conference on the Linguistics of Contemporary English. University of Edinburgh, Scotland, June, 23–26.

——— (2006). Flexibility in the face of incompatible English VOT systems, in L. Goldstein, D. H. Whalen, and C. T. Best (eds.), *Laboratory Phonology 8.* Berlin: Mouton de Gruyter, 367–92.

——— (2007a). Biological and social grounding of phonology: Variation as a research tool, in J. Trouvain and W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 225–8.

——— (2007b). Interface and overlap in phonetics and phonology, in G. Ramchand and C. Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press, 17–52.

——— Gibbon, Fiona, Hardcastle, William J., and Fletcher, Paul (2000). Covert contrast as a stage in the acquisition of phonetics and phonology, in M. B. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 194–207.

——— and Hewlett, Nigel (2008). Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English, in P. Avery, E. B. Dresher, and K. Rice (eds.), *Contrast: Perception and Acquisition: Selected Papers from the Second International Conference on Contrast in Phonology*. Berlin: Mouton de Gruyter, 87–113.

——— ——— and Lawson, Elinor (2008). Looking variation and change in the mouth: Developing the sociolinguistic potential of ultrasound tongue imaging. Final report to ESRC, Research Grant RES-000-22-2032.

——— ——— and Turk, Alice E. (1999). Standard English in Edinburgh and Glasgow: The Scottish vowel length rule revealed, in P. Foulkes and G. J. Docherty (eds.), *Urban Voices: Accent Studies in the British Isles.* London: Arnold, 230–45.

——— Turk, Alice E., and Hewlett, Nigel (1999). Morphemes, phonetics and lexical items: The case of the Scottish vowel length rule. *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 1617–20.

Scobbie, James M., Wrench, Alan, and van der Linden, Marietta (2008). Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement, in R. Sock, S. Fuchs, and Y. Laprie (eds.), *Proceedings of the 8th International Seminar on Speech Production*. Strasbourg, France.

Scott, Sophie K. (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication* 41: 23–34.

—— and Johnsrude, Ingrid S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences* 26: 100–7.

Scully, Celia (1979). Model prediction and real speech: Fricative dynamics, in B. Lindblom and S. Öhman (eds.), *Frontiers of Communication Research*. New York: Academic Press, 35–48.

—— Georges, Esther, and Castelli, Eric (1991). Fricative consonants and their articulatory trajectories. *Proceedings of the 12th International Congress of Phonetic Sciences*, Aix-en-Provence, vol. 3, 58–61.

—— —— —— (1992). Articulatory paths for some fricatives in connected speech. *Speech Communication* 11: 411–16.

Segui, Juan and Ferrand, Ludovic (2002). The role of the syllable in speech perception and production, in J. Durand and B. Laks (eds.), *Phonetics, Phonology, and Cognition*. Oxford: Oxford University Press, 151–67.

Seidl, Amanda and Buckley, Eugene (2005). On the learning of arbitrary phonological rules. *Language Learning and Development* 1: 289–316.

Selkirk, Elisabeth O. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

—— (1986). On derived domains in sentence phonology. *Phonology Yearbook* 3: 371–405.

—— (1990). On the nature of prosodic constituency: Comments on Beckman and Edward's paper, in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 179–200.

—— (1995). Sentence prosody: Intonation, stress and phrasing, in J. Goldsmith (ed.), *The Handbook of Phonological Theory*. Oxford: Blackwell, 550–69.

—— (1996). The prosodic structure of function words, in J. L. Morgan and K. Demuth (eds.), *Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah: Lawrence Erlbaum Associates, 187–213.

—— (2000). The interaction of constraints on prosodic phrasing, in M. Horne (ed.), *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Publishers, 231–61.

—— (2002). Contrastive FOCUS vs. presentational focus: Prosodic evidence from right-node raising in English. Speech Prosody 2002. Aix-en-Provence, France.

—— (2005). Comments on intonational phrasing in English, in S. Frota, M. Vigário, and M. J. Freitas (eds.), *Prosodies*. Berlin: Mouton de Gruyter, 11–58.

—— (2007). Contrastive focus, givenness and the unmarked status of "discourse-new," in C. Féry, G. Fanselow, and M. Krifka (eds.), *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure*, vol. 6. Potsdam: Universitätsverlag Potsdam, 125–46.

—— and Shen, Tong (1990). Prosodic domains in Shanghai Chinese, in S. Inkelas and D. Zec (eds.), *The Phonology-Syntax Connection*. Chicago: University of Chicago Press.

—— Shinya, Tahahito, and Kawahara, Shigeto (2004). Phonological and phonetic effects of minor phrase length on f0 in Japanese, in *Speech Prosody 2004: Proceedings of the 2nd International Conference on Speech Prosody*. Nara: Japan, 183–7.

Sereno, Joan A. and Jongman, Allard (1995). Acoustic correlates of grammatical class. *Language and Speech* 38: 57–76.

Serrurier, Antoine and Badin, Pierre (2008). A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data. *Journal of the Acoustical Society of America* 123: 2335–55.

Sevald, Christine A., Dell, Gary S., and Cole, Jennifer S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language* 34: 807–20.

Shademan, Shabnam (2006). Is phonotactic knowledge grammatical knowledge?, in D. Baumer, D. Montero, and M. Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 371–9.

Shadle, Christine H. (1990). Articulatory-acoustic relationships in fricative consonants, in W. J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Press, 187–209.

—— (1991). The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics* 19: 409–24.

—— (1997). The aerodynamics of speech, in W. J. Hardcastle and J. Laver (eds.), *Handbook of Phonetics*. Oxford: Blackwell, 33–64.

—— (2006). Acoustic phonetics, in K. Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn, vol. 9. Oxford: Elsevier, 442–60.

—— (2010). Aerodynamics of speech, in W. J. Hardcastle, J. Laver, and F. Gibbon (eds.), *Handbook of the Phonetic Sciences*, 2nd edn. London: Blackwell, 39–80.

—— and Mair, Sheila J. (1996). Quantifying spectral characteristics of fricatives, in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, Philadelphia, Oct. 1996, 1521–4.

—— —— and Carter, John N. (1996). Acoustic characteristics of the front fricatives [f, v, θ, ð]. *Proceedings of the 4th Speech Production Seminar*, Autrans, 193–6.

—— Proctor, Michael I., and Iskarous, Khalil (2008). An MRI study of the effect of vowel context on English fricatives. *Proceedings of Acoustics '08*, Paris, June 28–July 4, 5099–104.

—— and Scully, Celia (1995). An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics* 23: 53–66.

Shafer, Valerie L., Morr, Mara L., Datta, Hia, Kurtzberg, Diane, and Schwartz, Richard G. (2005). Neurophysiological indices of speech processing deficits in children with specific language impairment. *Journal of Cognitive Neurosciences* 17: 1168–80.

Shahin, Antoine J., Bishop, Christopher W., and Miller, Lee M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage* 44: 1133–43.

Shaked, Amit (2007). Competing syntactic and phonological constraints in Hebrew prosodic phrasing. *The Linguistic Review* (Special issue on Prosodic Phrasing, ed. S. Frota and P. Prieto) 24: 169–99.

Shannon, Claude E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27(July and October): 379–423, 623–56.

Sharma, Anu and Dorman, Michael F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America* 106: 1078–83.

Shattuck-Hufnagel, Stefanie (1979). Speech errors as evidence for a serial ordering mechanism in sentence production, in W. E. Cooper and E. C. T. Walker (eds.), *Sentence Processing*. New York: Halsted Press, 295–342.

——(1983). Sublexical units and suprasegmental structure in speech production planning, in P. F. MacNeilage (ed.), *The Production of Speech*. New York: Springer, 109–36.

——(1987). The role of word onset consonants in speech production planning: New evidence from speech error patterns, in E. Keller and M. Gopnik (eds.), *Motor and Sensory Processing in Language*. Hillsdale, NJ: Erlbaum, 17–51.

——(1992). The role of word structure in segmental serial ordering. *Cognition* 42: 213–59.

——(1995). The importance of phonological transcription in empirical approaches to "stress shift" versus "early accent": Comments on Grabe and Warren, and Vogel, Bunnell, and Hoskins, in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 128–40.

——(2000). Phrase-level phonology in speech production planning: Evidence for the role of prosodic structure, in M. Horne (ed.), *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Publishers, 201–29.

——(2011). The role of the syllable in speech production in American English: A fresh consideration of the evidence, in C. Cairns and E. Raimy (eds.), *Handbook of the Syllable*. Leiden: Brill, 195–204.

—— Demuth, Katherine, Hanson, Helen, and Stevens, Kenneth N. (2011). Acoustic cues to stop-coda voicing contrasts in the speech of American English 2–3-year-olds, in G. N. Clements and R. Ridouane (eds.), *Where Do Features Come From? The Nature and Sources of Phonological Primitives*. North-Holland Linguistics Series. Amsterdam: Elsevier, 327–42.

—— and Klatt, Dennis H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior* 18: 41–55.

—— Ostendorf, Marie, and Ross, Ken (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22: 357–88.

—— and Turk, Alice (2009). An experimental investigation of Abercrombian feet in American English. Talk presented at the 22nd Annual CUNY Conference on Human Sentence Processing.

—— and Veilleux, Nanette M. (2007). Robustness of acoustic landmarks in spontaneously-spoken American English. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 925–8.

Shaw, Jason and Gafos, Adamantios (2010). Quantitative evaluation of competing syllable parses, in J. Heinz, L. Cahill, and R. Wicentowski (eds.), *ACL SIGMORPHON, Proceedings of the 11th Meeting of the ACL special interest group in computational phonology and morphology*. Uppsala, July 11–16, Sweden.

—— —— Hoole, Philip, and Zeroual, Chakir (2009). Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation. *Phonology* 26: 187–215.

Sheldon, A. and Strange, Winifred (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3: 243–61.

Shen, Xiao-Nan (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics* 18: 281–95.

Shepard, Roger N. (1972). Psychological representation of speech sounds, in E. E. David and P. B. Denes (eds.), *Human Communication: A Unified View*. New York: McGraw-Hill, 67–113.

Shi, Feng, Shi, Lin, and Liao, Rongrong (1987). An experimental analysis of the five-level tones of the Gaoba Dong language. *Journal of Chinese Linguistics* 15: 335–61.

Shih, Chilin (1987). The phonetics of the Chinese tonal system. AT&T Bell Labs technical memo.

—— (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory 3: Stress, Tone and Intonation.* Ithaca: Cornell University, 83–109.

—— and Kochanski, Greg (2000). Chinese Tone Modeling with Stem-ML. Sixth International Conference on Spoken Language Processing (ICSLP 2000). Beijing, October 16–20.

—— —— and Yoon, Su-Youn (2007). The missing link between articulatory gestures and sentence planning, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 35–8.

—— and Sproat, Richard (1992). Variations of the Mandarin rising tone, in *Proceedings of the IRCS Workshop on Prosody in Natural Speech*. Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania, 193–200.

Shinn-Cunningham, Barbara G. (2008). Object-based auditory and visual attention. *TRENDS in Cognitive Sciences* 12: 182–6.

Shobbrook, Katherine and House, Jill (2003). High rising tones in Southern British English. *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, 1273–6.

Shockey, Linda (2008). Understanding casual English pronunciation: Poles apart. Presentation at the First Nijmegen Speech Reduction Workshop, MPI, Nijmegen, The Netherlands.

Shockley, Kevin, Richardson, Daniel C., and Dale, Rick (2009). Conversation and coordinative structures. *Topics in Cognitive Science* 1: 305–19.

—— Sabadini, Laura, and Fowler, Carol A. (2004). Imitation in shadowing words. *Perception and Psychophysics* 66: 422–9.

—— Santana, Marie-Vee, and Fowler, Carol A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance* 29: 326–32.

Shtyrov, Yury and Pulvermuller, Friedemann (2007). Language in the mismatch negativity design: Motivations, benefits and prospects. *Journal of Psychophysiology* 21(3–4): 176–87.

Shue, Yen-Liang, Keating, Patricia, and Vicenik, Chad (2009). VoiceSauce: A program for voice analysis. Poster presented at the Fall meeting of the Acoustical Society of America, San Antonio.

Silva, David (1992). The phonetics and phonology of stop lenition in Korean. Ph.D. dissertation, Cornell University.

Silverman, Daniel (2006). The diachrony of labiality in Trique, and the functional relevance of gradience and variation, in L. M. Goldstein, D. H. Whalen, and C. T. Best (eds.), *Laboratory Phonology 8*. New Haven: Mouton de Gruyter, 133–54.

Silverman, Kim E. A. (1986). F0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica* 43: 76–91.

—— (1990). The separation of prosodies: Comments on Kohler's paper, in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge: Cambridge University Press, 72–106.

—— and Pierrehumbert, Janet (1990). The timing of prenuclear high accents in English, in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 72–106.

Simon, Claude and Fourcin, Adrian J. (1978). Cross-language study of speech pattern learning. *Journal of the Acoustical Society of America* 63: 925–35.

Simon, Ellen, Escudero, Paola, and Broersma, Mirjam. (2010). Learning minimally different words in a third language: L2 proficiency as a crucial predictor of accuracy in an L3 word learning task. *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (New Sounds 2010).

—— and Van Herreweghe, Mieke (2010), The relation between orthography and phonology. *Language and Speech* 53(3): 303–6.

Simosa, Panagiotis G., Diehl, Randy L., Breier, Joshua I., Molis, Michelle R., Zouridakis, George, and Papanicolaou, Andrew C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cognitive Brain Research* 7: 215–19.

Singh, Leher, Morgan, James, and Best, Catherine T. (2002). Baby talk or happy talk? *Infancy* 3: 365–94.

—— —— White, Katherine (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language* 51(2): 173–89.

Sinnott, Joan M., Brown, Charles H., and Borneman, Melissa A. (1998). Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys. *Perception and Psychophysics* 60: 1032–43.

Sirotin, Yevgeniy B. and Das, Aniruddha (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature* 457: 475–9.

Sjölander, Kimmen (2003). An HMM-based system for automatic segmentation and alignment of speech, in *Proceedings of Fonetik 2003*, Umeå, Sweden, 93–6.

Skousen, Royal, Lonsdale, Deryl, and Parkinson, Dilworth (2002). *Analogical Modeling. An Exemplar-Based Approach to Language*. Amsterdam: John Benjamins.

Slaney, Malcom, Covell, Michele, and Lassiter, Bud (1996). Automatic audio morphing. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, GA, May 7–10, 1996, vol. 2, 1001–4.

Slevc, L. Robert and Miyake, Akira (2006). Individual differences in second language proficiency: Does musical ability matter? *Psychological Science* 17: 675–81.

Slifka, Janet (2000). Respiratory constraints at prosodic boundaries in speech. Ph. D. dissertation, Massachusetts Institute of Technology, Cambridge, MA. <http://hdl.handle.net/1721.1/29184>.

—— (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice* 20: 171–86.

Slobin, Dan I. (1973). Cognitive prerequisites for the development of grammar, in C. Ferguson and D. I. Slobin (eds.), *Studies of Child Language Development*. New York: Holt, Rhinehart and Winston, 175–208.

—— (1997). The origins of grammaticizable notions: Beyond the individual mind, in D. I. Slobin (ed.), *The Crosslinguistic Study of Language Acquisition: Expanding the Contexts*. Mahwah, NJ: Lawrence Erlbaum.

Sluijter, Agaath M. C. (1995). Phonetic correlates of stress and accent. Ph.D. dissertation, Leiden University.

—— and van Heuven, Vincent J. (1995). Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch. *Phonetica* 52: 71–89.

—— —— (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100: 2471–85.

Smiljanić, Rajka and Bradlow, Ann R. (2007). Clear speech intelligibility: Listener and talker effects. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany.

Smiljanić, Rajka and Bradlow, Ann R. (2009) Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass* 3: 236–64.

Smit, Ann B., Freilinger, John J., Bernthal, James E., Hand, Linda, and Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders* 55: 779–98.

Smith, Anne and Goffman, Lisa (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research* 41: 18–30.

Smith, Bruce L. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics* 6: 37–67.

—— (1992). Relationships between duration and temporal variability in children's speech. *Journal of the Acoustical Society of America* 91: 2165–74.

—— and Kenney, Mary Kay (1994). Variability control in speech production tasks performed by adults and children. *Journal of the Acoustical Society of America* 96: 699–705.

—— —— and Hussain, Sarmad (1996). A longitudinal investigation of duration and temporal variability in children's speech production. *Journal of the Acoustical Society of America* 99: 2344–9.

Smith, Caroline L. (1995). Prosodic patterns in the coordination of vowel and consonant gestures, in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence*. Cambridge: Cambridge University Press, 205–22.

Smith, E. A., Munson, Benjamin, and Hall, Kathleen C. (2008). Rethinking the meaning of Minnesotan [æ]: Sexual orientation or personal well-being? Oral presentation at the conference on New Ways of Analyzing Variation (NWAV), Houston, TX. <http://www.ling.ohio-state.edu/~kchall/Smith_Munson_Hall_NWAV_2008.pdf>, accessed June 17, 2009.

Smith, Jennifer, Durham, Mercedes, and Fortune, Liane (2007). "Mam, my trousers is fa'in doon!": Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19: 63–99.

Smith, Neilson V. (1973). *The Acquisition of Phonology: A Case Study*. London: Cambridge University Press.

Smith, Rachel. (2004). Fine acoustic detail and context effects in spoken word recognition. Ph.D. dissertation, University of Cambridge.

—— and Hawkins, Sarah (2000). Allophonic influences on word-spotting experiments, in A. Cutler, J. McQueen, and R. Zondervan (eds.), *Proceedings of the Workshop on Spoken Word Access Processes (SWAP)*. Nijmegen: Max Planck Institute, 139–42.

Smolensky, Paul (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27: 720–31.

—— and Legendre, Géraldine (2006). *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, MA: MIT Press/Bradford Books.

Snedeker, Jesse and Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48: 103–30.

Snider, Keith L. (1990). Tonal upstep in Krachi: Evidence for a register tier. *Language* 66(3): 453–74.

—— (1998). Phonetic realisation of downstep in Bimoba. *Phonology* 15: 77–101.

Snoeren, Natalie D., Gaskell, M. Gareth, and Di Betta, Anna Maria (2009). The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35(2): 542–9.

—— Segui, Juan, and Hallé, Pierre A. (2008). Perceptual processing of partially and fully assimilated words in French. *Journal of Experimental Psychology: Human Perception and Performance* 34(1): 193–204.

So, Lydia K. H. and Dodd, Barbara (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language* 22: 473–95.

Soames, Scott (1984). Linguistics and psychology. *Linguistics and Philosophy* 7: 155–80.

Soderstrom, Melanie, Mathis, Donald, and Smolensky, Paul. (2006). Abstract genomic encoding of Universal Grammar in Optimality Theory, in P. Smolensky and G. Legendre. *The Harmonic Mind*. Cambridge, MA: MIT Press, 403–71.

Solé, Maria-Josep (2007). Controlled and mechanical properties in speech, in M.-J. Solé, P. S. Beddor, and M. Ohala (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 302–21.

Sommers, Mitchell S., Nygaard, Lynne C., and Pisoni, David B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America* 96: 1314–24.

van Son, Rob J. J. H. and Pols, Louis C. W. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America* 88: 1683–93.

—— —— (1992). Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America* 92: 121–7.

Song, Jae Yung and Demuth, Katherine (2008). Compensatory vowel lengthening for omitted coda consonants: A phonetic investigation of children's early representations of prosodic words. *Language and Speech* 51: 385–402.

—— Sundara, Megha, and Demuth, Katherine (2009). Phonological constraints on children's production of English third person singular *–s*. *Journal of Speech, Language, and Hearing Research* 52: 623–42.

Sorace, Antonella and Keller, Frank (2005). Gradience in linguistic data. *Lingua* 115(11): 1497–1524.

Speer, Shari R. and Ito, Kiwako (2009). Prosody in first language acquisition: Acquiring intonation as a tool to organize information in conversation. *Language and Linguistics Compass* 3: 90–110.

Spelke, Elizabeth S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology* 15: 626–36.

Spencer, John, Blumberg, Mark, McMurray, Bob, Robinson, Scott, Samuelson, Larissa, and Tomblin, J. Bruce (2009). Short arms and talking eggs: Why we should no longer abide the nativist-empiricist debate. *Child Development Perspectives* 3(2): 79–87.

Spinelli, Elsa, Welby, Pauline, and Schaegis, A. L. (2007). Fine-grained access to targets and competitors in phonemically ambiguous spoken sequences: The case of French elision. *Language and Cognitive Processes*. 22(6): 828–59.

Spivey, Michael J. (2007). *The Continuity of Mind*. New York: Oxford University Press.

—— Grosjean, Marc, and Knoblich, Günther (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences* 102(29): 10393–8.

Sproat, Richard and Fujimura, Osamu (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21: 291–311.

Stager, Christine L. and Werker, Janet F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388: 381–2.

Stampe, David (1979). *A Dissertation on Natural Phonology*. New York: Garland Press [Ph.D. dissertation, University of Chicago, 1973].

Stark, Rachel (1980). Stages of speech development in the first year of life, in G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (eds.), *Child Phonology I: Production*. New York: Academic Press.

Starreveld, Peter. (2000). On the interpretation of onsets of auditory context effects in word production. *Journal of Memory and Language* 42: 497–525.

Staum Casasanto, Laura (2008). Experimental investigations of sociolinguistic knowledge. Ph.D. dissertation, Stanford University, Palo Alto, CA.

Steedman, Mark (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31: 649–89.

Steele, Shirley (1986a). Interaction of vowel F0 and prosody. *Phonetica* 43: 92–105.

—— (1986b). Nuclear accent F0 peak location: Effects of rate, vowel and number of following syllables. *Journal of the Acoustical Society of America* 80: s51.

Steels, Luc (1995). A self-organizing spatial vocabulary. *Artificial Life* 2(3): 319–32.

—— (1997). The synthetic modeling of language origins. *Evolution of Communication* 1: 1–34.

Stemberger, Joseph P. (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua* 56: 235–59.

—— (1991). Radical underspecification in language production. *Phonology* 8: 73–112.

—— (1992). A performance constraint on compensatory lengthening in child phonology. *Language and Speech* 24: 207–18.

—— and Stoel-Gammon, Carol (1991). The underspecification of coronals: Evidence from language acquisition and performance errors, in J. Paradis and J.-F. Prunet (eds.), *The Special Status of Coronals: Internal and External Evidence*. New York: Academic Press, 181–99.

Stephens, Joseph D. W. and Holt, Lori L. (submitted). A standard set of American-English voiced stop consonant stimuli from morphed natural speech.

Steriade, Donca (1987). Redundant values, in *CLS 23: Parasession on Autosegmental and Metrical Phonology*. Chicago: Chicago Linguistic Society, 339–62.

—— (1993). Closure, release, and nasal contours, in M. Huffman and R. Krakow (eds.), *Nasals, Nasalization, and the Velum (Phonetics and Phonology 5)*. San Diego: Academic Press, 401–70.

—— (1995). Underspecification and markedness, in John Goldsmith (ed.), *Handbook of Phonological Theory*. Cambridge, MA: Blackwell, 114–74.

—— (1997). Lexical conservatism, in *Linguistics in the Morning Calm, Selected Papers from SICOL 1997*. Linguistic Society of Korea, Hanshin Publishing House, 157–79.

—— (1999a). Alternatives to syllable-based accounts of consonantal phonotactics, in O. Fujimura, B. D. Joseph, and B. Palek (eds.), *Proceedings of LP '98: Item Order in Language and Speech*, vol. 1. Prague: Karolinum Press, 205–45.

—— (1999b). Phonetics in phonology: The case of laryngeal neutralization, in M. Gordon (ed.), *Papers in Phonology 3*, UCLA Working Papers in Linguistics 2. Los Angeles: Department of Linguistics, University of California, 25–145.

Steriade, Donca (2000). Paradigm uniformity and the phonetics/phonology boundary, in M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press.

——(2001). Directional asymmetries in place assimilation, in E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*. San Diego: Academic Press, 219–50.

Stevens, Kenneth N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America* 50(4/2): 1180–E92.

——(1972). The quantal nature of speech: Evidence from articulatory-acoustic data, in E. E. David and P. B. Denes (eds.), *Human Communication: A Unified View*. New York: McGraw-Hill, 51–66.

——(1989). On the quantal nature of speech. *Journal of Phonetics* 17: 3–45.

——(1992). Lexical access from features. Speech Communication Group Working Papers, Research Laboratory of Electronics, MIT 8, 119–44.

——(1994). Phonetic evidence for hierarchies of features, in P. A. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, 242–58.

——(1999). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

——(2002). Towards a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America* 111: 1872–91.

——and Bickley, Corine A. (1991). Constraints among parameters simplifying control of a Klatt Formant Synthesizer. *Journal of Phonetics* 19: 161–74.

——and Blumstein, Sheila E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* 64: 1358–68.

——and Keyser, Samuel (1989). Primary features and their enhancement in consonants. *Language* 65: 81–106.

————(2010). Quantal theory, enhancement and overlap. *Journal of Phonetics* 38: 10–19.

Stockmal, Verna, Moates, Danny R., and Bond, Zinny S. (2000). Same talker, different language. *Applied Psycholinguistics* 21: 383–93.

Stockman, Ida J. (2006). Alveolar bias in the final consonant deletion patterns of African American children. *Language, Speech, and Hearing Services in Schools* 37: 85–95.

——and Vaughn-Cooke, Fay (1989). Addressing new questions about black children's language, in R. W. Fasold and D. Schiffrin (eds.), *Language Change and Variation*. Amsterdam: John Benjamins, 274–300.

Stoel-Gammon, Carol and Buder, Eugene H. (1999). Vowel length, post-vocalic voicing and VOT in the speech of two-year olds. *Proceedings of the 13th International Conference of Phonetic Sciences* 3: 2485–8.

——Williams, Karen, and Buder, Eugene (1994). Cross-language differences in phonological acquisition: Swedish and American /t/. *Phonetica* 51: 146–58.

Stone, Maureen (1995). How the tongue takes advantage of the palate during speech, in F. Bell-Berti and L. Raphael (eds.), *Producing Speech: Contemporary Issues: A Festschrift for Katherine Safford Harris*. New York: American Institute of Physics, 143–53.

——(2005). A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics* 19(6–7): 455–502.

——and Davis, Edward P. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. *Journal of the Acoustical Society of America* 98(6): 3107–12.

————Douglas, Andrew S., Aiver, Moriel N., Gullipalli, Rao, Levine, William S., and Lundberg, Andrew J. (2001). Modeling tongue surface contours from Cine-MRI images. *Journal of Speech, Language, and Hearing Research* 44: 1026–40.

Stone, Maureen, Faber, Alice, Rafael, Lawrence, and Shawker, Thomas (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *Journal of Phonetics* 20(2): 253–70.

—— and Lundberg, Andrew (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America* 99: 3728–37.

—— and Vatikiotis-Bateson, Eric (1995). Trade-offs in tongue, jaw, and palate contributions to speech production. *Journal of Phonetics* 23(1–2): 81–100.

Storkel, Holly L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research* 44: 1321–38.

—— (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language* 29: 251–74.

—— (2004). The emerging lexicon of children with phonological delays: Phonotactic constraints and probability in acquisition. *Journal of Speech, Language, and Hearing Research* 47: 1194–212.

—— (2006). Do children still pick and choose? The relationship between phonological knowledge and lexical acquisition beyond 50 words. *Clinical Linguistics and Phonetics* 20: 523–9.

—— (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language* 36: 291–321.

—— Armbruster, Jonna, and Hogan, Tiffany P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research* 49: 1175–92.

—— and Morrisette, Michele L. (2002). The lexicon and phonology: Interactions in language acquisition. *Language, Speech, and Hearing in Schools* 33: 24–37.

Story, Brad H. (2005). Synergistic modes of vocal-tract articulation for American English vowels. *Journal of the Acoustical Society of America* 118: 3834–59.

—— (2008). Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. *Journal of the Acoustical Society of America* 123: 327–35.

Strand, Elizabeth A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18: 86–99.

—— (2000). Gender stereotype effects in speech processing. Ph.D. thesis, Ohio State University, Columbus, OH.

—— and Johnson, Keith (1996). Gradient and visual speaker normalization in the perception of fricatives, in D. Gibbon (ed.), *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference, Bielfelt, October 1996*. Berlin: Mouton de Gruyter, 14–26.

Strange, Winifred (1995). Cross-language studies of speech perception: A historical review, in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Baltimore: York Press, 3–45.

—— (2007). Cross-language phonetic similarity of vowel. Theoretical and methodological issues, in O.-S. Bohn and M. Munro (eds.), *Language Experience in Second-language Speech Learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 35–55.

Strange, Winifred, Akahane-Yamada, Reiko, Kubo, R., Trent, S. A., and Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America* 109: 1692–704.

—— and Dittmann, S. (1984). Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception and Psychophysics* 36: 131–45.

—— Jenkins, James J., and Johnson, Thomas L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America* 74: 695–705.

Strauss, Anselm and Corbin, Juliet (1998). *Basics of Qualitative Research*. Thousand Oaks, CA and London: Sage Publications.

Streeter, Lynn A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America* 64: 1582–92.

Strevens, Peter (1960). Spectra of fricative noise in human speech. *Language and Speech* 3: 32–49.

Strobl, Carolin, Malley, James, and Tutz, Gerhard (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14: 323–48.

Strogatz, Steven H. and Stewart, Ian (1993). Coupled oscillators and biological synchronization. *Scientific American* 269(6), 102–9.

Stuart-Smith, Jane (1999). Glasgow: accent and voice quality, in P. Foulkes and G. J. Docherty (eds.), *Urban Voices*. London: Arnold, 203–22.

—— (2003). The phonology of modern urban Scots, in J. Corbett, J. D. McClure, and J. Stuart-Smith (eds.), *The Edinburgh Companion to Scots*. Edinburgh: Edinburgh University Press, 110–37.

—— (2006). The influence of media on language, in C. Llamas, P. Stockwell, and L. Mullany (eds.), *The Routledge Companion to Sociolinguistics*, London: Routledge, 140–8.

—— (2007a). A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 1307.

—— (2007b). Empirical evidence for gendered speech production: /s/ in Glaswegian, in J. Cole and J. I. Hualde (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 65–86.

—— and Timmins, Claire (2006). "Tell her to shut her moof": The role of the lexicon in TH-fronting in Glaswegian, in G. Caie, C. Hough, and I. Wotherspoon (eds.), *The Power of Words*. Amsterdam: Rodopi, 171–83.

—— —— and Tweedie, Fiona (2007). "Talkin' Jockney"?: Accent change in Glaswegian. *Journal of Sociolinguistics* 11: 221–60.

Studdert-Kennedy, Michael and Goldstein, Louis (2003). Launching language: The gestural origin of discrete infinity, in M. H. Christiansen and S. Kirby (eds.), *Language Evolution: The States of the Art*. Oxford: Oxford University Press.

—— Shankweiler, Donald, and Pisoni, David (1972). Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology* 3: 455–66.

Stump, Greg and Finkel, Raphael (2009). Principal parts and degrees of paradigmatic transparency, in J. P. Blevins and J. Blevins (eds.), *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press, 13–53.

Sugahara, Mariko and Turk, Alice (2009). Durational correlates of English sublexical constituent structure. *Phonology* 26: 477–524.

Summerfield, Quentin (1981). Differences between spectral dependencies in auditory and phonetic temporal processing: relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America* 72: 51–61.

Summers, W. Van (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America* 82(3): 847–63.

Sundara, Megha, Demuth, Katherine, and Kuhl, Patricia (2011). Sentence-position effects on children's perception and production of English 3rd person singular *-s*. *Journal of Speech, Language, and Hearing Research* 54: 55–71.

—— and Polka, Linda (2008). Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction. *Cognition* 106: 234–58.

—— —— and Baum, Shari (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism, Language and Cognition* 9: 97–114.

—— —— and Genesee, Fred (2006). Language experience facilitates discrimination of / d /–/ D/ in monolingual and bilingual acquisition of English. *Cognition* 100: 369–88.

—— —— and Molnar, Monica (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peer. *Cognition* 108: 232–42.

Sundberg, Ulla and Lacerda, Francisco (1999). Voice Onset Time in speech to infants and adults. *Phonetica* 56: 186–99.

Suomi, Kari (2007). On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics* 35(1): 40–55.

—— McQueen, James M., and Cutler, Anne (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language* 36(3): 422–44.

—— Toivanen, Juhani, and Ylitalo, Riikka (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics* 31: 113–38.

—— —— —— (2008). *Finnish Sound Structure: Phonetics, Phonology, Phonotactics, and Prosody*. University of Oulu, Finland: Studia Humaniora Ouluensia 9.

Surprenant, Aimée M. and Goldstein, Louis (1998). The perception of speech gestures. *Journal of the Acoustical Society of America* 104: 518–29.

Sussman, Harvey M. and Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception and Psychophysics* 58(6): 936–46.

Svantesson, Jan-Olof and House, David (2006). Tone production, tone perception and Kammu tonogenesis. *Phonology* 23: 309–33.

Swadesh, Morris (1971). *Origin and diversification of language.* Chicago: Aldine Atherton.

Swartz, B. E. (1998). Timeline of the history of EEG and associated fields. *Electroencephalography and Clinical Neurophysiology* 106: 173–6.

Sweet, Henry (1874). *History of English Sounds*. London: Trübner.

Swerts, Marc and Krahmer, Emiel (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics* 36(2): 219–38.

—— —— and Avesani, Cinzia (2002). Prosodic marking of intonation status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30: 629–54.

Swingley, Daniel and Aslin, Richard N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition* 76: 147–66.

—— —— (2007). Lexical competition in young children's word learning. *Cognitive Psychology* 54(2): 99–132.

Swingley, Pinto, John P., and Fernald, Anne (1998). Assessing the speed and accuracy of word recognition in infants, in C. Rovee-Collier, L. Lipsitt, and H. Hayne (eds.), *Advances in Infancy Research*, vol. 12. Stamford, CT: Ablex.

Syrdal, Ann K., Hirschberg, Julia, McGory, Julie, and Beckman, Mary E. (2001). Automatic ToBI prediction and alignment to speed manual labeling of prosody. *Speech Communication* 33: 135–51.

Szczepek-Reed, Beatrice (2006). *Prosodic Orientation in English Conversation*. Houndmills, Basingstoke: Palgrave Macmillan.

Tabain, Marija (1998). Non-sibilant fricatives in English: Spectral information above 10 kHz. *Phonetica* 55: 107–30.

—— Breen, Gavan, and Butcher, Andrew (2004). VC vs. CV syllables: A comparison of Aboriginal languages with English. *Journal of the International Phonetic Association* 34: 175–200.

Tabak, Wieke, Schreuder, Robert, and Baayen, R. Harald (2010). Producing inflected verbs: A picture-naming study. *The Mental Lexicon* 5: 22–46.

Tabossi, Patrizia (1996). Cross-modal semantic priming. *Language and Cognitive Processes* 11: 569–76.

Tagliamonte, Sally (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambrige University Press.

—— and Baayen, R. Harald (2010). Forests and trees of York English: *Was/were* variation as a case study for statistical practice. MS submitted for publication.

Takane, Yoshio, Young, Forrest W., and de Leeuw, Jan (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika* 42: 7–67.

Tamariz, Mónica (2008). Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *Mental Lexicon* 3(2): 259–78.

Tanenhaus, Michael K., Spivey-Knowlton, Michael J., Eberhard, Kathleen M., and Sedivy, Julie C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268: 1632–4.

—— and Trueswell, John C. (2005). Eye movements as a tool for bridging the language-as-product and language-as-action divide, in J. C. Trueswell and M. K. Tanenhaus (eds.), *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*. Cambridge, MA: MIT Press.

—— —— (2006). Eye movements and spoken language comprehension, in M. Traxler and M. Gernsbacher (eds.), *Handbook of Psycholinguistics*, 2nd edn. New York: Academic Press, Elsevier, 863–900.

Teeranon, Phanintra (2007). The plausibility of tonal evolution in the Malay dialect spoken in Thailand: Evidence from an acoustic study. *Taiwan Journal of Linguistics* 5(2): 45–64.

Telkemeyer, Silke, Rossi, Sonje, Koch, Stefan P., Nierhaus, Till, Steinbrink, Jens, Poeppel, David, Obrig, Hellmuth, and Wartenburger, Isabell (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience* 29: 14726–33.

Thiessen, Erik D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language* 56: 16–34.

—— Hill, Emily A., and Saffran, Jenny R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7: 53–71.

Thomas, Erik R. (2002a). Sociophonetic approaches of speech perception experiments. *American Speech* 77: 115–47.

—— (2002b). Instrumental phonetics, in J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell, 168–200.

Thompson, Evan, Palacios, Adrian, and Varela, Francisco (1992). Ways of coloring: Comparative color vision as a case study in cognitive science. *Behavioral and Brain Studies* 15: 1–74.

Thorn, Annabel S. C. and Frankish, Clive R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31: 729–35.

Tiede, Mark, Boyce, S. E., Holland, Christy, and Choe, K. Ann. (2004). A new taxonomy of American English /r/ using MRI and ultrasound. *Journal of the Acoustical Society of America* 115: 2633–4.

Tilsen, Samuel (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science* 33: 839–79.

Timcke, Rolf, von Leden, Hans, and Moore, Paul (1958). Laryngeal vibrations: Measurements of the glottic wave. *AMA Archives of Otolaryngology* 68: 1–19.

Tomasello, Michael (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA and London: Harvard University Press.

Torreira, Francisco, Adda-Decker, Martine, and Ernestus, Mirjam (2010). The Nijmegen Corpus of Casual French. *Speech Communication* 52(3): 201–12.

Toscano, Joseph, and McMurray, Bob (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science* 34(3): 434–64.

Trager, George L. and Smith, Henry L. (1951). *An Outline of English Structure*. Norman, OK: Battenburg Press.

Traunmüller, Hartmut (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America* 69: 1465–75.

Trehub, Sandra E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development* 47: 466–72.

Treiman, Rebecca (1989). The internal structure of the syllable, in G. N. Carlson and M. T. Tanenhaus (eds.), *Linguistic Structure in Language Processing*. Dordrecht: Kluwer, 27–52.

—— and Kessler, Brett (1995). In defense of an onset-rime syllable structure for English. *Language and Speech* 38: 127–42.

—— —— Knewasser, Stephanie, Tincoff, Ruth, and Bowman, Margo (2000). English speakers' sensitivity to phonotactic patterns, in M. B. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 269–82.

Tricomi, Elizabeth, Delgado, Mauricio R., McCandliss, Bruce D., McClelland, Jay L., and Fiez, Julie A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of Cognitive Neuroscience* 18: 1029–43.

Trochim, William and Donnelly, James P. (2006). *The Research Methods Knowledge Base*, 3rd edn. Cincinnati: Atomic Dog Publishing.

Trubetzkoy, Nicholai S. (1939/1969). *Grundzuge der Phonologie*. Travaux du Cercle Linguistique de Prague 7. [English translation: C. A. M. Baltaxe (trans.), *Principles of Phonology*. Berkeley, CA: University of California Press. 1969.]

Truckenbrodt, Hubert (1995). Phonological phrases: Their relation to syntax, focus, and prominence. Ph.D. dissertation, MIT, Cambridge, MA.

—— (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30: 219–55.

——— (2002). Upstep and embedded register levels. *Phonology* 19: 77–120.
——— (2007a). Upstep on edge tones and on nuclear accents, in T. Riad and C. Gussenhoven (eds.), *Tones and Tunes*, vol. 2. Berlin: Mouton de Gruyter, 349–86.
——— (2007b). The syntax-phonology interface, in Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, 435–56.
Trudgill, Peter (1974). *The Social Differentiation of English in Norwich*. Cambridge Studies in Linguistics 13. Cambridge: Cambridge University Press.
——— (1986). *Dialects in Contact*. Oxford: Blackwell.
Tsao, Feng-Ming, Liu, Huei-Mei, and Kuhl, Patricia K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development* 75: 1067–84.
——— ——— ——— (2006). Perception of native and non-native affricate-fricative contrasts: Cross language tests on adults and infants. *Journal of the Acoustical Society of America* 120: 2285–94.
Tse, John (1978). Tone acquisition in Cantonese: A longitudinal case study. *Journal of Child Language* 5: 191–204.
Tsukada, Kimiko, Birdsong, David, Bialystok, Ellen, Mack, Molly, Sung, H., and Flege, James E. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics* 33: 263–90.
——— Burnham, Denis, Luksaneeyanawin, Sudaporn, Krachaikiat, Niratasai, and Rungrojsuwan, Sorabud (2004). The effect of tone on vowel duration in Thai: A developmental study. Presentation at the 9th Conference on Laboratory Phonology.
Tucker, Benjamin V. (2007). Spoken word recognition of the reduced American English flap. Ph.D. dissertation, University of Arizona.
Tuller, Betty (2004). Categorization and learning in speech perception as dynamical processes, in M. A. Riley and G. C. Van Orden (eds.), *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*. National Science Foundation. <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp>.
——— Case, Pamela, Ding, Mingzhou, and Kelso, J. A. Scott (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* 20: 3–16.
Turk, Alice E. (2010). Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *Laboratory Phonology* 1(2): 227–62.
——— and Dimitrova, Snezhina (2007). English phrasal stress targets multiple, optional lengthening sites, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 1177–80.
——— and Sawusch, James (1997). The domain of accentual lengthening in American English. *Journal of Phonetics* 25: 25–41.
——— and Shattuck-Hufnagel, Stefanie (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics* 28: 397–440.
——— ——— (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35(4): 445–72.
——— and White, Laurence (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics* 27: 171–206.
Turvey, Michael T. (1977). Preliminaries to a theory of action with reference to vision, in R. Shaw and J. Bransford (eds.), *Perceiving, Acting and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Lawrence Erlbaum, 211–65.

Turvey, Michael T. (1990). Coordination. *American Psychologist* 45: 938–53.

Tversky, Amos and Gati, Itamar (1982). Similarity, separability, and the triangle inequality. *Psychological Science* 89: 123–54.

Uldall, Elizabeth T. (1971). Isochronous stresses in RP, in L. L. Hammerich, R. Jakobson, and E. Zwirner (eds.), *Forms and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jorgensen*. Copenhagen: Akademisk Forlag, 205–10.

Urberg Carlson, Kari, Kaiser, Eden, and Munson, Benjamin (2008). Assessment of children's speech production 2: Testing gradient measures of children's productions. Poster presented at the 2008 ASHA Convention, Chicago, 20–22. <http://www.tc.umn.edu/~munso005/Urberg-CarlsonEtAl_Final.pdf>, accessed June 14, 2009.

Vallabha, Gautam K., McClelland, James L., Pons, Ferran, Werker, Janet F., and Amano, Shigeaki (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences* 104: 13273–8.

Vampola, Tomáš, Horáček, Jaromír, and Švec, Jan G. (2008). FE modeling of human vocal tract acoustics. Part 1: Production of Czech vowels. *Acta Acustica* 94: 433–47.

Van den Broeke, Marcel and Goldstein, Louis (1980). Consonant features in speech errors, in V. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press.

Van Engen, Kristin J., Baese-Berk, Melissa, Baker, Rachel E., Choi, Arim, Kim, Midam, and Bradlow, Ann R. (2010). The Wildcat Corpus of Native- and Foreign-Accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech* 53(4), 510–40.

—— and Bradlow, Ann R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America* 121(1): 519–26.

Van Guilder, Linda. (2007). Cross-language perception in foreign name transcription. Ph.D. dissertation, Georgetown University.

Van Orden, Guy C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition* 15: 181–98.

Van Turennout, Miranda, Hagoort, Peter, and Brown, Colin M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23: 787–806.

Van Wijngaarden, Sander, Steeneken, Herman, and Houtgast, Tammo (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America* 111: 1906–16.

Vance, Timothy (1977). Tonal Distinctions in Cantonese. *Phonetica* 34: 93–107.

Vanrell, Maria del Mar (2007). A tonal scaling contrast in Majorcan Catalan interrogatives. *Journal of Portuguese Linguistics* (special issue on Prosody of Iberian Languages, ed. G. Elordieta and M. Vigário) 6(1): 147–78.

Vasquez-Alvarez, Yolanda and Hewlett, Nigel (2007). The "trough effect": An ultrasound study. *Phonetica* 64: 105–21.

Vaux, Bert (2008). Why the phonological component must be serial and rule-based, in B. Vaux and A. Nevins (eds.), *Rules, Constraints, and Phonological Phenomena*. Oxford: Oxford University Press, 20–60.

Vecera, Shaun and O'Reilly, Randall (1998). Figure-ground organization and object recognition processes: An interactive account. *Journal of Experimental Psychology: Human Perception and Performance* 24(2): 441–62.

van de Velde, Hans and van Hout, Roeland (1999). The pronunciation of (r) in Standard Dutch. *Linguistics in the Netherlands* 16: 177–88.

Vennemann, Theo (1974). Phonological concreteness in Natural Generative Grammar, in R. Shuy and C.-J. N. Bailey (eds.), *Toward Tomorrow's Linguistics*. Washington DC: Georgetown University Press, 202–19.

Vesalius, Andreas (1543). *De humani corporis fabrica and [The] epitome [of Andreas Vesalius]*, Basel. *Epitome* trans. L. R. Lind. Cambridge, MA: MIT Press, 1969.

Vigário, Marina (2003). *The Prosodic Word in European Portuguese*. Berlin and New York: Mouton de Gruyter.

—— (2009). The Prosodic Word Group as a domain of prosodic hierarchy. Paper given at the Sixth Old Conference in Phonology (OCP6), University of Edinburgh.

—— and Frota, Sónia (2003). The intonation of Standard and Northern European Portuguese. *Journal of Portuguese Linguistics* (special issue on Portuguese Phonology, ed. W. L. Wetzels), 2(2): 115–37.

Vihman, Marilyn M. (1993). Variable paths to early word production. *Journal of Phonetics* 21: 61–82.

—— (1996). *Phonological Development: The Origins of Language in the Child*. Oxford: Blackwell.

—— and Croft, William. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics* 45: 683–725.

—— Macken, Marlys, Miller, Ruth, Simmons, Hazel, and Miller, Jim (1985). From babbling to speech: A reassessment of the continuity issue. *Language* 61: 397–445.

—— and Nakai, Satsuki (2003). Experimental evidence for an effect of vocal experience on infant speech perception, in M.-J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 1017–20.

—— Thierry, Guillaume, Lum, Jarrad, Keren-Portnoy, Tamar, and Martin, Pam (2007). Onset of word form recognition in English, Welsh and English-Welsh bilingual infants. *Applied Psycholinguistics* 28: 475–93.

—— and Velleman, Shelley L. (1989). Phonological reorganization: A case study. *Language and Speech* 32: 149–70.

—— —— (2002). The optimal initial state. Unpublished manuscript available on the Rutgers Optimality Archive, <http://roa.rutgers.edu/index.php3>.

Vio, Monique and Colas, Annie (2006). Pitch cues for the recognition of yes-no questions in French. *Journal of Psycholinguistic Research* 35(5): 427–45.

Vitevitch, Michael S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech* 40: 211–28.

—— (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 28(4): 735–47.

—— and Luce, Paul A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science* 9: 325–9.

—— and Rodríguez, Eva (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders* 3: 64–73.

Viviani, Paolo (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects, in K. Kowler (ed.), *Eye Movements and their Role in Visual and Cognitive Processes: Reviews of Oculomotor Research*, vol. 4. Amsterdam: Elsevier, 353–93.

Vogel, Irene, Bunnell, H. Timothy, and Hoskins, Steven (1995). The phonology and phonetics of the Rhythm Rule, in B. Connell and A. Arvaniti (eds.), *Phonology and*

Phonetic Evidence: Papers in Laboratory Phonology IV. Cambridge: Cambridge University Press, 111–27.

Volaitis, Lydia E. and Miller, Joanne L. (1992). Phonetic prototypes: influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America* 92: 723–35.

Vroomen, Jean, van Zon, Monique, and de Gelder, Beatrice (1996). Cues to speech segmentation: Evidence from juncture misperceptions and wordspotting. *Memory and Cognition* 24: 744–55.

Waals, Juliette (1999). *An Experimental View of the Dutch Syllable*. The Hague: Holland Academic Graphics (UiL OTS/Utrecht University Dissertation). [LOT International Series, 18.]

Wade, Travis and Holt, Lori. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *Journal of the Acoustical Society of America* 118: 2618–33.

Wagner, Michael (2002). The role of prosody in laryngeal neutralization, in A. Csirmaz, Z. Li, A. Nevins, O. Vaysman, and M. Wagner (eds.), *Phonological Answers (and their Corresponding Questions)*. MITWPL 42: 373–92.

Wahba, Grace (1990). *Spline Models for Observational Data*. Philadephia: Society of Industrial and Applied Mathematics.

Walker, Rachel (1999). Guaraní voiceless stops in oral versus nasal contexts: An acoustical study. *Journal of the International Phonetic Association* 29: 63–94.

—— Byrd, Dani, and Mpiranya, Fidèle (2008). An articulatory view of Kinyarwanda coronal harmony. *Phonology* 25: 499–535.

Wang, D. Marilyn and Bilger, Robert C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America* 54: 1248–66.

Wang, H. Samuel and Derwing, Bruce L. (1994). Some vowel schemas in three English morphological classes: Experimental evidence, in M. Y. Chen and O. C. L. Tzeng (eds.), *In Honor of Professor William S.-Y. Wang: Interdisciplinary Studies on Language and Language Change*. Taipei: Pyramid Press, 561–75.

Wang, Ying Wai (2006). Contextual tonal variation and pitch targets in Cantonese. *Proceedings of Speech Prosody 2006. Dresden, Germany*, 317–20.

Wang, Yue, Spence, Michelle, Jongman, Allard and Sereno, Joan (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America* 106: 3649–58.

Wanrooij, Karin (2009). Does distributional input improve the categorization of speech sounds? Neurobiological aspects and computer simulations. MA thesis, University of Amsterdam.

Ward, Gregory and Hirschberg, Julia (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61(4): 747–76.

Warner, Natasha (2003). Rapid perceptibility as a factor underlying universals of vowel inventories, in A. Carnie, H. Harley, and M. Willie (eds.), *Formal Approaches to Function in Grammar*. Amsterdam: John Benjamins, 245–61.

—— (2011). Reduction, in M. van Oostendorp, C. Ewen, E. Hume, and K. Rice (eds.), *The Blackwell Companion to Phonology*. Malden, MA & Oxford: Wiley-Blackwell.

—— and Arai, Takayuki (2001). The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America* 109: 1144–56.

——Brenner, Dan, Woods, Anna, Tucker, Benjamin V., and Ernestus, Mirjam (2009). Were we or are we? Perception of reduced function words in spontaneous conversations. *Journal of the Acoustical Society of America* 125: 2655 (abstract).

——Fountain, Amy, and Tucker, Benjamin V. (2009). Cues to perception of reduced flaps. *Journal of the Acoustical Society of America* 125: 3317–27.

——————(forthcoming). Cues to perception of reduced flaps. *Journal of the Acoustical Society of America*.

——Jongman, Allard, Sereno, Joan, and Kemps, Rachèl (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32: 251–76.

Warren, Donald W. (1982). Aerodynamics of speech, in N. Lass, L. McReynolds, J. Northern, and D. Yoder (eds.), *Speech, Language, and Hearing* 1. Philadelphia: W. B. Saunders, 219–44.

Warren, Paul (2005). Issues in the study of intonation in language varieties. *Language and Speech* 48(4): 345–58.

——Hay, Jennifer, and Thomas, Brynmor (2007). The loci of sound change effects in recognition and perception, in J. Cole and J. Hualde (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 87–112.

——Speer, Shari, and Schafer, Amy (2003). Wanna-contraction and prosodic disambiguation in US and NZ English. *Wellington Working Papers in Linguistics* 15: 31–50.

Warren, Richard M. (1999). *Auditory Perception: A New Analysis and Synthesis*. Cambridge: Cambridge University Press.

Wassink, Alicia B. (2001). Theme and variation in Jamaican vowels? *Language Variation and Change* 13(2): 135–59.

——Wright, Richard A., and Franklin, Amber D. (2007). Speaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics* 35: 363–79.

Waterson, Natalie (1971). Child phonology: A prosodic view. *Journal of Linguistics* 7: 179–211.

Watson, Duane G., Gunlogson, Christine A., and Tanenhaus, Michael K. (2006). Online methods for the investigation of prosody, in S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schlieer (eds.), *Methods in Empirical Prosody Research*. New York: Walter de Gruyter, 259–82.

——Tanenhaus, Michael K., and Gunlogson, Christine A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science* 32: 1232–44.

Watson, Ian (1990). Acquiring the voicing contrast in French: A comparative study of monolingual and bilingual children, in J. N. Green and W. Ayers-Bennett (eds.), *Variation and Change in French: Essays presented to Rebecca Posner on the Occasion of her Sixtieth Birthday*. London: Routledge, 37–60.

Watson, Kevin (2007). Liverpool English. *Journal of the International Phonetics Association* 37(3): 351–60.

Watt, Dominic and Fabricius, Anne (2002). Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1~F2 plane. *Leeds Working Papers in Linguistics and Phonetics* 9: 159–73.

——and Ingham, Catherine (2000). Durational evidence of the Scottish Vowel Length Rule in Berwick English, Leeds Working Papers in Linguistics and Phonetics 8, 205–28.

WATTS, DUNCAN J. (2002). A simple model of global cascades. *Proceedings of the National Academy of Sciences of the United States of America* 99(9): 5766–71.

—— and DODDS, PETER S. (2007). Infuentials, networks, and public opinion formation. *Journal of Consumer Research* 34: 441–58.

WAUQUIER-GRAVELINES, SOPHIE (2003). Troncation et reduplication: Peut-on parler de gabarits morphologiques dans le lexique précoce?, in B. Fradin, G. Dal, M. Hathout, F. Kerleroux, M. Roché, and M. Plénat (eds.), *Les unités morphologiques. Silexicales* 3. Lille: Université de Lille III.

WAYLAND, RATREE (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics* 18: 345–73.

—— and GUION, SUSAN (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report, *Language Learning* 54: 681–712.

—— —— LANDFAIR, DAVID, and LI, BIN (2006). Native Thai speakers acquisition of English word stress patterns. *Journal of Psycholinguistic Research* 35: 285–304.

—— and LI, BIN (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics* 36: 250–67.

WEBB, ANDREW (1999). *Statistical Pattern Recognition*. London: Arnold (Newnes).

WEBER, ANDREA (2008). What eye movements can tell us about spoken-language processing: A psycholinguistic survey, in C. M. Riehl (ed.), *Was ist linguistische Evidenz: Kolloquium des Zentrums Sprachenvielfalt und Mehrsprachigkeit*, November 2006. Aachen: Shaker, 57–68.

—— and CUTLER, ANNE (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50: 1–25.

WEDEL, ANDREW (2004). Self-organization and categorical behavior in phonology. Ph.D. thesis, University of California at Santa Cruz.

—— (2006). Exemplar models, evolution and language change. *The Linguistic Review* 23: 247–74.

—— (2007). Feedback and regularity in the lexicon. *Phonology* 24: 147–85.

—— and VOLKINBURG, HEATHER (2009). Modeling simultaneous convergence and divergence of linguistic features between differently-identifying groups in contact. MS.

WEIDE, ROBERT (1995). The Carnegie Mellon Pronouncing Dictionary (cmudict), technical report version 1.4, November 8, 1995.

VAN DE WEIJER, JOOST (1998). Language input for word discovery. Doctoral dissertation, Max Planck Institute for Psycholinguistics, Nijmegen.

WEINREICH, URIEL, LABOV, WILLIAM, and HERZOG, MARVIN (1968). Empirical foundations for a theory of language change, in W. Lehmann and Y. Malkiel (eds.), *Directions for Historical Linguistics*. Austin: University of Texas Press, 97–195.

WEISMER, GARRY, DINNSEN, DANIEL A., and ELBERT, MARY A. (1981). A study of the voicing distinction associated with omitted, word-final stops. *Journal of Speech and Hearing Disorders* 46: 320–7.

WELBY, PAULINE. (2003). The slaying of Lady Mondegreen, being a study of the association and alignment of French intonational rises and their role in speech segmentation. Ph.D. dissertation, Ohio State University.

—— (2004). The structure of French intonational rises: A study of text-to-tune alignment, in B. Bel and I. Marlien (eds.), *Proceedings of the Conference on Speech Prosody 2004*, Nara, Japan, March 23–26, 127–30. ISCA archive, <http://www.isca-speech.org/archive/sp2004>.

——— (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34(3): 343–71.

——— and Loevenbruck, Hélène (2006). Anchored down in Anchorage: Syllable structure and segmental anchoring in French. *Italian Journal of Linguistics* 18(1) (special issue: Current Issues in Tonal Alignment, ed. M. D'Imperio): 74–124.

Wells, John C. (1982). *Accents of English*, vol 1. Cambridge: Cambridge University Press.

Werker, Janet F., Cohen, Leslie B., Lloyd, Valerie, Casasola, Marianella, and Stager, Christine L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology* 34(6): 1289–309.

——— and Curtin, Suzanne (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development* 1(2): 197–234.

——— Fennell, Christopher T., Corcoran, Kathleen M., and Stager, Christine L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy* 3: 1–30.

——— Gilbert, J. V. H., Humphrey, K., and Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development* 52: 349–55.

——— and Polka, Linda (1993). Developmental changes in speech perception: New challenges and new directions. *Journal of Phonetics* 21: 83–101.

——— ——— and Pegg, Judith E. (1998). The conditioned head-turn procedure as a method for testing infant speech perception. *Early Development and Parenting* 6 (3–4): 171–8.

——— Pons, Ferran, Dietrich, Christiane, Kajikawa, Sachiyo, Fais, Laurel, and Amano, Shigeaki (2006). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition* 103(1): 147–62.

——— and Stager, Christine (2000). Developmental changes in infant speech perception and early word learning: Is there a link?, in M. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 181–93.

——— and Tees, Richard C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63.

——— ——— (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustic Society of America* 75(6): 1866–78.

——— ——— (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology* 50: 509–35.

West, Paula (1999). Perception of distributed coarticulatory properties in English /ℓ/ and /ɹ/. *Journal of Phonetics* 27: 405–26.

Westbury, John R. (1994). X-ray microbeam speech production database user's handbook, version 1.0, Madison, WI, <http://www.medsch.wisc.edu/~milenkvc/pdf/ubdbman.pdf>.

Westermann, Gert and Miranda, Eduardo R. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language* 89: 393–400.

Whalen, Doug H. (1990). Coarticulation is largely planned. *Journal of Phonetics* 18: 3–35.

——— Abramson, Arthur S., Lisker, Leigh, and Mody, Maria (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica* 47(12): 36–49.

——— ——— ——— ——— (1993). F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America* 93: 2152–60.

Whalen, Doug H., Best, Catherine T., and Irwin, Julia R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics* 25: 501–28.

—— Gick, Bryan, Kumada, Masanobu, and Honda, Kiyoshi (1999). Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *Journal of Phonetics* 27(2): 125–42.

—— Iskarous, Khalil, Tiede, Mark, Ostry, David, Lehnert-LeHouillier, Heike, Vatikiotis-Bateson, Eric, and Hailey, Donald (2005). The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research* 48(3): 543–53.

—— and Levitt, Andrea G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23(3): 349–66.

—— —— and Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language* 18: 501–16.

Wheeldon, Linda R. and Levelt, Willem J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language* 34: 311–34.

—— and Morgan, Jane L. (2002). Phoneme monitoring in internal and external speech. *Language and Cognitive Processes* 17: 503–35.

—— and Waksler, Rachelle (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language* 90(1–3): 401–12.

Wheeler, Max W. (2005). Voicing contrast: Licensed by prosody or licensed by cue? MS, University of Sussex, <http://roa.rutgers.edu/view.php3?roa=769> [ROA-769].

White, Katherine S. and Morgan, James L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language* 59: 114–32.

—— Peperkamp, Sharon, Kirk, Cecilia, and Morgan, James L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition* 107: 238–65.

White, Laurence (2002). English speech timing: A domain and locus approach. Ph.D. dissertation, University of Edinburgh.

—— and Mády, Katalin (2008). The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian, in P. A. Barbosa et al. (eds.), *Proceedings of the Fourth Conference on Speech Prosody 2008*, May 6–9, Campinas, Brazil, 363–6.

—— and Mattys, Sven L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35: 501–22.

—— and Turk, Alice E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics* 38: 459–71.

Wichmann, Anne (2000). *Intonation in Text and Discourse*. London: Longman.

—— (2008). Speech corpora and spoken corpora, in A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics: An International Handbook*. Berlin: Mouton de Gruyter, 187–206.

—— and Cauldwell, Richard (2003). Wh-questions and attitude: The effect of context, in A. Wilson, P. Rayson, and T. McEnery (eds.), *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*. Frankfurt am Main: Peter Lang.

Wickelgren, Wayne A. (1965). Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America* 38: 583–8.

—— (1966). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America* 39: 388–98.

Wiener, Norbert (1948). *Cybernetics.* New York: John Wiley and Sons.

Wightman, Colin W. and Ostendorf, Mari (1994). Automatic labeling of prosodic patterns. *IEEE Transactions on Audio, Speech and Language Processing* 2(4): 469–81.

——Shattuck-Hufnagel, Stefanie, Ostendorf, Mari, and Price, Patti (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91: 1707–17.

Willerman, Raquel (1994). The phonetics of pronouns: Articulatory bases of markedness. Ph.D dissertation, University of Texas, Austin.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics* 26: 95–105.

Williams, Sarah and Hammarberg, Bjorn (1998). Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics* 19: 295–333.

Wilshire, Carolyn E. and Nespoulous, Jean-Luc (2003). Syllables as units in speech production: Data from aphasia. *Brain and Language* 84: 724–47.

Wilson, Colin (2003). Experimental investigation of phonological naturalness, in *WCCFL 22*, ed. G. Garding and M. Tsujimura. Somerville, MA: Cascadilla Press, 533–46.

——(2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 3: 945–82.

——and Obdeyn, Marieke (2009). Simplifying subsidiary theory: Statistical evidence from Arabic, Muna, Shona, and Wargamay. MS, Johns Hopkins University.

Wilson, Ian (2007). The effects of post-velar consonants on vowels in Nuu-chah-nulth: Auditory, acoustic, and articulatory evidence. *Canadian Journal of Linguistics* 52(1/2): 43–70.

Winkler, István, Kujala, Teija, Alku, Paavo, and Näätänen, Risto (2003). Language context and phonetic change detection. *Cognitive Brain Research* 17: 833–44.

————Tiitinen, Hannu, Sivonen, Päivi, Alku, Paavo, Lehtokoski, Anne, Czigler, István, Csépe, Valéria, Ilmoniemi, Risto J., and Näätänen, Risto (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36: 638–42.

Wodzinski, Sylvie, Frisch, Stefan, and Stearns, Adrienne (2007). Update of research on velar stop consonant production. Talk presented at Ultrafest IV, New York University, September 2007. <http://jerome.linguistics.fas.nyu.edu/presentations/Ultrafest_IV_WFS_2007.pdf>, accessed March 15, 2009.

Wolfram, Walt (1991). *Dialects and American English*. Englewood Cliffs, NJ: Prentice Hall.

——and Thomas, Erik (2002). *The Development of African American English*. Oxford: Blackwell.

Wolfson, Nessa (1976). Speech events and natural speech: Some implications for sociolinguistic methodology. *Language in Society* 5: 189–209.

Wood, Sidney A. J. (1996). Assimilation or coarticulation: Evidence from the temporal coordination of tongue gestures for the palatalization of Bulgarian alveolar stops. *Journal of Phonetics* 24: 139–64.

Wrembel, Magdalena (2007). The impact of voice quality resetting on the perception of a foreign accent in third language acquisition, in A. S. Rauber, M. A. Watkins, and B. O. Baptista (eds.), *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*. Florianópolis, Brazil: Federal University of Santa Catarina, 481–91.

Wrench, Alan and Scobbie, James M. (2006). Spatio-temporal inaccuracies of video-based ultrasound images of the tongue, in H. C. Yehia, D. Demolin, and R. Laboissiere (eds.), *Proceedings of the 7th International Seminar on Speech Production*. São Paulo, Brazil.

————(2008). High-speed cineloop ultrasound vs. video ultrasound tongue imaging: Comparison of front and back lingual gesture location and relative timing, in R. Sock,

S. Fuchs, and Y. Laprie (eds.), *Proceedings of the 8th International Seminar on Speech Production*. Strasbourg, France, 57–60.

Wright, James T. (1986). The behavior of nasalized vowels in the perceptual vowel space, in J. Ohala and J. J. Jaeger (eds.), *Experimental Phonology*. Orlando: Academic Press, 45–67.

Wright, Richard A. (2001). Perceptual cues in contrast maintenance, in E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*. San Diego, CA: Academic Press, 251–77.

—— (2003). Factors of lexical competition in vowel articulation, in R. Local, R. Ogden, and R. Temple (eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 75–87.

—— (2004). A review of perceptual cues and cue robustness, in B. Hayes, R. Kirchner, and D. Steriade (eds.), *Phonetically Based Phonology*. Cambridge: Cambridge University Press, 34–57.

Wright, Susan and Kerswill, Paul (1989). Electropalatography in the analysis of connected speech processes. *Clinical Linguistics and Phonetics* 3: 49–57.

Xu, Yi (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America* 95(4): 2240–53.

—— (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61–83.

—— (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.

—— (1999). F0 peak delay: When, where and why it occurs, in J. Ohala (ed.), *Proceedings of the 14th International Congress of Phonetic Sciences*, 1881–4.

—— (2001). Fundamental frequency peak delay in Mandarin. *Phonetica* 58: 26–52.

—— (2002). Articulatory constraints and tonal alignment, in B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference*, Aix-en-Provence, Laboratoire Parole et Langage, 91–100.

—— (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics* 5: 757–97.

—— (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220–51.

—— and Sun, Xuejun (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111: 1399–413.

—— and Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33: 319–37.

Xu, Yi, and Xu, Ching X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159–97.

Yamada, Reiko A. (1995). Age and acquisition of second language speech sounds: Perception of American English /p/ and /l/ by native speakers of Japanese, in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press, 305–20.

—— Strange, Winifred, Magnuson, J. S., Pruitt, J. S., and Clarke III, W. D. (1994). The intelligibility of Japanese speakers' productions of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. *Proceedings of the International Conference of Spoken Language Processing*. Yokohama: Acoustical Society of Japan, 2023–6.

Yip, Moira (1989). Contour tones. *Phonology* 6: 149–74.

—— (1992). Tonal register in East Asian languages, in H. van der Hulst and K. Snider, *The Phonology of Tone: The Representation of Tonal Register*. Berlin: Mouton de Gruyter, 245–68.

——(1995). Tone in East Asian languages, in J. Goldsmith (ed.), *Handbook of Phonological Theory*. Oxford: Blackwell, 476–94.
——(2001). The complex interaction of tones and prominence, in M. Kim, and U. Strauss (eds.), *Proceedings of NELS 31*. U. Mass Amherst: G.L.S.A.
——(2002). *Tone*. Cambridge: Cambridge University Press.
——(2007). Tone, in P. de Lacy (ed.), *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, 229–52.
Yoon, Taejin (2007). A predictive model of prosody through grammatical interface: A computational approach. Ph.D. dissertation, University of Illinois.
Yoon, Yeo Bom and Derwing, Bruce L. (2001). A language without a rhyme: Syllable structure experiments in Korean. *Canadian Journal of Linguistics* 46: 187–237.
Young, Robert and Morgan, William (1980, 1987) *The Navajo Language*. Albuquerque: University of New Mexico Press.
Young, Steve, Evermann, Gunnar, Gales, Mark, Hain, Thomas, Kershaw, Dan, Liu, Xunying (Andrew), Moore, Gareth, Odell, Julian, Ollason, Dave, Povey, Dan, Valtchev, Valtcho, and Woodland, Phil (2002). *The HTK Book*. Cambridge: Cambridge University Engineering Department.
Yu, Alan C. L. (2004). Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80: 73–97.
——(2006). Tonal effects on perceived vowel duration. Presentation at the 10th Conference on Laboratory Phonology, Paris.
Yuan, Jiahong (2004). Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling. Doctoral dissertation, Cornell University, Ithaca.
——Brenier, Jason M., and Jurafsky, Dan (2005). Pitch accent prediction: Effects of genre and speaker, in *Proceedings of EUROSPEECH-05*. Lisbon, September, 1409–12.
——Shih, Chilin, and Kochanski, Greg (2002). Comparison of declarative and interrogative intonation in Chinese. *Speech Prosody 2002*, Aix-en-Provence, France, 711–14.
Zamuner, Tania S. (2009). Phonological probabilities at the onset of language development: Speech production and word position. *Journal of Speech, Language, and Hearing Research* 52: 49–60.
——Gerken, LouAnn, and Hammond, Michael (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language* 31: 515–36.
——and Ohala, Diane K. (1999). Preliterate children's syllabification of intervocalic consonants, in A. Greenhill, H. Littlefield, and C. Tano (eds.), *Proceedings of BUCLD 23*. Somerville: Cascadilla Press, 753–63.
Zangl, Renate and Fernald, Anne (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language Learning and Development* 3: 199–231.
Zanone, Pier-Giorgio and Kelso, J. A. Scott (1997). Coordination dynamics of learning and transfer: Collective and component levels. *Journal of Experimental Psychology: Human Perception and Performance* 23: 1454–80.
Zec, Draga (2007). The syllable, in P. de Lacy (ed.), *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, 161–94.
Zeng, Fan-Gang (2009). Phoneme perception experiment. University of California, Irvine, Speech and Hearing Research Laboratory, <http://www.ucihs.uci.edu/hesp/webtest/PHONEME/phonememain.htm>, accessed March 13, 2009.
Zerbian, Sabine (2007). Phonological phrasing in Northern Sotho (Bantu). *The Linguistic Review* (special issue on Prosodic Phrasing, ed. S. Frota and P. Prieto) 24: 233–62.

Zhang, Jialu, Lü, Shinan, and Qi, Shiqian (1982). A cluster analysis of the perceptual features of Chinese speech sounds. *Journal of Chinese Linguistics* 10: 190–206.

Zhang, Jie (2002). *The Effects of Duration and Sonority on Contour Tone Distribution: A Typological Survey and Formal Analysis*. New York: Routledge.

—— (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution, in B. Hayes, R. Kirchner, and D. Steriade (eds.), *Phonetically Based Phonology*. Cambridge: Cambridge University Press, 157–90.

Zhang, Jinsong and Hirose, Keikichi (2004). Tone nucleus modeling for Chinese lexical tone recognition. *Speech Communication* 42(3–4): 447–66.

Zhang, Qing (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society* 34: 431–66.

—— (2008). Rhotacization and the "Beijing Smooth Operator": The social meaning of a linguistic variable. *Journal of Sociolinguistics* 12: 201–22.

Zhang, Qingfang and Damian, Markus F. (2009). The time course of segment and tone encoding in Chinese spoken production: An event-related potential study. *Neuroscience* 163: 252–65.

—— —— and Yang, Yufang (2007). Electrophysiological estimates of the time course of tonal and orthographic encoding in Chinese speech production. *Brain Research* 1184: 234–44.

—— and Yang, Yufang (2007). Electrophysiological estimates of the time course of semantic and metrical encoding in Chinese speech production. *Neuroscience* 147: 986–95.

Zhang, Yang, Kuhl, Patricia K., Imada, Tokiashi, Iverson, Paul, Pruitt, John, Stevens, Erica B., Kawakatsu, Masaki, Tohkura, Yoh'ichi, and Nemoto, Iku (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage* 46: 226–40.

Zhang, Zaoyan, Mongeau, Luc, and Frankel, Steven H. (2002). Broadband sound generation by confined turbulent jets. *Journal of the Acoustical Society of America* 112(2): 677–89.

Zhao, Yuan and Jurafsky, Dan (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics* 37: 231–47.

Zharkova, Natalia and Hewlett, Nigel (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/. *Journal of Phonetics* 37: 248–56.

—— —— and Hardcastle, William (2008). An ultrasound study of lingual coarticulation in children and adults, in R. Sock, S. Fuchs, and Y. Laprie (eds.), *Proceedings of the Eighth International Seminar on Speech Production*. Strasbourg, France, December 8–12, 161–4.

Ziegler, Johannes C., Pech-Georgel, Catherine, George, Florence, Alario, F.-Xavier, and Lorenzi, Christian (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences* 102: 14410–15.

Zipf, George K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

Zsiga, Elizabeth C. (1995). An acoustic and electropalatographic study of lexical and postlexical palatalization in American English, in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 282–302.

——— (1997). Features, gestures, and Igbo vowels: An approach to the phonology/phonetics interface. *Language* 73: 227–74.
——— (2000). Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics* 28: 69–102.
——— and Nitisaroj, Rattima (2007). Tone features, tone perception, and peak alignment in Thai. *Language and Speech* 50: 343–83.
——— Gouskova, Maria, and Tlale, One (2006). On the status of voiced stops in Tswana: Against *ND, in C. Davis, A. R. Deal, and Y. Zabbal (eds.), *NELS 36: Proceedings of the 36th Annual Meeting of the North East Linguistic Society*. Amherst, MA: GLSA.
Zubritskaya, Katya (1997). Mechanism of sound change in Optimality Theory. *Language Variation and Change* 9: 121–48.
Zuraw, Kie (2000). Patterned exceptions in phonology. Ph.D. dissertation, UCLA.
——— (2007). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. *Language* 83: 277–316.
Zwitserlood, Pienie and Marslen-Wilson, William D. (1989). The locus of the effects of sentential-semantic contexts in spoken-word processing. *Cognition* 32: 25–64.

This page intentionally left blank

Index