Turbulent Sounds

Interface Explorations 21

Editors
Artemis Alexiadou
T. Alan Hall

De Gruyter Mouton

Turbulent Sounds

An Interdisciplinary Guide

edited by

Susanne Fuchs
Martine Toda
Marzena Żygis

De Gruyter Mouton

ISBN 978-3-11-022657-7
e-ISBN 978-3-11-022658-4
ISSN 1861-4167

Library of Congress Cataloging-in-Publication Data

Turbulent sounds : an interdisciplinary guide / edited by Susanne Fuchs, Martine Toda, Marzena Żygis.
p. cm. – (Interface explorations; 21)
Includes bibliographical references and index.
ISBN 978-3-11-022657-7 (hardcover : alk. paper)
1. Consonants. I. Fuchs, Susanne, 1969– II. Toda, Martine. III. Żygis, Marzena.
P235.5.I58 2010
414–dc22
2010010449

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Cover image: iStockphoto/Thinkstock
Printing: Hubert & Co. GmbH & Co. KG, Göttingen
∞ Printed on acid-free paper

Printed in Germany

www.degruyter.com

Preface

Turbulent chaotic behaviour occurs all around us in natural systems as a function of a number of various factors, and is often thought to show a system out of balance. In the physical world, turbulence makes one think of whirlwinds, hurricanes, thunderstorms and white water rapids. In the human world, what comes to mind are the global economic crisis at the beginning of the 21st century, revolutions, conflicts and other more personal moments when the otherwise smooth flow of life is thrown into turmoil. However, in this book we are not dealing with disasters or traumas, but rather with certain linguistic events that remind us that turbulence can be both positive and a type of balance or equilibrium. We focus on the characteristics of a class of frequently occurring speech sounds called obstruents. The term 'obstruent' refers to consonants that involve the partial or total obstruction of the vocal tract by the lips, teeth or tongue. Such obstructions disrupt a smooth airflow or block it entirely, and thereby generate a particular timbre of sound, one which lets obstruents stand out in speech and thereby function effectively, since they are audibly different from vowels and 'sonorant' consonants.

A total obstruction tends to cause a silence or a quasi-silence in the acoustic signal (with a glottal murmur in voiced segments). This occurs in oral stops, affricates, clicks, ejectives, and implosives. The former two are purely pulmonic sounds, where air comes from the lungs, and are made turbulent particularly on the release of the total closure of the vocal tract. The latter three are generally nonpulmonic sounds, where airflow is not dependent on the lungs.

A partial obstruction occurs throughout fricatives, in the slowed release phase of affricates, and during pre- or post- aspiration. The partial obstruction in the vocal tract and the glottal configuration together create the aerodynamic conditions that make a turbulent airflow arise near the tightest point of constriction, which is either at the glottis or at a supraglottal constriction. Noise realized at the glottis is commonly called aspiration, whereas noise at the supraglottal constriction is called frication.

In a given language it is generally well known which turbulent sounds are produced within which environments. However, the actual individual realisation of turbulent sounds is largely unpredictable, since a variety of changing factors come into play, making turbulent sounds one of the most challenging and interesting areas to study. On account of the difficulty in investigating turbulent sounds, considerable courage (or naivety) is called for on the part of the researcher, since it is necessary to apply knowledge from several different research disciplines. Moreover, the researcher has to face the fact that generally accepted acoustic parameters describing, for instance, fricatives, do not exist. This fact makes it rather difficult to know which acoustic parameters are perceptually relevant. None of the comparable classes of speech sounds requires so much basic knowledge of phonology (different sound inventories, places of articulation), acoustics (e.g. noise sources), aerodynamics (e.g. the relation between subglottal, intraoral, and atmospheric pressure), speech production (e.g. laryngeal-oral coordination) etc. as is required for obstruents. An optimal research team examining turbulent sounds should consist of a phonologist, a phonetician, two physicists (one from the area of acoustics and the other one from the area of fluid mechanics), a speech therapist (working on cleft palate speech, hearing impairment or neurological disorders), a psychologist (working on speech acquisition), a dentist (working on teeth size and different bites), and a mathematician (working on models). Since most of us are probably far from such an optimal research environment, this book is intended to bridge a gap by introducing the reader to the world of obstruents from a multidisciplinary perspective, with a particular focus on the phonetics and phonology of these sounds. Moreover, this multidisciplinary perspective involves the description of typological processes as well as detailed studies of various phenomena occurring in unrelated and even endangered languages: Germanic languages (German, Scottish English), Slavic languages (Polish and Slovak), Khoesan languages (N|uu-endangered), Caucasian languages (Avar, Georgian, Ingush, Bezhta-endangered, Tsez-endangered), Finno-Ugric languages (Hungarian), and Korean. Most of the chapters in this book that relate to a particular language use acoustic analyses in addressing the relevant research questions. Additionally, electropalatography, magnetic resonance imaging, ultrasound, and laryngography were chosen to investigate speech production phenomena of obstruents at the laryngeal and supralaryngeal levels. Aerodynamic characteristics were studied using the Rothenberg mask and a first attempt was made to do acoustic modelling of obstruent phenomena.

The book is structured as follows: It starts with a typological study by *Hall and Zygis,* featuring an overview of the phonology of obstruents in a variety of languages using the traditional features [sonorant], [continuant], and [strident]. It is shown how these features are (un)able to capture commonly occuring natural classes and phonological processes involving obstruents. Moreover, the authors take into account cross-linguistic tendencies with respect to frequent and infrequent processes.

A rich set of phonological processes occurring in obstruents are introduced from a radically different perspective in the chapter “Turbulence and phonology” by *Ohala and Solé*. The authors explain phonological processes as a consequence of aerodynamic principles, acoustic-auditory factors, interarticulatory timing as well as coordination, and summarize their own related work over the last decades.

Bárkányi and Kiss’ study approaches the puzzling phonological behaviour of the voiced labiodental fricative /v/ in Slovak and Hungarian by appealing to the phonetic properties of this sound. The authors pursue the question of whether the aerodynamic and acoustic properties of /v/ correlate with its double-faced (sonorant vs. obstruent) behaviour in voicing assimilation as well as its distribution in the lexicon.

Kim, Maeda, Honda and Hans tackle another issue, addressing the laryngeal characterization of the Korean alveolar lenis and fortis fricatives. Based on their acoustic and aerodynamic (air flow and intraoral pressure) experiment and a separate MRI study they conclude that the two fricatives are specified for [-spread glottis]. Moreover, the authors consider the aspiration noise occurring after the offset of the two fricatives as a consequence of the transition between a fricative and a vowel or a vowel and a fricative, regardless of the phonation types of the fricatives.

The next three chapters deal with languages and phenomena which are rather underrepresented in the scientific literature. Preaspiration is a rarely reported phonetic phenomenon occurring mostly in stops, not in fricatives. *Gordeeva and Scobbie* find evidence for preaspiration as a correlate of word-final voice in Scottish English fricatives. They explain it as a learnt dissociation of the lingual and supralaryngeal gestures in word-final voiceless fricatives, show how different speakers vary in the extent of this dissociation, and suggest that this variable (but quite persistent) inter-articulatory event helps to secure contrast in the Scottish English voice system with its tendency for final phonetic devoicing.

Based on their fieldwork data, *Grawunder, Simpson, and Khalilov* study phonetic characteristics of ejectives, and provide samples from Caucasian languages. These also include data from speakers of two endangered languages: Tsez and Bezhta. The authors discuss the (ir)relevance of a variety of acoustic parameters in the description of ejectives in comparison to secondary features or pulmonic stops in these languages.

Miller presents her fieldwork data on clicks from another endangered language: N|uu, spoken by only a few speakers in South Africa. Miller challenges current knowledge on clicks, in particular with respect to different air stream mechanisms, to front and back place of articulation and to vocalic environment, shaping the production of these sounds.

The last three chapters introduce selected phenomena occurring in obstruents from the perspective of sociophonetics, pathology, and physics / acoustic modelling.

It has often been shown that voiceless sibilants are realized differently by male and female speakers in a variety of languages. On the basis of a large articulatory, acoustic, and anatomical data set from English and German speakers, *Fuchs and Toda* investigate the following question: Do differences in male versus female /s/ reflect biological or sociophonetic factors? Their findings provide evidence that the sociophonetic factor plays a pivotal role, but results are also influenced by biology.

Obstruents are typically distorted in children or adults with a cleft palate, since the air can escape through the cleft and a high intraoral pressure cannot build up. In their chapter "Producing turbulent speech sounds in the context of cleft palate," *Gibbon and Lee* provide a comprehensive overview of compensatory mechanisms used by cleft palate speakers to satisfy the aerodynamic and perceptual goals in obstruent production. These compensatory mechanisms have a large impact on speech production, since they are often persistent, even after the subjects have undergone surgery to close the cleft palate.

Finally, *Toda, Maeda, and Honda* develop a combed portrait of the formant structure of sibilant fricatives and their affiliation to vocal tract cavities. The topic is introduced with simple vocal tract models, followed by recorded Polish /ç/ and /ṣ/ configurations derived from high-resolution, teeth-inserted MRI data. The chapter concludes with a note on the acoustic mechanisms involved in the phonemic contrast.

We would like to thank the German Research Council and the French-German University Saarbrücken for a grant to support the French-German cooperation, and also all the supporters of the conference "On turbulences," which took place in 2005 at ZAS in Berlin. This conference provided a fundamental basis for the ideas given in this book. We also would like to thank the editor in chief of the Interface Explorations Series, Tracy Alan Hall, for his engagement, the publisher Mouton de Gryuter, and all the reviewers who contributed to this book. This book is dedicated to our children Jolanda, Louise, Victor, Emilia, Weronika and Marysia, who not

only realize various turbulent sounds of German, French, and Polish, but also create lots of positive turbulences in our lives.

Susanne Fuchs, Martine Toda, Marzena Żygis

Contents

An overview of the phonology of obstruents[*]

T. A. Hall and Marzena Żygis

1. Introduction

Sounds in the languages of the world involving turbulent noise are referred to in generative phonology as ‘obstruents’, a natural class subsuming stops, affricates and fricatives. Sounds not belonging to the class of obstruents are traditionally considered to be ‘sonorants’, namely vowels, glides, liquids, nasals.

The purpose of this article is to provide an overview of the phonology of obstruents, concentrating on the three subcategories ‘stops’, ‘affricates’ and ‘fricatives’ and the rules in natural languages which change a sound from one of those categories to a sound in one of the other two. We begin in section 2 by discussing the distinctive feature used for capturing the sonorant vs. obstruent dimension ([sonorant]). In section 3 and 4 we turn to the features used to express the contrast between stops vs. affricates ([strident]) and stops vs. fricatives ([continuant]), respectively. In each section we discuss traditional definitions of the respective features as well as rules in a wide variety of languages which change the values of one of these features, e.g. obstruent → sonorant, stop → affricate, stop → fricative (and the respective mirror image rules). In section 5 we provide examples of processes which are problematic in terms of features but which can be explained if phonetic evidence (in addition to the phonological features) is taken into account. The aim of that section is to show the limits of feature theory – at least in its current state.

Although a number of works have appeared through the years which analyse the kinds of rules we discuss below, a number of controversial issues relating to feature theory remain. The purpose of this article is not to make new proposals but instead to discuss some of these open questions and to point to directions for further research, some of which are dealt with in other articles in the present volume (see e.g. Ohala and Solé).

2. The obstruent vs. sonorant dimension

In this section we discuss the definition of the distinctive feature [sonorant], posited in generative phonology to capture the distinction between obstruents and sonorants (section 2.1.) and rules which involve a change in the feature [sonorant] (section 2.2.). Some controversial questions involving the material we present below are discussed in section 2.3.

2.1. The definition of the feature [sonorant]

The two natural classes 'obstruent' and 'sonorant' are traditionally captured in phonological theory with the feature [sonorant], which Chomsky and Halle (1968; henceforth SPE) define as follows (p. 302): "Sonorants are sounds produced with a vocal tract cavity configuration in which spontaneous voicing is possible; obstruents are produced with a cavity configuration that makes spontaneous voicing impossible." SPE (p. 302) notes that spontaneous voicing is suppressed by "…narrowing the air passage to a point where the rate of flow is reduced below the critical value needed for the Bernoulli effect to take place." This occurs in segments whose constrictions are more radical than those found in glides like /j w/. This means that stops, affricates and fricatives – those sounds formed with more radical constrictions than the glides – are considered to be obstruents, i.e. [–sonorant], whereas vowels, glides, nasals and liquids are sonorant, i.e. [+sonorant].[1] The reader is referred to the Appendix for a list of the features we discuss in the present article.[2]

According to a second definition (Halle and Clements 1983: 6) sonorant sounds "are produced with a vocal tract configuration sufficiently open [so] that the air pressure inside and outside the mouth is approximately equal. Obstruent sounds are produced with a vocal tract configuration sufficient to increase the air pressure inside the mouth significantly over that of the ambient air." Apparently both the latter definition and the one from SPE make the same predictions concerning natural classes. Since the Halle and Clements's (1983) definition of [sonorant] seems to be the most widely accepted one among phonologists, it is the definition we adopt in the present study.[3]

It is usually assumed that the definition for obstruents correlates with their phonological behaviour in the sense that sounds like /p t k/, etc., are obstruents from the point of view of phonetics (i.e. they fit one or both of the definitions given above), but also from the point of view of phonology by patterning as [–sonorant] segments. A number of linguists have pointed to segments in various languages in which this correlation does not hold. One logical possibility involves sounds which are phonetically obstruents but which nevertheless pattern phonologically as if they were sonorants, while the mirror image situation obtains if phonetic sonorants pattern phonologically as obstruents. In the remainder of this section we briefly discuss examples illustrating the former case.

Examples of phonetic obstruents patterning with sonorants are discussed by Rice (1993), who dubs such sounds 'sonorant obstruents'. These are defined as obstruents (from the point of view of phonetics) that (a) take the place of sonorants in a system, (b) receive voicing from sonorants, or (c) alternate with sonorants. An example of the (c) case can be found in languages in which voiced stops like [b d g] alternate with nasals. (It is not clear whether (a), (b) or (c) alone is sufficient to define a sound as a 'sonorant obstruent', or whether all act together, and must all be met). Rice ultimately argues that the feature [sonorant] should be replaced with the feature [sonorant voice] (SV), which is underlyingly present in all sonorants and sonorant obstruents.[4]

Additional examples of surface obstruents which behave phonologically like sonorants involve the fricative [v] in a number of languages. See, for example, Barkaï and Horvath (1978) on Hebrew and Hungarian, Padgett (2002a) on Russian, Hall (2007) on German, and Bárkányi and Kiss (present volume) on Hungarian and Slovak. While it could be that the surface [v] in these languages really is a sonorant (e.g. the approximant [ʋ]), another possibility is that it is an obstruent (i.e. the fricative [v]), which needs to be analysed as a sonorant in the phonology. We therefore regard language-specific studies on the phonetics and phonology of [v] as a profitable area of future research because these studies have the potential of shedding light on features (e.g. [sonorant]) as well as the general interaction of phonetics and phonology. (Cf. also phonetic investigation of [v] in Padgett (2002a) and Bárkányi and Kiss (present volume)).

2.2. Rules altering [sonorant]

Rules altering [sonorant] subsume processes in which a sonorant becomes an obstruent (section 2.2.1.) or ones in which an obstruent becomes a sonorant (section 2.2.2.). We refer to these two types of processes as desonorizations and sonorizations, respectively. In contrast to some of the other changes discussed in sections 3 and 4, desonorizations and sonorizations typically affect individual segments rather than the entire natural class of obstruents or sonorants. As we note below, one general problem in investigating processes of this type is that they will often involve not only a change in sonorancy but also changes in other features (e.g. [consonantal], [continuant]). The theoretical relevance of this point is discussed in section 2.3.

Many of the processes we discuss in the present section (and in the subsequent sections) are referred to in the literature under the umbrella-category of lenitions – a term often meant to imply a reduction of ‘articulatory effort’. The literature on lenitions is vast, and we do not attempt to summarize it here. The reader is referred to Kirchner (2001) (including the references cited therein) for a recent phonetically-based analysis of lenitions, which also includes a cross-linguistic survey. Lenitions from the historical perspective are discussed in Hock (1986: 80ff.).

2.2.1. Sonorants becoming obstruents

The change from a sonorant into an obstruent will typically affect either a glide (e.g. /w/ or /j/) or a rhotic such as /r/.

Examples of desonorizations converting /r/ into a coronal fricative before high front vocoids (i.e. /j/ or /i/) are presented in (1a-b).[5] We refer to this process as the assibilation of a rhotic because the output is a sibilant (fricative), but also because the high, front vowel context is the same as the trigger for stop assibilations (see sections 3.2.1. and 4.2.1.).

(1) Assibilation of rhotics
a. r → s / __ j i   Jita (*)   Downing (2005)
b. r > s / __ i   Oceanic   Ross (1988)

In the Bantu language Jita in (1a), a stem-final /r/ (phonetically the flap [ɾ]) changes to [s] before /i j/ in certain morphological contexts. The change in (1b) transpired in a number of Oceanic languages, e.g. Proto-East Admiralties *[r] > [s] in Titan and Sori-Harengan (Ross 1988: 324).
Examples of the assibilation of rhotics which apply anywhere other than before high front vocoids are difficult to come by. One example to our

knowledge is the process in Tarasco, whereby /r r^h/ become [z] before /z/ (Swadesh 1969).

Desonorization rules with a glide (i.e. /w/ or /j/) as input are presented in (2-3). In the former examples we can observe the change from the input glide to a fricative; in the latter examples the output is a homorganic stop. The context for (2a) will be described below.[6]

(2) Desonorization of glides (in which the output is a fricative)

a. w → β / (V) __ i	Guayabero	Keels (1985: 75)
b. w → β / # __ C	Sawai	Whisler (1992: 13)
c. j → ʒ / $_{\sigma}$[__	Porteño Spanish	Harris (1983: 57-58)
d. w → v / __ i	Bagri	Gusain (2000: 10)

Keels (1985: 75) writes about the process in (2a): "The semivowel /w/ is realized as a voiced bilabial fricative…when it occurs following a front or high central vowel or preceding a stressed front vowel." Harris (1983: 57) gives alternations from Porteño Spanish (see 2c) like *convo[j]* ([j] = [i] in Harris's transcription) 'convoy' vs. *convo[ʒ]es* 'convoys'. In Bagri (see 2d) /w/ changes to [v] before front vowels, e.g. /wiwwa/ is pronounced as [vıwwa] 'marriage' (Gusain 2000: 10).

An anonymous reviewer points out that another example of the desonorization of glides might be Dutch [w], which is usually a labio-dental approximant, but which is typically realized as [v] before [r] (in onsets), e.g. *wreed* 'cruel' then becomes homophonous with *vreet* 'devour'. [w] may even become voiceless [f] in this position in some dialects.

We hypothesize that processes like the ones in (2) will typically occur in syllable-initial position, although this hypothesis needs to be tested by examining additional languages with this type of process. If correct, the syllable-initial context suggests that desonorizations occur because a less sonorant sound is preferred in the syllable onset. See Vennemann's (1988) 'Head Law', in which the author discusses segmental changes like the desonorizations in (2), which show a strengthening from glides to fricatives in the onset.

A final set of examples illustrating desonorization is presented in (3). These examples differ from the ones in (2) in that the input glide surfaces as a homorganic stop. In Pawnee (3a), /w/ shifts to a homorganic stop in word-initial position or after an obstruent. (/p/ is the only labial stop in the language). The example from Cypriot Greek in (3b) illustrates the change from a palatal glide (i.e. /j/) to a homorganic stop (transcribed in the original source as [k̯]) after certain consonants (i.e. /p f θ v ð/).

(3) Desonorization of glides (in which the output is a stop)

a. w → p / {#, [–son]} __V	Pawnee	Parks (1976: 50)
b. j → c / C __V	Cypriot Greek	Newton (1972a: 53)

In all of the examples provided by Parks (1976) and Newton (1972a) the respective desonorizations process takes place before a vowel. Hence, the two changes in (3) provide further support that desonorization tends to occur in syllable-initial position in conformance with the Head Law of Vennemann (1988).

In the examples presented in (1)-(3) we have only considered desonorizations of a rhotic or a glide, but other hypothetical desonorizations have not been presented, e.g. the desonorization of a nasal or a lateral to an obstruent. We have also only presented examples in (1)-(3) in which the output is a fricative or a stop, but not an affricate. We are not aware of processes like these excluding a few optional processes which we are not discussing in this article and tend to classify them as very infrequent or accidental.

2.2.2. Obstruents becoming sonorants

The output of sonorizations is typically a glide ([w]) or rhotic (e.g. [r] or [ɾ]) and the typical context is intervocalic or coda position. Representative examples of sonorizations in intervocalic position are provided in (4).

(4) Sonorizations (in intervocalic position)

a. β → w / V __ V	Kirgiz	Herbert and Poppe (1963)
b. v → w / V __ V	Turkish	Underhill (1976)
c. t d → ɾ / V __ V	Am. English	Kahn (1976)
d. k → ɾ / V __ V	Ibibio	Urua (2004)
e. z > r / V __ V	Latin	Catford (2001)
f. g → w / (u) __ (u)	Biri	Terrill (1998: 9)

In American English (see 4c) /t d/ also change to [ɾ] if the stops are preceded by a word boundary (as in *meeṯ Anne*). In the word-internal context in (4c) there is a condition that the second vowel cannot be stressed

(Kahn 1976). A similar process is found in Ibibio (4d) in which the intervocalic context also conditions the change of /k/ to the voiced velar approximant [ɰ] or optionally to the voiced uvula or tap (transcribed as [R] by Urua 2004: 106). In Biri /g/ changes to [w] before or after /u/ (Terrill 1998: 9).

Sonorizations in anything other than intervocalic position are presented in (5). In (5a-d) the context is either word- or syllable-final position and in (5e) it is before a consonant.

(5) Sonorizations (in other contexts)

a. p → w / __ {C, #}	Lama (*)	Ourso and Ulrich (1990)
b. ʒ → j / __ #	Haitian Creole	Tinelli (1981)
c. v → w / NUC __	Slovak	Short (1993)
d. ɖ → ˀɭ̰ / __]σ	Mambay	Anonby (2006)
e. t͡ʃ ʃ (s) → j / __C	Arbore	Hayward (1984)

In Lama, spoken in northern Togo, a morphologically-conditioned process converts /p/ to [w] in coda position before a consonant and word-finally, e.g. kpa/p/su is realized as kpa[w]su 'to reconcile' and ya/p/ as ya[w] 'buy, imp.' (Ourso and Ulrich 1990: 136). In the Slovak examples in (5c) /v/ changes to [w] after a syllable nucleus (NUC), which can be a vowel or a sonorant consonant, e.g. pra/v/da is pronounced as pra[w]da 'truth' and kr/v/ný as kr[w]ný 'blood, adj.' (Short 1993: 536).[7] In Mambay (Niger-Congo), the retroflex /ɖ/ changes to a preglottalized, creaky voiced [ˀɭ̰] in syllable-final position (Anonby 2006). In the Arbore language sibilants change to a glide before a consonant, e.g. wara/t͡ʃ/té surfaces as wara[j]té 'hyena, fem.' (Hayward 1984: 65).

A number of languages are attested in which obstruents (or some subset thereof) are realized as laryngeals (i.e. either [h] or [ʔ]) in coda position, e.g. in Spanish dialects /s/ surfaces as [h] in the rhyme (Harris 1983). Such processes are usually referred to as debuccalizations because they involve the loss of an oral constriction. If laryngeals like /h/ are analysed as [+sonorant] (recall note 1), then debuccalizations like the one from Spanish could be added to the list of sonorizations in (5). However, as we pointed out in that note, the characterization of sounds like /h/ as [+sonorant] is not clear.

We hypothesize that the context for sonorizations in (5) makes sense from the point of view of syllable markedness: A number of linguists have observed that languages prefer sonorants to obstruents in the syllable coda (see, for example, Vennemann's 1988 'Coda Law' and much subsequent work in Optimality Theory). In our view the processes in (5) are triggered by a requirement that codas contain sonorant sounds. This analysis of the processes in (5) begs the question of what motivates the sonorizations in (4), since these processes occur in the onset and not in the coda. This is a question that exceeds the goals of the present study because it would require that one takes a closer look at the phonology of the individual languages. Our preliminary hypothesis is that onset sonorizations like the ones in (4) are triggered not by a syllable-based requirement, but instead by a requirement that intervocalic sounds weaken; see, for example, the approach taken by Kirchner (2001), which relies on the constraint LAZY.

2.3. Some open questions

Although all of the processes in (1-5) look superficially like they change a [αsonorant] sound into the corresponding [–αsonorant] one, it is not always clear whether or not this is the correct analysis of the respective rule because it is usually the case that other features are changing as well. With respect to the assibilation of rhotics in (1), it could be that the change to [–sonorant] is a consequence of the addition of the feature [+strident] (see section 3.1.), which itself is assigned if the change from /r/ to [s] has its phonetic origin in a brief period of frication which occurs at the point when a rhotic is released into a front vocoid (see our explanation for stop assibilations in section 5). Consider now (3a) and (3b). Since glides like /w j/ are [+continuant] and stops like /p c/ are [–continuant], one could argue that (3a-b) involve a change in continuancy alone and that the change in sonorancy is not a part of this rule at all, but instead is the consequence of an independent default rule saying that [+continuant] segments are [+sonorant].[8]

Since the correct answer to the questions discussed above depends on language-internal arguments for distinctive features and default rules, we do not take a stand here and instead leave the issue open. These featural issues are further complicated by the fact that they need to be evaluated with respect to phonetically based treatments of sonorizations and desonorizations which eschew traditional features like [sonorant]. See, for example the treatments proposed in Lavoie (2001) and Kirchner (2001), which suggest that the traditional features are inadequate.

A related question is whether or not there are clear examples in which the feature [sonorant] assimilates in some process. One might argue that some of the processes in (4) involve the assimilation of an input obstruent to a sonorant in the neighbourhood of [+sonorant] vowels, but some of these changes involve features in addition to [sonorant]. For example, the process in (4f) seems also to involve a change in [continuant] and [consonantal]. A more obvious example of the type of assimilatory process we are describing here would have a segment in the input (an obstruent [X]) and output (a sonorant [Y]) which differ in sonorancy alone. The process in (4a) appears to be such an example. To confirm this analysis one would need to examine the phonology of the input segment /β/ to determine if it is in fact a [–sonorant] segment. This is a question we leave open for further study.

Should there be no examples at all of rules which involve the spreading of [sonorant], then one might want to implement the suggestion proposed by McCarthy (1988) that [sonorant] (like [consonantal]) is a part of the root node itself and that it therefore cannot display autosegmental behaviour such as spreading. The analysis of [consonantal] in this fashion (i.e. as a part of the root node) stands in contrast to the proposal made by Kaisse (1992) that [consonantal] can spread.

3. The stridency dimension

In this section we discuss first (in section 3.1.) the distinctive feature [strident], posited in generative phonology to capture the distinction between sibilant vs. non-sibilant fricatives (e.g. /s/ vs. /θ/) and stops vs. affricates (e.g. /t/ vs. /t͡s/). In section 3.2. we consider rules which manipulate the feature [strident].

3.1. The feature [strident]

We employ the feature [strident] to capture the contrast between two sets of fricatives: (a) non-sibilant (coronal) fricatives (i.e. /θ ð/) vs. sibilant (coronal) fricatives (e.g. /s z/) and (b) palatals (e.g. /ç ʝ/) from postalveolars (e.g. /ʃ ʒ/). The use of the feature [strident] for these fricatives is illustrated in the matrices in Table 1. Note that all sounds traditionally described as 'sibilants' are [+strident] (including all sibilant affricates; see below).[9] We follow earlier writers who see [strident] as being only distinctive for coronal sounds (e.g. Lahiri and Evers 1991, Shaw 1991, Hall 1997).

Table 1. Features for strident and non-strident fricatives.

	ɸ	f	θ	s	ʃ	ç	x	χ
	β	v	ð	z	ʒ	ʝ	ɣ	ʁ
[sonorant]	–	–	–	–	–	–	–	–
[continuant]	+	+	+	+	+	+	+	+
[CORONAL]			✓	✓	✓	✓		
[strident]			–	+	+	–		
[LABIAL]	✓	✓						
[DORSAL]							✓	✓

On this view labials, velars and uvulars are not specified for [strident] at all. We leave open which feature is necessary to contrast bilabials from labiodentals. For approaches to [strident] in which non-coronals are specified for [strident] as well, the reader is referred to Jakobson, Fant, and Halle (1952), SPE and Hume (1992).[10]

A commonly assumed definition (SPE, p. 329) says that "strident sounds are marked acoustically by greater noisiness than their non-strident counterparts." Jakobson, Fant and Halle (1952 [1967: 23]) state that strident sounds are represented "by a random distribution of black areas" in spectrograms. This is in contrast to non-strident sounds which show a more regular waveform distribution. (For a more detailed phonetic specification of [strident] and its relevance for perceptually driven processes see Żygis submitted and Padgett and Żygis 2007).

We adopt the view that [strident] distinguishes oral stops from the corresponding affricates, e.g. /t/ vs. /t͡s/. For a defence of the analysis of affricates as 'strident stops' the reader is referred to Jakobson, Fant and Halle (1952), LaCharité (1993), Rubach (1994), Clements (1999) and Kehrein (2002). According to all of these authors both affricates and stops have in common that they are [–sonorant, –continuant] and differ in terms of stridency: stops are [–strident] and affricates [+strident]. This approach to [strident] is captured in Table 2.

Table 2. Strident stop analysis of affricates.

	t	t͡s
[sonorant]	–	–
[continuant]	–	–
[strident]	–	+

An apparent problem for the features in Table 2 is that they cannot account for non-strident affricates (e.g. /t͡θ/ in Tahltan; see section 3.2.3.). However, as Kehrein (2002) shows, no language contrasts a non-strident fricative with a plain stop, i.e. /t/ and /t͡θ/. Significantly, in languages with /t/ and /t͡θ/ or /p/ and /p͡f/, there is always a minor place contrast which distinguishes the two sounds. For example, in German /p/ is bilabial and /p͡f/ is labiodental and in Shilluk /t/ is lamino-alveolar and /t͡θ/ is apico-dental (Kehrein 2002: 27). This means that in a language like Shilluk, /t/ and /t͡θ/ are both [–sonorant, –continuant, –strident] and the feature [distributed] distinguishes the stop from the affricate. We do not commit ourselves here to the feature(s) required to distinguish /p/ from /p͡f/ and /x/ from /k͡x/, but we assume that it is a (minor) place feature. By contrast, a place feature cannot distinguish coronal stops like /t/ from coronal affricates like /t͡s/ because coronal affricates uncontroversially belong to the natural class of strident sounds.[11]

There is yet another issue concerning the feature [strident] which needs to be addressed at this point. While both values of [sonorant] and [continuant] define natural classes, this is not so clear with the feature [strident]. The class of [+strident] sounds is uncontroversial (i.e. this is the natural class which conditions the selection of the plural allomorph [ɪz] in English), but is the natural class of [–strident] sounds as well documented? Since this is a question that can only be answered by taking an in depth examination of the phonology of individual languages, we leave it open for further study.

3.2. Rules altering [strident]

Rules altering [strident] subsume processes in which a stop becomes an affricate (e.g. /t/ → [t͡s]; section 3.2.1.), or an affricate becomes a stop (e.g. /t͡s/ → [t]; section 3.2.2.). Other possible changes involving stridency are discussed in section 3.2.3.

3.2.1. Stops becoming affricates

Processes which convert a stop to a strident affricate (or fricative) (i.e. /t/ → [t͡s] or /t/ → [s]) are referred to henceforth as stop assibilations. Stop assibilations with an affricate as the output are provided in (6). Note that this type of process is typically triggered before a high vocoid such as [j i u].

(6) Stop assibilations (with an affricate as the output)

a. t d → t͡s d͡z/ __ i y j y ɥ	Quebec French	Kim (2001)
b. t d > t͡s d͡z/ __ j	Old Polish	Carlton (1991)
c. t > t͡s/ __ j	Latin	Pope (1952: 129)
d. t t^h→ t͡s t͡s^h/ __ (h) i	Korean (*)	Kim (2001)
e. t → t͡s / __ i u	Maori	Bauer and Parker (1993)

In Korean (i.e. 6d), stem-final plosives /t t^h/ are affricated when followed by derivational or inflectional suffixes beginning with /(h)i/, e.g. before the nominative suffix /i/: /mat+i/ → [ma.d͡zi] 'first child', and before the causative suffixes /i/ and /hi/: /mut+hi+ta/ → [mut͡s^{h}i.da] 'to be buried' (Kim 2001: 89). In Maori (i.e. 6e) /t/ is produced with frication before a devoiced final /i/ or /u/ in unstressed syllables.[12]

It is not at all clear how the processes in (6) should be analysed in terms of the features posited in Table 1 (see Clements 1999 for discussion). While it is certainly possible to say that the input coronal stop becomes [+strident], this approach cannot explain why this particular feature is assigned in the high vowel context. What this suggests to us is that a complete explanation for stop assibilations like the ones in (6) requires reference to various phonetic parameters; see section 5 for discussion.

We refer here to processes which convert a stop (or a fricative) to an affricate in a context other than before a (high) front vocoid as affrications. Examples are presented in (7). Process (7b) applies phrase-initially and -finally and in slow speech. The diachronic process in (7c) affected the three voiceless stops in Bavarian and Alemannic dialects of Old High German. (7c) also affected /p t k/ in other environments, e.g. after a nasal.[13]

(7) Affrication (of stops)

a. p p^h b → p͡ɸ p͡ɸ^h b͡β / __ u	Lahu	Luschützky (1992)
b. p k → p͡ɸ k͡x / # __	Kunimaipa	Pence (1966: 60)
c. p t k > p͡ɸ t͡s k͡x / # __	Old H. German	Braune (1967)

Another example of affrications can be found in Polish (Rubach 1994). In that language the coronal stops /t d/ surface as [t͡s d͡z] before strident fricatives and affricates (i.e. [s z t͡s d͡z]). Note that this type of process is easily analysable phonologically as the spreading of [+strident] from [s z t͡s d͡z] to a preceding /t/.

The examples in (8) illustrate the process of palatalization, by which coronal or dorsal stops surface as postalveolar affricates.

(8) Palatalizations

a. t d → t͡ʃ d͡ʒ / __ front V	Tera	Tench (2007)
b. t d → t͡ʃ d͡ʒ / __ i ɪ	Braz. Portug.	Barbosa and Albano (2004)
c. t → t͡ʃ / i __	Basque	Hualde (1991)
d. t d → t͡ʃ d͡ʒ / __ j	Ilocano	Rubino (1997)
e. k g → t͡ʃ d͡ʒ / __ e i	Cassubian (*)	Stone (1993)

In Tera, spoken in Nigeria, coronal stops /t d/ are palatalized to [t͡ʃ d͡ʒ] before front vowels, e.g. /ti/ 'to stir' is pronounced as [t͡ʃi] and /kudi/ 'chief' as [kud͡ʒi] (Tench 2007: 227). According to Hualde (1991: 46) the process in (8c) holds in the Arbizu dialect of Basque. The same rule shifts /l n/ to [ʎ ɲ]. In Ilocano, /t d/ convert to [t͡ʃ d͡ʒ] before /j/, e.g. /diak/ is pronounced as [d͡ʒak] 'I don't' and /buttiog/ as [but.t͡ʃog] 'large abdomen' (Rubino 1997: 15). Process (8e) takes place in the declension of nouns and adjectives if the declensional suffixes start with /i/ or /e/, e.g. [rek] / [ret͡ʃi] 'crab' nom sg. / nom.pl or [długó] / [dłud͡ʒi] fem.nom.sg. / masc.nom.sg. (Stone 1993: 767).

A typical output for palatalizations is a (laminal) postalveolar sound, i.e. [t͡ʃ d͡ʒ], although many languages are attested in which the output is alveolopalatal [t͡ɕ], e.g. Mandarin (Duanmu 2000), Polish (Rubach 1984). In some languages the output is alveolar, e.g. the sound change k g > t͡s d͡z/_e i in Proto-Slavic (Carlton 1991). However, it is sometimes not clear whether or not changes of this type involved a single step, i.e. from velar to alveolar, or two steps, i.e. k g > t͡ʃ d͡ʒ/ __ e i followed by t͡ʃ d͡ʒ > t͡s d͡z/_e i (recall note 5).

In phonological theory, palatalizations as in (8a-d) require the spreading of the feature [–anterior] from front vowels to the coronal stop (e.g. Hume 1992). That the output is an affricate might be captured in this approach with a default rule which assigns [+strident] to a non-anterior coronal stop. The same approach can account for velar palatalizations as in (8e) in terms of the spreading of [CORONAL] (and [–anterior]) from a front vowel to the coronal stop. The default rule referred to above would then assign [+strident] to the output.[14]

3.2.2. Affricates becoming stops

We refer here to processes which convert affricates into stops (or fricatives) as deaffrications. Representative examples of the former are presented in (9):

(9) Deaffrications (with a stop as the output)
a. t͡ʃ → c / __ {N, #} Anejom Lynch (2000)
b. t͡ʃ → t / __ $]_{\sigma}$ Korean Kim-Renaud (1974: 113)
c. t͡s t͡ʃ → t / __ s ʃ Sahaptin Rigsby and Rude (1996: 671)

In Anejom the affricate /t͡ʃ/ changes to the palatal stop before a nasal, e.g. /nit͡ʃman/ is pronounced as [nɪcman] ‘his hand’ (Lynch 2000: 14). The same process is optional in final position, i.e. /nat͡ʃat͡ʃ/ ‘flatfish’ is pronounced either as [naʲd̥ʒaʲt͡ʃ] or [naʲt͡ʃaʲc] (Lynch 2000: 14). Deaffrication in Korean is usually analysed as a very general phenomenon which also includes the loss of aspiration and the change from /s/ to /t/, also in syllable-final position (see Kim-Renaud 1974 for discussion). Kehrein (2002: 39, note 44) mentions Guajiro and Maidu as languages in which coda /t͡ʃ/ shifts to [t], but we were unable to confirm this with the original sources.

It is interesting to note that the processes in (9) all involve the deaffrication of *strident* affricates. In fact, Sahaptin in (9c) is a language with lateral affricates and non-strident affricates (/p͡ɸ t͡θ/), but neither of these undergoes a shift to stops before /s ʃ/. This type of example suggests that the process in (9c) is triggered by an avoidance of adjacent [+strident] sounds; recall that we are adopting the position that affricates are strident stops, cf. Table 2. The reader is referred to Kehrein (2002: 44ff.) for discussion.

3.2.3. Other changes involving stridency

In addition to the changes described in sections 3.2.1. and 3.2.2., there are several other types of changes which involve a change in stridency. The changes in (10) have in common that the manner of articulation of the input and output remains constant; hence, the strident fricative in (10a) remains a fricative in the output, although it becomes non-strident. The input affricate

in (10b) similarly stays an affricate in the output, although stridency changes.
Consider the possibilities listed in (10):

(10) Changes in stridency
  a. s → θ...
  b. t͡s → t͡θ...
  c. θ → s...
  d. t͡θ → t͡s...

We are aware of one language (Tahltan), which seems to have all of the changes in (10) together in the process of coronal harmony. (10a) is illustrated by the first person singular subject marker /s/, which is realized as [θ] if followed anywhere in the string to the right by an interdental fricative or affricate [d͡ð t͡θ t͡θ’ θ ð]. Thus, e.g. /ɛdɛdɛsduːθ/ is realized as [ɛdɛdɛθduːθ] ‘I whipped myself’ (Shaw 1991: 145). In this process /s/ is also realized as [ʃ] if /s/ is followed by [d͡ʒ t͡ʃ t͡ʃ’ ʃ ʒ j] in the string and as [s] elsewhere. Tahltan also provides an example for (10c): the initial underlying /θ/ of the first-person dual subject prefix surfaces as [s] or [ʃ] in appropriate harmonic contexts, e.g. de/θ/idzel is pronounced as de[s]idzel ‘we shouted’ (Shaw 1991: 145). Shaw stresses that the triggers and targets of the processes discussed above are composed of any member of the following series: [d͡ð t͡θ t͡θ’ θ ð], [d͡z t͡s t͡s’ s z] or [d͡ʒ t͡ʃ t͡ʃ’ ʃ ʒ j]. Since these natural classes also include affricates, Tahltan also provides evidence for the processes in (10b, d).

4. The stop vs. fricative dimension

In this section we discuss the distinctive feature [continuant], posited in generative phonology to capture the distinction between stops and fricatives (section 4.1.), and rules which alter the same feature (section 4.2.).

4.1. The definition of [continuant]

The two classes ‘stops’ and ‘fricatives’ are traditionally captured with the two features [sonorant] and [continuant]. According to Halle and Clements (1983: 7) “Continuants are formed with a vocal tract configuration allowing the airstream to flow through the midsagittal region of the oral tract." According to this definition [+continuant] includes fricatives, vowels and glides, while [–continuant] describes stops and nasals and (because of the clause "midsagittal region") lateral approximants like /l/, which are realized in such a way that the air escapes along one side of the tongue.[15]

The natural classes of stops and fricatives are captured by referring to two feature values, i.e. stops (and affricates) are [–continuant, –sonorant] and fricatives are [+continuant, –sonorant]. We follow the now uncontroversial view that affricates are [–continuant] (recall section 3.1.); hence, the natural class [–continuant, –sonorant] refers to both stops and affricates. The class of stops only can be referred to as [–continuant, –sonorant, –strident].

4.2. Rules altering [continuant]

Rules altering [continuant] subsume processes in which a stop becomes a fricative (section 4.2.1.), a fricative becomes a stop (section 4.2.2.), a fricative becomes an affricate (section 4.2.3.) or an affricate becomes a fricative (section 4.2.4.).

4.2.1. Stops becoming fricatives

We refer to processes which change stops to fricatives as spirantizations. In (11) we have listed several examples of spirantizations in intervocalic position.

(11) Spirantizations (in intervocalic position)

a. b d g → β ð ɣ / V __ V	Tzeltal (*)	Kaufman (1971)
b. d d^y → ð ʒ / V __ V	Badimaya	Dunn (1988)
c. t → ð/ i e __ V	Tümpisa Shoshone	Dayley (1989)
d. p b → β / V __ V	Ibibio	Urua (2004)
e. b → β / V __ V	Af Tunni	Tosco (1997: 27)

The spirantization process in (11a) also applies in other contexts as well, i.e. after a vowel and before a consonant (V__C) as well as postvocalically if the vowel occurs at a specific morphemic juncture (V__+); see Kaufman (1971: 11). In Badimaya (in 11b) the dental /d/ is frequently realized as the fricative [ð] in intervocalic position whereas the palatal stop (transcribed by the author as /d^{y}/) is pronounced as [ʒ] (Dunn 1988: 29). In Tümpisa Shoshone (in 11c) /t/ becomes [ð] intervocalically if the preceding vowel is front, but it changes to [θ] if the preceding vowel is front and the following vowel is devoiced (Dayley 1989: 407). In Ibibio /p b/ change to [β] in intervocalic position. This context also conditions other changes of obstruents in Ibibio, as presented in section 2.2.2. In Af Tunni, the Southern Somali dialect, /b/ becomes [β] in intervocalic position, e.g. the last word in ána kikí yaa/b/ə 'I was surprised at you' is pronounced as yaa[β]ə (Tosco 1997: 27).

One cross-linguistic generalization is that spirantizations with a voiceless stop as the input are considerable less common than ones in which the input is voiced. For additional discussion see Lavoie (2001: 32ff.).

Spirantizations in contexts other than intervocalic position are presented in (12). According to Keels (1985) the process in (12a) only occurs in stressed syllables. In Af Tunni (12c) /q/ is realized as [ɣ] in the preconsonantal position, e.g. /áqli/ is pronounced as [áɣli] 'wisdom' Tosco (1997: 26). The process of g-spirantization in (12d) only applies after /ɪ/ in Standard German, but in other varieties it applies after all front vowels.

(12) Spirantizations (in other contexts)

a. d → θ/ __ $]_{\sigma}$	Guayabero	Keels (1985: 72)
b. t^{h} > θ / V__	Laconian	Bubeník (1983)
c. q→ ɣ/ __ C	Af Tunni	Tosco(1997: 26)
d. g → ç / ɪ __ $]_{\sigma}$	German	Hall (1992: 228)

There are also languages like Greek, in which the change from stop to fricative is motivated as a dissimilation to avoid sequences of adjacent stops; see Newton (1972b: 106). According to that source a stop surfaces as a fricative before a stop (and a fricative surfaces as a stop before a fricative, with the exception of /fs/, which surfaces as [fs]). By contrast, sequences of stop plus fricative and fricative plus stop surface without any changes.

Processes of spirantization as in (11-12) appear to obey the following cross-linguistic generalizations: A spirantized velar surfaces as [ɣ], while a spirantized bilabial (/b p/) and coronal /t d/ will surface as [β ɸ] and [θ ð] respectively. According to Kirchner (2001) spirantizations like the ones in (11) and (12), i. e. those which he sees as lenitions, will never surface as [f v] or [s z]. See below for discussion on this point.

A special type of spirantization is illustrated with the examples in (13), which have in common that the processes are triggered by a following high (and in some cases front) vocoid. We refer henceforth to processes like the ones in (6) and (13) collectively as stop assibilations (Clements 1999, Kim 2001, Hall and Hamann 2006, Hall, Hamann and Żygis 2006 and references cited therein). The crucial difference between the processes in (6) and the ones in (13) is that the output of the former ones are affricates and the latter ones are fricatives.

(13) Stop Assibilation (in which the output is a fricative)

a. t → s / __ i	Finnish (*)	Kiparsky (1973)
b. t > s / __ i u e o	Woleaian	Tawerilmang and Sohn (1984)
c. t d → s / V __ i V	Ancient Greek	Sommerstein (1973: 15-16)
d. *t > s / __ i e	Kosraean	Lee and Wang (1984)

The process in (13a) is known from the literature as a derived environment rule (Kiparsky 1973). For example, stem-final /t/ changes into [s] before /i/ across a morpheme boundary, e.g. /turpot+i/ → [turposi] ‘swelled’. The process also applies if the underlying stem-final /e/ changes into /i/ by an independent process in word-final position: /vete/ → veti → [vesi] ‘water, essive’.

Stop assibilations in other contexts other than the ones in (13) are rare and often morphologically conditioned, e.g. in Nez Perce the morpheme /c/, a palatal stop, changes to the fricative [s] before [n] or [w], e.g. /yúʔcne/ is pronounced as [yúʔsne] ‘poor, object case’ (Aoki 1970: 39). In Iranian Azari a stem-final /t/ changes to [s] in the conditional form, e.g. [atmax] ‘throw’ inf. is pronounced as [assa] ‘throw, conditional’ and [yatmax] ‘sleep’ inf. as [yassa] ‘sleep, conditional’ (Dehghani 2000: 50).

The correct featural analysis of the processes in (13) is not at all obvious. Recall the discussion above with respect to the similar processes in (6). Our conclusion is that a featural analysis of processes like the ones in (13) in terms of phonological features alone (i.e. the addition of [continuant] and [strident]) is unsatisfactory because it does not account for the fact that the rules are triggered only by certain vowels and not by others.

An alternative approach to the processes described in (11-13) is proposed by Kirchner (2001), who argues that the processes are driven by a phonetic tendency to minimize articulatory effort, i.e. a more effortful set of gestures is substituted by a less effortful set of gestures. In particular, spirantization – a change of a stop to a fricative is viewed as reduction of the magnitude of a stop gesture which eventually leads to a constriction typical for fricatives. The fact that assibilations (and affrications) occur more frequently before high front vowels and /j/ is attributed to a strong stop frication resulting from the closeness of the tongue blade to the hard palate. The higher and fronter the vowel, and the greater the coarticulation of the stop and the following vowel, the stronger is the frication (Jäger 1978, Ohala 1983, Kirchner 2001).

Kirchner also differentiates between spirantization processes which result in sibilants and nonsibilants. He argues that the former are less frequent as output cross-linguistically because they require more articulatory effort, i.e. the articulatory settings of sibilants are more precise due to a relatively long and more controlled constriction (Kirchner 2001: 85f). The author also states that the acoustic properties of the input stop are decisive for the output in the sense that stops without any friction never spirantize to sibilants. Along these lines, the author claims that the input stop in spirantizations in (11) is not accompanied by any frication. By contrast, processes of assibilation as in (13) have stops accompanied with frication in the input. Whether or not this claim is always correct should be proved experimentally. To his credit, Kirchner (2001) discusses several apparent counterexamples from Lavoie (2001) and shows that they are compatible with his approach.

4.2.2. Fricatives becoming stops

We refer to processes which involve the change from fricatives to stops as occlusivizations; in terms of features, a fricative becomes [–continuant]. Four examples are provided in (14). The process in (14a) is the replacement of word-initial French /v/ into [b] in Antillean Creole French. The Old High German example in (14b) is context-free. A frequent context of occlusivization is in post-nasal position. Three examples are provided in (15). The reader is referred to Lavoie (2001: 42) for discussion and additional examples.

In the Lumasaaba example in (15a) voiced continuants surface as homorganic stops following a nasal prefix, e.g. /zi+N+li/ → [zin̩di] 'roots'. In Eastern Cheremis (in 15b) /β/ changes to [b] after /m/, e.g. /komβo/ is pronounced as /kombo/ 'goose'. The same process converts the phoneme /ð/ to [d] after nasals and liquids, e.g. /komðek/ changes to [komdek] ‘supine’ and /ɣ/ to [g] after /ŋ/, e.g. /koŋɣa/ is pronounced as [koŋga] ‘stove’.

(14) Occlusivization

a. v > b/ #__	Creole French	Goodman (1964)
b. ð > d	Old High German	Braune (1967)
c. ɸ s → p t / __ $]_{\sigma}$	West Tarangan	Nivens (1992: 145)
d. s s^{ʔ} → t / __ $]_{\sigma}$	Korean	Kim-Renaud (1974)

(15) Occlusivization (after a nasal):

a. β l j → b d ɟ / N __	Lumasaaba (*)	Brown (1972)
b. β ð ɣ → b d g/ nasals __	Eastern Cheremis	Ristinien (1960)

4.2.3. Fricatives becoming affricates

In the following examples (sibilant) fricatives convert to homorganic affricates:

(16) Affrication (of fricatives)

a. s → t͡s/ $_{U}$[__	Samoan	Mosel and Hovdhaugen (1992)
b. z → ^{d}z/ { #, N} __	Japanese	Okada (1999: 118)
c. s → t͡s / C __	It. dialects	Vincent (1988: 281)

In Samoan /s/ is pronounced as [t͡s] not only in utterance-initial position (abbreviated as ‘$_{U}$’) but also before or after stressed vowels under emphasis. In Japanese (in 16b) underlying /z/ tends to be realized with [^{d}z] initially or after the moraic nasal /N/. In the central and southern Italian dialects in (16c), /s/ neutralizes to [t͡s] (transcribed by Vincent as [^{t}s]) after consonants, although Vincent’s only examples involve nasals or laterals, e.g. *falso* [fal͡tso] ‘false’. This process has spread to labials, especially in southern dialects, i.e. /v f/ surface as [b͡v p͡f] ([^{b}v ^{p}f] for Vincent 1988: 281); We interpret affrication after laterals and nasals (as in Italian), as the addition of an intrusive stop; cf. Clements (1987), Ohala (1997), Ohala and Sole (this volume) for a discussion on intrusive stops.

4.2.4. Affricates becoming fricatives

Processes which change affricates to fricatives (deaffrications) are presented in (17). Featurally, a deaffrication means that a strident stop loses [–continuant] and becomes [+continuant].

(17) Deaffrication of sibilant affricates (with fricatives as the output)

a. t͡ʃ d͡ʒ → ʃ ʒ / V __V	Flor. Italian	Giannelli and Savoia (1979)
b. d͡z → z	Jukun	Shimizu (1980)
c. t͡s t͡ʃ → s ʃ / __ C	Pochutla Aztec	Campbell (1974: 62)
d. t͡s→ s/ __ [–cont]	Basque	Hualde (1991: 128)
e. t͡s → z / nonfrontV __	Shoshoni	McLaughlin (1989)

In Florentine Italian the affricates /t͡ʃ d͡ʒ/ become fricatives intervocalically and optionally word-initially, e.g. /t͡ʃerkalo/ ‘cercalo’ is pronounced as [t͡ʃerkalo] or [ʃerkalo] ‘look for it’ (Giannelli and Savoia 1979: 46). The Junkun rule in (17b) is optional, e.g. /d͡z/o ‘put into’ is pronounced as [d͡z]o or [z]o (Shimizu 1980: 57). In Shoshoni and Panamint, Central Numic languages /t͡s/ changes to [z] after a nonfront vowel and to [ʒ] after a front vowel, e.g. in Western Shoshoni /tat͡sɨʔɨmpi/ ‘star’ is realized as [tazɨʔɨmbi] and /kwait͡soi/ ‘wash’ as [koiʒoi] (McLaughlin 1989: 243). Additional examples of deaffrications are discussed by Kehrein (2002), who mentions Luiseño as a language which changes /t͡ʃ/ to [ʃ] word-finally or preceding another stop. Similar examples from Basque can be found in van de Weijer (1994).

The output of deaffrications is virtually always a sibilant fricative (see also Kehrein 2002). One exception seems to be Gosiute Shoshoni, in which the output of deaffrication is a dental [ð] after nonfront vowels, e.g. /tat͡sɨʔɨmpi/ ‘star’ is pronounced as [taðɨʔɨmbi] (McLaughlin 1989: 243). ([ʒ] appears after front vowels, as in Shoshoni and Panamint).

Deaffrications virtually always have a coronal (strident) affricate as the input. One of the rare examples of a deaffrication with a non-coronal affricate as the input (i.e. /p͡f/) is the change in colloquial speech in Modern German from /p͡f/ to [f] in word-initial position, e.g. [p͡f]ennig → [f]ennig ‘penny’.

5. Stop assibilations and affrications in light of phonetic evidence

As suggested at various points in the previous sections, features in rules might adequately describe the relevant processes, but they are not always able to provide a satisfactory explanation for them. In the search for new explanatory possibilities several new approaches have been proposed. These alternatives either interpret such processes from the acoustic/-perceptual perspective (Flemming 1995, Hayes 1999, Padgett 2002b, 2003, Steriade 2001, Wright 1996) or they provide articulatory explanations different from the ones we offered above (Boersma 1998, Kirchner 2001, Lavoie 2001). In this section we concentrate on certain phonetic explanations for stop assibilations with an affricate as the output (e.g. t → t͡s / __ i u as in 6) and palatalizations (e.g. k g → t͡ʃ d͡ʒ / __ e i, as in 8).

From the point of view of the feature theory described in sections 3.1. and 3.2.1., it is not clear why [strident] is assigned to the output of stop assibilations and palatalizations in the (high) vowel context. Indeed, the assignment of this feature seems to be arbitrary and does not offer a principled answer to why this particular feature and not some other one is inserted.

An explanation for the assignment of [strident] to the output of stop assibilations is proposed by a number of phonetic studies. According to this view, stop assibilations are triggered in the context of high vocoids because the cross-sectional area of the supraglottal constriction of vocoids triggering the process is small and because the airflow passing through this constriction therefore causes a turbulent noise, which eventually is reinterpreted as a fricative part of the affricate; cf. Bhat (1974), Jäger (1978), Ohala (1983), Stevens (1993), Clements (1999), and Chang, Plauché and Ohala (2001). It has also been shown experimentally that the (phonetic) aspiration of stops like /t/ is longer before high vowels than before low vowels (Kim 2001, Hall, Hamann and Żygis 2006); this fact can potentially explain why stops undergo stop assibilation preferably before front vocoids (see examples in 6). On the basis of the aforementioned studies, one could speculate that the coronal stops undergoing stop assibilation must always have a certain aspiration length (i.e. positive Voice Onset Time (VOT)).

Żygis, Recasens and Espinosa (2008) pose the hypothesis that the palatalization of velars (as in 8e) can be triggered either perceptually or articulatorily depending on whether or not the stop is aspirated. If the input velar is nonaspirated, it changes to [t͡ʃ] via an intermediate stage, i.e., a palatal stop [c] (e.g. /k/ (Latin) > [c] > [t͡ʃ] > [(t)s] Majorcan Catalan). This articulatorily based hypothesis has its roots in articulatorily based phonetic tradition (Rousselot 1924-1925, Antilla 1972) and has been recently confirmed for Majorcan articulatory data (Recasens and Espinosa 2009). If, on the other hand, the input velar stop is aspirated, it changes to [t͡ʃ] via perceptual reinterpretation (see also Guion 1998). Further experimental studies are needed in order to explore the dichotomy between articulatorily vs. perceptually based palatalization processes.

6. Conclusion

Given that the focus of the papers in the present volume is on the phonetics and phonology of turbulent sounds (i.e. obstruents) the purpose of this article was to provide an overview of the phonological behaviour of stops, affricates and fricatives. We posited a set of traditional features (i.e. [sonorant], [continuant], [strident]) and showed how these features are able to capture commonly occurring natural classes and phonological processes. In several cases we also pointed out whether the processes are frequent or occur rather sporadically which sheds light on cross-linguistic tendencies. Finally, we point out new explanations for phonological processes in the literature which open new perspectives in phonological research. We hope that the open questions posited in the course of the article will spark future work in the phonology of obstruents and especially in the relationship between phonology and phonetics with respect to stops, affricates and fricatives.

Appendix

Table 3. Features for selected consonants.

[–voice]	p	t	c	k	q	t͡s	t͡ʃ	ɸ	f	θ	s	ʃ	ç	x	χ
[+voice]	b	d	ɟ	g	ɢ	d͡z	d͡ʒ	β	v	ð	z	ʒ	ʝ	ɣ	ʁ
[sonorant]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
[continuant]	–	–	–	–	–	–	–	+	+	+	+	+	+	+	+
[strident]						+	+			–	+	+	–		
[LABIAL]	✓							✓	✓						
[round]	–							–	–						
[CORONAL]		✓	✓			✓	✓			✓	✓	✓	✓		
[anterior]		+	–			+	–			+	+	–	–		
[DORSAL]				✓	✓									✓	✓
[back]				+	+									+	+
[high]				+	–									+	–

The matrices include the three features [sonorant], [continuant] and [strident] which we discussed in this article, as well as some of the other features we referred to in our treatment. We do not commit ourselves to the feature necessary to distinguish bilabials from labiodentals. All segments listed above are assumed to be [+consonantal] as well. The '✓' indicates the presence of a privative node in the feature geometry framework (see McCarthy (1988) and Clements and Hume (1995) for two approaches in that framework and Hall (2007) for a discussion of distinctive feature theory in general). Definitions of the features listed in the matrices above which are not discussed in the present article can be found in the preceding works.

Notes

[*] This article has benefitted greatly from the comments by two anonymous reviewers. All disclaimers apply. The study was supported by a grant from the German Research Council (DFG) GWZ-4/11-1-P2.

1. The category 'liquid' is intended to subsume only 'unmarked' liquids, i.e., lateral approximants like /l/, while lateral fricatives are [–sonorant]. Among r-sounds, trills like /r/ and central approximants like English /ɹ/ are usually assumed to be [+sonorant], but some languages are attested with an r-sound which is probably [–sonorant] (e.g. Czech *ř*). We do not investigate the properties of liquids in this article.

   The features for laryngeals (/h ʔ/) are controversial. Chomsky and Halle (1968) and Halle and Clements (1983) consider them to be [+sonorant], but it might be possible to interpret either definition in such a way that /h ʔ/ are

[–sonorant] (see Gussenhoven and Jacobs 1998: 67, note 1). We do not pursue this question here.

Selkirk (1984) argues that [sonorant] can be eliminated as a binary feature and that the natural classes described above can be captured given sonority indices on a Sonority Hierarchy. See Clements (1990), who defends the binary feature [sonorant] as one of the features which defines the concept of ‘sonority’ in syllable structure.

2. The feature [sonorant] is just one of the two features which involve voicing, the other being [voice]. Significantly, the latter feature is argued to be distinctive for obstruents only, while the feature which expresses spontaneous voicing ([+sonorant]) is only relevant for vowels, glides, liquids and nasals.
3. We point out here another approach to [sonorant], which predates the definition proposed by Halle and Clements (1983). Kenstowicz and Kisseberth (1979: 21), write concerning the distinction between obstruents vs. sonorants (and consonants vs. non-consonants): “There are no truly satisfactory articulatory or acoustic definitions for the bases of these two different partitions. Nevertheless, they are crucial for the description of the phonological structure of practically every language.” Thus, on this view, the feature [obstruent] simply has a classificatory function. It is interesting to note that Kenstowicz himself seems to have abandoned his earlier scepticism (in Kenstowicz 1994) by proposing a definition of [sonorant] which incorporates aspects of both the SPE definition and the one proposed by Halle and Clements (1983).
4. Another example of sonorant obstruents is the *r* found in the German dialects spoken near Düsseldorf (Hall 1993). In these dialects the *r* in words like *Hirsch* [hɪχʃ] ‘stag’ is unquestionably an obstruent from the point of view of phonetics because it is non-distinct from the dorsal fricative in words like *Bach* [baχ] ‘stream’, which itself is uncontroversially an obstruent. However, as Hall (1993) shows, there are phonotactic reasons for believing that the surface obstruent [χ] in words like *Hirsch* is underlyingly the [+sonorant] /ʀ/.
5. Synchronic examples are indicated in (1a) and below with the arrow ‘→’ and sound changes with the wedge ‘>’ (see 1b). All processes which are morphologically conditioned are indicated with ‘(*)’ after the language. All examples in this article are presented in the IPA and therefore non-IPA symbols employed by the authors we cite have been changed accordingly.

   An anonymous reviewer points out that sound changes (A > B) will often pass through an intermediate stage (e.g. A > C > B), therefore obscuring the change from A to B. To the best of our knowledge neither (1b), nor any of the other sound changes we discuss below had such an intermediate stage.
6. As pointed to us by a reviewer, the symbols /β ð γ/ do not always indicate fricatives in language descriptions (and in some version of the IPA), but that they can instead denote approximants. We have attempted to choose processes below in which /β ð γ/ are realized as fricatives according to the source

provided. If a process shows variation between a fricative and an approximant, then we indicate this below. We omit processes in which sounds transcribed as /β ð ɣ/ are realized as approximants.

7. If the underlying /v/ is followed by a sonorant consonant then it shows a free variation in its realization, i.e. as [v] or [w], e.g. slá/v/ny is pronounced as slá[v]ny or slá[w]ny ‘famous’ (Short 1993: 536).
8. Note too that (3a-b) involve a change from [–consonantal] /w j/ to [+consonantal] [p c]. For discussion on the status of [consonantal] see Kaisse (1992) and Hume and Odden (1996).
9. That all palatals (including /ç/) are [CORONAL] is defended by Hume (1992). By contrast, Hall (1997) argues that palatal fricatives like /ç/ are [DORSAL] and not [CORONAL], while Robinson (2001) contends that they are [CORONAL] *and* [DORSAL]. We do not commit ourselves to either analysis here because this issue is peripheral to the goals of the present study. Note, however, that if we were to adopt the position endorsed by Hall (1997), that /ç/ would be unmarked for stridency.
10. SPE also uses [strident] to distinguish ([–strident]) bilabials (e.g. /ɸ β/) from ([+strident]) labiodentals (e.g. /f v/).
11. A number of linguists have argued that affricates also have a [+continuant] component. Among these proposals, Sagey (1986) sees [–continuant] and [+continuant] in affricates as being linearly ordered. By contrast, Lombardi (1990) argues that [–continuant] and [+continuant] are linearly unordered. Other approaches to affricates can be found in van de Weijer (1993, 1994) and Schafer (1995).

    We do not discuss the aforementioned approaches to affricates below; our assumption is that affricates do not have a [+continuant] component, although this is clearly an issue that needs to be left open for further study.
12. Stop assibilations like the ones in (6) are also attested as a pronunciation of /t d/ when those stops are secondarily palatalized, e.g. /tʲ dʲ/ → [t͡sʲ d͡zʲ] in Belorussian (Mayo 1993).
13. We were unable to confirm process (7a) with the original source (Crothers 1979).
14. Flemming (1995) proposes an alternative account of processes like the ones in (8), focusing on perceptual distinctiveness. He argues that palatalization involves enhancing the difference in F2 at the release of consonants preceding front and back vowels, e.g. /ka/ vs. /ki/. F2 is naturally increased at the consonantal release in the latter case as it is preceded by a vowel with higher F2. Due to the shift of the primary constriction from velar to palatal, a certain degree of frication emerges (palatal sounds are always accompanied by frication). A difference between a palatal and plain stop is further enhanced by increasing the loudness of frication which results in an acoustically prominent sibilant affricate. In order to account for affrication in palatalization processes,

Flemming proposes Minimal Distance constraints selecting candidates with optimal F2 distance and differences in burst energy.

In his discussion on frequency differences between velar and labial stops Flemming argues that velar palatalizations occur more frequently than labial palatalizations because the palatalization of labials does not naturally result in frication which itself is necessary for assibilation. If the assibilation of labials does occur the process is independent of the place of articulation of the stop. In contrast, the assibilation of velar is more often attested because co-articulation naturally gives rise to frication which leads to assibilation.

15. The controversial questions involving [continuant] pertain primarily to [+sonorant] sounds and will therefore not be discussed, e.g. whether or not lateral approximants are plus or minus [continuant]. See Mielke (2005) for discussion. Another open question concerns how to capture the connection between rules assimilating [continuant] and the place features as a unit (see Padgett 1991 and van de Weijer 1992).

References

Anonby, Erik John
2006 Mambay. *International Journal of American Linguistics* 35: 221–233.

Anttila, Raimo
1972 *An Introduction to Historical and Comparative Linguistics*. New York: MacMillan.

Aoki, Haruo
1970 *Nez Perce Grammar*. Berkeley: University California Press.

Barbosa, Plínio, and Eleonora C. Albano
2004 Brazilian Portuguese. *International Journal of American Linguistics* 34: 227–232.

Barkaï, Malachi, and Julia Horvath
1978 Voicing assimilation and the sonority hierarchy. Evidence from Russian, Hebrew and Hungarian. *Linguistics* 212: 77–78.

Bárkányi, Zsuzsanna, and Zoltan Kiss
A phonetically-based approach to the phonology of v: A case-study from Hungarian and Slovak. This volume.

Bauer, Winifred, and William Parker
1993 *Maori*. New York: Routledge.

Bhat, D. N. S.
1974 A general study of palatalization. In *Phonology. Universals of Human Language*, Joseph H. Greenberg (ed.), vol.2, 247–292. Stanford: Stanford University Press.

Boersma, Paul
1998 *Functional Phonology.* Den Haag: Holland Academic Graphics.

Braune, Wilhelm
1967 Reprint. *Althochdeutsche Grammatik.* 12th edition. Tübingen: Niemeyer. Original edition, Halle: Niemeyer, 1886.

Brown, Gillian
1972 *Phonological Rules and Dialect Variation: A Study of the Phonology of the Lumasaaba*. Cambridge: Cambridge University Press.

Bubeník, Vit
1983 *The Phonological Interpretation of Ancient Greek: A Pandialectal Analysis*. Toronto: University Toronto Press.

Campbell, Lyle
1974 Phonological features: problems and proposals. *Language* 50 (1): 52–65.

Carlton, Terence R.
1991 *Introduction to the Phonological History of Slavic Languages*. Columbus: Slavica.

Catford, J. C.
2001 On r's, rhotacism and paleophony. *Journal of the International Phonetic Association* 31: 171–185.

Chang, Steve S., Madelaine C. Plauché, and John J. Ohala
2001 Markedness and consonant confusion asymmetries. In *The Role of Speech Perception Phenomena in Phonology*, Elizabeth Hume, and Keith Johnson (eds.), 79–101. San Diego, CA: Academic Press.

Chomsky, Noam, and Morris Halle
1968 *The Sound Pattern of English.* New York: Harper and Row.

Clements, G. N.
1987 Phonological feature representation and the description of intrusive stops. In *Parasession on Autosegmental and Metrical Phonology*, Anna Bosch, Barbara Need, and Eric Schiller (eds.), 29–50. Chicago: Linguistic Society.

Clements, G. N.
1990 The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, John Kingston, and Mary Beckman (eds.), 283–333. Cambridge: Cambridge University Press.

Clements, G. N.
1999 Affricates as noncontoured stops. In *Item, Order in Language and Speech*, Osamu Fujimura, Brian D. Joseph, and Bohumil Palek (eds.), 271–299. Prague: Charles University Press.

Clements, G. N., and Elizabeth Hume

1995 The internal organization of speech sounds. In *The Handbook of Phonological Theory*, John Goldsmith (ed.), 245–307. Oxford: Blackwell.

Crothers, John (ed.)

1979 *Handbook of Phonological Data from a Sample of the World's Languages. A Report of the Stanford Phonology Archive*. Department of Linguistics: Stanford University.

Dayley, John P.

1989 *Tümpisa (Panamint) Shoshone Grammar*. Publications in Linguistics vol. 115. University of California: UC press.

Dehghani, Yavar

2000 *A Grammar of Iranian Azari.* München: Europa.

Downing, Laura J.

2005 Jita Causative Doubling and Paradigm Uniformity. In *Paradigms in Phonological Theory*, Laura J. Downing, T. A. Hall, and Renate Raffelsiefen (eds.), 122–144. Oxford: Oxford University Press.

Duanmu, San

2000 *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Dunn, Leone

1988 Badimaya, a western Australian language. *Papers in Australian Linguistics* 17: 19–149.

Flemming, Edward S.

1995 *Auditory Representations in Phonology*. Ph.D. diss., University of California Los Angeles. [Published 2002. London: Routledge]

Giannelli, Luciano, and Leonardo M. Savoia

1979 Indebolimento Consonantico in Toscana. *Rivista Italiana di Diallettologia* 2: 23–58.

Goodman, Morris F.

1964 *A Comparative Study of Creole French Dialects*. The Hague: Mouton.

Guion, Susann G.

1998 The role of perception in the sound change of velar palatalization. *Phonetica* 55: 18–52.

Gusain, Lakhan

2000 *Bagri*. München: Lincom.

Gussenhoven, Carlos, and Haike Jacobs

1998 *Understanding Phonology*. London: Arnold.

Hall, T. A.

1992 *Syllable Structure and Syllable-Related Processes in German*. Tübingen: Niemeyer.

Hall, T. A.
1993 Phonology of German /R/. *Phonology* 10 (1): 83–105.

Hall, T. A.
1997 *The Phonology of Coronals*. Amsterdam: Benjamins.

Hall, T. A.
2007 German glide formation and its theoretical consequences. *The Linguistic Review* 24: 1–31.

Hall, T. A., and Silke Hamann
2006 Towards a typology of phonological stop assibilation. *Linguistics* 44: 1195–1236.

Hall, T. A., Silke Hamann, and Marzena Żygis
2006 The phonetic motivation for phonological stop assibilation. *Journal of the International Phonetic Association* 36: 59–81.

Halle, Morris, and G. N. Clements
1983 *Problem Book in Phonology*. Cambridge, Mass.: Cambridge University Press.

Harris, James W.
1983 *Syllable Structure and Stress in Spanish. A Nonlinear Analysis.* Cambridge, Mass.: MIT Press.

Hayward, Dick
1984 *The Arbore Language: A First Investigation*. Hamburg: Helmut Buske Verlag.

Hayes, Bruce
1999 Phonetically driven phonology: The role of optimality theory and inductive grounding. In *Surface Syllable Structure and Segment Sequencing*, Michael Darnell, Edith Moravscik, Frederick Newmeyer, Michael Noonan, and Kathleen Wheatley (eds.), 243–285. Leiden: Holland Institute of Linguistics.

Herbert, Raymond, and Nicholas Poppe
1963 *Kirgiz Manual*. Bloomington: Indiana University Press.

Hock, Hans H.
1986 *Principles of Historical Linguistics*. Berlin: de Gruyter.

Hualde, José Ignacio
1991 *Basque Phonology*. London: Routledge.

Hume, Elizabeth
1992 *Front Vowels, Coronal Consonants and their Interaction in Non-Linear Phonology*. Ph.D. diss., Cornell University. [Published 1994. New York: Garland.]

Hume, Elizabeth, and David Odden
1996 Reconsidering [consonantal]. *Phonology* 13 (3): 345–376.

Jäger, Jeri J.
1978 Speech aerodynamics and phonological universals. *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society*, Christine Chiarello, Henry Thompson, Farrell Ackerman, Orin Gensler, John Kingston, Eve C. Sweetser, Anthony C. Woodbury, Kenneth Whistler, and Jeri J. Jäger (eds.), 311–329. Berkeley, CA: Berkeley Linguistics Society.

Jakobson, Roman C., Gunnar M. Fant, and Morris Halle
1967 Reprint. *Preliminaries to Speech Analysis*. 6th ed. Cambridge, Mass.: The MIT Press, Original edition, Technical Report 13, MIT Acoustics Laboratory, 1952.

Kahn, Daniel
1976 Syllable-based generalizations in English phonology. Ph.D. diss., MIT.

Kaisse, Ellen M.
1992 Can [consonantal] spread? *Language* 68: 313–332.

Kaufman, Terrence
1971 *Tzeltal Phonology and Morphology*. Berkeley: University of Califormia Press.

Keels, Jack
1985 Guayabero: phonology and morphophonemics. In *From Phonology to Discourse: Studies in Six Colombian Languages*, Ruth Mary Brend (ed.), 57–87. Dallas: Summer Institute of Linguistics.

Kehrein, Wolfgang
2002 *Phonological Representation and Phonetic Phrasing: Affricates and Laryngeals*. Tübingen: Niemeyer.

Kenstowicz, Michael
1994 *Phonology in Generative Grammar*. Oxford: Blackwell.

Kenstowicz, Michael, and Charles Kisseberth
1979 *Generative Phonology. Description and Theory*. New York: Academic Press.

Kim, Hyunsoon
2001 A phonetically based account of phonological stop assibilation. *Phonology* 18: 81–108.

Kim-Renaud, Young-Key
1974 Korean consonantal phonology. Ph.D. diss., University of Hawaii.

Kiparsky, Paul
1973 Abstractness, opacity and global rules. In *Three Dimensions of Linguistic Theory*, Osamu Fujimura (ed.), 57–86. Tokio: Taikusha.

Kirchner, Robert
2001 *An Effort Based Approach to Consonant Lenition*. New York: Routledge.

LaCharité, Darlene

1993 The internal structure of affricates. Ph.D. diss., University of Ottawa.

Lahiri, Aditi, and Vincent Evers

1991 Palatalization and coronality. In *The Special Status of Coronals: Internal and External Evidence*, Carole Paradis, and Jean-Francois Prunet (eds.), 79–100. San Diego: Academic Press.

Lavoie, Lisa M.

2001 *Consonant Strength. Phonological Patterns and Phonetic Manifestation*. New York: Garland Publishing.

Lee, Kee-Dong, and Judith Wang

1984 Kosraean reflexes of Proto-Oceanic phonemes. In *Studies in Micronesian Linguistics*, Byron W. Bender (ed.), 403–422. Canberra: The Australian National University.

Lombardi, Linda

1990 The nonlinear organization of the affricate. *Natural Language and Linguistic Theory* 8: 375–425.

Luschützky, Hans Christian

1992 *Zur Phonologie der Affrikaten*. Frankfurt am Main: Hector.

Lynch, John

2000 *A Grammar of Anejom*. Pacific Linguistics 507. Canberra: The Australian National University.

Mayo, Peter

1993 Belorussian. In *The Slavonic Languages*, Bernard Comrie, and Greville G. Corbett (eds.), 887–946. London: Routledge.

McCarthy, John J.

1988 Feature geometry and dependency: A review. *Phonetica* 43: 84–108.

McLaughlin, John E.

1989 A note on the change of strident to nonstrident in Gosiute Shoshoni. *International Journal of American Linguistics* 55: 240–247.

Mielke, Jeff

2005 Ambivalence and ambiguity in laterals and nasals. *Phonology* 22: 169–203.

Mosel, Ulrike, and Even Hovdhaugen

1992 *Samoan Reference Grammar*. Oslo: Scandinavian University Press.

Newton, Brian

1972a *Cypriot Greek: Its Phonology and Inflections*. The Hague: Mouton.

Newton, Brian

1972b *The Generative Interpretation of Dialect. A Study of Modern Greek Phonology*. The Hague: Mouton.

Nivens, Richard
1992 A lexical phonology of West Tarangan. In *Phonological Studies in Four Languages of Maluku*, Donald A. Burquest, and Wyn D. Laidig (eds.), 127–227. Arlington, Texas: Summer Institute of Linguistics and University of Texas at Arlington Publications.

Ohala, John J.
1983 The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, Peter F. MacNeilage (ed.), 189–216. New York: Springer Verlag.

Ohala, John J.
1997 Emergent stops. *Proceedings of the 4th Seoul International Conference on Linguistics* [SICOL], 84–91. Seoul: The Linguistic Society of Korea.

Ohala, John J., and Maria-Josep Solé
Turbulence and phonology. This volume.

Okada, Hideo
1999 Japanese. In *Handbook of the International Phonetic Association: A guide to the usage of the International Phonetic Alphabet*, 117–119. Cambridge, England: Cambridge University Press.

Ourso, Meterwa, and Charles H. Ulrich
1990 Sonorant strengthening in Lama. *Studies in the Linguistic Sciences* 20: 135–147.

Padgett, Jaye
1991 Stricture in feature geometry. Ph.D. diss., University of Massachusetts at Amherst.

Padgett, Jaye
2002a Russian voicing assimilation, final devoicing, and the problem of [v]. Ms. University of Santa Cruz.

Padgett, Jaye
2002b Contrast and post-velar fronting in Russian. *Natural Language and Linguistic Theory* 21: 39–87.

Padgett, Jaye
2003 The emergence of contrastive palatalization in Russian. In *Optimality Theory and Language Change*, Eric D. Holt (ed.), 307–335. Dordrecht: Kluwer Academic Press.

Padgett, Jaye, and Marzena Żygis
2007 The evolution of sibilants in Polish and Russian. *Journal of Slavic Linguistics* 15 (2): 291–324.

Parks, Douglas R.
1976 *A Grammar of Pawnee.* New York: Garland.

Pence, Alan
1966 Kunimaipa phonology: hierarchical levels. *Pacific Linguistics,* Ser. A, 7: 49–67.

Pope, Mildred K.
1952 *From Latin to Modern French with Special Consideration of Anglo-Norman. Phonology and Morphology*. Manchester: Manchester University Press.

Recasens, Daniel, and Aina Espinosa
2009 Acoustics and perception of velar softening for unaspirated stops. *Journal of Phonetics* 37: 189–211.

Rice, Keren
1992 A reexamination of the feature [sonorant]: the status of 'sonorant obstruents'. *Language* 69 (2): 308–344.

Rigsby, Bruce, and Noel Rude
1996 Sketch of Sahaptin, a Sahaptin language. In *Handbook of North American Indian Languages*, Ives Goddard (ed.), vol. 17: 666–692. Washington: Smithsonian Institute.

Ristinien, Elaine Katherine
1960 An East Cheremis phonology. In *American Studies in Uralic Linguistics*, 249–287. Bloomington: Indiana University.

Robinson, Orrin
2001 *Whose German? The ach/ich Alternation and Related Phenomena in 'Standard' and 'Colloquial'*. Amsterdam: Benjamins.

Rospond, Stanislaw
1971 *Gramatyka historyczna języka polskiego*. [Historical Grammar of Polish] Warszawa: Państwowe Wydawnictwo Naukowe.

Ross, Malcom D.
1998 Proto Oceanic and the Austronesian Languages of Western Melanesia. *Pacific Linguistics Series* C-No. 98. Canberra: Australian National University.

Rousselot, Jean Pierre
1924-1925 *Principes de phonétique expérimentale*. Leipzig: Welter.

Rubach, Jerzy
1984 *Cyclic and Lexical Phonology: The Structure of Polish*. Dordrecht: Foris.

Rubach, Jerzy
1994 Affricates as strident stops in Polish. *Linguistic Inquiry* 25: 119–143.

Rubino, Carl Ralph Gavalez
1997 *A Reference Grammar of Ilocano*. UMI: Ann Arbor.

Sagey, Elizabeth
1986 *The Representations of Features and Relations in Nonlinear Phonology*. Ph.D. diss., MIT.

Schafer, Robin
1995 Headedness and the representation of affricates. *The Linguistic Review* 12: 61–87.

Selkirk, Elisabeth O.
1984 On the major class features and syllable theory. In *Language Sound Structure: Studies in Phonology Presented to Morris Halle by his Teacher and Students*, Mark Aronoff, and Richard T. Oehrle (eds.), 107–136. Cambridge, Mass.: MIT Press.

Shaw, Patricia A.
1991 Consonant harmony systems: the special status of coronal harmony. In *The Special Status of Coronals. Internal and External Evidence*, Carole Paradis, and Jean-François Prunet (eds.), 125–157. San Diego: Academic Press.

Shimizu, Kiyoshi
1980 *Jukun. A Grammar*. Wien: Afro-Pub.

Short, David
1993 Slovak. In *The Slavonic Languages*, Bernard Comrie, and Greville G. Corbett (eds.), 533–592. London: Routledge.

Sommerstein, Alan H.
1973 *The Sound Pattern of Ancient Greek*. Oxford: Blackwell.

Steriade, Donca
2001 Directional asymmetries in place assimilation: A perceptual account. In *The Role of Speech Perception Phenomena in Phonology*, Elisabeth Hume, and Keith Johnson (eds.), 219–250. San Diego, CA: Academic Press.

Stevens, Kenneth N.
1993 Modelling affricate consonants. *Speech Communication* 13: 33–43.

Stone, Gerald
1993 Cassubian. In *The Slavonic Languages*, Bernard Comrie, and Greville G. Corbett (eds.), 759–794. London: Routledge.

Swadesh, Morris
1969 *Elementos del Tarasco Antigua*. México: Universidad Nacional Autónoma de México.

Tawerilmang, Anthony F., and Ho-min Sohn
1984 Proto-Oceanic reflexes in Woleaian. In *Studies in Micronesian Linguistics*, Bayon W. Bender (ed.), Pacific Linguistics Series C-No. 80: 171–199. Canberra: The Australian National University.

Tench, Paul
2007 Tera. *International Journal of American Linguistics* 37: 227–234.

Terrill, Angela
1998 *Biri.* München: Lincom.

Tinelli, Henri
1981 *Creole Phonology*. The Hague: Mouton.

Tosco, Mauro
1997 *Af Tunni.* Köln: Rüdiger Köppe Verlag.

Underhill, Robert
1976 *Turkish Grammar*. Cambridge, Mass.: MIT Press.

Urua, Eno-Abasi
2004 Ibibio. *Journal of the International Phonetic Association* 34, 105–109.
Vennemann, Theo
1988 *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Berlin: Mouton de Gruyter.
Vincent, Nigel
1988 Italian. In *The Romance Languages*, Martin Harris, and Nigel Vincent (eds.), 279–313. London: Routledge.
Weijer, Jeroen van de
1992 Basque affricates and the manner-place dependency. *Lingua* 88: 129–147.
Weijer, Jeroen van de
1993 The manner-place dependency in complex segments. *Linguistics* 31: 87–110.
Weijer, Jeroen van de
1994 *Segmental Structure and Complex Segments*. Ph.D diss., University of Leiden. [Published 1996. Tübingen: Niemeyer.]
Whisler, Ronald
1992 Phonology of Sawai. In *Phonological Studies in Four Languages of Maluku*, Donald A. Burquest, and Wyn D. Laidig (eds.), 7–32. Dallas: The Summer Institute of Linguistics.
Wright, Richard
1996 Consonant clusters and cue preservation in Tsou. Ph.D. diss., UCLA.
Żygis, Marzena
submitted *Contrast optimisation in Slavic sibilant systems*. Amsterdam: John Benjamins.
Żygis, Marzena, Recasens, Daniel, and Aina Espinosa
2008 Acoustic characteristics of velar stops and velar softening in German, Polish and Catalan. In *Proceedings of the 8th International Seminar on Speech Production*, Rudolph Sock, Susanne Fuchs, and Yves Laprie (eds.), 97–100. Strasbourg: Inria.

Turbulence and phonology*

John J. Ohala and Maria-Josep Solé

0. Introduction

In this paper we aim to provide an account of some of the phonological patterns involving turbulent sounds, summarizing material we have published previously and results from other investigators. In addition, we explore the ways in which sounds pattern, combine, and evolve in language and how these patterns can be derived from a few physical and perceptual principles which are independent from language itself (Lindblom 1984, 1999) and which can be empirically verified (Ohala and Jaeger 1986). This approach should be contrasted with that of mainstream phonological theory (i.e. phonological theory within generative linguistics) which primarily considers sound structure as motivated by "formal" principles or constraints that are specific to language, rather than relevant to other physical or cognitive domains.

For this reason, the title of this paper is meant to be ambiguous. The primary sense of it refers to sound patterns in languages involving sounds with turbulence, e.g. fricatives and stops bursts, but a secondary meaning is the metaphorical turbulence in the practice of phonology over the past several decades. We shall treat the latter topic first.

1. Turbulence in phonology

Anyone familiar with the history of phonological science in the 20th century will have to concede that there has been considerable turbulence in the theoretical domain. To be sure, there were controversies in phonology in the 19th century, too, for example, the dispute as to whether Sanskrit should be taken as the oldest ancestor of what became known as the Indo-European language family or whether an attempt should be made to reconstruct a parent language of which even Sanskrit was an off-shoot. Schleicher (1861-62) the advocate of the latter view, eventually won that dispute. There were also disputes as to the causes and mechanisms of sound change; these disputes have not been satisfactorily resolved even to this day. Nevertheless, this domain of phonological science – historical linguistics – made steady and remarkable progress from its beginning in the 18th century (e.g. ten Kate 1723, des Brosses 1765).[1] The methodology, the "comparative method", has been refined and proven itself. Beginning around the turn of the 19th to 20th centuries, a new program developed, spurred by the writings of Kruszewski, Baudouin de Courtenay, Saussure, Sapir and others. This was to be an account of the psychological aspect of language, i.e. how distinctive speech sounds and their contextual variants arise and are maintained and managed in the mind of the speaker-hearer. Turbulence in linguistic theory arose when patterns arising from physical and physiological factors, that had previously been judged to be extra-linguistic, were claimed by generative grammarians in the mid-20th century to be incorporated into the posited psychological lexicon and grammar of the speaker. The dust still has not settled on this controversy. In form these supposed psychological representations of language and the methods used to discover them were largely identical to the descriptive entities and methods of the historical phonologist.

Against this historical background, we declare that the explanations we give below for sound patterns involving noise (turbulence) are strictly physical phonetic. In our view there is no mystery about how physical phonetic factors can become phonologized and manifested in languages' sound patterns: the variation in the speech due to these physical constraints can lead to the listener's misperception, misparsing, and misconstruing of the speaker's target pronunciation. Listener error, then, can lead to a change in the pronunciation norms just as manuscript copyists' errors led to different variants of ancient texts in the time before printing (Ohala 1981b, 1989).

The ultimate purpose of the experimental approach to phonology illustrated in this paper is to demonstrate that phonological theory must be based on mechanisms and principles coming from the subsystems involved in speech production and speech perception (Ohala 1981a, 1981b, 1992).

For ease of presentation, we group the sound patterns according to aerodynamic factors (sections 2, 3), acoustic-auditory factors (sections 4, 5), and timing factors (section 6), but most of the patterns involve the interaction between a number of these factors.

2. Aerodynamic factors relevant to turbulence. Basic principles

From an aerodynamic point of view, we can think of the vocal tract as two air cavities, the lung cavity and the supraglottal cavity, ultimately connected to the atmosphere. The two cavities are connected by the glottis which allows pulmonic air to flow into the oral cavity as pulmonic forces (from muscular activity or passive recoil) compress the lungs. The supraglottal cavity is connected to the atmosphere by the mouth (and the nose) which can impede the air flowing out with changes in articulatory constriction of the lips and tongue. Thus the two main valves (along with the nasal valve) that regulate the airflow used in speech are the glottis and the oral constriction, and turbulent noise can be generated at both of these constrictions.

The generation of audible turbulence, i.e. noise, in the vocal tract is necessary for the production of fricatives, the fricative release of affricates and the burst of stops. However, audible turbulence may also be associated with the production of vowels and sonorants in certain conditions. Although there is some degree of low-level air flow turbulence even for the most open of speech sounds, i.e. something like [ɛ] (because the air flowing into the vocal tract acquires some turbulence upon passing through the vibrating vocal cords) it is only when the turbulence reaches a level to become audible that it can play some role in speech.

Turbulent airflow is determined by a multiplicity of factors, including roughness and length of the channel, shape of the orifice, and whether the air downstream of the constriction is already turbulent, but it is the *speed* of air flowing through the constriction which is the main factor. The speed of air (or ‘particle velocity’, v), in turn, depends on the volume of air flowing through the constriction (called the ‘volume velocity’, U) and on the cross-sectional area of the constriction (A), as indicated in (1). Thus the larger the volume of air per unit time and the smaller the constriction, the higher the velocity, and the more intense the frication noise.

(1) $v = U/A$

v is the particle velocity in cm/sec, U is the volume velocity in cm^3/sec, A is the cross-dimensional area of the constriction in cm^2.

The volume of air, U, that will flow through a constriction depends on the size of the aperture, A, and the pressure difference across the aperture, that is, the difference between the upstream pressure and downstream pressure, $P_{upstream}$-$P_{downstream}$, as shown in (2) (Warren and DuBois 1964). In the case of an oral constriction, this will be the difference in pressure between the oral cavity and the atmosphere (P_{oral}-$P_{atmospheric})^2$, and in the case of a glottal constriction, the difference in pressure between the subglottal (or pulmonic) cavity and the oral cavity ($P_{subglottal}$-P_{oral}). The greater the difference in pressure and the larger the area of the constriction, the larger the rate of flow. The exponent, *a*, varies between .5 and 1, depending on the nature of the flow; it is 1 when the flow is smooth or laminar, and .5 when the flow is turbulent. In the conditions found in speech production, i.e. what is called 'nozzle flow', this number may vary continuously between these two extremes (Jaeger and Matthys 1970). Naturally, the direction of air flow will always be from the cavity with greater pressure to that with lesser pressure.

(2) $U = A\,(P_{upstream} - P_{downstream})^{a}\,c$ or $U = A\,(\Delta P)^{a}\,c$

P is the pressure in cm H_2O and c is a constant. As mentioned, traditionally, the critical velocity at which the change from laminar to turbulent flow occurs is determined by a number of factors, including particle velocity, the diameter and roughness of the channel the air passes through, etc. The relative contribution of these factors for certain flow conditions is quantified in the Reynolds number. When the Reynolds number exceeds a certain threshold the airflow is supposed to change from smooth or “laminar” to turbulent. However, it is the case that in irregularly-shaped channels like the vocal tract and with airflow that usually has a turbulent entry into the vocal tract (certainly the case as the air passes several “rough” surfaces – narrow alveoli in the lungs, tracheal rings, vocal cords, ventricular folds, epiglottis, etc.), some turbulence, even audible turbulence can occur in conditions where the Reynolds number is far below the ideal threshold between laminar and turbulent flow. So it is simplest just to state the relation between air velocity and noise in a purely qualitative way: the intensity (i.e. loudness) and centre frequency (i.e. pitch) of frication noise varies monotonically with the particle velocity of the air flow, as given in (3a), below (Catford 1977: ch. 3; Stevens 1971; Flanagan and Ishizaka 1976; Flanagan, Ishizaka, and Shipley 1975, 1980; Shadle 1990). A variant of this relation, stating that intensity of frication at a supraglottal constriction increases with increasing oral pressure and decreasing aperture of constriction (therefore, conflating principles 1 and 2), is given in (3b) (Stevens 1971).

(3a) $I_{frication} \sim v$
(3b) $I_{frication} \sim P_o^{3/2} A^{1/2}$

The articulatory constriction for vowels, glides and sonorants is not typically narrow enough to cause a pressure difference across the constriction (in other words, P_{oral} is not much higher than $P_{atmospheric}$), so that particle velocity, *v*, through the constriction is kept low and does not reach a level sufficient to generate audible frication (but see below for exceptions when these are voiceless). Obstruents such as fricatives, stops and affricates, on the other hand, are produced with a narrow or complete constriction which causes the P_{oral} to rise substantially over $P_{atmospheric}$; upon release the particle velocity is high and the airflow becomes more turbulent.

Additional turbulence can be also generated when an air jet that has passed through the major oral cavity constriction encounters any sharp discontinuity: either an abrupt enlargement of the channel or the opposite, i.e. an additional barrier or “baffle”. The former occurs (a) when air passes through the vocal cords during voicing (“voicing” consists of periodic short-term noise bursts occurring at a rate equal to the fundamental frequency) and (b) when the air flows past the sharp-edged constriction at the teeth in a labio-dental fricative such as [f]. The latter occurs when an air jet emerging from an apical-alveolar constriction is directed at the upper and lower incisors. (It is this factor which accounts for the somewhat impoverished apical fricatives made by children who have lost their incisors as part of the change from baby teeth to permanent dentition at approximately age 6 and on.)

A final factor needs to be mentioned regarding the acoustic amplitude of the noise which the turbulence generates: other things being equal, the intensity of the noise is greater, the larger is the resonating cavity downstream of the point where the turbulence occurs. For this reason palatal and velar fricatives have more intense noise than labial and labio-dental fricatives.

3. Generalizations on phonetic and phonological universals deduced from aerodynamic principles

In the following sections we review a number of sound patterns, involving the emergence or extinction of turbulence, which can be deduced in part from variations in glottal flow, U_g (section 3.1.), changes in area of oral constriction, A_o (3.2), and changes in oral pressure, P_o (3.3.).

3.1. Variations in glottal flow

It is known that the glottis regulates the flow of air from the lungs into the oral cavity. Vibration at the vocal folds for voiced obstruents causes a diminished rate of flow through the glottis and a significantly lower oral pressure vis-à-vis voiceless obstruents which, in contrast, have a large glottal opening and continuous flow. A relatively low oral pressure is necessary to maintain a sufficient pressure differential across the glottis so that there will be continuous transglottal flow and thus voicing during the obstruent.[3]

Since by principles (2) and (3), intensity of turbulence is dependent on the pressure difference across the oral constriction, a lower oral pressure for voiced obstruents will result in a lower intensity of high frequency noise during the fricative constriction or at stop release vis-à-vis voiceless obstruents. In addition, due to the reduced transglottal flow, voiced obstruents take longer to build up oral pressure behind the oral constriction, which results in a delayed onset of audible frication for fricatives (Solé 2002b) and a weaker burst for stops compared to their voiceless counterparts. Thus, the characteristic cues for obstruency – abrupt amplitude discontinuities and high intensity noise cues – are enhanced in voiceless obstruents due to the larger rate of flow through the glottis. In sum, for aerodynamic and auditory-acoustic reasons *voicelessness favours or enhances obstruency* (i.e. high intensity frication and release burst). The following phonological generalizations can be derived from this principle.

3.1.1. Voicelessness favours obstruency

3.1.1.1. Sonorants (glides, laterals and nasals) become fricatives when devoiced

As stated in section 2, the common description of the articulatory difference between an approximant and a fricative is that they have different degrees of constriction (e.g. Clark and Yallop 1990: 81; Laver 1994: 134-135). A difference based on constriction degree is endorsed by the present structure of the IPA phonetic chart. While this is generally an adequate description, there are cases with considerable phonological interest where this is not completely true.

In general, approximants have a constriction that is large enough to allow the airstream to flow through it without causing turbulence. Nevertheless, by equation (1), $v = U/A$, they may cross the threshold into obstruents if the constriction (A) narrows further or if a higher rate of flow (U) passes through the same constriction. Approximants, e.g. [ʋ j w l ɹ ɻ], and nasals, which by definition are non-obstruents, are usually voiced. When voiceless, however, without any variation in the configuration of the oral articulators, they can become fricative (and thus obstruents), e.g. [f ç ʍ ɸ ɬ ɹ̥ ʂ] and [m̥ n̥ ɲ̥ ŋ̊], respectively. This happens simply due to the increased airflow passing through the constriction created by these consonants – the increased airflow being caused by the greater opening (and thus lesser resistance to airflow) at the glottis (see Catford 1977: 120ff.). Thus, it is the higher rate of flow through the open glottis which creates the higher particle velocity of the airflow through the supraglottal constriction (velocity being directly proportional to flowrate, for a given aperture, $v = U/A$) and leads to turbulence, hence the obstruent character of the voiceless (former) sonorants. Some phonological consequences of the frication of devoiced sonorants are the following.

Laterals and /r/s

First, there are cases where a fricative and an approximant alternate and there is co-variation between frication and voicelessness, such that the fricative is voiceless and the approximant is voiced. For example, in Kwakiutl (Boas 1947) there are morphophonemic alternations between "plain" (voiceless), "hardened" (ejectives), and "weakened" (voiced) obstruents and laterals. Of special interest is the fact that the lateral alternates also in manner: the plain voiceless is a fricative /ɬ/ and the weakened form is the voiced approximant /l/, as illustrated in (4). In Welsh (Ball and Williams 2001) there are morphophonemic alternants known as 'soft mutation', involving the alternation of voiced-voiceless pairs. Part of these alternations involves the alternation of voiceless and voiced laterals and trills, with the voiceless counterpart being fricative [ɬ] and [r̥] (spelled 'll' and rh', respectively) and the voiced, simple sonorants [l], [r], see example (5).

(4) Kwakiutl (transcription simplified and converted to IPA) (Boas 1947)

/t͡s’ōɬ/	‘to be black’
/t͡s’ōlato/	‘black-eared’

(5) Welsh (Ball and Williams 2001)

llyfr [ɬ]	‘book’
ei lyfr [l]	‘his book’

Glides

In Northern and Central Standard Swedish preaspiration of voiceless stops following stressed vowels has commonly been observed. However, fricativization rather than aspiration is produced between long high vowels, which are diphthongal, and the voiceless stop (Millardet 1911; Rositzke 1940; Helgason 2002: 88). No frication, however, is found for the non-high vowels, see (6). Helgasson (2002) notes another factor which may contribute to the observed friction: the tendency to produce friction noise at the end of long, close vowels, regardless of whether or not a consonant follows (e.g. *bi* [bij̝] ‘honey bee’; *gud* [gʉ̟βd] ‘God’). Particularly, in a sequence of a long, close vowel and a voiceless stop, the early glottal abduction for the stop (i.e. preaspiration) during the preceding high vowel will enhance the tendency for friction by increasing the velocity of air across the oral constriction. Similar patterns have been reported for the Jutland dialect of Danish by Andersen (1972).

(6) Fricated glides in Swedish (Millardet 1911; Rositzke 1940)

bit [biçt]	‘a bit, a bite’
kut [kʉ̟ɸt]	‘seal puppy’

BUT:

tack [tahk:]	‘thanks’
peka [pehkɐ̞]	‘point’

Bauer (1982) reports fricativization of syllable-initial /w/ into either a voiceless bilabial fricative [ɸ] or a voiced labio-dental fricative [v] in Hong-Kong Cantonese. The voiceless realization – with increased flow of air through the open glottis – may well give rise to turbulence at the oral constriction, without changing the articulatory configuration. However, considering that a voiced realization may also give way to frication, in the pronunciation [v], we cannot dismiss the possibility that a closer articulatory constriction syllable-initially is responsible for the turbulence and spirantization.

Some varieties of American English retain the older pronunciation [ʍ], a labial velar fricative, in words such as *which, whether*, and *white*, rather than the voiced approximant [w]. The correspondence of the voiceless fricative and the voiced glide illustrates the covariation between voicing and frication.

Nasals

In voiced nasals, vocal fold vibration resonates in the oral and nasal cavities resulting in the low frequency resonances and zeroes characteristic of nasals. Voiceless nasals, on the other hand, have an open glottis for most of the oral closure giving rise to turbulence generated primarily at the nostrils (the point of maximum constriction) no matter what the place of articulation of the nasal. Since the turbulence generated at the nostrils is not very intense and is not amplified and shaped by a downstream cavity[4], the frication has a weak intensity and there will not be much spectral difference between [m̥ n̥ ɲ̥ ŋ̊] during the consonant constriction, though of course they can be still be differentiated by their transitions in adjacent vowels. Although Maddieson (1983) reports that spectral differences may be found during the voiceless portion, nevertheless, as noted, place distinctions in voiceless nasals are obscure. Even so, many languages with distinctive voiceless nasals usually have them at more than one place of articulation. These voiceless nasals usually have a brief voiced period in the last portion of the oral closure, thus they are phonetic sequences [m̥m], [n̥n] etc. In this way different places of articulation can be differentiated – both by the distinctive resonances of the voiced nasal and the transitions in adjacent vowels.

Evidence that voiceless nasals are obstruent-like, specifically fricatives, is provided by Ohala and Ohala (1993). First, distinctive voiceless nasals frequently derive from original /s/+nasal clusters. This is the case for Burmese, where present-day /n̥a/ ‘nose’ stems from Proto-Burmese-Loloish **sna* (Bradley 1979), and corresponds to orthographic *sna* in Tibetan. Parallel cases are found in Primitive Greek and Old Irish, where /m̥/ and /n̥/ seem to derive from Indo-European *sm, *sn. Second, children learning English sometimes produce target #*sm*- and #*sn*- as [m̥] and [n̥], e.g. [m̥æk] *smack* and [n̥id] *sneeze* (references in Ohala and Ohala 1993: 233). These cases suggest that the voiceless nasal is an adequate auditory substitute for a fricative.

3.1.1.2. *Emergence of 'buccal' fricatives due to /h/ coarticulating with high vowels*

A related phonological pattern, due to the interaction of aerodynamic and acoustic factors, is the emergence of supraglottal fricatives when /h/ is coarticulated with high vowels. For example, the glottal fricative /h/ in Japanese has distinct palatal, [ç], and labial, [ɸ], variants before high front and high back vowels, respectively, and [h] before more open vowels. This is illustrated in (7) below. (Note that high vowels are allophonically devoiced between voiceless sounds and optionally devoiced when following a voiceless consonant or utterance-finally):

(7) Japanese variants of /h/

hikaku /hikakɯ/	[çi̥kakɯ]	'comparison'
futa /hɯta/	[ɸɯ̥ta]	'lid'
BUT:		
happyaku /happjakɯ/	[hapʲ:akɯ]	'eight hundred'

In present-day English, /h/, which is the voiceless version of the following vowel or glide (Lehiste 1964, ch. 5), has the fricative allophone [ç] before /j/ and /i:/, *Hugh* (a name) [çju:], *heal* [çi:ɫ]. Similarly, in Fante /h/ is phonetically [ç] before /ɪ/ (Schachter and Fromkin 1968). The emergent fricatives in Swedish, illustrated in (6) above, are a further example.[5]

In segments involving a glottal and a supraglottal constriction, and turbulent noise generated at both of these constrictions – as is the case for voiceless high vowels and sonorants – acoustic factors are also relevant. In general terms, in sounds with more than one constriction, the major noise source is contributed by the constriction with a smaller cross-sectional area, other things being equal (e.g. smoothness of the channel surface, presence of an obstacle). The area of supraglottal constriction for high vowels like [i] and [u] is about 0.2 to 0.3 cm^2 (Chiba and Kajiyama 1941; Fant 1960; Baer et al. 1991), and hence comparable to the area of the open glottis, approximately 0.3 cm^2 (Stevens 1998: 37). Thus, if the two constrictions are of about the same size ($A_o = A_g = 0.3$ cm^2) the relative noise level at each constriction is about the same. However, the transfer function of the two sources is not the same, the major noise source is contributed by the outer constriction (Stevens 1998: 442-3). This is because the cavity anterior to the outer constriction enhances the amplitude of the frequencies of the front-cavity resonance for [i] (F3 peak), and thus palatal frication dominates over glottal frication. Similarly, for [ɯ] and [u] the amplitude of the labial or velar frication (F2 peak) is greater. Hence the distinct 'buccal' fricatives result from voiceless high vowels.

3.1.1.3. Stop releases engender frication on adjacent glides, high vowels and sonorants

A stop release is in itself a brief period of turbulence noise – due to the high rate of airflow which arises from the high back pressure developed during the consonantal closure. However in some cases the turbulence noise is prolonged at the release of a stop followed by a high vowel or a glide and this leads to the emergence of a fricative (Stevens 1971, Ohala 1983b). When a stop is followed by segments involving a high tongue position, such as high vowels and glides (and to a certain extent liquids /r, l/), the air pressure build-up behind the stop constriction is released through a narrow channel (A) which offers a high resistance to exiting air and thus increases the particle velocity and turbulence (by equation (1)). It can take a few tens of milliseconds for the P_{oral} to approach $P_{atmospheric}$ and during this time the air will be forced through the constriction at a higher rate. Hence the initial portions of the vowel or glide can be fricated. The phonologization of the stop as an affricate is due to the listener parsing the prolonged frication with the stop, not the vowel. Such emergent affricates do not develop before more open vowels with a wider constriction.

In addition, onset of vocal fold vibration after a stop is delayed in high vowels and glides vis-à-vis open vowels (Ohala 1976, 1981a, 1983b, Chang 1999), due to the slower release of the oral pressure through the narrow constriction and the longer time needed to achieve the pressure differential for voicing. This results in a longer period of turbulent flow, which contributes to the percept of frication. This is illustrated in Figure 1 which shows that the oral pressure impulse from the stop decays more slowly for /tjɑː/, /twɑː/, /trɑː/ vis-à-vis /tɑː/ due to the exiting air encountering greater resistance. For these sequences the high velocity airflow passes through a narrow constriction, enhancing turbulence, for a relatively long time creating the percept of an affricated stop release.

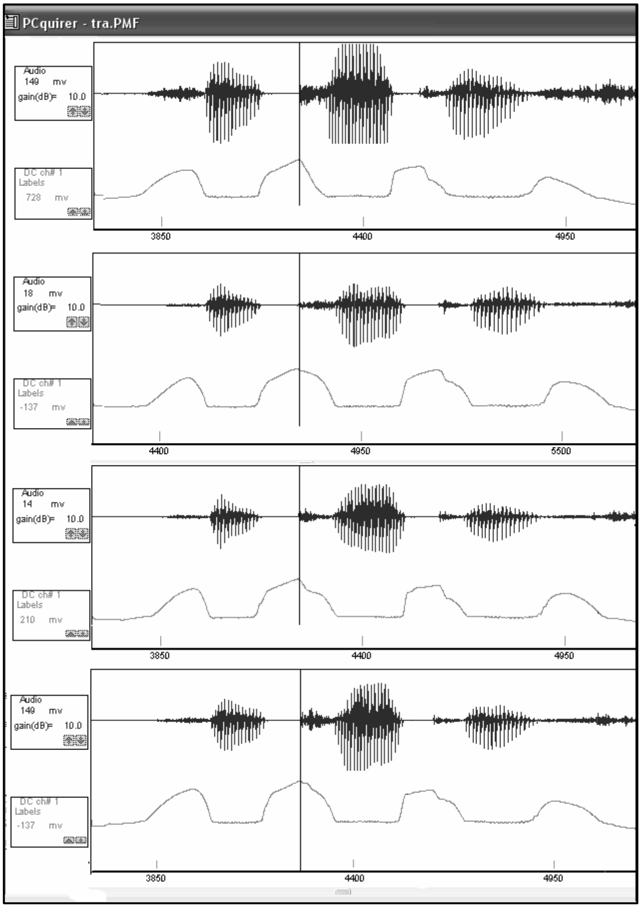


Figure 1. Audio and oral pressure for aspirated /tɑː/, /tjɑː/, /twɑː/, /trɑː/ in the carrier sentence ‘Say__twice’. The vertical line is placed at the [t] release. (See text).

Although most phonological descriptions describe this process as the *stop* becoming affricated, synchronic phonetic evidence (illustrated below) suggests that it is the initial part of the close *vowel* or *glide* that becomes (devoiced and) fricated immediately after the stop release. Frication cannot be attributed merely to the narrower aperture of the close vowel or glide immediately after the stop release because (af)frication is not found in words such as *canyon, million, saviour,* or *burial*, where the same glide or vowel apertures are involved but there are different preceding consonants. As there is no considerable pressure build up during the production of C1 in /nj, lj, vj, rj/ sequences, the air velocity is not high enough to cause frication noise. Thus, it is the joint effect of a vowel or glide with a narrow aperture and the pressure build up for the stop which are responsible for the resulting frication.

Most likely the same principles are responsible for the historical fricativization of high front vowels following sibilants in the transition from Middle Chinese to Modern Mandarin, for example, */si/→[sz̩] 'poetry', */ʂi/→[ʂʐ̩] 'lion' (Chen 1976). The high pressure build-up for the voiceless sibilant is released through the narrow constriction for the high vowel, generating frication.

The affrication of stops followed by high vowels and glides (i.e. segments impeding the free flow of air) is at the origin of a number of cross-linguistic sound patterns and present-day synchronic variation. Some of these are illustrated in (8) and (9) (see Hall and Hamann 2006 for a typological study of alveolar assibilation in 45 languages):

(8) Affrication of alveolars/dentals
Historical

a. English /tj dj/ > /t͡ʃ d͡ʒ/
/t͡ʃ/ *actual, nature, mature, picture*
/d͡ʒ/ or /dj/ *residual, soldier, remedial*

b. Latin /tj dj/ > Old Catalan /t͡s d͡ʒ/ > Mod. Catalan /s (d)ʒ/
Latin *petia* > *peça* ['pesə] 'piece'
Latin *diurnum* > *jorn* [ʒorn] ~ [d͡ʒorn] 'day'

c. Japanese /t d/ > [t͡ɕ d͡ʑ]/__ /i/; /t d/ > [t͡s d͡z]/__/u/ (=[ɯ])

/tja/ > /t͡ɕa/ [t͡ɕa] *cha*	'tea'
/tii/ [t͡ɕii] *chii*	'social status'
/di/ [d͡ʑ] [d͡ʑiɽemma]	'dilemma'
/katu/ [kat͡sɯ̥]	'win' (pres.)

d. Ikalanga (Mathangwane 1996, cited in Ohala 1997a), stop frication before the high front vowel /i/, and distinctive aspiration before high vowels /i, u/, but not before the next lower vowels /ɪ ʊ/. The high and high-mid vowels have now merged.

Proto Bantu
*/tima/ > Ik. /t͡sʰima/‘well’

BUT:
PB*/tɪma/ > Ik. /tima/ ‘heart’

Proto Bantu
*/tudi/ > Ik. /tʰud͡zi/‘shoulder’

BUT:
PB */tʊndʊ/ > Ik. /tundu/ ‘basket’

Synchronic

e. English
[tᶴj] or [t͡ʃ] *Tuesday, tune, got you*
[tᶴr] *train, truck*
[tˢi:] *tea*

f. Brazilian Portuguese (Albano 1999)[6]
/di/ [gə̃nd͡ʒi] *Ghandi*
/ti/ [internɛt͡ʃi] *Internet*

g. Italian dialects (Tuttle 1997)

Standard Italian /tj/ *tieni* ‘hold’(imperative)	Venetian [t͡ʃeŋ]
Standard Italian /ti/ *alti* (plural -i) ‘tall’ (pl.)	Ticino [alt͡ʃ], [ɛlt͡ʃ]

(9) Velar softening [k] /[g] + [i ɪ e y j] > [t͡ʃ t͡s ʃ s] / [d͡ʒ d͡z ʒ z]
Historical

a. English
O.E. *ciele, cele* [k] > *chill* [t͡ʃ] (but ‘cold’ [k])
O.E. *cirice* [k] > *church* [t͡ʃ] (cf Scots ‘kirk’ [k])
Gk. *gymnasion* [g] > *gymnasium* [d͡ʒ]

b. Italian
Lat. *cena(m)* [k] ‘dinner’ > *cena* [ˈt͡ʃe:na]
Lat. *regia(m)* [g] ‘palace’ > *regia* [ˈredd͡ʒa]

c. Tai (Li 1977: 221)

Lungchow	Po-ai	
kjau	*t͡ʃau*	‘head’, ‘knot of hair on top of the head’
kjaa	*t͡ʃaa*	‘rice seedlings’
kjooŋ	*t͡ʃooŋ*	‘drum’

Synchronic

d. Paduan affrication (Krämer 2004)

Standard Italian		Paduan	
ghiaccio	['gʲat͡tʃo]	['d͡ʒaso]	‘ice’
chiama	['kʲama]	['t͡ʃama]	‘he/she calls’

The ‘fronting’ of /k g/+/i j e/ sequences into alveolar or palatal affricates or fricatives has been the focus of much investigation (Grammont 1933, Bhat 1978, Guion 1998, to name a few). In contrast to articulatory accounts of the sound change, in terms of coarticulatory palatalization of the velar, Ohala (1989, 1992) provides an acoustic-auditory motivation for velar fronting. Chang, Plauché and Ohala (2001) present evidence that if the characteristic mid-frequency spectral peak of the burst in [ki] is degraded (and consequently perceptually missed), an alveolar sequence [ti] is reported by listeners, in line with confusion studies (e.g. Winitz, Scheib and Reeds 1972). Moreover, the [ki] tokens with the mid-frequency peak filtered out received better /t͡ʃ/ goodness scores than unfiltered tokens. Since variation of the acoustic cues of the stop burst influenced the direction of the consonant confusion – and paralleled the direction of the sound change – they argue that acoustic-auditory factors underlie ‘velar fronting’ (see also Ohala 1985, 1993). See Ohala (1983a, 1997c) and Plauché, Delogu and Ohala (1997) for asymmetries in the direction of confusion patterns and sound change, i.e. ki > ti but not the reverse.

A corollary to the generalization that stops tend to engender frication on adjacent glides and high vowels is 3.1.1.4.

3.1.1.4. Voiceless stops plus high vowels, glides and sonorants tend to be affricated more often than voiced stops

Another generalization that can be accounted by aerodynamic factors is that frication especially emerges after voiceless stops, though it may also emerge after voiced stops (Ohala 1976), as illustrated in (8) and (9) above.

The higher incidence of affrication in voiceless than voiced stops has been noted cross-linguistically (Bhat 1978 for velar stops; Hall and Hamann 2006 for dental-alveolar stops) and is illustrated in the English example (8a) above. While historically the /tj/ sequence in *nature, actual*, and *picture* has been lexicalized as /t͡ʃ/, in comparable voiced sequences, e.g. *soldier, medial, individual*, the /dj/ > /d͡ʒ/ change has not lexicalized as often, and these words may be pronounced either [dj] or [d͡ʒ]. Similarly, in German the sequence /tj/ developed into [t͡s] whereas /dj/ did not affricate (e.g. *nation* [na't͡sjo:n] vs *indianisch* [ɪn'dja:nɪʃ] 'indian'). The effect of the voicing of the stop on "stop assibilation" in Latin and Romance languages has been observed by a number of investigators. For example, Pope (1952: 129, 131) notes that, in Latin, [t͡sj] is attested for /tj/ in the 4th century (e.g. iusti[t͡sj]a 'justice') but this process did not affect /dj/. Affrication of /dj/ is not reported till Late Latin when palatalization and affrication of both /tj/ and /dj/ are attested. Hall and Hamann (2006) on the basis of observed assibilation of /tj/ and /dj/ sequences in 45 languages posit that "voiced stops cannot undergo assibilation unless voiceless ones do", such that there is an implication relationship /dj/ assibilation ⊃ /tj/ assibilation.

The differential effect of voicing in the stop was addressed at the beginning of section 3.1. To recall, the vibrating vocal folds for voiced stops constitute a relatively high resistance to air flowing from the lungs, and allow less air pressure to build up behind the stop constriction vis-à-vis voiceless stops, with an open glottis and a large and unimpeded flow. Consequently, there is a lower oral pressure at the release of a voiced stop, and thus less turbulence is generated (by equations 2 and 3). Empirical data corroborate that a higher amount of airflow and a longer duration of the release phase is found in /tjV/ than in /djV/ sequences (Hamann and Velkov 2005).

Since affricates frequently derive from stops with a long noisy release, a corollary to the diminished glottal flow and lesser affrication for voiced vis-à-vis voiceless stops, is the observation that voiced affricates are less frequent in languages of the world than voiceless affricates (ratio 1:3; Maddieson 1984: 38-39).

In sum, the generalization that "stops engender frication on adjacent glides, high vowels and sonorants" is moderated by voicing effects, with voiceless stops tending to be affricated more often than voiced stops.

3.1.2. Voicing impairs obstruency

For the same reasons that increased glottal flow during voiceless sounds favours supraglottal turbulence, the reduced glottal flow due to voicing, impairs the high intensity noise (and abrupt spectral discontinuities) characteristic of obstruency.

3.1.2.1. Voiced fricatives are hard to make; if voicing is strong, there is a tendency to de-fricate; if frication is achieved, there is a tendency to devoice

Voiced fricatives are relatively difficult to produce due to the antagonistic aerodynamic requirements for frication at the supraglottal constriction (high oral pressure) and voicing (low oral pressure) (note that this does not apply to the glottal fricative, [ɦ], for which both turbulence and voicing are generated at the vocal folds). Voiced fricatives require a pressure difference (ΔP) across the oral constriction sufficient to generate turbulence. This implies high oral pressure. That same high oral pressure, however, tends to impair the transglottal flow required for voicing. Thus, voiced fricatives involve very finely tuned aerodynamic conditions so that a pressure drop is maintained across both the glottal and the supraglottal constrictions[7] (Ohala, 1983b; Solé 2002b).

During the production of voiced fricatives, if voicing is present, the reduced transglottal flow due to the vibrating vocal folds tends to impair strong frication (as intensity of turbulence is proportional to rate of flow), and if strong frication is achieved, the high oral pressure will tend to impair vocal fold vibration. Thus, voiced fricatives tend to devoice or to defricate, as evidenced synchronically and diachronically in the patterns below.

Voiced fricatives are relatively rare cross-linguistically

The difficulty to produce simultaneous voicing and frication is reflected in segment inventories. Overall, voicing contrasts in fricatives are much rarer than in plosives, and they are found only in about a third of the world's languages as compared to 60 percent for plosive voicing contrasts (some languages, however, have voiced fricatives without corresponding voiceless fricatives emerging from weakened stops or fortition of initial approximants; Maddieson 2005). Furthermore, considering languages that utilize voicing with one of the obstruent types but not the other, the probability of vocal fold vibration being absent on fricatives is double than found for stops (Ohala 1983b: 201).

Voiced fricatives tend to defricate

For the same magnitude of the oral constriction, voiced fricatives have a lower intensity of friction than voiceless fricatives, which makes them more likely to be perceptually heard as frictionless continuants (e.g. glides, rhotics, approximants) or missed altogether. The reason is diminished airflow through the glottis for voiced vis-à-vis voiceless fricatives due to vocal fold vibration (i.e. increased glottal resistance), and the need to keep oral pressure low for voicing. In addition, voiced fricatives are known to be shorter than voiceless fricatives, thus they allow less time for air to accumulate behind the constriction and create a high pressure build-up. These mechanisms – lowered rate and duration of transglottal flow – are responsible for a lower oral pressure, and a lower intensity of noise vis-à-vis voiceless fricatives (by equations (2) and (3)). As a phonological consequence voiced fricatives resemble more closely the so-called "frictionless continuants" such as [j w ɹ], and, indeed, diachronically this is often their ultimate fate, as illustrated in (10a) and (10b). Approximants or frictionless continuants are the common phonetic manifestation of /v ð/ in Danish, e.g. [mað̞] *mad* 'food' (cf. OE *mete*, Swedish and Norwegian *mat*, Icelandic *matur*). In Spanish and Catalan, the medial voiced stops /b d g/ are "spirantized" to [β̞ ð̞ ɣ̞][8], see (10c), these latter sounds being more adequately described as approximants rather than fricatives (Martínez Celdrán 1991, 2004; Romero 1995). In Spanish these approximants may even disappear in some cases, for example, in past participles ending in -*ado*, e.g. *hablado* [a'β̞lao] 'talked' (cases of "hypercorrection", i.e. insertion of non-etymological [ð̞] in similar sequences, e.g. [β̞aka'lað̞o] for *bacalao* 'cod', are common).

(10) Defrication of voiced fricatives[9]

a. Gliding and vocalization

In Middle English the voiced velar fricative allophone of /g/ (a voiced stop in OE) became /w/ or /u/ and the palatal allophone either became /j/ ~ /i/ or was lost (Mossé 1968).

[ɣ] > [w]

OE *swelgan* > ME *swolwen* 'swallow'

OE *boga* > ME *bow* 'bow'

OE *sorg* > ME *sorow* ‘sorrow’

[ʝ] > [j], [i]
OE *genog* > ME *inough* ‘enough’
OE *mægden* > ME *maiden* ‘maid’
OE *sægde* > ME *said* ‘said’

b. S-rhotacism (Solé 1992)
Latin *cerasea* > Catalan *cirera* ‘cherry’
Prelit. Catalan *Tolosanu* > Catalan *Tolrà*, *Toldrà* (place name)
English *was – were* (< O.E. *wesan*)
English *lost – forlorn* (< O.E. *forleosan*)
Yurak *fire* ‘nest’ cf. Finnish *pesä*
Yurak *kuro*- ‘to cough’ cf. Lappish *gossâ*-

c. Spirantization
Spanish *sabe* /sabe/ ['saβ̞e] ‘(s)he knows’
Spanish *cada* /kada/ ['kað̞a] ‘each’
Spanish *pega* /pega/ ['peɣ̞a] ‘(s)he hits’

Voiced fricatives tend to be weakened or lost earlier than voiceless fricatives

Interestingly, historical data indicate that voiced fricatives tend to be weakened or lost earlier than voiceless fricatives. This is illustrated in fricative weakening[10] in Gallo-Romance. Preconsonantal /s/ was voiced before a voiced consonant, and [z] was weakened (into vowel, glide or tap) and lost as early as the 11th century, whereas voiceless [s] was pronounced well into the 13th century. Thus, for example, in Old French /s/ weakening and loss is found earlier in *blâmer* < *blasmer* < Lat. **blastemare* ‘blame’ and *mêler* < *mesler, medler* [ðl] < Lat. *misculare* ‘meddle’ than in *fête* < *feste* < Lat. *festa* ‘holiday’ and *epuzer* < *espozer* < Lat. **sponsare* ‘to marry’ (Pope 1952: 151, 449). Further, the different fate of etymological /s/ in the English words in (11) shows that [z] but not [s] had been lost at the time of the Norman Conquest, when the words were borrowed into English (Pope 1952: 151). The aerodynamic and acoustic differences for voiced as opposed to voiceless obstruents have phonological significance in a number of patterns below (3.2.1.; 3.3.1.2.).

(11)	[z] + voiced C	[s]+ voiceless C
	dine < O.Fr. *disner*	*feast* < O.Fr. *feste*
	hideous < O.Fr. *hisdos*	*espouse* < O.Fr. *espouser*
	male < O.Fr. *masle*	*esquire* < O.Fr. *esquier*

3.2. Changes in magnitude of the constriction

Intensity of turbulence is dependent on the shape and area of the constriction through which the air has to pass. In this section we will review sound patterns involving the generation or impairment of turbulence due to variations in the area of constriction. Such variations in the cross-dimensional area of the constriction may result from coarticulation with adjacent sounds or position in the syllable.

3.2.1. Lingual fricatives tend to weaken when followed by consonants involving conflicting tongue configurations

Lingual fricatives exhibit highly constrained articulatory, aerodynamic and time requirements (Bladon and Nolan 1977). They require articulatory positioning to form a constriction within a certain critical range (approximately $0.1cm^2$; Stevens 1998: 47) and creating sufficient pressure difference across the oral constriction to generate frication. This requires sufficient rate of flow through the glottis and sufficient time to build up oral pressure behind the oral constriction. Precisely because they are highly constrained, lingual fricatives allow less articulatory and aerodynamic variation – in magnitude and time – than other segment types.[11] Solé (2002a, b) has shown that apical trills have similar, if not more constrained positional, shape, aerodynamic and elasticity requirements, and that they do not allow much articulatory variation if trilling is to be present.

When lingual fricatives are followed by apical trills, involving conflicting positional requirements of the tongue-tip/dorsum – raised and advanced tongue dorsum and a central groove for /s z ʃ ʒ/ vs. predorsum lowering and postdorsum retraction with a lax tongue-tip touching the alveolar ridge for the trill (see Figure 2) –, anticipatory tongue gestures for the trill may perturb the critical articulatory configuration (i.e. cross-sectional area of constriction) and/or temporal requirements for the generation of turbulence, and the fricative may be weakened or lost.

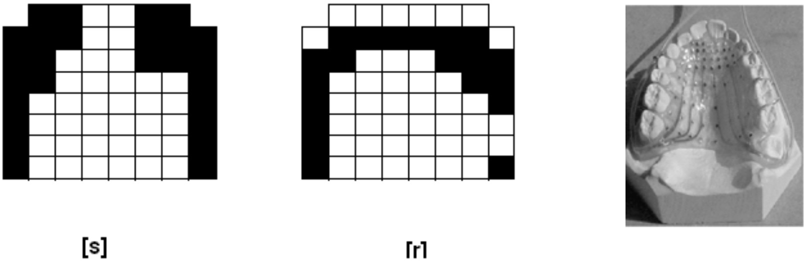


Figure 2. Left. EPG linguopalatal contact configurations for /s/ and /r/ between low vowels. Right. Artificial palate with the electrodes. Note that in the artificial palate the distance between the electrodes in the first four rows is much smaller than in the back rows. This difference in scale is not preserved in the grid on the left.

Palatographic and aerodynamic evidence for fricative-trill sequences suggests that early onset of movements for the trill – within 30ms from the onset of lingual movements for the fricative – perturbs the articulatory trajectory and the critical constriction area for friction. In cases of lesser overlap, motor commands for the trill may arrive after the articulator attains the cross-sectional area for frication, but within the time needed to build up sufficient pressure difference to create audible frication (approximately 50ms from onset of oral pressure rise for voiced fricatives, with increased glottal resistance, and 30ms for voiceless fricatives. In such cases turbulent noise will not be generated (Solé 2002b).

S-weakening (into a tap or assimilated to the following sound) is also common before the conflicting lingual fricative /θ/ in Spanish (for example, /sθ/ > [ɾθ], [θ:] in *ascenso* 'promotion', *piscina* 'swimming-pool', Navarro Tomás 1980: 111). Examples of fricative weakening and loss in lingual fricative–trill sequences[12] are illustrated in (12).

(12) Fricative weakening in lingual fricative – trill sequences

a. Iberian Spanish

/sr/ *dos-reales* [ˌd̪ore'ales]	'halfpenny' (Navarro Tomás 1980)
/sr/ *Osram* ['oram]	

/θr/ *voz ronca* [ˌbo'roŋka]	‘hoarse voice’
/θr/ *Cruz Roja* [ˌkru'roxa]	‘Red Cross’

BUT:

/sl/ *desleal* [d̪ezle'al]	‘disloyal’
/sb/ *esbozar* [ezβ̞o'θaɾ]	‘to sketch’
/sk/ *asco* ['asko]	‘disgust’
/θl/ *hazlo* ['aðlo]	‘do it’
/θm/ *voz melodiosa* [ˌboð melo'ð̞josa]	‘pleasant voice’
/θk/ *mezcla* ['meθkla]	‘mixture’

b. Catalan

/sr/ *les Rambles* [lə'rambləs]	(Recasens 1993)
/ʃr/ *mateix rotllo* [məˌ t̪e(j)'rɔd̠ʎu]	‘same story’
/ʒr/ *boig rematat* [ˌbɔ(j)d̠rəmə't̪at̪]	‘real crazy’

BUT:

/sl/ *fes-li* ['fezli]	‘do it’ (for him)
/sb/ *les bledes* [ləz'β̞lɛð̞əs]	‘the chard’
/st/ *costellada* [kust̪ə'ʎað̞ə]	‘barbecue’
/ʃʎ/ *mateix llit* [məˌt̪ eʒ'ʎit̪]	‘same bed’
/ʃp/ *mateix poble* [məˌ t̪eʃ 'pɔplə]	‘same town’
/ʒb/ *boig valent* [ˌbɔdʒβ̞ə'len]	‘brave crazy man’
/ʒp/ *boig per tu* [ˌbɔtʃpər't̪u]	‘mad about you’

c. Portuguese

/sr/ [ʒR] [ʒr] *dos reis* [du 'Rɐjʃ], [du 'rɐjʃ]	‘of the kings’
Israel [iRɐ'ɛɫ], [irɐ'ɛɫ]	

Note that in the languages illustrated in (12) coda fricatives are voiced before a trill due to regressive voice assimilation, and thus they are more likely to be defricated (see 3.1.2.1. above). Only the single trill realization is exemplified for the Spanish and Catalan data, but a long trill [r:] or a sequence [ɹ̝r] ([ɹ̝] = fricative r) are also possible. The Portuguese data in (12c) illustrate that the uvular and apical variants of trills – the former involving the tongue dorsum – assimilate predorsal fricatives. In Italian, on the other hand, the fricative is preserved (e.g. /s # r/ *autobus rosso* [ˌaut̪obus^{ə} 'rɔsɔ] ‘red bus’, *Israele* [iz^{d}ra'ɛlɛ], [iz^{d}ɾa'ɛlɛ]) most probably due to the widespread insertion of epenthetic sounds at consonant release which allows the sequencing (i.e. lack of overlap) of the gestures for the fricative and the trill. Whereas fricative to trill assimilation is probably an articulatorily gradient process (Solé 2002b), due to varying amounts of consonant overlap, the perceptual result is mostly categorical, that is, no frication is produced and the percept is commonly that of a trill or a long trill. Indeed, this process has led to reinterpretation in some place names in Catalan, where the fricative has disappeared in the lexical form, e.g. *Purroi* < etym. *Puigroig* /tʃ+r/ (Alcover and Moll 1979), *Puigreig* [pu'retʃ] < [ˌpudʒ'retʃ].

Additional evidence for the claim that the competing lingual requirements for trills impact on the generation of turbulence in preceding fricatives comes from (i) electropalatographic and acoustic evidence that lingual fricatives lose their frication and are lost more commonly before trills (and also laterals and nasals) than before voiced stops and fricatives (e.g. in Majorcan Catalan, Recasens 2006), and (ii) the more common and historically earlier weakening of lingual fricatives before trills (and also before laterals and nasals; see section 4) vis-à-vis other voiced consonants in Romance (Pope 1952: 151 footnote, 449; Rohlfs 1949; Torreblanca 1976; Recasens 2002: 352, 360).

3.2.2. Fricativization of syllable initial glides

Whereas the emergence of frication in glides in 3.1.1.1. and 3.1.1.3. above was attributed to an increased rate of flow through the glottis when devoiced or when coarticulating with adjacent voiceless consonants, the frication of [j] word and syllable-initially in dialects of Spanish, illustrated in (13a–13c), is most likely due to a narrower oral constriction (principle 2). (Such narrowing of the constriction may give rise to an affricate [dʒ], see (13b) and (13c)). There are two main reasons for this interpretation. One is that fortition (narrowing of a constriction) is common utterance and word-initially (Keating, Wright, and Zhang 1999; Keating et al. 2003). Second, that fricativization of [j] also takes place in voiced contexts (after a voiced consonant in Iberian Spanish, and intervocalically in Argentinian Spanish, as illustrated in (13b) and (13c)) and, therefore, cannot be attributed to increased transglottal flow. It is the case, though, that these emergent fricatives may be devoiced in Argentinian Spanish, i.e. [ʃ]. Most likely, the devoiced variant results from increased narrowing of the constriction area and passive devoicing (see Ohala's 'aerodynamic voicing constraint' 1983b). Similar cases of fortition are found during the development of the Romance and Germanic languages.

The labial velar glide [w] is also fricativized or stopped word-initially in Spanish being pronounced [ɣw], [gw] and, less commonly, [βw] (see 13d).

(13) Frication of glides ([j̝] = voiced palatal fricative)

a. *buey* [bwej] 'ox' vs *bueyes* ['bwej̝es], ['bweʒes] 'oxen'
ley [lej] 'law' vs *leyes* ['lej̝es], ['leʒes] 'laws'

b. Iberian Spanish (Navarro Tomás 1980: 127ff.)

yo [jo], [j̝o], [dʒo]	'I'
yeso ['jeso], ['j̝eso],['dʒeso]	'chalk'
cónyuge ['koɲj̝uxe], ['koɲdʒuxe]	'spouse'
subyugar [suβj̝u'ɣ̞aɾ], [subʤu'ɣ̞aɾ]	'subjugate'

c. Argentinian Spanish (Colantoni 2006)

ayuda [a'j̝uð̞a], [a'ʒuð̞a], [a'dʒuð̞a], [a'ʃuð̞a]	'help'
calle ['kaj̝e], ['kaʒe], ['kadʒe], ['kaʃe]	'street'

d. Iberian Spanish (Navarro Tomás 1980: 64)

huevo ['weβ̞o], ['ɣweβ̞o], ['gweβ̞o]	'egg'
huelga ['welɣ̞a], ['ɣwelɣ̞a], ['gwelɣ̞a]	'a strike'

3.3. Changes in oral pressure

Changes in oral pressure, such as those brought about by opening the velopharyngeal valve, will impact on the intensity of the resulting frication or stop burst. We review here a number of patterns illustrating that strong frication or high intensity noise bursts are difficult to achieve with concurrent or coarticulatory nasalization. Specifically, patterns showing that nasalization induces defricativization, and that buccal fricatives and stops do not tolerate nasalization. This leads to the generalization that nasalization impairs obstruency.

3.3.1. Nasalization and obstruency do not mix

Obstruents are characterized by high intensity frication and/or a noisy release burst. Since intense turbulence is pressure dependent (by equation (3b)), obstruents require a high build-up of air pressure behind the constriction in order to create audible turbulence when the pressure is released. If the obstruent constriction is downstream of the velopharyngeal port (i.e. “buccal”, labial to uvular), a tightly sealed velum is necessary to build up oral pressure. A lowered velum for nasality would vent the airflow through the nasal cavity, thus reducing or eliminating the required pressure difference across the oral constriction for intense turbulence. As a consequence, an open velopharyngeal port for nasality impairs high amplitude turbulence in buccal obstruents. In the case of glottal and pharyngeal fricatives and stops, for which the build up of pressure takes place further upstream than the velic valve, a lowered velum would not affect the pressure build up, and thus they can be nasalized (Ohala 1975, Ohala and Ohala 1993). Nasalized glottal fricatives, /ɦ̃/, have been widely reported in languages (Ladefoged and Maddieson 1996: 131–134), and they occur phonetically in American English, e.g. *home* [ɦ̃õʊ̃m]. Thus the requirement of a raised velum, and consequently the incompatibility of nasalization and obstruency, applies exclusively to obstruents articulated in front of the point of velic opening (buccal obstruents).

The antagonistic requirements of nasalization and obstruency, in addition to being predictable from aerodynamic and acoustic-auditory principles, have been demonstrated empirically. First, studies where oral pressure during the production of speech sounds was varied with a pseudo-velopharyngeal valve (a tube inserted at the side of the mouth via the buccal sulcus), simulating different degrees of nasalization (Ohala, Solé, and Ying 1998), show that in producing a fricative there can be some opening of the velic valve, but the resistance at the velum has to be high relative to the resistance in the oral constriction so that the air will mostly escape through the aperture with lower resistance and create friction at the consonantal constriction. If resistance at the velopharyngeal port is lower than that at the oral constriction the air will escape through the nose (i.e. the fricative will be nasalized), but supraglottal frication will be impaired. Ohala, Solé, and Ying (1998) argue that velic openings which do not impair the build up of pressure for audible turbulence would be insufficient to create the percept of nasalization in the consonant or even adjacent vowels. Shosted (2006) obtained similar results with a mechanical model of the vocal tract with which he generated fricatives with different degrees of velopharyngeal opening.

Second, studies on velopharyngeal impairment (e.g. as presented in clinical cases of cleft palate, see Gibbon and Lee this volume) suggest that a velic opening of less than 10mm^2 during the production of oral stops exhibits normal aerodynamic values and can be tolerated without any perception of nasality, but velic openings of 10-20 mm^2 show diminished pressure and airflow values and the perception of abnormal nasal resonance (Warren, Dalston and Mayo 1993).[13]

Third, studies on coarticulatory nasalization in obstruents (e.g. Rothenberg 1968, Cohn 1990, Ohala and Ohala 1991, Basset et al. 2001) show that (mostly voiced) obstruents exhibit coarticulatory velic leakage preceding and following nasal vowels and nasal consonants, but that the velum may close before the release, allowing pressure to build up behind the constriction so as to produce frication or an oral burst, or that the velum may be slightly lowered throughout the obstruent, resulting in a relatively weak frication or burst.

The evidence presented so far suggests that obstruency and nasality do not mix. To the extent that an obstruent is a good obstruent perceptually (i.e. with intense frication or noisy release burst), it cannot be a good nasal (i.e. with perceptible nasal coupling); to the extent that it is perceptibly nasalized, it does not have the high amplitude noise cues for obstruency. We review the phonological consequences of nasalization on fricatives and stops separately. We will address fricatives and nasalization first.

3.3.1.1. The rarity of nasal buccal fricatives

Languages of the world have sonorant nasal stops, nasal taps, nasal approximants, nasal glides and nasal vowels but no nasal fricatives. Segments reported as nasalized fricatives are more adequately described as (i) frictionless continuants or approximants, due to the lack of high frequency aperiodic noise (e.g. in Umbundu, Schadeberg 1982; Coatzospan Mixtec, Gerfen 1996; and Waffa, Stringer and Hotz 1973) or (ii) as sequences of nasal and fricative segments, i.e. prenasalized fricatives (in Bantu languages, Kwa languages, and Igbo; Welmers 1973: 70-73) (Ohala 1983b; Ohala and Ohala 1993; Ohala, Solé, and Ying 1998). As shown above, if turbulence is created further upstream the point of velic opening – as in glottal and pharyngeal fricatives – velic lowering is of no consequence to the pressure build up, and the fricative can be simultaneously nasalized (see Ladefoged and Maddieson 1996: 131–134 for examples).

3.3.1.2. Nasalization is associated with defricativization. The effect of voicing

Nasalized fricatives, though rare, have been reported to occur in languages and it has been observed that they tend to be defricated if voiced – evidencing the difficulty to produce simultaneous frication and nasalization with reduced transglottal flow for voicing – and to lose their nasality if voiceless. For example, *voiced* nasalized fricatives are phonetically nasalized frictionless continuants (e.g. Waffa /β̃ / [β̞̃], Stringer and Hotz 1973; Umbundu /ṽ/ [ʋ̞̃], Schadeberg 1982), and voiced fricatives tend to lose their friction due to spreading nasalization and become nasalized approximants (e.g. [v ɣ] ~ [ʋ̞̃ ɰ̞̃] in Guaraní, Gregores and Suarez 1967). In contrast, nasalized *voiceless* fricatives retain frication but do not differ much auditorily from non-nasalized fricatives, that is, the acoustic cues for nasalization are hardly detectable (Ohala 1975; Cohn 1993; Ladefoged and Maddieson 1996: 132).

The loss of frication in voiced but not voiceless nasalized fricatives follows from the aerodynamic factors reviewed. For the same degree of velopharyngeal opening, frication is more severely impaired in voiced than in voiceless fricatives.[14] This is so because voiced nasalized fricatives have two additional mechanisms, other than nasal venting, impairing strong frication: (i) increased glottal resistance – which results in a lower oral pressure and inhibits the air vented through the nasal passage to be resupplied from the lungs (as it is the case for voiceless fricatives with an open glottis) – and (ii) the need to keep oral pressure low for voicing. The differential effect of voicing in nasalized fricatives further illustrates the tendency for voicing to disfavour frication (section 3.1.2.1.). Other sound patterns illustrating the principle that nasalization induces defricativization are given in 3.3.1.3. and 3.3.1.4. below.

3.3.1.3. Failure of frication to emerge in a nasal context

Ohala (1983b: 205–207) and Ohala and Ohala (1993: 228) provide the following examples of frication failing to emerge in a nasal context.

1. In the development from Middle Chinese to Mandarin, high vowels become fricatives when preceded by a sibilant fricative (e.g. */ʂi/ → [ʂʐ̩] ‘lion’), as stated in section 3.1.1.3. above. However, vowel assibilation fails to occur when the vowel is followed by a nasal consonant and is, consequently, nasalized. For example, */ʂiəm/ → [ʂən] ‘forest’ but not *[sz̩n] (Chen 1976).

2. In English, /h/ has the allophone [ç] before /j/, as in *huge* [çju:d͡ʒ], as noted in section 3.1.1.2. above, but frication at the supraglottal constriction is not present if there is coarticulatory nasalization, e.g. *inhuman* [ĩn'ɦ̃ j̃ũmə̃n], not *[ĩn'çj̃ũmə̃n].

3. In Yuchi, voiceless fricatives appear predictably between all vowels and following lingual stops, but fail to occur if the vowel is nasalized (Wagner 1933).

3.3.1.4. Fricatives are weakened or lost more often when followed by nasal than by non-nasal segments

Coarticulation – the overlap of the articulatory configurations of contiguous segments – is well known and arises because it takes some minimum time to move articulators from one position to another. The antagonistic requirements of turbulence generation (a tightly closed velum to allow turbulent airflow in the oral tract) and nasal coupling (a lowered velum) in contiguous fricatives and nasals severely constrain the timing of velic movements if both segments are to be preserved. The relative phasing of velic and oral gestures in fricative + nasal (N or Ṽ) sequences have resulted in several sound changes, including (i) fricative weakening and loss, (ii) stop epenthesis, and (iii) vowel epenthesis (the latter two outcomes will be dealt with in section 4). Of interest here is fricative weakening or loss when followed by a nasal segment. Aerodynamic and acoustic data for fricative–nasal sequences shows that there might be anticipatory velopharyngeal opening for nasality during the acoustic duration of the fricative. Such nasal leakage diminishes the oral pressure build-up behind the fricative constriction, and attenuates the amplitude of frication, which may lead to fricative weakening or loss (Solé 2007a).

Although fricative weakening may also occur before non-nasals (e.g. Latin *misculare* ‘to mix’ > O.Fr. *mesler* > *mêler;* Germanic **bruzdon* ‘to embroider’ > Old Occitan *broidar* (where the ‘i’ is the result of the weakening process); Latin *festa* > French *fête* ‘holiday’), a number of scholars have noted that this process is favoured by a following voiced consonant and, in particular, by a following [n], [m], [r] or [l] (Pope 1952: 151 footnote, 449; Rohlfs 1949; Torreblanca 1976; Recasens 2002: 352, 360). Whereas /s/ weakening before [r] may be attributed to antagonistic positional requirements of the tongue-tip and blade (see section 3.2.1.), the weakening of fricatives before nasals (and laterals) may result from anticipatory velum (or tongue sides) lowering, thus affecting the aerodynamic requirements for the generation of turbulence.

Examples of fricative weakening due to coarticulatory nasalization resulting in vocalization or gliding (see 14a), rhotacism (exemplified in 14b), nasal assimilation (illustrated in 14c), and elision (see 14d) are found in historical sound change, morphophonological alternations and dialectal-stylistic variation.

(14) Examples of prenasal fricative weakening and loss

a. [ɣn] > [jn], [wn]
Latin *agnu* ‘lamb’, *ligna* ‘line’ > S. Italian dialects ['ajənə], ['lɛwna] (Recasens 2002)

b. [zn], [zm] > [ɾn], [ɾm]
Latin **dis(ju)nare* ‘to eat breakfast’ > Old Occitan *dirnar/disnar* (cf. Cat. *dinar*) (Grandgent 1905: 53)
S. Spanish *mismo* ['miɾmo] ‘same’ (Recasens 2002)

c. [zn̩] > [nn̩], [jn]
isn’t [ɪnn̩t], *ain’t* [eɪnt], *doesn’t* [dʌnn̩t], *wasn’t* [wɒnn̩t]
(Gimson 1962)

BUT:
is there? [ˈɪzðər], *is she?* [ˈɪʒʃi], *does he?* [ˈdʌzi]

d. [sm], [sn] > [m], [n]
IE **gras-men* > Latin *grāmen* ‘fodder’, English *grama, gramineous*

BUT:
IE **gras-ter* > Greek *gāster* ‘stomach’, English *gastric, epigastrium* (Watkins 1985)

IE **dhus-no* > Welsh *dwn* ‘dull, brown colour’
OE *dun(n)* ‘dark brown’

BUT:
IE **dhus-ko* > Latin *fuscus*, OE *dox,* English *dusk* (Watkins 1985)

3.3.1.5. Lower transitional frequency of fricatives followed by nasals

Fricatives combine less frequently with following nasals than with non-nasals. Solé (2007b) found a lower lexical frequency of word-medial fricative + N sequences than of comparable fricative + C sequences in English, German and Dutch (the languages available in the CELEX database). Similarly, Rossato (2004) reports a bias against fricative + N sequences in a cross-linguistic count in 14 languages. Thus the transitional probabilities in the sequencing of sounds reflect the constraint against fricatives followed by nasal segments that endanger their high airflow requirements.

In sum, the data in 3.3.1.1. to 3.3.1.5. show that fricatives tend to lose their frication more often and earlier when they are nasalized, when they occur before a nasal vis-à-vis an oral sound, and that they combine less frequently with following nasals than with non-nasals, illustrating the generalization that nasalization bleeds obstruency or, put another way, that fricatives do not tolerate nasalization.

3.3.1.6. In languages with nasal harmony, obstruents, including fricatives, block spreading nasalization

The incompatibility between obstruents and velic opening is evident in languages with nasal harmony. In such languages a nasal segment precipitates the spreading of nasalization to all following segments unless blocked by an oral obstruent. However, only buccal obstruents (labial to uvular), requiring a sealed velum, block spreading nasalization. The glottal obstruents [h ʔ], which do not require a raised velum since in their case frication is generated further upstream of the velic valve, do not block nasalization. This is captured by Schourup’s (1972) scale of permeability of segment types to nasalization (which ranks laryngeal obstruents low in the scale, next to vowels), but not by other hierarchies (e.g. Walker 1998).

Schourup (1972), Ohala (1983b), and Ohala and Ohala (1993) have pointed to a number of languages showing this patterning. The case of Sundanese, illustrated in (15a), shows that nasalization, following a nasal consonant, spreads until blocked by a buccal obstruent such as [k] or [s] or a sonorant[15]; however, it passes through the glottal obstruents [h ʔ]. Another example comes from Capanahua (a Panoan language of South America), where the leftward spreading nasal harmony from nasal consonants is blocked by buccal obstruents – such as [p], [b] or [s] – and /r/, but not by glottal stops, see (15b).

(15) Nasal harmony

a. Sundanese (Robins 1957)

ɲãh̃õkɤn	‘to inform’
mĩʔ̃ãsih	‘to love’
ŋãtur	‘to arrange’
mõlohok	‘to stare’
kumãh̃ã	‘how?’
bɤŋh̃ãr	‘to be rich’

b. Capanhua (Loos 1969)

t͡ʃipõnki	‘downdriver’
bãw̃ĩn	‘catfish’
t͡ʃĩʔ̃ĩn	‘by fire’
wɯrãnjasãʔ̃nwɯ	‘push it sometime’

3.3.1.7. *Consonants relying on high intensity noise cues, such as voiceless stops, do not tolerate nasalization*

When stops occur in a nasal context, partial or incomplete velopharyngeal closure during the stop constriction may vent the pressure necessary for a strong fricative release burst. Such nasal leakage would have a larger perceptual effect on voiceless than on voiced stops, as high intensity noise is a perceptual cue for voiceless stops (Ali, Daniloff, and Hammarberg 1979). In line with this, it has been noted that phonetically voiceless stops tend to inhibit coarticulatory nasalization, that is, they show a shorter temporal extent of velum lowering preceding and following nasalized vowels and nasal consonants vis-à-vis voiced stops (Rothenberg 1968: 7.4; Cohn 1990: 108; Ohala and Ohala 1991; Basset et al. 2001). Ohala and

Ohala (1991) provide an acoustic-auditory explanation for voiceless stops having less tolerance for nasalization than voiced stops in terms of nasal leakage undermining the stop or voiceless character (i.e. the spectral and amplitude discontinuity, and noisy release burst) of voiceless stops while voiced stops can meet their auditory requirements with a partially lowered velum. Such phonetic motivation is the basis for Pater's (1999) *NC̥ constraint.

The lower tolerance of voiceless stops to coarticulatory nasalization is evident in sound patterns showing (i) that if the nasal is preserved, the voiceless (buccal) obstruent is impaired (see (16)), and (ii) that if the obstruent is preserved, the nasal tends to be lost (see 17)).

(16) Nasals impair voiceless obstruents

a. Loss of voiceless but not voiced stops in a nasal context, e.g. Indonesian (Halle and Clements 1983)

/məN+bəli/ [mə**mb**əli]	'to buy'
/məN+dapat/ [mə**nd**apat]	'to get, to receive'
/məN+ ganti/ [mə**ŋg**anti]	'to change'

BUT:

/məN+pilih/ [mə**m**ilih]	'to choose, to vote'
/məN+tulis/ [mə**n**ulis]	'to write'
/məN+kasih/ [mə**ŋ**asih]	'to give'

b. Assimilation of nasality in /nt/ [ɾ̃] – but not /nd/ – clusters in American English, resulting in *winter* and *winner* being pronounced the same, e.g.

*ce**nt**er* [ɾ̃] vs *se**nd**er* [nd]
*i**nt**ernational* [ɾ̃] vs *i**nd**icational* [nd]

c. Glottalization of voiceless stops. In German and many dialects of English, a /t/ is realized as a glottal stop or irregular glottal pulsing when followed by a nasal, as exemplified below, whereas /d/ is preserved in the same context. In such contexts the voiceless stop would be nasally released and would lack the strong noisy release burst, which is a perceptual cue for voiceless stops. A glottal stop (with a constriction and build-up of pressure further upstream than the velic opening) allows velic lowering while showing a discontinuity in amplitude and a release burst characteristic of a stop (Kohler 2001).

German (Kohler 2001)
*zwei**t**en* [ˈt͡svaɪ̰**ʔ**n̩], [ˈt͡svaɪ̰n̩] ‘second’ vs
*lei**d**en* [ˈlaɪ**d**n̩] ‘pain’,‘to suffer’
American English
*Clin**t**on* [ˈkl̥ĩn**ʔ**n̩] vs *Bran**d**on* [ˈbræ̃n**d**ə̃n]
*cap**t**ain* [ˈkʰæp**ʔ**m̩] vs *Og**d**en* [ˈɑg**d**n̩]

Along the same lines, a tendency for /t/ to be more likely to be deleted than /d/ before a nasal (e.g. *sweeten* vs *Sweden*) in American English, due to the lack of a release burst in this environment, is reported by Zue and Laferriere (1979).

d. Postnasal voicing[16]

Phonological evidence of the tendency of voiceless stops to become voiced after a nasal is provided by languages with post-nasal voicing, as illustrated in (d.1) for Japanese; phonological alternations between voiceless stops and prenasalized voiced stops (e.g. Terena, where nasalization is affixed at the beginning of the word and spreads until an obstruent blocks it, and the obstruent becomes voiced in the process, see (d.2)); progressive voicing assimilation in stops following nasals, see (d.3); and historical sound change, for example, in the development from Classical Armenian to the Armenian language New Julfa, exemplified in (d.4).

Examples of post-nasal voicing

d.1. Japanese (Itô, Mester and Padgett 1995)

root	-te ‘gerundive’	-ta ‘past’
mi- ‘see’	*mi*+***t**e* ‘seeing’	*mi*+***t**a* ‘saw’
yom- ‘read	*yon*+***d**e* ‘reading’	*yon*+***d**a* ‘read’
root + root		
fumu + ***k**iru*		*fuŋ**g**iru* ‘give up’
fumu + ***h**aru* (from **paru*)		*fum**b**aru* ‘resist’

d.2. Terena alternations voiceless stops – prenasalized voiced stops (Bendor-Samuel 1966, cited in Ohala and Ohala 1993)

***p**iho* ‘I went’	ᵐ***b**iho* ‘he went’
*i**s**o* ‘I hoed’	*ĩ*ⁿ***z**o* ‘he hoed’
*owo**k**u* ‘my house’	*õw̃õ*ᵑ***g**u* ‘his house’

d.3. Progressive voicing assimilation in nasal+stop clusters (Rohlfs 1949: 88-89; Rohlfs 1970)

South. Italian dialects	
santo [ˈsan**d**ə]	‘saint’
pampano [ˈpam**b**anə]	‘hopscotch’
bianco [ˈjɛŋ**g**ə]	‘white’
Gascon	
candar [kanˈ**d**a]	‘to sing’ from Lat. *cantare*

d.4. Historical change (Vaux 1998: 506)

Classical Armenian	New Julfa
*ən**k**anel*	*ən**g**aniel* ‘fall’
*ajn**t**eɬ*	*ən**d**ieʁ* ‘there’
*t͡ʃan**t͡ʃ***	*t͡ʃan**d͡ʒ*** ‘fly’

Whereas the examples above illustrate that voiceless buccal obstruents do not tolerate nasalization, and that they tend to be lost, replaced or changed in a nasal environment, the examples in (17) illustrate that voiceless obstruents may preserve their spectral integrity (i.e. a strong release burst) by inhibiting coarticulatory nasalization, which accounts for the different fate of nasals in a voiced or a voiceless context.

(17) Nasals do not emerge or are lost next to voiceless but not voiced obstruents.[17]

a. Nasals occur before voiced but not before voiceless obstruents in the Kenlantan dialect of Malay (Teoh 1988), and in a number of African languages, such as Venda, Swahili and Maore (cited in Pater 1999: 319).

b. Nasals are deleted before voiceless but not voiced stops, e.g. in Mandar (Mills 1975).
/maN+tunu/ *ma**t**tunu* ‘to burn’
/maN+dundu/ *ma**nd**undu* ‘to drink’

A similar process is found in American English whereby nasals are lost before tautosyllabic voiceless stops but not before voiced stops.

American English nasal loss (Malécot 1960)
tent /tent/ [t^{h}ẽ**t**] *can't* /kænt/ [k^{h}æ̃**t**] *camp* /kamp/ [k^{h}æ̃**p**]

BUT:
tend /tend/ [t^{h}ẽ**nd**] ~ [t^{h}ẽ**n**]

c. In Hindi, nasals emerge between a nasalized vowel and a voiced but not a voiceless stop (Hindi - Ohala and Ohala 1991).

Sanskrit	Old Hindi	Modern Hindi
čhandra	*čhã:dra*	[t͡ʃã**nd**] 'moon'
danta	*dã:ta*	[dã**t**] 'tooth'

4. Acoustic-auditory factors. Basic acoustics of fricative production

As mentioned above in section 2, flowing air can have one of two states: laminar or turbulent. Turbulence will occur even in a smooth bore conduit if the air flows at a particular critical speed or even at lower speeds if it encounters anything which induces eddies in the flow, e.g. a barrier, rough surfaces, or another air jet – in short, any substantial resistance to smooth flow. Turbulence will also be created when airflow expands suddenly on exiting a narrow constriction – as is the case of fricatives. If the motion of air is sufficiently turbulent, i.e. intense, an audible sound is generated. This, essentially, is how stop bursts and fricative noises are produced and can be exploited to create different speech sounds. Fricative noise can be combined with periodic sound as in voiced fricatives but in this case the noise is pulsed at the same rate as that of the vocal cords (with each glottal pulse a puff of air is released and it is this higher-than-normal airflow which creates a momentary peak in the noise generated).

The intensity and thus audibility of the turbulent noise is determined by the degree of turbulence which in turn is determined by the velocity and random motion of the air stream and by the resonance cavities excited by the turbulent noise. In most cases of stops and fricatives, it is primarily the *downstream* cavities, if any, which have resonances that can affect the turbulent noise. An exception to this occurs in the case of non-speech whistling where the turbulence occurs at the pursed lips but it is the upstream cavities which resonate; thus different whistled frequencies are controlled by modifying the shape of the upstream, i.e. the buccal, cavity. The case of whistling is special also because there is a coupling between the resonator and the source, i.e. the resonator dictates, as it were, the frequencies dominant in the source.[18]

Given that turbulence noise is predominantly high frequency, higher output intensity usually results when the downstream cavity has high resonant frequencies and, therefore, amplifies the noise in the high-frequency range. This is the case with the anterior sibilant fricatives, such as [s] and [ʃ], with a relatively short downstream cavity, since, other things being equal, the resonant frequencies of a tube are inversely proportional to its length. Thus, apical to palatal articulations have resonances ideally matched to the inherent high frequencies of fricative sources; fricative articulations made further back in the vocal tract, e.g. velars, uvulars, etc. do not and this no doubt contributes to their relative infrequency vis-à-vis more forward fricatives. The above generalization included the qualifier "other things being equal"; among the "other things" that may not be equal is an additional constriction or narrowing in the downstream cavity, e.g. added lip constriction tends to lower resonant frequencies.

5. Generalizations on phonetic and phonological universals derived from acoustic-auditory principles

The following cases illustrate the role of acoustic-auditory factors in phonological patterns.

5.1. The effect of intensity

5.1.1. Voiceless stops

Among the voiceless pulmonic stops, the labial /p/ has the weakest release burst (and spectrally most diffuse) due to lack of a downstream resonator, and is thus less auditorily salient. /p/ is missing in many languages' sound inventories even though these languages may have pairs of voiced and voiceless stops at other places of articulation, e.g. Arabic and other Afro-Asian languages (Sherman 1975; Maddieson 1984: 35). /p/ can also be unstable, often changing to a voiceless fricative such as [ɸ]. In Japanese, for example, /p/ has a highly asymmetrical distribution. Unlike stops at other places of articulation, a voiceless bilabial stop is limited to loanwords (/pan/ 'bread'), onomatopoeic words (/pat͡ʃiŋko/ 'pin ball game'), and medial geminates (/tepːan/ 'iron plate'). Morphophonemic alternations reveal that there was an original /p/: reduplicated forms such as /hitobito/ 'people' (< /hito/ (now [çito] 'man, person')). The word-initial */p/ changed to /h/ (philological evidence reveals this to have the following path: /p/ > /ɸ/ > /h/) whereas the intervocalic /p/ changed to its voiced counterpart /b/ (as happens in general with voiceless obstruents that end up in such an environment due to morphemic concatenation, e.g. /tokidoki/ 'sometimes' < /toki/ 'time, hour').

5.1.2. Voiceless fricatives

Similarly, among voiceless fricatives, the bilabial [ɸ] and labial-dental [f] are less frequent in languages of the world in comparison to the more common and louder sibilants [s] and [ʃ] (Maddieson 1984: 45). As mentioned earlier, sibilants have high intensity partly because they have some downstream resonating cavity in the space between the point of constriction and the teeth and because the air jet passing through the apical-palatal groove also strikes the incisors which produces added turbulence (known as 'obstacle turbulence').

5.1.3. Voiced fricatives

The aerodynamic factors responsible for a reduced intensity and lesser perceptibility of frication in voiced vis-à-vis voiceless fricatives, and the phonological consequences, were addressed in 3.1.2.1. above.

5.1.4. Non-pulmonic obstruents

But the physical intensity cannot be the only factor determining the frequency of occurrence of obstruents in languages' segment inventories. If it were then more languages would have ejectives (glottalic egressives) or clicks (velaric ingressives) than is the case. Since the magnitude of the pressure differential between the oral cavity and the atmosphere is what determines the intensity of the turbulence, these segment types, clicks in particular, would have wider incidence. Here the concept of *Maximum Utilization of Available Features* (MUAF; Ohala 1979; Schwartz, Boë, and Abry 2007) comes into play. The MUAF principle posits that sound systems are not only shaped by perceptual-motor factors, such as maximization of perceptual dispersion and perceptual contrast (Lindblom 1986, 1990) or quantal effects (Stevens 1972, 1989), which would predict the use of multiple contrastive features. Instead, systems tend to limit their use of phonetic features, such that a given feature tends to combine systematically with the existing features in the system, thus maximizing the use of the available features. New features or segments can arise via sound change as modifications of existing segments. For example, ejectives and glottalized stops may arise from sequences of pulmonic stop + glottal stop (see Ohala 1995). But this point needs further research.

Nevertheless, though non-pulmonic obstruents are less frequent than pulmonic, within the ejective stops, labial ejectives have a lesser incidence that non-labial ejectives, paralleling the distribution in voiceless pulmonic stops (Greenberg 1970: 127, Maddieson 1984: 103).

5.2. When generating noise, labiovelars behave as labials

The labiovelars [w], [ʍ], [k͡p], [ɡ͡b] and [ŋ͡m] are doubly-articulated consonants with two simultaneous primary constrictions, labial and velar. In spite of their two constrictions, in certain cases these sounds pattern as labial, and in other cases as velar (Ohala and Lorentz 1977; Ohala 2005). Of interest here is that when generating noise (frication or stop bursts) labiovelars tend to behave as labials.[19] For example, in Sentani (Cowan 1965) /h/ is realized as [s] after certain sounds (the vowel /i/, nasals and glides); however, after the labiovelar glide /w/ it is optionally realized as labial [f] or [s], e.g. *kɛwfike* or *kɛwsik*e, but not **kɛwhike* 'he threw away' (aorist). In Tenango Otomi the /h/ before /w/ is realized as the voiceless labial fricative [ɸ] (Blight and Pike 1976). The labiovelar glide in the borrowed French word *lieutenant* is pronounced as a labial fricative in British English, [lefˈtenənt]. Similar evidence is found in a wide variety of languages (see Ohala and Lorentz 1977: 587 for a list of languages and sources). In addition, auditory impressions of the perceptual dominance of the frication produced at the labial constriction over that produced at the velar constriction are reported by Pike (1943: 132) and Heffner (1964: 160) among others.

The reason why in a fricativized labiovelar, with turbulent airflow produced at each of the two strictures, the noise generated at the labial constriction dominates is provided by acoustic factors. It is known that the intensity of a sound is a function of its inherent intensity and its transfer function, that is, the way the resonating cavities the sound passes through modify the intensity at various frequencies. Since turbulent noise is inherently high-frequency, the noise at the velar constriction would be attenuated by the low-pass filtering effect of the downstream resonator (Fant 1960; Stevens 1971, 1998), whereas the labial noise source would not have such attenuation.

6. Timing of the articulators

Stops emerging in the transition between nasals or laterals and adjacent consonants have been extensively studied (see references in Ohala 1995). Ohala (1983b, 1997b) proposed a unified account of emergent stops in terms of variation in interarticulatory timing, specifically, when the articulatory configurations for adjacent segments overlap, they may result in transitional stops.[20]

6.1. At the junction of nasal and laterals, stops emerge

The vocal tract has two major exit valves for the pulmonic airflow, the oral and the nasal passage. For a nasal consonant, the oral passage (controlled by the tongue or lips) is closed and the nasal passage (controlled by the velum) is open; for an oral consonant it is the reverse. Stops, by definition, have all exit valves closed. Laterals and apical fricatives may also be considered as having two independent exit valves: the tongue sides and the apex, see Figure 3, bottom.

For a lateral, the lateral valve (controlled by one or both sides of the tongue) is open and the apical valve is closed. For the central fricatives like [θ s ʃ] (and to some extent for trills) it is the reverse, the tongue sides are raised and help to channel the air through the midline opening. As shown in the figure, sequences such as [mt] and [ls] require a simultaneous and opposite change of state of the two exit valves (cf. Figure 3a and 3c). If, in the transition between these segments, both exit valves are closed (Figure

3b), then air flowing from the lungs accumulates in the oral cavity, oral pressure rises, and when the oral constriction is released it causes a burst and an obstruent is created.

The place of articulation of the epenthetic – or better, the "emergent"[21] – stop, i.e. its release, will be at the valve which is the first to open. In NC clusters, any epenthetic stop will be homorganic with the nasal since the oral constriction for the nasal is released first. In /ls/ and /lr/ clusters, the first valve to be released after the transitional stop is that of the second member of the cluster and the emergent stop will be homorganic with C2. (It may be difficult to evaluate this latter claim since the [l], the first member of the cluster, is necessarily homorganic or near-homorganic to the second member).

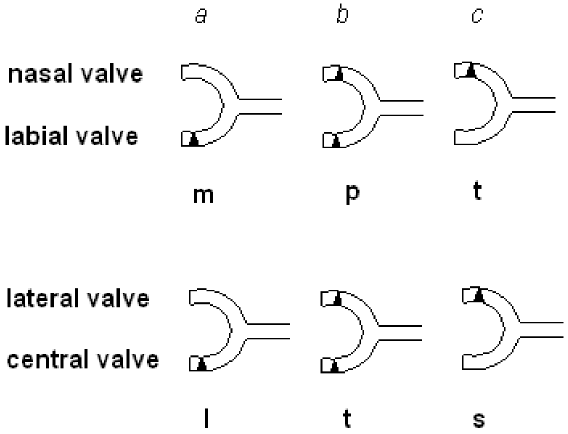


Figure 3. Schematic representation of the vocal tract with independent airways. Top: valvular configuration for a labial nasal (a), for a heterorganic oral obstruent (c), and for the emergent stop, with both airways closed (b). Bottom: valvular configuration for a lateral (a), for a central fricative (c), and for the transitional stop (b).

6.1.1. Emergent stops in nasal clusters

As shown in Figure 3b (top), such transitional stops involve denasalization of the latter portion of the nasal due to anticipatory velic closure when the oral constriction for the nasal has not yet been released. Such early velic closure may be required for (i) aerodynamic reasons or for (ii) acoustic-auditory reasons. (i) If an obstruent like [θ s ʃ], a heterorganic stop or a trill follows, the velopharyngeal valve must be closed in order to build up pressure behind the oral constriction and generate continuous noise in the case of fricatives, transient noise in the case of stops, or to set the tongue tip into vibration for trills. An early velic raising (and anticipatory glottal abduction) during the oral constriction for the nasal, will ensure sufficient time and rate of flow to create a pressure differential across the oral constriction for turbulence or tongue-tip vibration (Ali, Daniloff, and Hammarberg 1979). Indeed the requirement of a closed velum will only apply to "buccal" obstruents, as explained in section 3.3.1., and hence these are the only segments that will trigger emergent stops. Examples of epenthetic stops emerging in nasal–obstruent and nasal–trill sequences (or nasal–rhotic; see below for other variants of rhotics) are given in (18).

(ii) A segment may require a closed velopharyngeal valve due to acoustic-auditory factors since nasal coupling would distort the acoustic characteristics of the sound. This is the case for any distinctively oral segment, including oral (vs nasal) vowels, and any segment, distinctively oral or not, that has a low first formant – such as [l w i u] – since this formant would be most distorted by nasal coupling, the effect of which is seen primarily in the low frequencies (Fant 1973; Fujimura and Lindqvist 1971; Bell-Berti 1993); accordingly it is vowels and sonorants with a low F1 which tend to trigger an early velic raising and emergent stops (Ohala 1975). Taps may have the same acoustic motivations as [l] for remaining oral if they are to maintain contrast with [n] (for nasalized laterals alternating with nasals, see Cohn 1993 and Ohala 1975; examples of nasalized taps alternating with nasals are the sound change /n/ > /ɾ̃/ in Middle Indo-Aryan (Skt. *manah* > dial. MIAr. *mano* [maɾ̃o] (Hock, 1986: 82), and /n/ > /ɾ/ in Rumanian, presumably through [ɾ̃] (Rosetti 1978; Sampson 1999)). Examples of segments like taps, [l], high vowels, and distinctive oral vowels promoting emergent stops in adjacent nasals are given in (18c), (19a) and (19b).

(18)

a. Nasal – fricative sequences

English	*Hampstead, Hampshire* < Old English *hām + stede*, *sćīr* *once, sense, prince* [nts], *Banff* [mpf], *warmth* [mpθ], *strength, length* [ŋkθ]
Eastern Catalan	*anxova* [əɲ'ʃɔβ̞ə] ~ [əɲ't͡ʃɔβ̞ə] 'anchovy'

Eastern Catalan	*menjar* [məɲ'ʒa] ~ [məɲ'd͡ʒa] ‘to eat’
Central Italian	*penso* ['pɛn͡t͡so] ‘I think’ (Busà 2007)
Dutch	*langs* [laŋs] ~ [laŋks] ‘along’ (Warner and Weber 2001)

b. Heterorganic Nasal – stop sequences

English	*empty* < Old English *ǣmtiǵ*
	peremptory < Middle French *peremtoir*
	Hampton < O.E. *Hamtun*
	dreamt [dremt] ~ [drempt]
Dutch	*hangt* [haŋt] ~ [haŋkt] ’hangs’ (Warner and Weber 2001)
Catalan	*comte* ['komtə] ~ ['komptə] ‘count, earl’
Latin	*prom-p-tus* < past participle of *promere*
	exem-p-tus < *eximo* ‘take away’ (Meillet and Vendryes 1924: 82)

c. Nasal – rhotic sequences

Spanish	*vendrá* < Latin *ven(i)re* ‘he will come’
	Alhambra < Arabic *al hamra* ‘the red’
Catalan	*cogombre* < Latin *cucumere* ‘cucumber’
	cendra < Latin *cinere* ‘ash’
French	*chambre* < Latin *cam(e)ra* ‘chamber’
	gendre < Latin *genere* ‘gender’
Cl. Greek	*andros* < *an(e)ros* ‘man’
English	*thunder* < O.E. *þunor*
	slumber < Middle English *slumeren*, O.E. *sluma*
Pali	*amba* < *ambra* < Sanskrit **āmra*
	tamba < *tamb(r)a* < Sanskrit **tāmra* (Oberlies 2001)
Swedish dialects	Pernå /semberi/ for Standard Swedish /semre/ ‘worse’ (Ivars 1996, cited in Engstrand et al. 1998)

(19)

a. Nasal – lateral sequences

Swedish dialects	Lappfjärd /sa:mblast/ for Standard Swedish /samlades/ ‘gathered’ (Ivars 1996, cited in Engstrand et al. 1998)
English	*spindle* < O.E. *spinel, bramble* < O.E. *brēmel, humble* < O.Fr. *humble* < Latin *hum(i)lis* (Mossé 1968)
Spanish	*temblar* < Latin *trem(u)lus* ‘to shiver’
Catalan	*semblar*, French *sembler* < Latin *sim(u)lare* ‘to seem, appear’
Latin	*templum* <**tem-lo* ‘a section’ *exemplum* <**ex-em-lo* ‘a sample’ (Meillet and Vendryes 1924: 83ff)

b. Nasal – high vowel or nasal – oral vowel sequences (see Ohala 1983b for citations to the source data)

Ulu Muar Malay	ban ~ ban^{d}ʊ ‘doorsill’
Korean	mul ~ m^{b}ul ‘water’

BUT:

Korean	mal ‘language’

Telefol	/su:m/ [su:^{b}m] ‘banana’
Tenango Otomi	/mohi/ [m^{b}ohi] ‘plate’ /nĩne/ [nĩn^{d}e] ‘your mouth’
Parintintin	/õmoapɨ/ [õ^{m}boapɨ] ‘he cooks’ /ɲãnu/ [ɲãndu] ‘spider’

In the case of reverse sequences, specifically fricative-nasal sequences, variations in the relative timing of velic opening for the nasal results in several outcomes historically, including (i) fricative weakening and loss (that we suggest, results from anticipatory velic opening as reviewed in section 3.3.1.4.), or (ii) preservation of the fricative but the emergence of an epenthetic stop or an epenthetic vowel. A stop emerges due to a delayed velic opening, i.e. a prolonged velic occlusion of the fricative during the oral constriction for the nasal (e.g. Middle English *listen* < O.E. *hlysnan*; Sanskrit *kṛṣṇā* > *Krishna* ~ *Krish**t**na, grīṣma-* ~ *grīṣ**p**ma-* ‘heat’; Ohala 1997b). Similarly, the insertion of an epenthetic schwa in /sm̰/ > [s^{ə}m̰] and

/sn̰/ > [sᵊn̰] sequences in Montana Salish (Ladefoged and Maddieson, 1996: 109–110) reflects a delayed velic lowering and oral closure for the nasal relative to the end of the fricative. In both cases, the delayed opening of the velic valve preserves frication. Parallel patterns in the timing of the velum and the oral articulators in fricative–nasal sequences are found phonetically (Solé 2007a).

6.1.2. Emergent stops in lateral clusters

As mentioned, laterals and fricatives have opposite requirements for the lateral and apical valves (Figure 3 bottom). The relative phasing of the exit valves (closure of the lateral valve and release of the central valve) for these segments may result in a transitional state where both valves are closed, oral pressure builds up and a stop burst is produced. Stops emerging from laterals require that the fricative is homorganic or nearly so, that is, that the two sounds in sequence have complementary exit valves. Emerging stops in homorganic lateral-fricative clusters are common synchronically and diachronically, as exemplified in (20).

Emergent stops from reverse lingual fricative-lateral sequences, resulting in a laterally-released stop, have been attested in a variety of languages (see Ohala 1997b, 2005 and references therein) and are illustrated in (21).

(20) Lateral – fricative sequences
  a. English
    false [fɒɫᵗs], *else* [ɛɫᵗs], *pulse* [pʰʌɫᵗs], *Elsie* ['ɛɫᵗsi]

  b. Eastern Catalan
    àlgebra ['aɫʒəβ̞ɾə] ~ ['aɫd͡ʒəβ̞ɾə] 'algebra'
    àlgid ['aɫʒit] ~ ['aɫd͡ʒit] 'culminating' (adj.)

  c. Kwakiutl (Boas 1947)
    k!wēɬt͡soᵋ < k!wēɬ-soᵋ 'to be feasted'
    lɛg.wi´ɬ͡tsa gōkᵘ < lɛg.wi´ɬ-sa gōkᵘ 'the fire of the house'
    maᵋɬ͡tsɛ´m < maᵋɬ-sɛ´m 'two round ones'

(21) Fricative – lateral sequences
  a. English
    hustle < Dutch *husseln, hutselen*
    wrestle < O.E. **wræstlian*, Cf. N.Fris. *wrassele*

  b. Greek
    hestlos < heslos (Wetzels 1985)

  c. Italian
    schiavo[22], French *esclave,* Spanish *esclavo* 'slave' < **stlavo* <
    Late Latin *slavo* < Old Slavonic *sloveninu* 'a Slav'

(22) Lateral – rhotic sequences
  a. Middle English
    alderbest 'best of the all' < OE *ealra* 'of all those'

  b. Spanish
    saldrá < Latin *sal(i)re+ha* 'he'll leave'
    medrar < *meldrar* < *melrar* < Latin *meliorare* 'to prosper, to succeed'

  c. Catalan
    doldre (dial. *dolre*) < Latin *dolere* 'to hurt'
    moldre (dial. *molre*), French *moudre* (O.F. *moldre*) < Latin *molere* 'to grind'

  d. Swedish dialects
    Lappfjärd (west Findland) /ldr/ ~ Standard Swedish /lr/
    (Ivars 1996, cited in Engstrand et al. 1998)

(23) Fricative – rhotic sequences
  a. Spanish
    sidra, French *cidre* < **sizra* < Latin *sic(e)ra* 'cider'

  b. French
    être 'to be' < Old French *estre* < Latin *essere*
    ancêtre 'ancestor'< O.Fr. *ancestre* < Latin *antecessor*
    (Millardet 1910: 88; Wetzels 1985)

c. Italian
Israele [izdra'ɛlɛ], [izdɾa'ɛlɛ]

In the case of /lr/ > /ldr/, illustrated in (22), the emergent stop may also be attributed to variability in the phasing of the lateral and central valves, as not only the trill but also the tap and fricative varieties of the rhotic require elevated tongue sides and an open central valve. The case of /sr/ > /str/, exemplified in (23) cannot be explained in the same terms as both segments have a central release. However, both the tap and the trill involve an initial momentary central closure (of 25–30ms, Lindau 1985) which, on release, creates abrupt amplitude changes and may convey enough auditory cues for a stop (this is especially the case for trills, the first closure period of which involves a longer duration and a higher pressure build up than subsequent contacts in order to set the tongue tip into vibration, Solé 2002a). See Ohala (1995, 1997b, 2005) for the emergence of non-pulmonic stops and other epenthetic sounds.

The cases presented so far illustrate that variations in interarticulatory timing (i.e. denasalization and delateralization, principally), and associated aerodynamic and acoustic effects, account for the emergence of transient stop bursts which may be reinterpreted by the listeners as intended stops. In fact, failure to distinguish whether the stop was intended or not is probably at the origin of (i) the loss of an etymological /t/ in fricative–/t/–nasal and fricative–/t/–lateral sequences, e.g. *soften* (but *softer*), *christen* (but *Christianity*), *hasten*; *castle* (but *-chester*), *thistle*, *wrestle*, and (ii) the emergence of a non-etymological /t/ in *listen* and *hustle* in English. Warner and Weber (2001) provide perceptual data supporting the proposed articulatory and perceptual account; they found that listeners perceive stops the speaker did not intend, mostly in environments where the articulatory explanation predicts epenthetic stops to occur. As detailed above, such epenthetic stops originate in variations in the timing of pre-existing articulatory events and have been reported phonetically in a variety of languages (e.g. Catalan, Recasens and Pallarès 2001; American English, Solé 2007a). However, language-specific timing habits may avoid the transitional overlap of articulatory closures leading to emergent stops, thus epenthetic stops are not found phonetically in South African English (Fourakis and Port 1986).

7. Conclusions

We have reviewed a number of sound patterns where audible turbulence in the form of frication or a brief stop burst appears where it had not been present before, or fails to appear in contexts where it is expected to occur. We have argued that turbulence may *arise* from variations in the aerodynamic conditions due to interaction of articulatory gestures (i) within a segment (e.g. devoiced approximants becoming fricatives), or (ii) coarticulation with adjacent segments (e.g. glides or close vowels becoming fricativized following stops). Turbulence may also arise from changes in interarticulatory timing across segments (e.g. emergent stops) and perception of turbulence may be boosted due to auditory-acoustic factors (e.g. in anterior vs back fricatives or the emergence of buccal frication).

Audible turbulence for fricatives or stops may also be *diminished* and go undetected. This may be due to failure to create a pressure differential across the point of constriction sufficient to generate audible turbulence owing to (i) gestural interaction (e.g. an open velum or vocal fold vibration, as in voiced nasalized fricatives or stops in a nasal context), (ii) coarticulation with conflicting consecutive segments (e.g. lingual fricatives and trills), or (iii) prosodic conditions (e.g. syllable-final or utterance-final position). Turbulence may also be perceptually missed because of strictly auditory-acoustic factors, such as the lack of a downstream resonator to amplify the weak intensity noise created at the constriction (e.g. in labial stops or fricatives without a downstream cavity), or the attenuation of friction noise produced in the velar region (in labiovelar segments) due to low-pass filtering by the immediately downstream cavity.

The phonological patterns reviewed above illustrate that if sound patterns in language are to be explained, it is necessary to refer to details about the physical phonetic content of speech production and perception and, where relevant and where possible, factors that are not exclusively limited to the domain of speech. The alternative is mere stipulation: e.g. "obstruents tend to be voiceless"–even if re-coded using shorthand notations like [+obstruent]→[-voice] or [0voice]→[-voice]/[__,+obstruent]. In addition, the patterns reviewed illustrate the dependency between frication and voicelessness, between defrication and nasality, between the constriction location of the obstruent (buccal or non-buccal) and nasality, etc. due to aerodynamic, acoustic-auditory factors or the timing of articulatory events. As argued by Ohala (2005), dependency relations between features due to speech aerodynamics, acoustics or perception cannot be adequately captured by models such as Feature Geometry or Optimality Theory. For example, the aerodynamic interaction between voicelessness and frication demonstrates that what happens at the glottis can influence the generation of turbulence at the oral constriction. Such dependency relations cannot be accounted for in a model where the laryngeal feature is at a different branch from the supralaryngeal features and, therefore, cannot specify supralaryngeal frication. Finally, we would like to emphasize that what we have attempted to do here is part of a long tradition of noting common cross-language sound patterns –often referred to as "phonological universals"– and seeking explanations for them in the physical, physiological, acoustic, and perceptual domains (e.g., Bindseil 1838, Key 1852, Rosapelly 1876, Passy, 1890, Rousselot 1891, Grandgent 1896, Phelps 1937, Greenberg 1970, to mention just a few). As stated at the start of this paper, we do not claim and do not believe that these patterns are psychological or innate nor did these predecessors. Common cross-language sound patterns arise from the universality of physical phonetic constraints. Currently, many phonologists are also making interesting generalizations on common cross-language sound patterns but they claim that these arise from constraints that are part of the mental grammar or of the innate human language faculty. How can it be that the generalizations they note are of the same type as those made more than a century ago but are attributed to completely different causes?

The current situation, we believe, is comparable to that in the children's story of "Chicken Little" (borrowing the analogy from Ohala 1996). In the version we allude to, Chicken Little experiences a large blow on her head and then sets the entire barnyard into a panic with her claim that "the sky is falling". Skipping a lot of the intermediate plot, the dénouement is that it was questioned why Chicken Little made the claim that the sky was falling. The empirical evidence was a swelling on the top of her head. Where was she when the injury happened? Under an oak tree. And violà! A large acorn was found on the very spot! To those who situate natural sound patterns in the synchronic (psychological) grammar and in the human genome, we respectfully submit: the acorn has been found.

Notes

* Work supported in part by grant HUM2005-02746 from the Ministry of Science and Technology, Spain, and by the research group 2009-SGR-3 of

the Catalan Government. The suggestions of Daniel Recasens, who reviewed an earlier version of this paper, are gratefully acknowledged.

1. Not to neglect worthy precursors even before that, e.g. van Boxhorn (1647).
2. Although atmospheric pressure (absolute) is roughly 1033 cm H_2O at sea level, it is common to take atmospheric pressure as zero (0) and express pressure in the vocal tract as x cm H_2O with respect to that. Nothing is lost with such a mathematical convention.
3. Other factors have been reported to contribute to keeping a low oral pressure for voiced obstruents, for example, an increased volume of the oral cavity (Ewan and Krones 1973; Kent and Moll 1969; Bell-Berti 1975; Westbury 1983) and/or intentional relaxation of vocal tract muscles resulting in more passive expansion of the walls (Svirsky et al. 1997).
4. The sound source excites mostly the cavities anterior to the constriction where the sound is generated, whereas the back cavities do not contribute much acoustically (Fant 1960, Stevens 1998). Since, in the case of frication generated at the nostrils, the oral and nasal cavities posterior to the nostrils do not contribute much to the spectrum of the noise, the different voiceless nasals do not differ much acoustically.
5. A related phenomenon is the preservation of Latin word-initial /f/ only before /w/ in Spanish (e.g. *fuente* ['fwen̪t̪e] < *fonte* ‘fountain’). /f/ became /h/ and was later lost preceding all other vowels (e.g. *hablar* [a'β̞laɾ] < *fabulare* ‘to talk’, *hierro* ['jero] < *ferru* ‘iron’, *hondo* ['on̪d̪o] < *fundu* ‘deep’) (Menéndez Pidal 1968: 122).
6. Palatalization and affrication of dental stops is a geographical dialect marker in Brazilian Portuguese. Dental stops are palatalized and affricated when followed by a high vowel [i], which may be lexical or inserted after a word-final dental stop in borrowed words (as BP only allows /s, r, n, l/ syllable-finally).
7. Solé (2002b) estimated the allowable range of aerodynamic variation for voiced fricatives from aerodynamic data. For one of her subjects, she estimated a subglottal pressure (*Ps*) of 7.6 cmH_2O during fricative production. Since transglottal flow to maintain voicing requires a pressure drop across the glottis (*Ps-Po*) of at least 1-2 cmH_2O (and higher values to initiate voicing, 2–3 cmH_2O), that leaves a *Po* of approximately 5.6 cmH_2O. Generation of turbulence for voiced fricatives ceases when the transoral pressure drops to about 3 cmH_2O (Ohala, Solé, and Ying 1998; Catford, 1977: 124; Stevens 1998: 480), which means that *Po* may vary between a rather narrow range of 5.6–3 cmH_2O in order to sustain voicing and frication.
8. Spirantization of voiced stops may be a manoeuvre to lower oral pressure in order to facilitate voicing (Ohala 1983b).
9. It has been reported that voiced stops may become voiced approximants without an intermediate fricative stage (see, e.g. Villafana Dalcher 2006 for

Florentine Italian). In such cases defrication of voiced fricatives may not be necessarily at work.

10. The term fricative “weakening” is used here to indicate attenuation of the high frequency noise which characterizes fricatives, due to gestural reduction or aerodynamic factors. Fricative loss is considered the endpoint of the weakening continuum, i.e. extreme attenuation leading to the segment becoming inaudible. In perceptual terms gradient attenuation of the friction noise may result in identification of a discrete segment (e.g. a frictionless continuant, a vowel, a tap, an assimilated segment, or /h/) or in the perceptual loss of the segment (i.e. deletion).
11. For example, because the constriction shape and area of lingual fricatives is critical, they tolerate less coarticulation with neighbouring sounds (Recasens, Pallarès, and Fontdevila 1997), they are less overlapped (Byrd 1996), and they show lesser articulatory reduction in magnitude (Byrd and Tan 1996) than other segment types (e.g. stops); and because temporal factors are also critical, fricatives are less susceptible to temporal reduction vis-à-vis other segments (Klatt 1976; Byrd and Tan 1996).
12. In reverse sequences, /rs/, where the trill is in coda rather than onset position, the trill is commonly detrilled and may assimilate to the fricative, e.g. Latin *bursa, morsicare* > Catalan *bossa, mossegar* ‘bag’, ‘to bite’ (Badía 1951: 202); /rs/ > [ʂ] (retroflex fricative) in Scots English (Bähr 1974: 132ff) or Standard Swedish.
13. This contrasts with the requirements for nasal consonants which require a velic opening greater than 20mm^2 and typically between 50 and 100mm^2.
14. Thus, Ohala, Solé, and Ying (1998) report that when voiced and voiceless fricatives are vented with a pseudo-velopharyngeal valve – a tube inserted at the sides of the mouth via the buccal sulcus and the gap behind the molars – simulating different degrees of nasalization, when the valve has a similar impedance to that at the oral constriction (and as a result air is flowing out both through the nose and the mouth) voiced fricatives become frictionless continuants while voiceless fricatives retain their frication (though the intensity of friction is attenuated).
15. The unusual blocking of nasalisation by sonorants might be due to morphological factors or the fossilisation of historical forms.
16. Post-nasal voicing is an aerodynamically and perceptually-based process by which voiceless stops become voiced after nasals. When a voiceless stop is preceded by a nasal, voicing into the stop closure is prolonged, vis-à-vis postvocalic stops, by nasal leakage before full velic closure is achieved and continued velic raising even after velic closure has occurred, thus expanding the volume of the oral cavity. Nasal leakage and oral cavity expansion lower the oral pressure which accumulates in the oral cavity and thus prolong transglottal flow for voicing (Rothenberg 1968; Westbury 1983; Ohala and Ohala 1991; Bell-Berti 1993; Hayes and Stivers 2000). These factors lead to

postnasal voiceless obstruents being phonetically partially voiced and with a weaker stop burst, which leads to being reinterpreted as voiced.

17. For a perceptual account of the loss of nasals before fricatives (but not other segment types), as in English *goose* vis-à-vis German *Gans* which preserves the nasal, see Ohala and Busà (1995).
18. This is not usually the case. In normal voiced speech the impedance of the sound source, i.e. the vibrating vocal cords, is so much greater than the impedance (inertia) of the resonances of the vocal tract that there is a negligible amount of coupling between the source and the tract.
19. In the case of doubly articulated labiovelar stops, [k͡p] [g͡b], the auditory impression from the release is that of a labial, rather than a velar. This seems to be due to the timing of the two closures. The dorsal closure leads the labial closure, which is released at a later stage. Hence there is an acoustic similarity of the [k͡p] release to the [p] release (Ladefoged and Maddieson 1996: 336-339).
20. Not covered here are cases of apparent clicks arising from partial temporal overlap of buccal articulations; see Marchal (1987), Ohala (1995).
21. Although the term “epenthetic” is more common to describe stops such as the [p] in *warm*[p]*th*, the term “emergent” is preferred: “epenthetic”, by its etymology implies that the stop was simply “inserted”, i.e. it came out of nowhere, whereas the term “emergent” correctly implies that the stop emerged from pre-existing pre-cursors, e.g. a temporal overlap of the pre-existing closures at the velum and in the oral cavity.
22. The origin of the Italian greeting *ciao*, i.e. a much abbreviated form of the old formula meaning ‘your servant’.

References

Albano, Eleonora Cavalcante
1999 O português brasileiro e as controversias da fonética atual: Pelo aperfeiçoamento da fonología articulatória. *D.E.L.T.A.: Documentação de Estudos em Lingüística Teórica e Aplicada* 15: 23–51.

Ali, Latif, Ray Daniloff, and Robert Hammarberg
1979 Intrusive stops in nasal-fricative clusters: An aerodynamic and acoustic investigation. *Phonetica* 36 (2): 85–97.

Alcover, Antoni M., and Francesc de B. Moll
1978–1979 *Diccionari Català-Valencià-Balear*. Barcelona: Gràfiques Instar.

Andersen, Henning
1972 Diphthongization. *Language* 48: 11–50.

Baer, Thomas, John C. Gore, L.C. Gracco, and Patrick W. Nye
1991 Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *Journal of the Acoustical Society of America* 90 (2): 799–828.

Bähr, Dieter
1974 *Standard English und seine Geographischen Varianten*. München: Wilhelm Fink.

Ball, Martin, and Briony Williams
2001 *Welsh Phonetics*. Lewiston, NY: Edwin Mellen Press.

Basset, Patricia, Angélique Amelot, Jacqueline Vaissière, and Bernard Roubeau.
2001 Nasal flow in French spontaneous speech. *Journal of the International Phonetic Association* 31: 87–99.

Bauer, Robert S.
1982 Cantonese sociolinguistic patterns: Correlating social characteristics of speakers with phonological variables in Hong Kong Cantonese. Ph. D. diss., Department of Linguistics, University of California, Berkeley.

Bell-Berti, Fredericka
1975 Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America* 57: 456–461.

Bell-Berti, Fredericka
1993 Understanding velic motor control: Studies of segmental context. In: Marie K Huffman, and Rena A. Krakow (eds.), *Phonetics and Phonology: Nasals, Nasalization, and the Velum*, 87–116. San Diego: Academic Press.

Bendor-Samuel, John T.
1966 Some prosodic features in Terena. In: Charles E. Bazell, John C. Catford, Michael A.K. Halliday, and Robert H. Robins (eds.), *In Memory of J.R. Firth*, 30–39. London: Longmans.

Bhat, D.N.S.
1978 A general study of palatalization. In: Joseph H. Greenberg (ed.), *Universals of Human Language*, 47–91. Vol. 2. Stanford, Ca.: Stanford University Press.

Bindseil, Heinrich E.
1838 *Abhandlungen zur Allgemeinen Vergleichenden Sprachlehre*. Hamburg: F. Perthes.

Bladon, R.A.W., and Francis Nolan
1977 A videofluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics* 5: 185–193.

Blight, Richard C., and Eunice V. Pike
1976 The phonology of Tenango Otomi. *International Journal of American Linguistics* 42: 51–57.

Boas, Franz

1947 Grammar of Kwakiutl (with a glossary of the suffixes). *Transactions of the American Philosophical Society* 37 (3): 201–377.

van Boxhorn, Marcus Zuerius

1647 *Bediedinge van de tot noch toe onbekende afgodinne Nehalennia*. Leyden: Willem Christiaens van der Boxe.

Bradley, David

1979 *Proto-Loloish*. London/Malmö: Curzon.

des Brosses, Charles

1765 *Traité de la Formation Méchanique des Langues, et de Principes Physiques de l'Étymologie*. Paris: Chez Saillant, Vincent, Desaint.

Byrd, Dani

1996 Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24: 263–282.

Byrd, Dani, and Cheng Cheng Tan

1996 Saying consonant clusters quickly. *Journal of Phonetics* 24: 263–282.

Busà, Maria Grazia

2007 Coarticulatory nasalization and phonological developments: Data from Italian and English nasal-fricative sequences. In: Maria-Josep Solé, Patrice S. Beddor, and Manjari Ohala (eds.), *Experimental Approaches to Phonology*, 155–174. Oxford: Oxford University Press.

Catford, John C.

1977 *Fundamental Problems in Phonetics*. Bloomington: Indiana University Press.

Chang, Steve C.

1999 Vowel-dependent VOT variation. *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco. Vol. 2. 1021–1204.

Chang, Steve, Madelaine C. Plauché, and John J. Ohala

2001 Markedness and consonant confusion asymmetries. In: Elizabeth Hume, and Keith Johnson (eds.), *The Role of Speech Perception in Phonology*, 79–101. San Diego: Academic Press.

Chen, Matthew Y.

1976 From Middle Chinese to Modern Peking. *Journal of Chinese Linguistics* 4 (2/3): 113–277.

Chiba, Tsutomu, and Masato Kajiyama

1941 *The Vowel: Its Nature and Structure*. Tokyo: Kaiseikan.

Clark, John, and Colin Yallop

1990 *An Introduction to Phonetics and Phonology*. Oxford: Blackwell.

Cohn, Abigail C.

1990 Phonetic and phonological rules of nasalization. *UCLA Working Papers in Phonetics* 76: 1–224.

Cohn, Abigail C.

1993 The status of nasalized continuants. In: Marie K. Huffman, and Rena A. Krakow (eds.), *Nasals, Nasalization and the Velum*, 329–367. San Diego: Academic Press.

Colantoni, Laura.

2006 Micro and macro sound variation and change in Argentine Spanish. In: Nuria Sagarra and Almeida Jacqueline Toribio (eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, 91–102. Somerville, MA: Cascadilla Proceedings Project.

Cowan, H.K.J.

1965 *Grammar of the Sentani Language*. Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde, 47. ’S-Gravenhage: Martinus Nijhoff.

Engstrand, Olle, Sven Björsten, Gösta Bruce, and Anders Eriksson

1998 Phonetic preconditions for historical sound change – evidence from the dialects. *Proceedings of FONETIK 98*, 24–27. The Tenth Swedish Phonetics Conference. Stockholm University. http://www.ling.su.se/staff/olle/f98_dial2.html

Ewan, William G., and Robert Krones

1973 A study of larynx height in speech using the thyro-umbrometer. *Journal of the Acoustical Society of America* 53: 345.

Fant, Gunnar

1960 *Acoustic Theory of Speech Production*. The Hague: Mouton.

Fant, Gunnar

1973 *Speech Sounds and Features*. Cambridge: MIT Press.

Flanagan, James L., and Kenzo Ishizaka

1976 Automatic generation of voiceless excitation in a vocal cord-vocal tract speech synthesizer. *IEEE-ASSP* 24 (2): 163–170.

Flanagan, James L., Kenzo Ishizaka, and K. L. Shipley

1975 Synthesis of speech from a dynamic model of the vocal cords and vocal tract. *Bell System Technical Journal* 54: 485–506.

1980 Signal models for low bit-rate encoding of speech. *Journal of the Acoustical Society of America* 68: 780–791.

Fourakis, Marios, and Robert Port

1986 Stop epenthesis in English. *Journal of Phonetics* 14: 197–221.

Fujimura, Osamu, and Jan Lindqvist

1971 Sweep-tone measurements of vocal tract characteristics. *Journal of the Acoustical Society of America* 49: 541–558.

Gerfen, Henry (Chip) J.
1996 Topics in the phonology and phonetics of Coatzospan Mixtec. Ph. D. diss., Deparment of Linguistics, University of Arizona.
Gimson, Alfred C.
1962 *An Introduction to the Pronunciation of English*. London: Arnold.
Grammont, Maurice
1933 *Traité de Phonétique*. Paris: Delagrave.
Grandgent, Charles H.
1896 Warmpth. *Publications of the Modern Language Association* 11 (New Series 4): 63–75.
Grandgent, Charles H.
1905 *An Outline of the Phonology and Morphology of Old Provençal*. Boston: Heath.
Greenberg, Joseph H.
1970 Some generalizations concerning glottalic consonants, especially implosives. *International Journal of American Linguistics* 36 (2): 123–145.
Gregores, Emma, and Jorge Suarez
1967 *A Description of Colloquial Guaraní*. The Hague: Mouton.
Guion, Susan. G.
1998 The role of perception in the sound change of velar palatalization. *Phonetica* 55: 18–52.
Hall, T. A., and Silke Hamann
2006 Towards a typology of phonological stop assibilation. *Linguistics* 44 (6): 1195–1236.
Halle, Morris, and George N. Clements
1983 *Problem Book in Phonology*. Cambridge: MIT Press.
Hamann, Silke, and Hristo Velkov
2005 Airflow in stop-vowel sequences of German. *ZAS Papers in Linguistics* 42: 1–12.
Hayes, Bruce, and Tanya Stivers
2000 Postnasal voicing. Manuscript.
Heffner, Roe-Merrill S
1964 *General Phonetics*. Madison: University of Wisconsin Press.
Helgason, Pétur
2002 Preaspiration in the nordic languages. Synchronic and diachronic aspects. Ph. D. diss., University of Stockholm.
Hock, Hans Heinrich
1986 *Principles of Historical Linguistics*. New York: Mouton.
Itô, Junko, Armin Mester, and Jaye Padgett
1995 Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26: 571–614.

Ivars, Ann-Marie
1996 Stad och bygd. Finlandssvenska stadsmål i ett regionalt och socialt perspektiv. *Folkmålsstudier* 37.

Jaeger, M. J., and H. Matthys
1970 The pressure flow characteristics of the human airways. In: A. Bouhuys (ed.), *Airway Dynamics, Physiology and Pharmacology*, 21–32. Springfield, IL: Charles C. Thomas.

ten Kate Hermansz, L.
1723 *Aenleidning tot de Kennisse van het verhevene Deel der nederduitsche Sprake*. 2 vols. Amsterdam: Rudolph en Gerard Wetstein.

Keating, Patricia, Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu
2003 Domain-initial strengthening in four languages. In: John Local, Richard Ogden, and Rosalind Temple (eds.), *Papers in Laboratory Phonology 6*, 143–163. London: Cambridge U.P.

Keating, Patricia, Richard Wright, and Jie Zhang
1999 Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics* 97: 157–173.

Kent, Raymond D., and Kenneth L. Moll
1969 Vocal tract characteristics of the stop cognates. *Journal of the Acoustical Society of America* 46: 1549–1555.

Key, T. Hewitt.
1852 On vowel assimilation, especially in relation to Professor Willis' experiment on vowel sounds. *Transactions of the Philological Society*. London: The Philological Society 5: 191-204.

Klatt, Dennis H.
1976 Linguistic use of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59: 1208–1221.

Kohler, Klaus
2001 Plosive-related glottalization phenomena in read and spontaneous speech. A stød in German? In: Nina Grønnum, and Jorgen Rischel (eds.), *To Honour Eli Fischer-Jørgensen*, 174–211. (Travaux du Cercle Linguistique de Copenhague 31). Copenhagen: Reitzel

Krämer, Martin
2004 Affricates and the phonetic implementation of laryngeal contrast in Italian. Presentation at the *Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*. Mainz.
http://www.hum.uit.no/a/kraemer/dgfs04.pdf

Ladefoged, Peter, and Ian Maddieson
1996 *The Sounds of the World's Languages*. Oxford: Blackwell.

Laver, John
1994 *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lehiste, Ilse
1964 *Acoustical Aspects of Selected English Consonants*. Bloomington: Indiana University Press.

Li, Fang Kuei
1977 *A Handbook of Comparative Tai*. Honolulu: The University Press of Hawaii.

Lindau, Mona
1985 The story of /r/. In: Victoria A. Fromkin (ed.), *Phonetic Linguistics. Essays in Honor of Peter Ladefoged*, 157–168. Orlando, FL: Academic Press.

Lindblom, Björn
1984 Can the models of evolutionary biology be applied to phonetic problems? *Proceedings of the 10th International Congress of Phonetic Sciences*, Utrecht, 67–81.

Lindblom, Björn
1986 Phonetic universals in vowel systems. In: John Ohala and Jeri Jaeger (eds.), *Experimental Phonology*, 13–44. Orlando: Academic Press.

Lindblom, Björn
1990 On the notion of possible speech sound. *Journal of Phonetics* 18: 135–152.

Lindblom, Björn
1999 Emergent phonology. Paper delivered at the 25th Annual Meeting, *Berkeley Linguistics Society*. Full text: http://www.ling.su.se/fon/perilus/1999_14.pdf

Loos, Eugene E.
1969 *The Phonology of Capanahua and its Grammatical Basis*. University of Oklahoma: Summer Institute of Linguistics publication no. 20. (Austin, University of Texas dissertation, 1967).

Maddieson, Ian
1983 The analysis of complex phonetic elements in Bura and the syllable. *Studies in African Linguistics* 14: 285–310.

Maddieson, Ian
1984 *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian
2005 Voicing in plosives and fricatives. In: Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.), *The World Atlas of Language Structures*, 26–29. Oxford: Oxford University Press.

Malécot, André
1960 Vowel nasality as a distinctive feature in American English. *Language* 36: 222–229.

Marchal, Alain
1987 Des clics en français? *Phonetica* 44: 30-37.

Martínez Celdrán, Eugenio
1991 Sobre la naturaleza fonética de los alófonos de /b, d, g/ en español y sus distintas denominaciones. *Verba. Anuario Galego de Filoloxía* 18: 235–253.

Martínez Celdrán, Eugenio
2004 Problems in the classification of approximants. *Journal of the International Phonetic Association* 43 (2): 201–210.

Mathangwane, Joyce T.
1996 Phonetics and phonology of Ikalanga: A diachronic and synchronic study. Ph. D. diss., Department of Linguistics, University of California, Berkeley.

Menéndez Pidal, Ramón
1968 Reprint. *Manual de Gramática Histórica Española*. 13th ed. Madrid: Espasa Calpe. Original edition, 1904.

Meillet, Antoine, and Joseph Vendryes
1924 *Traité de Grammaire Comparée des Langues Classiques*. Paris: Librairie Ancienne Edouard Champion.

Millardet, Georges
1910 *Étude de Dialectologie Landaise. Le Dévelopement des Phonèmes Additionels*. Toulouse: Librairie Edouard Privat.

Millardet, Georges
1911 Insertions de consonnes en suédois moderne. *Revue de Phonétique* 1: 309–346.

Mills, Roger F.
1975 Proto South Sulawesi and Proto Austronesian phonology. Ph. D. diss., University of Michigan.

Mossé, Ferdinand
1968 Reprint. *A Handbook of Middle English*. 5th ed. Baltimore: Johns Hopkins Press. Original edition, 1952.

Navarro Tomás, Tomás
1980 Reprint. *Manual de Pronunciación Española*. 20th edition. Madrid: CSIC. Original edition, 1918.

Oberlies, Thomas
2001 *Pali: A Grammar of the Language of the Theravada Tipitaka, with a Concordance to Pischel's Grammatik der Prakrit-Sprachen*. Berlin: Walter de Gruyter.

Ohala, John J.
1975 Phonetic explanations for nasal sound patterns. In: Charles A. Ferguson, Larry M. Hyman, and John J. Ohala (eds.), *Nasálfest: Papers from a Symposium on Nasals and Nasalization*, 289–316. Stanford: Language Universals Project.

Ohala, John J.
1976 A model of speech aerodynamics. *Report of the Phonology Laboratory*, University of California, Berkeley, 1: 93–107.

Ohala, John J.
1979 Phonetic universals in phonological systems and their explanation. [Summary of symposium moderator's introduction.] *Proceedings of the 9th International Congress of Phonetic Sciences*, 5–8.

Ohala, John J.
1981a Articulatory constraints on the cognitive representation of speech. In: Terry Myers, John Laver, John Anderson (eds.), *The Cognitive Representation of Speech*, 113–124. Amsterdam: North-Holland.

Ohala, John J.
1981b 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,* 178–203. Chicago: Chicago Linguistic Society.

Ohala, John J.
1983a The phonological end justifies any means. In: Shiro Hattori and Kazuko Inoue (eds.), *Proceedings of the 13th International Congress of Linguists*, 232–243. Tokyo: distributed by Sanseido Shoten.

Ohala, John J.
1983b The origin of sound patterns in vocal tract constraints. In: P. F. Mac Neilage (ed.), *The Production of Speech*, 189–216. New York: Springer-Verlag.

Ohala, John J.
1985 Linguistics and automatic speech processing. In: Renato De Mori and Ching Y. Suen (eds.), *New Systems and Architectures for Automatic Speech Recognition and Synthesis,* 447–475. Berlin: Springer-Verlag.

Ohala, John J.
1989 Sound change is drawn from a pool of synchronic variation. In: Leiv Egil Breivik and Ernst Hakon Jahr (eds.), *Language Change: Contributions to the Study of its Causes*, 173–198. Berlin: Mouton de Gruyter.

Ohala, John J.
1992 What's cognitive, what's not, in sound change. In: Günter Kellermann, and Michael D. Morrissey (eds.), *Diachrony within Synchrony: Language History and Cognition*, 309–355. (Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 14.) Frankfurt: Lang.

Ohala, John J.
1993 Coarticulation and phonology. *Language and Speech* 36 (2, 3): 155–170.

Ohala, John J.
1995 A probable case of clicks influencing the sound patterns of some European languages. *Phonetica* 52: 160–170.

Ohala, John, J.
1996 Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America* 99: 1718–1725.

Ohala, John J.
1997a Aerodynamics of phonology. *Proceedings of the 4th Seoul International Conference on Linguistics* [SICOL], Seoul: Linguistic Society of Korea, 92–97.

Ohala, John J.
1997b Emergent stops. *Proceedings of the 4th Seoul International Conference on Linguistics* [SICOL], Seoul: Linguistic Society of Korea, 84–91.

Ohala, John J.
1997c Comparison of speech sounds: Distance vs. cost metrics. In: Shigeru Kiritani, Hajime Hirose, and Hiroya Fujisaki (eds.), *Speech Production and Language. In honor of Osamu Fujimura*, 261–270. Berlin: Mouton de Gruyter.

Ohala, John J.
2005 Phonetic explanations for sound patterns. Implications for grammars of competence. In: William J. Hardcastle & Janet M. Beck (eds.), *A Figure of Speech. A Festschrift for John Laver*, 23–38. London: Erlbaum.

Ohala John J. and Maria Grazia Busà
1995 Nasal loss before voiceless fricatives: A perceptually-based sound change. *Rivista di Linguistica* 7: 125-144.

Ohala, John J., and Jeri J. Jaeger (eds)
1986 *Experimental Phonology*. Orlando: Academic Press.

Ohala, John J., and James Lorentz
1977 The story of [w]: An exercise in the phonetic explanation for sound patterns. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 3: 577–599.

Ohala, John J., and Manjari Ohala
1993 The phonetics of nasal phonology: Theorems and data. In: Marie K. Huffman, and Rena A. Krakow (eds.), *Nasals, Nasalization, and the Velum*, 225–249. (Phonetics and Phonology Series Vol. 5). San Diego: Academic Press.

Ohala, John J., Maria-Josep Solé, and Goangshiuan Ying
1998 The controversy of nasalized fricatives. *Proceedings of the 135th Meeting of the ICA/ASA*, Seattle, Washington, 2921–2922.

Ohala, Manjari, and John J. Ohala
1991 Nasal epenthesis in Hindi. *Phonetica* 48: 207–220.

Passy, Paul
1890 *Étude sur les Changements Phonétiques et leurs Caractères Généraux*. Paris: Firmin-Didot.

Pater, Joe
1999 Austronesian nasal substitution and other NC effects. In: René Kager, Harry van der Hulst & Wim Zonneveld (eds.), *The Prosody-Morphology Interface*, 310–343. Cambridge: Cambridge University Press.

Phelps, J.
1937 Indo-european initial sl. *Language* 13: 279–284.

Pike, Kenneth L.
1943 *Phonetics*. Ann Arbor: University of Michigan Press.

Plauché, Madelaine C., Cristina Delogu, and John J. Ohala
1997 Asymmetries in consonant confusion. *Proceedings of Eurospeech '97*. Fifth European Conference on Speech Communication and Technology 4: 2187–2190.

Pope, Mildred K.
1952 *From Latin to Modern French with Special Consideration of Anglo-Norman*. Manchester: Manchester University Press.

Recasens, Daniel
1993 *Fonètica i Fonologia*. Barcelona: Enciclopèdia Catalana.

Recasens, Daniel
2002 Weakening and strengthening in Romance revisited. *Rivista di Linguistica* 14 (2): 327–373.

Recasens, Daniel
2006 Gradient weakening of syllable-final /s, r/ in Majorcan Catalan consonant clusters. *Proceedings of the 7th International Seminar on Speech Production*. CEFALA, Belo Horizonte, 11–18.

Recasens, Daniel, and Maria Dolors Pallarès
2001 Coarticulation, assimilation and blending in Catalan consonant clusters. *Journal of Phonetics* 29 (3): 273–301.

Recasens, Daniel, Maria Dolors Pallarès, and Jordi Fontdevila
1997 A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America* 102 (1): 544–561.

Robins, Robert H.
1957 Vowel nasality in Sundanese: A phonological and grammatical study. *Studies in Linguistic Analysis* (Special volume of the Philological Society), 87–103. London: Basil Blackwell.

Rohlfs, Gerhard
1949 *Historische Grammatik der Italienischen Sprache und ihrer Mundarten*. Vol. 1. Bern: A. Francke AG. Verlag.

Rohlfs, Gerhard

1970 *Les Gascons. Études de Philologie Pyrénéenne*. Tübingen: Max Niemeyer Verlag.

Romero, Joaquín

1995 Gestural organization in Spanish. An experimental study of spirantization and aspiration. Ph. D. diss., Department of Linguistics, University of Connecticut.

Rosapelly, Charles L.

1876 Inscriptions des mouvements phonétiques. Physiologie Expérimentale. *Travaux du Laboratoire de M. Marey* 2: 109–131. Paris: G. Masson.

Rosetti, Alexandru

1978 *Istoria limbii române*. 2nd ed. Bucharest: Ed. St. Si. Encicl.

Rositzke, Harry A.

1940 Epenthetic consonants in Swedish. *Journal of English and Germanic Philology* 39: 473–485.

Rossato, Solange

2004 Quelques statistiques sur la nasalité vocalique et consonantique. *Actes des 25èmes Journées d'Études sur la Parole*, Fès, Maroc, 433–436.

Rothenberg, Martin

1968 *The Breath-Stream Dynamics of Simple-Released-Plosive Production*. Basel: Karger.

Rousselot, Pierre-Jean

1891 *Les Modifications Phonétiques du Language Étudiées dans le Patois d'une Famille de Cellefrouin (Charante)*. Paris: H. Welter. (Extrait de la Revue de Patois Gallo-Romans, 1891).

Sampson, Rodney

1999 *Nasal Vowel Evolution in Romance*. Oxford: Oxford University Press.

Schachter, Paul, and Victoria Fromkin

1968 A phonology of Akan: Akuapem, Asante, and Fante. *UCLA Working Papers in Phonetics* 9.

Schadeberg, Thilo C.

1982 Nasalization in Umbundu. *Journal of African Languages and Linguistics* 4: 109–132.

Schleicher, August.

1861-1862 *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. (Vol. 1 u. 2). Weimar: Hermann Böhlau.

Schourup, Lawrence

1972 Characteristics of vowel nasalization. *Papers in Linguistics* 5: 530–548.

Schwartz, Jean-Luc, Louis-Jean Boë, and Christian Abry
2007 Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action-Control Theory (PACT). In: Maria-Josep Solé, Patrice S. Beddor, and Manjari Ohala (eds.), *Experimental Approaches to Phonology*, 104-124. Oxford: Oxford University Press.

Shadle, Christine H.
1990 Articulatory-acoustic relationships in fricative consonants. In: William. H. Hardcastle, and Alain Marchal (eds.), *Speech Production and Modelling*, 187–209. Dordrecht: Kluwer.

Sherman, Donald
1975 Stop and fricative systems: A discussion of paradigmatic gaps and the question of language sampling. *Stanford Working Papers on Language Universals* 17: 1–31.

Shosted, Ryan.
2006 The aeroacoustics of nasalized fricatives. Ph. D. diss., Department of Linguistics, University of California, Berkeley.

Solé, Maria-Josep
1992 Experimental phonology: The case of rhotacism. In: Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer, and John R. Rennison (eds.), *Phonologica 1988*, 259–271. Cambridge: Cambridge University Press.

Solé, Maria-Josep
2002a Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics* 30: 655–688.

Solé, Maria-Josep
2002b Assimilatory processes and aerodynamic factors. In: Carlos Gussenhoven, and Natasha Warner (eds.), *Papers in Laboratory Phonology 7*, 351–386. New York: Mouton de Gruyter.

Solé, Maria-Josep
2007a The stability of phonological features within and across segments. The effect of nasalization on frication. In: Pilar Prieto, Joan Mascaró, and Maria-Josep Solé (eds.), *Segmental and Prosodic Issues in Romance Phonology*, 41–65. Amsterdam: John Benjamins.

Solé, Maria-Josep
2007b Compatibility of features and phonetic content: The case of nasalization. *Proceedings of the 16th Congress of Phonetic Sciences*. Saarbrücken, Germany, 261–266.

Stevens, Kenneth N.
1971 Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America* 50: 1180–1192.

Stevens, Kenneth N.
1972 The quantal nature of speech: evidence from articulatory-acoustic data. In: Edward E. Davis Jr., and Peter.B.Denes (eds.), *Human Communication: A Unified View*, 51–66. New York: Mc Graw-Hill.

Stevens, Kenneth N.
1989 On the quantal nature of speech. *Journal of Phonetics* 17: 3–45.

Stevens, Kenneth N.
1998 *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Stringer, Mary D., and Joyce M. Hotz
1973 Waffa phonemes. In: H. McKaughan (ed), *The Language of the Eastern Family of the East New Guinea Highland Stock*, 523–529. Seattle: University of Washington Press.

Svirsky, Mario A., Kenneth N. Stevens, Melanie L. Matthies, Joyce Manzella, Joseph S. Perkell, and Reiner Wilhelms-Tricarico
1997 Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America* 102(1): 562–571.

Teoh, Boon Seong
1988 Aspects of Malay phonology revisted. A non-linear approach. Ph. D. diss., University of Illinois at Urbana-Champaign.

Torreblanca, Máximo
1976 *Estudio del Habla de Villena y su Comarca*. Alicante: Diputación Provincial de Alicante.

Tuttle, Edward
1997 Palatalization. In: Martin Maiden, and Mair M. Parry (eds.), *The dialects of Italy*, 26–31. London: Routledge.

Vaux, Bert
1998 *The Phonology of Armenian*. Oxford: Clarendon Press.

Villafana Dalcher, Christina
2006 Consonant weakening in Florentine Italian: An acoustic study of gradient and variable sound change. Ph. D. diss., Georgetown University.

Wagner, Günter
1933 Yuchi. In: Franz Boas (ed.), *Handbook of American Indian Languages 3*, 291-383. Hamburg/New York: Columbia University Press.

Walker, Rachel
1998 Nasalization, neutral segments, and opacity effects. Ph. D. diss., University of California, Santa Cruz.

Warner, Natasha, and Andrea Weber
2001 Perception of epenthetic stops. *Journal of Phonetics* 29: 53–87

Warren, Donald W., and Arthur B. DuBois
1964 A pressure-flow technique for measuring velopharyngeal orifice area during continuous speech. *Cleft Palate-Craniofacial Journal* 1(1): 52–71.

Warren, Donald W., Rodger M. Dalston, and Robert Mayo
1993 Hypernasality and velopharyngeal impairment. *Cleft Palate-Craniofacial Journal* 31(4): 257–262.

Watkins, Calvert (ed.)
1985 *The American Heritage Dictionary of Indo-European Roots*. Boston: Houghton Mifflin.

Welmers, William
1973 *African Language Structures*. Berkeley: University of California Press.

Wetzels, W. Leo
1985 The historical phonology of intrusive stops. A nonlinear description. *Canadian Journal of Linguistics* 30: 285–333.

Westbury, John R.
1983 Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America* 73: 1322–1336.

Winitz, Harris, M. E. Scheib, and James A. Reeds
1972 Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *Journal of the Acoustical Society of America* 51(4): 1309–1317.

Zue, Victor W. and Martha Laferriere
1979 Acoustic study of internal /t, d/ in American English. *Journal of the Acoustical Society of America* 66: 1039–1050.

A phonetic approach to the phonology of *v*: A case study from Hungarian and Slovak*

Zsuzsanna Bárkányi and Zoltán Kiss

1. Introduction

The seemingly odd phonological behaviour of the voiced labiodental fricative *v* in many unrelated languages like Russian (Padgett 2002), Hebrew (Barkaï and Horvath 1978) and Hungarian (Siptár 1996; Siptár and Törkenczy 2000), just to mention a few, has attracted considerable attention in the phonological literature. With respect to Hungarian, most attention has been focused on its two-fold patterning in voicing assimilation, namely, that it patterns with obstruents in being targeted by voicing assimilation, but it behaves like a sonorant in that it does not trigger the process. Related to this dynamic aspect, its static phonotactic distribution has also been of interest, as it also displays asymmetrical properties. In this paper we would like to present evidence that there is a direct link between the surface phonetic properties of *v* (in particular its articulatory and aerodynamic properties) and its phonological patterning. We will concentrate on the double-faced phonological behaviour of *v* in Hungarian and Slovak. The study is backed up by two phonetic experiments examining the acoustic properties of *v* in these two languages.

In the next section we introduce the relevant phonological data. In section 3.1. we present a phonetic analysis that pays special attention to the aerodynamic properties of *v*. Section 3.2. puts forth the results of the acoustic experiments and checks the validity of the proposed predictions.[1] Section 4 concludes the study.

2. The phonology of *v* in Hungarian and Slovak

2.1. *v* and the "Voicing Requirement"

In both Hungarian (H.) and Slovak (S.), two obstruents standing next to each other may not differ in voicing, that is, either both are voiceless, or both are voiced. This requirement embraces the whole Hungarian and Slovak obstruent phonology as it applies morpheme-internally as well as across morpheme and word boundaries. We will refer to this property of Hungarian and Slovak phonology as the Voicing Requirement (VR) to cover both the static and the dynamic aspects of this phenomenon, which the traditional literature calls "voicing agreement" morpheme-internally and "voicing assimilation" across morpheme and word boundaries.

The dynamic aspect of the VR is a regressive process: if two obstruent segments with different values for voicing are adjacent, it is always the second segment that determines the voicing of the first, thus it acts as the trigger of regressive voicing assimilation. Clusters flanking a boundary between (i) a stem and a suffix and (ii) two words (in compound words as well as in phrases) are affected as long as no pause intervenes. The phenomenon is iterative, that is, it can apply to its own output. (1) illustrates all this with a few examples; for further details on Hungarian voicing assimilation, see among others Vago (1980), Siptár and Törkenczy (2000), and Kenstowicz, Abu-Mansour, and Törkenczy (2003); Pauliny (1979) and Rubach (1993) discuss Slovak.

(1) Regressive voicing assimilation in Hungarian and Slovak
  a. voiced obstruents voice preceding voiceless obstruents:
    i. Hungarian
      /tb/ → [db]: e.g., *hát-ba* 'back-ill.'; *két#barát* 'two friends'
      /ʃb/ → [ʒb]: e.g., *has-ba* 'stomach-ill.'; *hús#bolt* 'meat shop'
    ii. Slovak
      /tb/ → [db]: e.g., *kliat-ba* 'curse'; *brat#bol* 'brother was'
      /sb/ → [zb]: e.g., *pros-ba* 'request'; *nos#babke*
      'carry for grandmother'

  b. voiceless obstruents devoice preceding voiced obstruents:
    i. Hungarian
      /bt/ → [pt]: e.g., *láb-tól* 'foot-abl.'; *láb#torna* 'foot exercise'
      /zt/ → [st]: e.g., *víz-től* 'water-abl.'; *víz#torony* 'water tower'
    ii. Slovak
      /bk/ → [pk]: e.g., *Srb-ka* 'Serbian.fem.'; *Srb#ktorý*
      'Serbian who'
      /zk/ → [sk]: e.g., *z#kina* 'from cinema'; *obraz#ktorý*
      'picture which'

c. voicing assimilation is right-to-left iterative:
  i. Hungarian
    /gdh/ → [kth]: e.g., *smaragd-hoz* ‘emerald-allat’
  ii. Slovak
    /stb/ → [zdb]: e.g., *most#bol* ‘bridge was’

Obstruents do not devoice word-finally in Hungarian, but they do in Slovak, as shown in (2).

(2) Word-final obstruents
  a. Hungarian (do not devoice)
    láb-ak [b] ‘foot-pl.’ ~ *láb* [b] ‘foot’; *láp-ok* [p] ‘marshland-pl.’
    láp [p] ‘marshland’; *méz-ek* [z] ‘honey-pl.’ ~ *méz* [z] ‘honey’;
    mesz-ek [s] ‘limestone-pl.’ ~ *mész* [s] ‘limestone’

  b. Slovak (devoice)
    dub-a [b] ‘oak-gen.sg.’ ~ *dub* [p] ‘oak’;
    zväz-u [z] ‘union-gen.sg.’ ~ *zväz* [s] ‘union’

Crucially, the VR does not apply to obstruent–sonorant/sonorant–obstruent clusters. In monomorphemic words, both voiced and voiceless obstruents can stand next to a sonorant, as shown in (3).

(3) Obstruent – sonorant sequences
  a. Hungarian
    plakát ‘poster’, *blöki* ‘doggy’, *sróf* ‘screw’, *zrí* ‘fuss’

  b. Slovak
    plagát ‘poster’, *blesk* ‘lightening’, *smiech* ‘laughter’,
    zmija ‘viper’

Sonorants are not devoiced before voiceless obstruents or word-finally in either of the two languages, and sonorants do not trigger voicing in Hungarian. Sonorants in Slovak exhibit a peculiar behaviour. They (including vowels) trigger voicing in a specific environment, which may be roughly identified as the word boundary. That is, sonorants in Slovak do not voice within words and over “weak” morpheme boundaries, but they do across “strong” boundaries.

Turning to *v*, we can state that it – apparently – behaves asymmetrically with respect to the VR: it undergoes devoicing before all voiceless obstruents in both languages (4a), but does not trigger voicing in Hungarian at all or in Slovak monomorphemic words (4b), while it does in Slovak across word boundaries (4c):

(4) *v* and the VR
- a. *v* devoices before voiceless obstruents:
  - i. Hungarian
    /vt/ → [ft]: e.g., *sav-tól* ‘acid-abl.’ (vs. *savak* ‘acid-pl.’ [v])
    /vs/ → [fs]: e.g., *sav-szerű* ‘acid-like’
  - ii. Slovak
    /vt/ → [ft]: e.g., *v tom* ‘in that’ (vs. *v okne* ‘in window’ [v])
    /vs/ → [fs]: e.g., *v skrini* ‘in cupboard’
- b. *v* does not trigger voicing:
  - i. Hungarian
    /tv/ → [tv] (*[dv]): *két vár* ‘two castles’
    /sv/ → [sv] (*[zv]): *kész vár* ‘finished castle’
  - ii. Slovak (monomorphemic forms)
    /tv/ → [tv] (*[dv]): *tvár* ‘face’
    /sv/ → [sv] (*[zv]): *svet* ‘world’
- c. *v* triggers voicing in Slovak (across strong boundaries):
  /tv/ → [dv] *brat vám (zavolá)* ‘brother (calls you)’
  /sv/ → [zv] *pes vyje* ‘dog howls’

Thus in Hungarian, preobstruent/target *v* behaves as an obstruent, while postobstruent (and prevocalic)/trigger *v* patterns with sonorants. In Slovak, prevocalic onset *v* behaves as a sonorant, and preconsonantal onset *v* is an obstruent.

2.2. The distribution and realizations of *v* in Hungarian

Let us look at the basic facts concerning the distribution of *v* in Hungarian monomorphemic two-member consonant clusters. Table 1 below displays the logical possibilities of *v*’s clustering ability in three environments:

(i) intervocalically, (ii) word-finally (before a pause), and (iii) word-initially (after a pause).

Table 1. The distribution of *v* in monomorphemic words (CC clusters) in Hungarian. ● marks clusters that exist in the language, whereas ○ shows clusters of low frequency (in other words, it only occurs in a handful of words). For some clusters, there are no monomorphemic examples, but since they are included in the experiment below, they have also been included in the table; these non-monomorphemic clusters are placed in parentheses. Detailed charts for CC clusters with v and illustrative examples can be found in Kiss and Bárkányi (2006) for Hungarian and Bárkányi and Kiss (2007) for Slovak.

		p	t	c	k	b	d	ɟ	g	ʦ	ʧ/ʤ/f	s	ʃ	v	z	ʒ	m	n	ɲ	l	r	j	h
V_V	VvCV						○		○											●	○	○	
	VCvV	●	●	○	●	○	●	○	○			○	●		○		○	●	●	●	○	○	
_#	VvC#						(•)															(•)	
	VCv#						○										○	○	○	○	●	○	
#_	#vCV																			○	○		
	#CvV		○		●		○		○	○		○	●										

It is of course the blank cells of this table that represent the most interesting cases, that is, the clusters that are missing in the language, as well as those whose type frequency is low. The distribution of a single *v* is not restricted intervocalically, or word-initially; word-finally we do not find *-uv* and *-üv* in Hungarian, so this context is somewhat restricted. As soon as the position on either the left- or the right-hand side is occupied by a consonant, however, distributional restrictions do occur – with more severe restrictions cropping up preconsonantally, as indicated by low type frequency. Similar observations can be made with respect to the word-final as well as the word-initial position: *v*'s distribution is limited in the context of an adjacent consonant. These distributional effects are summed up in Table 2.

Word-initial CC clusters constitute a subcase where the VR is stricter: in this position, obstruent clusters in Hungarian are always voiceless (see, for example, Siptár and Törkenczy 2000: Chapter 5). If we consider *v* a voiced fricative (as the traditional approach does), then this segment is the only regular exception to this generalization, as we do find voiceless/voiced obstruent plus *v* clusters in this position (*tviszt* 'twist', *gvardián* 'guardian', etc.). In this position, thus, *v* patterns with sonorant consonants, which are free to occur here.

Table 2. The effect of the immediate environment on the distribution of *v* in Hungarian.

Left environment	*v*	Right environment	Example	Effects on *v*'s distribution
V	*v*	V	*kavics*	no restrictions
C	*v*	V	*medve*	few restrictions
V	*v*	C	*bóvli*	restricted
V	*v*	#	*sav*	few restrictions
C	*v*	#	*kedv*	restricted
#	*v*	V	*vas*	no restrictions
#	*v*	C	[v]*rangler*	restricted

Glosses: *kavics* 'pebble', *medve* 'bear', *bóvli* 'trash', *sav* 'acid', *kedv* 'mood', *vas* 'iron', *Wrangler* 'Wrangler jeans'.

In word-final position, however, *v* patterns with obstruents as it can cluster with sonorants as the second consonant. Sonorants do not normally occupy such a position.

(5) Word-final C*v* clusters in Hungarian (complete list)
[mv]: *hamv* 'ash'; [nv]: *ellenszenv* 'aversion', *rokonszenv* 'sympathy'; [ɲv]: *könyv* 'book', *enyv* 'glue'; [lv]: *elv* 'principle', *nyelv* 'language'; [rv]: *terv* 'plan', *szerv* 'organ', *érv* 'argument', *konzerv* 'tinned food', *ismérv* 'criterion', *keserv* 'sorrow', *mérv* 'extent', *orv* 'vile', *örv* 'guise', *sérv* 'hernia', *szarv* 'horn'; [jv]: *ölyv* 'hawk'; [dv]: *kedv* 'mood', *nedv* 'fluid', *üdv* 'salvation'

The generalizations regarding *v*'s behaviour that we discussed above can thus be summarized as follows: prevocalic *v* in a syllable onset behaves like a *sonorant*, while a *v* syllabified in a coda patterns as an *obstruent*. This is manifested both in *v*'s distribution in CC clusters and in its patterning with respect to the Voicing Requirement:

(6) Sonorant behaviour of Hungarian *v*
a. *v* can stand with obstruents word-initially, like the sonorants: (*tviszt* 'twist', *kvarc* 'quartz' ~ *tréfa* 'joke', *klarinét* 'clarinet')

b. “trigger” (postobstruent/prevocalic) *v* fails to satisfy the VR (even though it is voiced), like the sonorants: *hatvan* ‘sixty’ *[dv] ~ *hátra* ‘backwards’ *[dr]

(7) Obstruent behaviour of Hungarian *v*

a. *v* can stand after sonorants word-finally, like the obstruents:
könyv ‘book’, *terv* ‘plan’ ~ *vonz* ‘attract’, *torz* ‘distorted’

b. “target” (preobstruent) *v* satisfies the VR, like the obstruents:
/vt/ → [ft]: *sav-tól* ‘acid-abl.’ ~ /zt/ → [st]: *láz-tól* ‘fever-abl.’

The traditional literature on this “Janus-faced” behaviour of Hungarian *v* often draws a parallel between its two-fold patterning and its phonetic manifestation (see, e.g., Barkaï and Horvath 1978; Vago 1980; Siptár 1996; Siptár and Törkenczy 2000). Specifically, when *v* is realized with frication, it behaves like the obstruents; when it is realized without frication, it displays sonorant-like behaviour. Furthermore, these works suggest that *v* in C__# (e.g., *terv* ‘plan’, *könyv* ‘book’, etc.) preserves its voicing and also maintains frication. This aspect of *v*’s phonetics (voicing vs. frication) will be crucial in the phonetically based phonological analysis to be presented in this paper.

2.3. The distribution and realizations of *v* in Slovak

Let us now turn to the distribution of *v* in Slovak monomorphemic two-member clusters. The task of tabulating the distribution of *v* in Slovak is not as straightforward as it is in Hungarian. At this point the question arises of whether to regard a nominal monomorphemic only in the nominative case. If we do, we are left with almost no word-final C*v* and *v*C clusters. In Table 3 below, we make no formal distinction between singular nominative forms and other (oblique) cases: all are considered here as “monomorphemic”. For example, while *rv* and *vk* clusters can be exemplified by singular nominative forms (*nerv* ‘nerv’ and *huriavk* ‘hubbub’), the other clusters are represented by plural genitive forms (e.g., *rozplávb* ‘qualifying round in swimming’, *právd* ‘truth’, *sálv* ‘salvo’). Similarly, there are no intervocalic *vl* clusters in Slovak, unless we take cases other than the nominative into account (*Pavol~Pavla* ‘Paul.sing.nom.~sing.gen.’). We have also considered “monomorphemic” those forms in which the morphological boundary is completely obscured or non-transparent, like *bankovka* ‘bank note’, where the word *bank(a)* ‘bank’ is recognizable but the presence of the adjectival suffix *-ov* and the diminutive *-ka* are semantically rather implausible, or *advokát*, a loanword that in the source language contains the prefix *ad-*, but this morphological information is not available for most speakers of Slovak. The picture in Slovak is further complicated by the fact that a lone *v* is a preposition (meaning ‘in’, ‘into’) and a perfective verbal prefix, which means that it can form a word-initial cluster with almost any consonant (cluster).

Table 3 presents the distribution of Slovak C*v* and *v*C clusters in word-initial, intervocalic and word-final position.

Table 3. The distribution of *v* in CC clusters in Slovak (words with weak-boundary “+” affixes are included). As in the case of Hungarian in Table 1, clusters that occur monomorphemically are marked with •, and rare clusters with ○. Clusters that only occur in morphologically complex forms are indicated by +, and clusters of low type frequency of this latter kind are marked with †.

	p	t	c	k	b	d	ɟ	g	ʦ	ʧ	ʤ	f	s	ʃ	v	z	ʒ	m	n	ɲ	l	r	j	x	ɦ
VvCV				●	○	●	+	○	●	●			+	●		+	+	+	+	+	○	●	○	+	+
VCvV	+	●		●	†	○		●	+	○			+	●		●	●	○	●		●	●	○	○	+
VvC#				○	○	○																			
VCv#																					○	○			
#vCV	+	+	+	+	+	○	+	+	+	+	+		+	+		●	+	+	●	●	●	●	+	+	+
#CvV		●		●		●			●	●			●	●		●	○					●		●	●

Considering monomorphemic words in Slovak, we can state that the effect of the immediate environment on the overall distribution of *v* can be summarized along very similar lines as in Hungarian. Slovak *v* is restricted in positions where it does not stand next to a vowel: (i) in the immediate vicinity of consonants (especially when it occurs *before* consonants) and (ii) word-finally/before a pause after a consonant. The only difference between the distributional properties of Hungarian vs. Slovak *v* is that in Slovak, the postvocalic word-final position does not restrict the distribution of *v*. Actually – as we will discuss later – it is one of the positions in which the realization of *v* typically differs in the two languages: in Hungarian it is pronounced as a fricative, whereas in Slovak it is vocalized.

The phonetic realizations of *v* in Slovak have been claimed to be even more variable than in Hungarian. According to Kráľ (1974), *v* has four regular pronunciations: (i) it becomes [f] if it undergoes devoicing due to a following voiceless obstruent; (ii) in non-prevocalic onset position (and before nasals) it is a voiced fricative; (iii) in prevocalic position (and before non-nasal sonorants) it appears as an approximant; and (iv) it is vocalized elsewhere in the coda.

Table 4. The effect of the immediate environment on the distribution of *v* in Slovak.

Left environment	*v*	Right environment	Example	Effect on *v*'s distribution
V	*v*	V	*káva*	no restrictions
C	*v*	V	*vetva*	few restrictions
V	*v*	C	*ovca*	restricted
V	*v*	#	*názov*	no restrictions
C	*v*	#	*červ*	restricted
#	*v*	V	*voda*	no restrictions
#	*v*	C	*vdova*	restricted*

*If weak-boundary verbal prefixes are not considered. Glosses: *káva* 'coffee', *vetva* 'branch', *ovca* 'sheep', *názov* 'name', *červ* 'worm', *voda* 'water', *vdova* 'widow'.

In the remainder of the paper, we will seek to answer the following questions: (i) what acoustic phonetic features characterize the realizations of *v* in the various phonetic contexts in Hungarian and Slovak? (ii) do these properties significantly correlate with *v*'s double-faced behaviour in voicing assimilation and its distribution in the lexicon?

3. A phonetically based approach to *v*

There is a growing body of evidence that functional factors previously thought to be external to grammar can nevertheless exert a direct influence on it. These factors include such "low level effects" as speech production (articulation) and speech perception (Hume and Johnson 2001; Hayes, Kirchner and Steriade 2004). The basic idea that we pursue in this paper is that the phonetic, specifically, the aerodynamic and acoustic properties of sounds can regulate their phonological patterning, i.e., segmental distribution, allophony, and assimilation.

3.1. The phonetics of voiced fricatives

We begin our phonetically grounded analysis of Hungarian and Slovak *v* with a description of the phonetics of voiced fricatives, in particular, their aerodynamic properties.[2] For the articulatory system to target voicing and frication (turbulent noise) at the same time, an uneasy and "compromized" balance needs to be maintained (see Ohala 1983 on this point). High-amplitude turbulent noise requires a relatively high volume velocity of the airflow as it blows out from a constriction. In order to achieve this condition (i) the glottis is widely abducted so that the intraoral pressure equals or approaches the subglottal pressure, and (ii) the oral cavity is relatively constricted, creating a pressure drop across the supraglottal constriction (on the phonetic details of frication, see among others Shadle 1985; Stevens et al. 1992; Stevens 1998; Jesus 2001; Johnson 2003: 120–133; Krane 2005).

In contrast, for vocal fold vibration to be initiated, the following basic phonetic conditions are necessary: (i) the vocal folds must be set into modal phonation mode: they must be adducted or nearly so; (ii) subglottal air pressure must build up below the adducted vocal folds, forcing the lower part of the folds to blow apart (with the consequence that subglottal pressure drops close to zero relative to atmospheric pressure); and (iii) the negative pressure that occurs as air passes between the folds must suck the elastic folds together again (Bernoulli effect). These conditions constitute the physical bases of what is referred to as passive/modal voicing (Jansen 2004: 36). The most important feature of passive voicing is therefore that the air pressure below the folds must exceed the pressure above the folds. If the pressure above the folds builds up so that the pressure difference drops across the glottis, phonation ceases. If passive voicing cannot be achieved (such as during the closure phase of stops), sounds are said to be passively devoiced. To overcome passive devoicing a number of articulatory gestures, which aim at preserving a transglottal difference of pressure, need to be implemented to enlarge the oral cavity volume, e.g., raising the soft palate, advancing the tongue root so that there is an outward movement of the neck surfaces, lowering the larynx, expanding the pharyngeal volume, decreasing the stiffness of the vocal tract walls (reducing vocal tract compliance), or a combination of these gestures (see Stevens 1998: 465–486).

Based on the above, it has to be admitted that the concurrent production of frication noise and voicing involves *conflicting aerodynamic* *requirements* and *complex articulatory gestures*. Due to the inherently contradictory conditions of turbulence and voicing, their simultaneous preservation can only be achieved if the vocal tract is "reconfigured", thereby inhibiting the buildup of intraoral pressure as the supraglottal constriction area becomes narrow (so voicing can be maintained). This is what we will refer to as *active voicing*. In coarticulation-based theories, actively voiced sounds are thought to propagate their voicing to neighbouring sounds (especially those preceding them) more easily, as the actively set phonetic gestures can be maintained for a longer period of time, and they can more readily spill over to other gestures (see Farnetani 1997, Browman and Goldstein 1992, Jansen 2004, and Harris 2009, among others).

Our starting point in the analysis of *v*'s behaviour is based on the *aerodynamic conflict of the targets of v*: we assume that the simultaneous maximal realization of *both* voicing and frication at a labiodental place is compromised. Due to the aerodynamic conflict of *v*'s targets, if voicing is to be maintained, noise must be less turbulent (because of a necessarily wider constriction). This compromised (and rather uneasy) balance is best achieved between and before sonorants (especially vowels and wide approximants) because the oral cavity is relatively open for both some noise and voicing to be produced simultaneously. The sound which is thus expected to be realized between/before vowels is a passively voiced, moderately fricated sound due to a constriction which is presumably wider than that of voiceless and noisy fricatives but narrower than that of frictionless approximants. Building on the insights of Padgett (2002) concerning the phonetically grounded analysis of *v* in Russian, we thus also propose that *v* in Hungarian and Slovak (and in many other languages) is like a sonorant with respect to its voicing qualities but like a fricative in possessing (some) turbulent noise, too. Padgett (2002) calls this sound a *labiodental narrow approximant* (and transcribes it as [ʋ̝]), a term (and symbol) that we will henceforth also adopt for Hungarian and Slovak *v*. Padgett (2002) classifies Russian *v* as [+sonorant, -wide], and provides an OT-based formalized phonological analysis of *v*'s behaviour in voicing assimilation in Russian (employing a constraint system and binary feature representation). We believe that Padgett's analysis can be extended to Hungarian and Slovak. In this paper, however, we only wish to focus on the phonetic underpinnings of *v*'s behaviour in these languages.

With respect to its noise qualities, narrow approximant *v* is assumed to stand between actively voiced and noisy [v] and passively voiced approximant [ʋ]. Between two sonorants, [ʋ̝] is expected to display more vowel-like formant structure and less turbulent noise than any other fricative, but less formant structure and more turbulent noise than any other (“wide”) approximant.

In other contexts (not between or before (wide) sonorants) the balance between *v*’s targets is likely to shift. In accordance with the aerodynamic premise presented above, the prediction is that – unless active voicing strategies are employed – *v* cannot maintain its targets simultaneously. Two logical routes are predicted: (i) the noise/turbulence target is implemented, but voicing is not – in this case the sound that is produced is a devoiced labiodental fricative ([v̥]), or (ii) the noise/turbulence target (narrow constriction) is not implemented, but voicing is – in this case the sound that is produced is a passively voiced (sonorant) labiodental wide approximant/glide ([ʋ]). As we will see, Hungarian takes the first route, whereas Slovak takes the second. In other words, in the positions where *v* does not stand between/before vowels – for example, postconsonantally and word-finally – *v* is expected to be realized as a devoiced [v̥] (the “Hungarian route”), or as [ʋ] (the “Slovak route”). [ʋ̝]’s fricativization under devoicing is in line with the behaviour of other approximants; for instance, [l] and [j] (with wide “constriction”) also show frication (thus occur as [ɬ] and [ç]) when they occur in positions where they devoice.[3] In our approach, a devoiced narrow [ʋ̝] should thus show more frication than devoiced wide approximants do.

Our predictions as to the (likely) realization of [ʋ̝] in the various contexts are shown in Table 5.

Table 5. Likely realizations of *v* in various contexts. R stands for sonorants (especially wide sonorants like vowels and glides); O stands for obstruents.

		predicted likely realization	
context		Hungarian route	Slovak route
#__O; O__O; O__#	⇒	[v̥]	[ʋ]
R__O; R__#	⇒	[v̥]	[ʋ]
#__R; O__R	⇒	[ʋ̝]	[ʋ̝]
R__R	⇒	[ʋ̝]	[ʋ̝]

Thus, *v* is likely to occur as [ʋ̝] only before/between sonorants (especially vowels); in all other contexts, it is likely to devoice and obstruentize, or lose frication and vocalize. We must note that these predictions regarding [ʋ̝]'s realizations are founded on the phonetic (aerodynamic) premise alone; as we will see, other factors such as the possible coarticulatory effect of the active voicing of a following obstruent, or higher level functional factors such as morphological, analogical or paradigmatic effects may modify these expectations.

[ʋ̝]'s phonological behaviour with respect to the VR can be explained on the phonetic grounds that have been laid down here. In presonorant position, as we argued, [ʋ̝] is realized as a passively voiced (narrow) approximant. As such, it is expected to pattern with other sounds that bear similar voicing characteristics, like sonorants. It is for this reason that presonorant [ʋ̝] is hypothesized not to voice a preceding obstruent. The question fundamentally boils down to the phonetics of active/passive (de)voicing and voicing assimilation, which is also tackled in section 3.2.2. We see the core of the problem as being related to the coarticulatory properties of voicing targets (in this we assume the ideas put forth in Farnetani 1997, Jansen 2004, and Browman and Goldstein 1992, among others). Only actively voiced and devoiced sounds are assumed to participate in voicing assimilation as their voicing/devoicing-enhancing gestures can "spill over" into neighbouring segments (mainly those preceding them). Passively voiced sounds, on the other hand, do not possess voice-enhancing gestures, and so they "can have no co-articulatory effect on the voicing control of neighboring obstruents: […] there is simply nothing to spill over into flanking sounds" (Jansen 2004: 108).

When [ʋ̝] is not followed by a sonorant but by a voiceless obstruent (as in H. *savtól* 'acid-abl.', S. *vpád* 'fall'), the coarticulation-based voicing assimilation model (together with the aerodynamically driven approach we propose) predicts that [ʋ̝] should appear as a truly voiceless noisy fricative. This is because voiceless obstruents are claimed to be actively devoiced in Hungarian and Slovak with devoicing gestures that can spill over into [ʋ̝]. We assume that the active devoicing gestures of a following voiceless obstruent only enhance the voicelessness of [v̥], and so we expect a sound very close to [f] to be realized, with the possible consequence that the [f]-[ʋ̝] contrast will be neutralized in this context.

Consequently, the model hypothesizes that if *v* is followed by a voiced consonant, it will more easily receive voicing from it, and depending on the aperture qualities of that consonant, *v*'s realization will gradually move between stages of (i) a weakly fricated voiced narrow approximant [ʋ̝] (before vowels and wide sonorant consonants), (ii) a more turbulent voiced fricative, thus a sound close to [v] (before voiced obstruents), and (iii) a very noisy devoiced fricative [v̥]. For Slovak this is true only in cases where the "vocalization" route is not available (#*v*C clusters) – for more details see 3.2.2.

3.2. The realizations of *v* in Hungarian and Slovak – Acoustic experiments

3.2.1. Method

In the experiments presented here, we focused on the acoustic realization of *v* next to a consonant. (The list of test words is included in the Appendix.) We examined *v*C and C*v* clusters word-initially, word-finally and in intervocalic position (see Tables 1 and 3); the experiment included V*v*V sequences as a point of reference since we consider the realization of *v* in this position as the "prototypical" manifestation of what we described as the narrow labiodental approximant [ʋ̝].

Ten native speakers (six female and four male) of Standard Hungarian and five native speakers of Standard Slovak (three female and two male) were asked to read out the test sentences twice (containing words with all the clusters in Tables 1 and 3) at a normal speech rate in a soundproof cabin. The sentences all had neutral prosody (no test words occurred in contrastive topic or focus position). All initial clusters were in utterance-initial words, and all word-final clusters occurred in utterance-final position.[4] The age of the speakers ranged between 22 and 65. They were ignorant of the purpose of the experiment, were not trained phoneticians and were not paid for their participation in the experiment. The data were recorded with a Sony ECM-MS907 microphone onto a Sony MDMZ0710 minidisk in the case of Hungarian and onto a laptop through an M-Audio MobilePre USB preamplifier in the case of Slovak, digitized at 44100 Hz and resampled at 22050 Hz. The acoustic measurements were analysed using Praat (Boersma and Weenink 2005).

The experiment aimed to measure the following parameters:

(8) a. voicing
    b. harmonics-to-noise ratio (HNR; harmonicity median)

Voicing was measured on the basis of periodicity in the waveform, f_0 in the spectrogram, the presence/absence of voice striations in the spectrogram and Praat's voice report ("unvoiced frames percentage"). We used Praat's default settings (pitch range: 75 Hz-500 Hz, and with the following advanced pulses settings, maximum period factor: 1.3, maximum amplitude factor: 1.6, pitch setting was optimized for voice analysis – see the Praat manual).

To compare the relation of voicing to frication in the various realizations of *v,* we adopted Hamann and Sennema's (2005) method of finding what they call the *harmonicity median*, i.e., the degree of acoustic periodicity. The harmonicity median was determined by calculating the average of the harmonics-to-noise ratio with time steps of 0.01 s, a minimum pitch of 75 Hz, a silence threshold of 0.1 and 1 period per window. The interpretation of the median values is the following (see Boersma 1993). A harmonicity median of 0 dB means that there is equal energy in the harmonics and in the noise signal, whereas a median approximating to 20 dB indicates that almost 100% of the energy of the signal is in the periodic part. Based on this, a *v* with a harmonicity median around 15 dB suggests that it has a non-turbulent (glide) realization.

3.2.2. Results

Figures 1 and 2 exhibit boxplots of unvoiced frames (%) across subjects for Hungarian and Slovak *v* in the contexts listed in Tables 1 and 3. In cases where it was relevant, the consonantal group was split up into voiceless obstruents, voiced obstruents, and sonorant consonants. The environment "VvC#" is in parentheses in the Hungarian results because, as we said previously, we focused on monomorphemic clusters in the Hungarian experiment, but there are no monomorphemic V*v*C# clusters in Hungarian; we nevertheless included two morphologically complex words to test this context, too (*hívd* 'call.2sg.def.imp.' and *hívj* 'call.2sg.indef.imp.').

According to Figure 1, Hungarian *v* is realized with more than 50% of unvoiced frames in three contexts: after a consonant word-finally (mean unvoiced frames: 81%), before a voiceless obstruent (mean unvoiced frames: 67%), and word-finally after a vowel (mean unvoiced frames: 57%). *v* in this language was almost always voiced before voiced sounds, especially sonorants and vowels.

These results support our hypothesis that Hungarian *v* is likely to lose its phonation when it does not occur before sonorants, especially vowels. According to the results of two-tailed t-tests, the differences between the mean unvoiced frame values of the preconsonantal vs. non-preconsonantal groups were always statistically significant (with p always being less than or equal to 0.007).

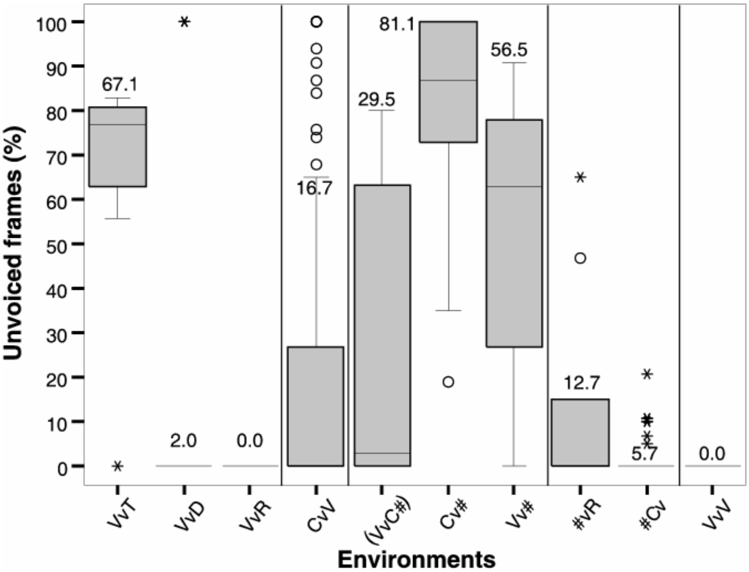


Figure 1. Boxplots (with outliers ○ and extreme cases ∗) of unvoiced frames (%) of Hungarian v for all speakers in various contexts. "C" = consonant, "T" = voiceless obstruent, "D" = voiced obstruent, "R" = sonorant consonant, "V" = vowel, "#" = word boundary. Numbers next to the boxplots refer to means.

As far as the Slovak results are concerned (Figure 2), the mean percentage of unvoiced frames was below 14% for all contexts, except for #*v*T, in which *v* was almost always devoiced (mean unvoiced frames: 89%) and for V*v*T (mean unvoiced frames: 44%).

Thus, in accordance with the expectations of the aerodynamic model introduced in the previous section, Slovak *v* remains voiced in all contexts (even in phonetically "unfavourable" ones), except when a voiceless obstruent follows it word-initially or a morpheme boundary intervenes between the preceding vowel and the *v*T sequence (i.e., V + *v*T). When there is a vowel before *v*T, however, the variation as to the voicing of *v* is greater. The difference in unvoiced frames between #*v*T/V*v*T and all the other contexts was always statistically significant (two-tailed t-tests, p's < 0.001).

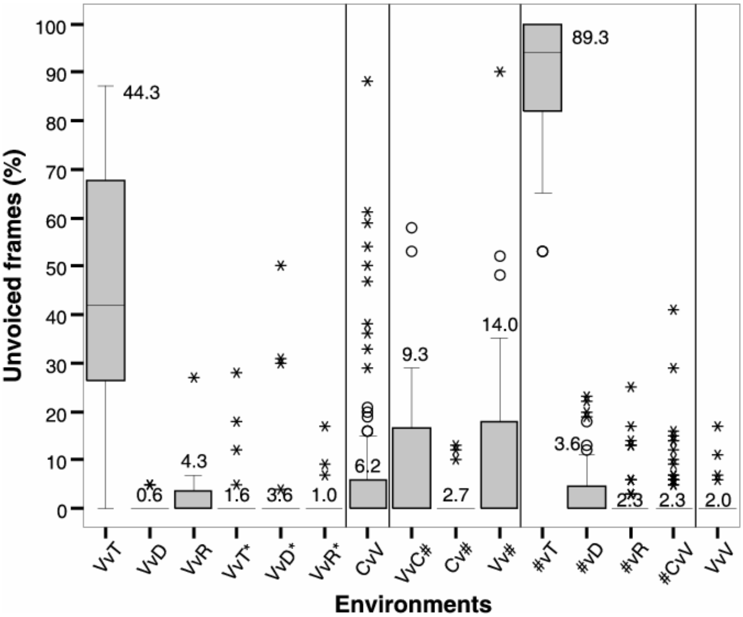


Figure 2. Boxplots (with outliers ○ and extreme cases ∗) of unvoiced frames (%) of Slovak *v* for all speakers in various contexts. "C" = consonant, "T" = voiceless obstruent, "D" = voiced obstruent, "R" = sonorant consonant, "V" = vowel, "#" = word boundary. "*" indicates those consonants that are not separated from *v* by a morphological boundary (i.e., the *v*C cluster belongs to the same morpheme; see Table 3 and also section 3.2.6.). Numbers next to the boxplots refer to mean values.

We will further discuss the details of the various contexts in the following sections. Figure 3 displays boxplots of the mean values for the harmonicity median of *v* that were computed for all subjects in various contexts in Hungarian. The same parameter for Slovak *v* is shown in Figure 4.

Figure 3 indicates that Hungarian *v* had the smallest harmonicity median (it was the *least* periodic) (i) word-finally (after a consonant or a vowel, means: 1.18 dB and 4.74 dB, respectively), and (ii) before a voiceless obstruent (mean: 3.27 dB). This result is a further indication that *v* in this language tends to be noisy/fricative-like in these contexts.

In contrast, when it stands before sonorants, *v* is much more periodic, with a harmonicity median above 10-12 dB (the highest level, 18.00 dB on average, was measured for intervocalic *v*'s). This result can again be interpreted as an indication that presonorant *v* is more a sonorant glide than a fricative.

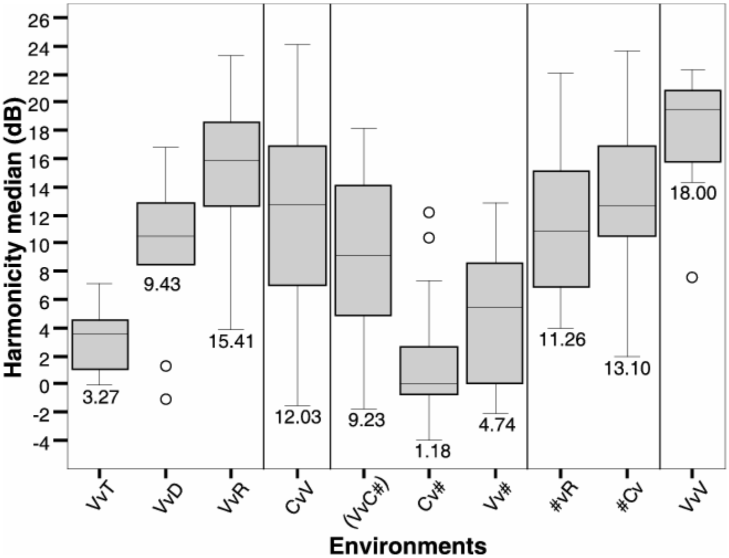


Figure 3. Boxplots (with outliers ○) of the harmonicity median (dB) of Hungarian *v* for all speakers in various contexts. Numbers next to the boxplots refer to means.

Lastly, *v* before a voiced obstruent occupies an intermediate position in this respect, suggesting that it is more noisy/fricative-like than the presonorant *v*'s. We observed considerable individual variation in some contexts in both languages. This issue will be touched upon in sections 3.2.3. – 3.2.5.

As far as the Slovak results are concerned (see Figure 4) two environments stand out with respect to the harmonicity median of *v*: (i) #_T (mean: -0.33 dB), and (ii) V_T (mean: 6.03 dB). It is in these contexts that Slovak *v* can be considered to be rather noisy. In all other positions, *v*'s harmonicity median showed higher values.

We also investigated the correlation between the harmonicity median and the percentage of unvoiced frames for both languages. The scatterplots that graph the harmonicity median values against the corresponding value of the unvoiced frame percentages are shown in Figure 5 (for Hungarian) and Figure 6 (for Slovak).

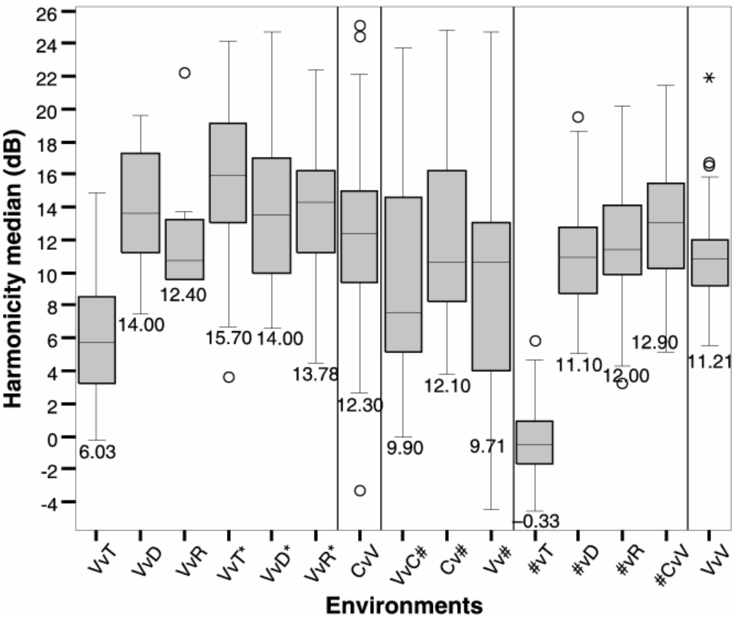


Figure 4. Boxplots (with outliers ○ and extreme cases *) of the harmonicity median (dB) of Slovak *v* for all speakers in various contexts. Numbers next to the boxplots refer to means.

To make viewing easier, we have collected the various environments into three groups, based on the predictions concerning the realization of *v* that we discussed in section 3.1. (see especially Table 5). Accordingly, circles (○) are meant to represent *v*-tokens that are expected to be realized as voiced approximants (in the case of Slovak, they also include contexts where the "vocalized" [ʋ] realizations are expected to arise). Triangles (Δ) stand for *v*-tokens that are expected to be realized as (partially or fully) devoiced and noisy, while horizontal lines (_) represent voiced and noisy *v*'s. In the case of fully voiced tokens (unvoiced frames = 0%), we have graphed *v*-tokens marked with "Δ" and "_" slightly offset to the left in both scatterplots so that they can be visually separated from the tokens marked by "○" better.

Similarly, in the case of Hungarian in Figure 5, fully devoiced tokens (unvoiced frames = 100%) represented as "○" have been graphed slightly offset to the right in order to distinguish them from the tokens marked with "Δ". At this point, however, our aim is to present a general overview of the correlation between harmonicity and voicing over all contexts; the detailed discussion of the individual environments will be tackled in the following sections.

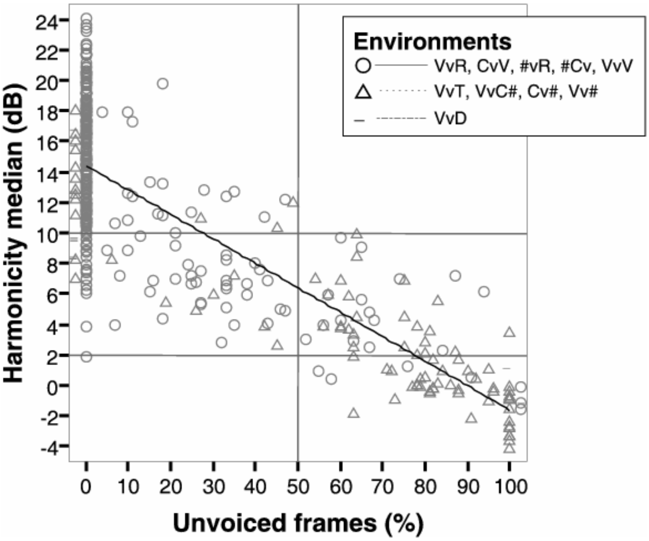


Figure 5. Scatterplot of the harmonicity median (dB) and unvoiced frames (%) of Hungarian v for all contexts.

In the case of Hungarian *v,* the harmonicity median significantly correlates with unvoiced frames (Pearson's correlation coefficient r = -.825, p < 0.001). This result indicates that the higher the harmonicity median, the lower the percentage of unvoiced frames (periodicity negatively correlates with voicing), and the higher the percentage of unvoiced frames, the lower the harmonicity median (devoicing negatively correlates with noise). As the scatterplot shows, there are no cases with a harmonicity median above 10 dB and a percentage of unvoiced frames above 50% (i.e., there are no periodic and voiceless *v*'s). More importantly, we do not find *v*'s whose harmonicity median is below 2 dB and whose percentage of unvoiced frames is below 50% (i.e., there are no noisy and voiced *v*'s). If the harmonicity median is between 2-10 dB, then there are both voiced and voiceless *v* tokens (whose percentage of unvoiced frames is between 0-50% and between 50-100%, respectively).

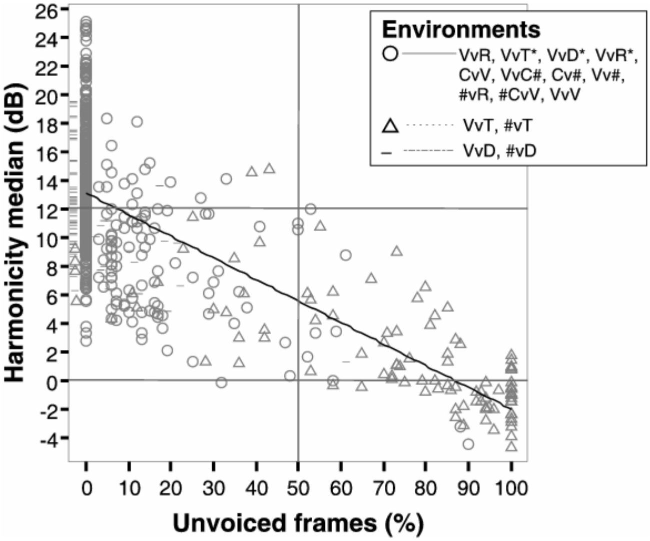


Figure 6. Scatterplot of the harmonicity median (dB) and unvoiced frames (%) of Slovak *v* for all contexts.

Figure 5 indicates that there are two large groups of *v* realizations in Hungarian: (i) voiced *v*’s, and (ii) noisy and (partially or fully) voiceless *v*’s. The separation of *v* tokens into these groups has also been verified by a k-means cluster analysis. All this suggests that earlier claims about the allophonic realizations of Hungarian *v* according to which there is a voiced fricatival realization and a voiced sonorant-like realization do not hold. As far as the *voiced v* realizations are concerned, in accordance with our hypotheses presented in section 3.1., they occupy a continuum that ranges from (i) weakly fricated approximants to (ii) more noisy fricatives. A k-means cluster analysis separated two subgroups within the fully voiced *v*-tokens, with the following harmonicity median cluster centres: 10.21 dB vs. 17.78 dB (the cluster centre for the voiceless tokens was at 1.50 dB). This we take as an indication of the fact that in Hungarian there are two – partially overlapping – groups of voiced *v*’s: (i) narrow approximant sonorant *v*’s ([ʋ̝]) and (ii) more noisy fricative *v*’s ([v]). Overall, the scatterplot confirms our predictions that Hungarian *v* is a passively voiced sound and it is strongly influenced by the neighbouring consonants.

The scatterplot of the harmonicity median and unvoiced frames of Slovak *v* (Figure 6) also suggests a correlation between the two parameters. This correlation, just like in the case of Hungarian, turned out to be significant (Pearson's correlation coefficient r = -.739, p < 0.001). We can conclude that periodicity negatively correlates with voicing, and devoicing negatively correlates with noise in the case of Slovak *v*, too.

A comparison of the scatterplots in Figures 5 and 6 indicates that in Slovak, unlike in Hungarian, there are more voiced tokens (whose unvoiced frames are below 50%) when the harmonicity median is below 10-8 dB. In other words, Slovak displays more voiced and (somewhat) noisy *v*'s than Hungarian does. Most of these tokens arise (as we will discuss below) in #_D, #_R, V_D, V_R and C_V. A comparison of the scatterplots in Figures 5 and 6 also suggests that there are more *v* tokens above 18-20 dB in Slovak than in Hungarian. We take this as an indication of the fact that Slovak also has *v* tokens that contain very little noise, namely, a very sonorous [ʋ]. Just as in the case of Hungarian, a k-means cluster analysis separates Slovak *v*'s into two large groups: (i) noisy and (partially or fully) devoiced [v̥]'s (with a harmonicity median below 2 dB) and (ii) a voiced group. Voiced *v*'s in Slovak form a continuum with a larger harmonicity median interval than in Hungarian, ranging from (i) wide approximant tokens ([ʋ]), with a harmonicity median around 18 dB, to (ii) narrow approximant *v*'s ([ʋ̝]), with a harmonicity median between around 10-18 dB, to (iii) somewhat noisy and voiced tokens ([v]), with a harmonicity median between 2-10 dB. The acoustic parameters we employed in the analysis (voicing and harmonicity median) cannot categorically differentiate between the main realizations of voiced Slovak *v* that the traditional literature has established (cf., for example, Kráľ 1974), but we can confirm that all these realizations exist. Further research is needed (e.g., formant analysis, intensity measurements) to separate the (perceptually clearly different) vowel-like wide approximant [ʋ] from the narrow approximant [ʋ̝] within the voiced *v* realizations.

Let us next examine the individual contexts in more detail. We will start with the word-final position because this best illustrates the two strategies predicted by the aerodynamic model.

3.2.3. Word/utterance-final v

In the previous sections we suggested that if the targets of *v* are unrealizable in a certain position, two strategies can be assumed to be available: either frication is preserved and enhanced and voicing is lost, or voicing is preserved, in which case the noise element is lost. Hungarian chooses the first route ([v̥]), while Slovak opts for the second, where *v* in an unfavourable position appears as a labial wide approximant/"offglide" [ʋ]/[u̯].

Our results supported our assumptions. In C_#, Hungarian *v* was almost always devoiced (mean unvoiced frames: 81%). Rounds of two-tailed t-tests showed that the voicing of *v* in this context differed to a statistically significant extent (with $p < 0.002$) from its voicing in all the other contexts except in V_T, i.e., before a voiceless obstruent. This suggests that the voicing of *v* in postconsonantal/word-final position and before voiceless obstruents is not significantly different. We got similar results concerning the voicing of *v* word-finally, after a vowel. Here, the mean percentage of unvoiced frames was 57%. Again, two-tailed t-tests showed that the voicing of *v* in this environment is significantly different (p was always smaller than 0.042) from all the other contexts, except V_T. We must note that we also observed the deletion of final *v* in one case.

As far as the harmonicity median is concerned, Hungarian *v* in C_# was produced with a fair amount of noise (mean harmonicity: 1.18 dB). Two-tailed t-tests showed that the harmonicity median of *v* here was statistically different ($p < 0.001$) from that of the other *v*'s in all the other contexts, except in (i) V_T and (ii) V_#. This indicates that postconsonantal/word-final, postvocalic/word-final and preobstruent *v*'s (when the obstruent is voiceless) are all rather noisy (as predicted). Similarly, *v* in V_# also turned out to be rather noisy (mean harmonicity median: 4.74 dB). Two-tailed t-tests showed that the *v*'s in V_# were significantly different ($p < 0.005$) from those in all other positions except: (i) V_T, (ii) C_#, and (iii) V_D. That is, the frication of *v* in V_# is similar to those *v*'s that occur before voiceless or voiced obstruents or those that stand after a consonant word-finally.

We can conclude that word-final *v* tends to be devoiced and noisy in Hungarian, just as the aerodynamic model predicted, which is a result that has not been reported in the literature on this consonant before.

In Slovak, on the other hand, deletion was one of the main strategies in C_# (*v* in this position was deleted in 17 cases out of 30). The other main strategy was the same as in the case of a postvocalic word-final *v*: vocalization. The "Hungarian route" (devoicing and frication) was also observed (marginally, in 3 cases), and in some cases *v* was realized as a voiceless bilabial stop with very short closure duration and an actual burst. Both the type frequency and the token frequency of *v* in this position are very low in Slovak, and this is precisely the context where we observed the most vacillation and variation in *v*'s realization. Based on these observations, we can state that speakers of Slovak do not have a consistent strategy for preserving *v* in this position.

As far as Slovak postvocalic, word-final *v*'s are concerned, they were typically articulated voiced (there were no V*v*# tokens with unvoiced frames above 50%, mean: 14%) and sonorous, without much frication (mean harmonicity median: 9.71 dB; however, with a large standard deviation (7.67 dB), showing that there was a large amount of variation with respect to the harmonicity median of *v* in V_#, just like in the case of C*v*#). These results are in accordance with our expectations: postvocalic, word-final *v* is realized as a voiced wide approximant or fricationless glide in Slovak.

In Figure 7 we provide wide band spectrograms (with corresponding waveforms) and narrow band FFT spectra, taken from the middle portion of *v*, in order to illustrate the typical realization of word-final postvocalic *v* in the two languages under discussion.

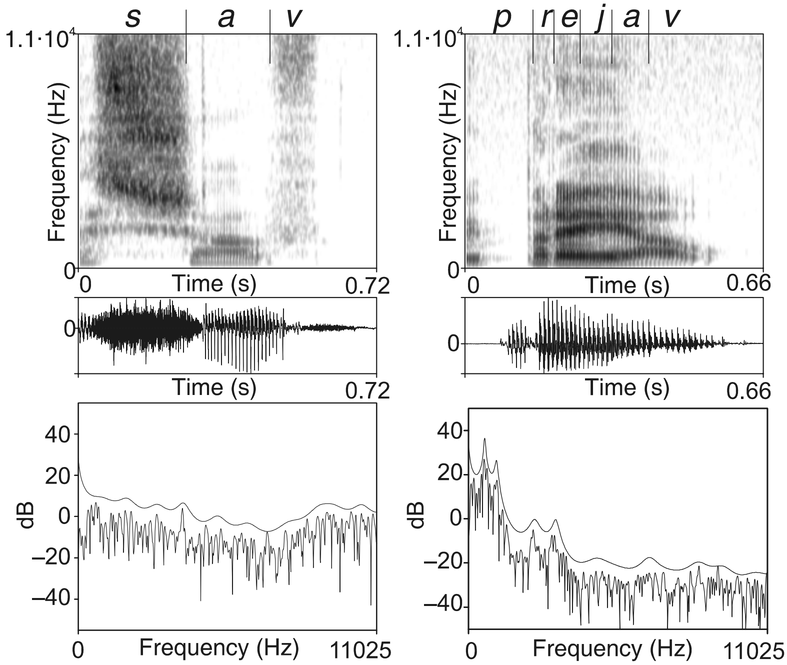


Figure 7. Realizations of Hungarian (left) and Slovak (right) *v* in *sav* 'acid' and *prejav* 'manifestation' (subject 7 and subject 1).

Linear predictive coding (LPC) smoothed spectra are superimposed on the FFT spectra for easier viewing. In Hungarian *sav, v* is not voiced (see the lack of glottal pulses and low frequency energy in the spectrograms and the aperiodicity on the waveforms), and there is no formant structure (see the abrupt start/finish of formant transitions at the neighbouring sonorant). The sound shows a flat/peakless spectrum, typical of (diffuse) labial/labiodental voiceless fricatives with energy spread over a large frequency range. In the Slovak word *prejav*, in contrast, the spectrogram and spectrum exhibit a typically approximant-like realization – consider the scarcity of noise at high frequencies, the clear formant structure throughout *v* and the preceding vowel, and the vertical striations in the spectrogram corresponding to glottal pulses, as well as the formant peaks at low frequencies in the spectrum. Thus, in accordance with earlier observations (Kráľ 1974), this segment is a frictionless glide.

Word-final *v*C clusters are very rare in both languages. While in Hungarian there are no such monomorphemic words, in Slovak most such words are plural genitives (e.g., *krívd* 'unfairness.pl.gen.', *právd* 'truth.pl.gen.'). In Slovak, as expected, *v* in this position was realized in our experiment as a wide approximant, an offglide.[5] Note that the consonant following *v* is itself in a phonetically unfavourable context; obstruents in this position are always devoiced in Slovak. In Hungarian we measured voicing and frication in two words (*hívd* 'call.2sg.def.imp.', *hívj* 'call.2sg.indef.imp.') and we found that word-finally, after another consonant, *d* and *j* tend to become devoiced: both were devoiced in 9/10 cases (*j* actually also became strongly fricated when it was devoiced, just like *v*). When these segments were devoiced, *v* was devoiced/fricated in five out of nine cases.

3.2.4. VvCV, VCvV and #CvV clusters

In these contexts, the effect of the adjacent consonants was particularly relevant for the realization of *v* (and, in the case of Slovak, morphology played an important role, too). Let us look at V*v*CV clusters in Hungarian first. We examined the following words: *felhívtam* 'I called', *bovden* 'V-shaped belt', *bóvli* 'trash', *Chevrolet* (car brand), *szovjet* 'Soviet'. 9/10 subjects produced *v* in *felhívtam* with more than 50% of the segment being devoiced (mean unvoiced frames percentage for all subjects: 67%). If we disregard the only extreme case (one subject pronounced a fully voiced *v* before voiceless *t*; see Figure 1), then the mean percentage of unvoiced frames for V*v*T jumps up to 75%. At the same time, the average harmonicity median of *v* was also very low (3.27 dB) here. In this position, Hungarian *v* thus turned out to be (partially) devoiced and rather noisy, as expected. Two-tailed t-tests showed that both the voicing and the harmonicity median of *v* in V_T were significantly different from all the other contexts except (i) C_# and (ii) V_#, that is, when *v* is word-final.

Let us turn now to the V_D position. Two subjects pronounced *v* fully devoiced in *bovden* (mean unvoiced frames: 100%). Actually, these subjects even devoiced the following *d*, too. Whenever *v* was devoiced, it also had a low harmonicity median (-1.05 dB and 1.26 dB, respectively), suggesting, again, that *v* here is voiceless and noisy. It was also in this context that *v* had a relatively low level of harmonicity median *when it was voiced at the same time*. Further investigations are necessary, but this result corroborates the active voicing analysis of voiced obstruents in Hungarian, saying that both (some) noise and voicing before a non-sonorant are only likely to be preserved when an actively voiced consonant follows. Two-tailed t-tests indicated that the difference between the voicing of *v* before voiced obstruents versus other contexts was only statistically significant in the case of (i) V_T, (ii) C_#, and (iii) V_#.

Before sonorant consonants, especially *j*, *v* had a rather high mean harmonicty median (above 14 dB), while it also preserved voicing. These facts suggest an approximant-like *v*-realization. Among the words we investigated, it looks as if Hungarian *v* is most unstable before voiceless *t*, then before the voiced obstruent *d*, and the most stable before *j* ("stable" referring to the fact that *v* can preserve all its articulatory targets). *l* and *r* occupy an intermediate position in affecting *v* in this way.

The strategy for Slovak *v* in this environment is similar to the case in word-final position: *v* is vocalized and non-turbulent. However, morphology in these cases plays an important role; see section 3.2.6. for more discussion.

Let us now focus on the *postobstruent* (and prevocalic) position. This context also showed a similar two-fold variation in both languages as the preconsonantal context did (in Hungarian). *v* either preserved its voicing and had a relatively high harmonicity median (suggestive of wider constriction, weak frication, periodicity) or it was devoiced and had a low harmonicity (suggestive of narrower constriction, stronger frication, aperiodicity). Again, however, the nature of the consonant (voicing, place, manner) does seem to play an important role: some consonants are more likely to cause devoicing of *v* than others. For example, in V*tv*V clusters in Hungarian, *v* kept its voicing more often than in V*pv*V clusters. In V*pv*V 6/10 subjects pronounced *v* with more than 60% of it devoiced and with low harmonicity (below 4 dB), while we got similar results in V*tv*V for only one subject (out of ten). Progressive coarticulatory effects similar to those caused by *p* were also observed for the voiceless sibilants (as in S. *Bečva* (name of a river) and H. *fösvény* 'miser', *köszvény* 'arthritis'), and the trilled rhotic *r*. The masking effect of the turbulence of the release noise of the stops and fricative noise is, we hypothesize, the reason behind the devoicing/frication of *v* in some of the cases, but further research is needed in this area.

Figure 8 shows an example of the two strategies in the pronunciation of Slovak *Bečva*. (See Kiss and Bárkányi 2006 and Bárkányi and Kiss 2007 for further illustrations of the two strategies in Hungarian and Slovak.)

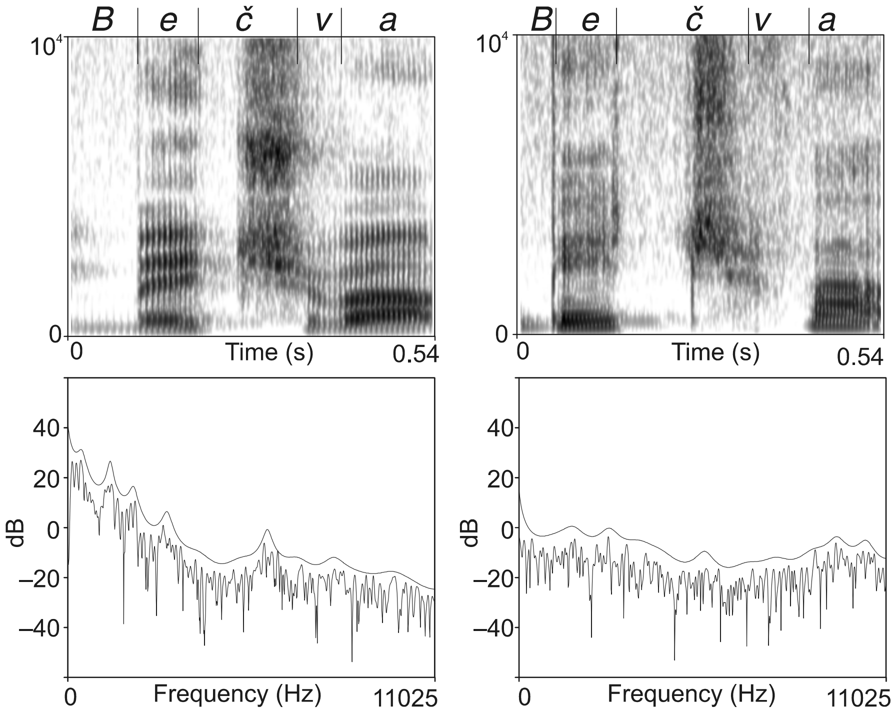


Figure 8. Realizations of Slovak *v* in *Bečva* (name of a river) (subject 1 vs. 2).

The realizational patterns of *v* can thus explain the apparent puzzle of why there is no voicing assimilation in voiceless obstruent–*v* clusters: when *v* is realized as a narrow approximant, its passive voicing cannot spill over to the preceding obstruent; when, however, it devoices (and becomes noisy), being voiceless, there is no voicing target to spread.

The influence of surrounding consonants on *v* is once again evident in voiced obstruents/sonorant–*v* clusters: in such cases *v* almost never got devoiced; in words like H. *dugvány* 'cutting', *medve* 'bear' and S. *advokát* 'lawyer', *jazva* 'scar', etc., all *v*'s were pronounced voiced with a relatively high harmonicity median (mean: 15.55 dB, standard deviation: 5.66 dB). These *v*'s were – as predicted – all sonorous and voiced.

Word-initial C*v* clusters showed a pattern somewhat similar to intervocalic C*v* clusters in both languages. Two-tailed t-tests indicated that the harmonicity median of *v* in the two environments did not differ significantly (most *v*-tokens in #C_V were produced with a harmonicity median above 10 dB, similarly to those in VC_V). However, this was not the case concerning the percentage of unvoiced frames: the two groups differed significantly with respect to this variable: only one *v*-token was pronounced with a ratio of unvoiced frames over 50%; all the other tokens were below 50% (and most were fully voiced, i.e., with 0% unvoiced frames). Word-initially, postconsonantal *v* (regardless of the voicing of the preceding consonant) is thus typically sonorous and voiced. We leave it open for future research to determine what the asymmetry between #C*v* and VC*v*V is due to.

3.2.5. #__C

This context, according to the hypotheses of our analysis, is also a highly infelicitous environment aerodynamically when it comes to preserving *v*'s targeted noise/voice. This phonetic markedness is also reflected by the rarity of such clusters in Hungarian. Their scarcity as well as the fact that they occur in foreign proper names made their testing rather difficult and therefore conclusions are hard to draw for Hungarian. It was nonetheless precisely in these tokens that we observed the highest variability in the realization of *v* across subjects (similarly to C*v*# clusters in Slovak). The following four "strategies" were noticeable: (i) some subjects attempted to pronounce these words rather slowly/carefully (as if putting them into "phonological quotes"); in these cases, both voicing and some noise were preserved; (ii) a few subjects pronounced these *v*'s devoiced and with a turbulent noise (as predicted by our model); (iii) in the case of *Wrangler* (name of a brand of jeans), *v* was pronounced by some subjects as if it was an English [w] (even though, of course, no [w] occurs in this word in English, and there is no labiovelar glide in Hungarian); and lastly (iv) we also observed *v*-deletion in the word *Vlach* (proper name). Actually, whenever *v* was preserved as voiced, it was extremely short in all cases (30-35 ms), which made segmentation very difficult (illustrative waveforms, spectra and spectrograms can be found in Kiss and Bárkányi 2006).

In Slovak, on the other hand, the #*v*C sequence is very frequent since in this language a lone *v* can be a verbal prefix as well as a preposition, as mentioned in 2.3.; contrast preservation is very important in these cases since word-initial *v* has a high functional load. The expected strategy for Slovak would be to lose frication and preserve voicing so as to realize *v* as a wide approximant. To achieve this, however, *v* must rely on a preceding vowel, so that it can be realized as an offglide. In word-initial #*v*C clusters no such vowel is available, thus this is the position where the only available strategy is the "Hungarian route"; therefore, this is precisely where the phonetic properties of the neighbouring consonants really gain importance in Slovak. This means that when *v* is followed by an actively voiced consonant, it tends to be actively voiced with a fair amount of frication ([v]) as in *vbehnúť* 'to run in', and when *v* is followed by a voiceless consonant, *v* itself is devoiced, realized as [v̥]/[f], as in *vpád* 'fall'. Furthermore, *v* in this position is typically short in Slovak.

One of the acoustic correlates of active voicing is the presence of prevoicing (word-initially). In many cases we observed prevoicing before sonorants, too; as we mentioned in section 2.1. above, sonorants in Slovak trigger voicing assimilation across certain morphological boundaries.[6] Figure 9 illustrates Subject 1's pronunciation of *vmietla* 'cast sth in sb's face': *v* in this position is a true voiced fricative, with voicing and turbulent noise present at the same time (which also suggests that the following sonorant is actively voiced, too). Simultaneous frication and voicing can be achieved by actively voicing the whole consonant cluster; note that there is prevoicing, too (highlighted by a circle). The intensity graph superimposed on the spectrogram shows intensity before the *v* constriction, also indicating the presence of voicing (this initial intensity can also be seen at the beginning of the waveform below the spectrogram).

According to the results of the experiment, the range of duration of the prevoicing phase for *v* in Slovak was between 10 to 60 ms overall, but typically between 10 to 30 ms. There was a correlation between the harmonicity median, voicing and prevoicing: whenever *v* was voiced *and* had a relatively low harmonicity median (below 10 dB), *v* was predominantly prevoiced by at least 10 ms. Actually, when *v* was voiced and had a harmonicity median below 6 dB, *v* was always prevoiced, mostly by 20 to 40 ms (Pearson's correlation coefficient r = -.493, p < 0.001).

All this indicates that initial Slovak *v* tends to be prevoiced when it is realized with *both* low harmonicity (that is, when it is produced with narrow constriction, hence with turbulence) and voicing during its constriction phase. It is perhaps not surprising then that this sound can actively participate in voicing assimilation in this position, just like other actively (pre)voiced obstruents.

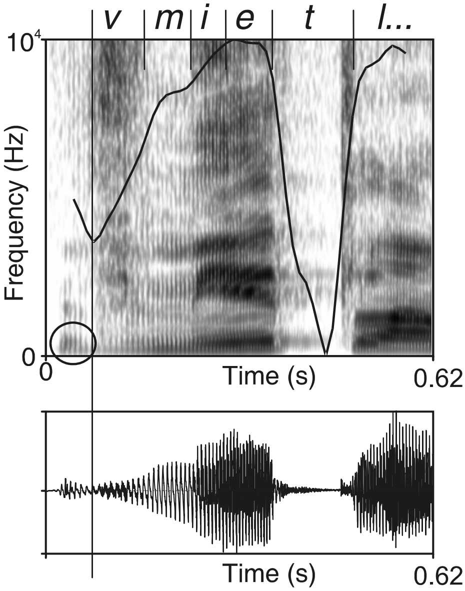


Figure 9. Spectrogram and waveform of the production of Slovak *v* in *vmietla* 'cast sth in sb's face' by subject 1. The spectrogram also contains an intensity graph. The prevoicing in the spectrogram is highlighted by a circle.

3.2.6. The role of morphology in Slovak

The role of morphology in Slovak with respect to *v*C and C*v* clusters can be considered from two perspectives. Firstly, one must look at whether a certain cluster exists monomorphemically or only in morphologically complex forms – we have touched upon this issue in section 2.3. Secondly, one must also consider whether the morphological structure of the word influences the phonetic realization of *v*. It is this latter case that we will elaborate on now.

We mentioned earlier that in Slovak, a lone *v* is a preposition and a verbal prefix, which means that it can form a word-initial cluster with almost any consonant (cluster), and the preservation of this lone *v* is essential due to its high functional load. We discussed word-initial *v*C clusters in the previous section. Morphology in those cases does not influence *v*'s phonetic realizations; it can be accounted for on aerodynamic grounds. It is in word-internal *v*C clusters that the morphological composition of the form determines the actual realization of *v*. Let us look at some examples. In the word *zavčas* 'early', *v* is realized as a voiceless fricative [f], while our model – in the light of the "Slovak route" (preserving voicing and losing frication) – predicts a wide approximant. This word, however, is clearly decomposable into *za*+*v*+*čas* 'prefix of circumstance + in + time' morphologically as well as semantically, so *v* is realized as it is realized in the expression *včas* 'in time', where it is word-initial and where it is pronounced [f] as predicted by our model – *včas* and *zavčas* are, of course, in a close morphological and semantic relationship; we consider the fact that *v* in *zavčas* is realized in the same way as in *včas* to be a kind of output-output identity effect.

In the same phonological environment, with no morphological boundary intervening between *v* and the preceding vowel, *v* is realized as an offglide (*dievča* 'girl'), as predicted by the aerodynamic model. The same (predicted) wide approximant appears if a morphological boundary occurs between *v* and the following consonant, as in *bravčový* 'pork adj.', which is decomposable into *brav* 'pork' and the adjectival suffix -(*č*)*ový*. The *v* in *brav* is realized as a wide approximant, as expected in word-final position, and *brav* and *bravčový* are in a close morphological and semantic relationship.

The same difference is observed in the realization of *v* in the words *nevbehol* 'did not run in' and *stavba* 'building site', for instance. In the first case, *v* is generally realized as a voiced fricative, which apparently contradicts the aerodynamic model, while in the second case it appears as a wide approximant, as expected. It is again the morphological structure of the two words that is responsible for the difference. In the first example, a morphological boundary intervenes between *v* and the preceding vowel *ne+vbehol* ‘not+ran in’ (*vbehol* itself is decomposable into *v+behol* ‘in ran’), so *v* is realized in the same way as in *vbehol*, where it is word-initial preceding an actively voiced obstruent. The word *stavba*, on the other hand, is decomposable into *stav+ba* (*stavať* means ‘to build’ and *-ba* is a deverbalizing nominal suffix), so *v* in this case appears as an offglide as expected in our model.

In some cases, however, the relevant morphological boundary is obscured. The word *návšteva* ‘visit’ provides an example of such a case, which is sometimes, for some speakers, pronounced with a wide approximant, as predicted, and sometimes with [f], as if it were composed of *ná+všteva*.[7] Figure 10 illustrates the “morphologically simple” and “morphologically complex” realizations of *v* in *návšteva*.

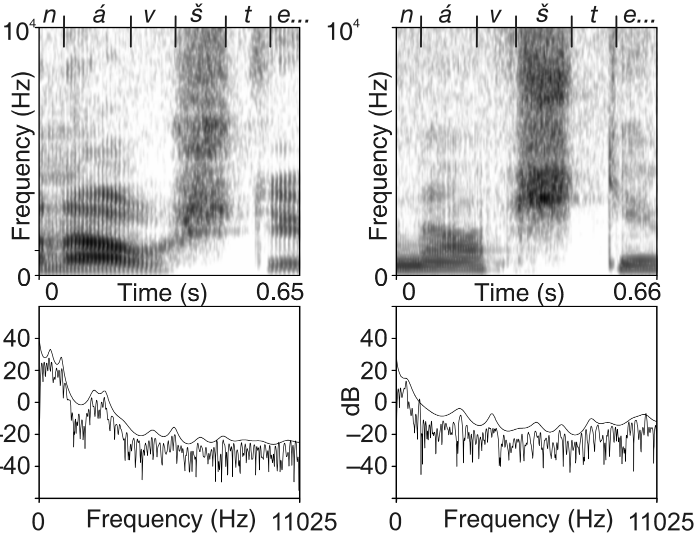


Figure 10. The two realizations of Slovak *v* in *návšteva* ‘visit’ (subject 1 (left) vs. subject 8 (right)).

4. Conclusion and remaining issues

This paper has put forth a unified analysis of Hungarian and Slovak *v*, in which its phonological patterning can be explained in a model based on the phonetic properties of this segment and its linear context. The most important claim has been that the phonetic targets of *v* are inherently contradictory on aerodynamic grounds (Ohala 1983) and can only be maintained in phonetically favourable positions. In such "beneficial" environments, the model predicts the emergence of a passively voiced narrow approximant [ʋ̝]. The articulatory targets of this consonant have been proposed and found to be preservable between sonorants (primarily vowels and (wide) approximants). In other positions, [ʋ̝] is predicted to give up one of its articulatory targets (either voicing or narrow constriction, i.e., turbulent noise). As a result of this, two realizations are possible: when *v* devoices, it becomes a strongly fricated, noisy sound (narrow constriction and wide abduction of the vocal folds); when its voicing target is kept, it loses much of its frication (wider constriction). We have argued that languages are free to choose which of the two routes they will follow. In a series of acoustic experiments that inspected voicing and the harmonicity median, Hungarian has been shown to be a language which prefers the devoicing strategy in aerodynamically unfavourable positions, while Slovak opts for "denoising". As a consequence of the phonetic realizations (owing to specific, well-definable and purely *phonetic* factors), *v*'s phonological behaviour, including both its role in what we call the "Voicing Requirement" and its phonotactic distribution, can also be explained. Neither a prevocalic [ʋ̝], being passively voiced, nor a devoiced [v̥] will induce voicing assimilation as a potential trigger. *v* has also been found to be highly dependent on the phonetic properties of surrounding sounds.

In the surface-oriented model we have proposed, all cases of voicing assimilation involving *v* are explained, including cases where other functional factors are taken into consideration, like the morphological structure of words, as we have showed in the case of Slovak.

This analysis is a first step in a phonetically rooted phonological analysis of *v* in Hungarian and Slovak. The analysis, we assume, can be extended to a perceptually based account of *v*'s phonotactics (and the phonotactics of consonant clusters in general), the details of which, however, must be worked out in future research.

In our view, the aerodynamic model presented here can be extended to the analysis of *v*'s patterning in other languages as well. The labiodental fricative/approximant – obviously – does not behave the same way in all languages. For example, in Polish, as Padgett (2002) reports, voiceless obstruents assimilate to *v,* and English *v* (just like the rest of the voiced fricatives) has also been reported to increase the voicing of a preceding obstruent (Jansen 2004). As Padgett (2002) discusses in detail, it seems that the functional factors of contrast and contrast dispersion, in addition to the phonetic factors, play an important role in languages where *v* voices: apparently, *v* only triggers voicing assimilation when it is actively voiced (like Polish and English *v*, and unlike Hungarian *v*). Preliminary data suggest that if a language has a contrastive actively voiced [v], it also has a contrastive passively voiced approximant counterpart, [ʋ/w]. It is as if active voicing and frication ([v]) are employed so that the contrast can be maximally salient and distinct from passively voiced and weakly fricated or frictionless [ʋ/w]. Such a two- or three-way contrast is reported for a few languages in Ladefoged and Maddieson (1996) as well as Padgett (2002): crucially, the non-approximant [v] in these languages is actively voiced and fricated. The tentative prediction is thus that we should find the Hungarian way of *v*-patterning in those languages in which *v* does not contrast with other voiced labial/labiodental fricatives/approximants. No doubt, much further research is needed in this area, too; nevertheless, in this paper we have found some preliminary evidence that functional factors can indeed play a role in both strong frication (narrow constriction) and vocal fold vibration being preserved during the production of *v* in word-initial position in Slovak.

Future research is also needed to answer such questions as: why do certain consonants trigger devoicing of *v* to a higher degree than others? How exactly can the link between final obstruent devoicing and sonorant voicing be explained in a phonetics-based model? How can the potential phonological neutralization of *v* be described (is it partial or complete) – especially in the case of word-final C*v* clusters in Hungarian? If neutralization is partial, what phonetic parameters help maintain/perceive the contrast? How do speakers generally perceive and interpret the devoicing of *v*? How can the analysis be extended to other consonants (especially voiced fricatives and approximants)?[8] Future research into the phonetics of Hungarian and Slovak *v* must also incorporate other parameters (like intensity measurements).

Even though our aim was not to give a complete account of how a non-formalist phonological framework is built up, nor to provide a formalized constraint-based phonological account of *v*'s challenging behaviour with respect to the Voicing Requirement in Hungarian and Slovak, we see the current analysis as a useful contribution to the growing body of work on surface-based phonology, according to which synchronic grammar (or grammatical change) is directly influenced by such low-level functional factors as the aerodynamics of articulation, as well as a contribution to the phonetic and phonological literature on the description and typology of *v*.

Appendix

Hungarian words used in the acoustic experiment with their glosses:
bovden 'V-shaped belt', *bóvli* 'trash', *Chevrolet* (car brand), *cvekedli* 'pasta with cabbage', *dugvány* 'cutting', *Dvorzsák* (proper name, Dvořák), *fegyveres* 'armed', *felhívtam* 'I called', *fösvény* 'miser', *Guatemala* (country name), *gvardián* 'guardian', *hamvas* 'blooming', *hatvan* 'sixty', *hívd* 'call.2sg.def.imp.', *hívj* 'call.2sg.indef.imp.', *jókedv* 'good mood', *konkvisztádor* 'conquistador', *kotyvaszt* 'concoct', *könyv* 'book', *köszvény* 'arthritis', *kvarckristály* 'quartz crystal', *likvid* (*tőke*) 'liquid (capital)', *lopva* 'furtively', *medve* 'bear', *nyelv* 'language', *orvos* 'doctor', *ölyv* 'hawk', *özvegyasszony* 'widow', *sátorponyva* 'tent canvas', *sav* 'acid', *svédcsepp* 'Swedish drops' (medicine name), *szerv* 'organ', *szovjet* 'Soviet', *szubvenció* 'subsidy', *szvetter* 'cardigan', *tolvaj* 'thief', *tviszt* 'twist', *Udvaros* (proper name), *vlach* 'Vlachian' (also: proper name), *Wrangler* 'Wrangler jeans'.

Slovak words used in the acoustic experiment with their glosses:
balvan 'idol', *bankovka* 'banknote', *Bečva* (river name), *búlv* 'eyeball.gen.pl', *chváliť* 'praise', *cvičiť* 'exercise', *červ* 'worm', *čvirikať* 'chirp', *dievča* 'girl', *dovtedy* 'till then', *dôvtipný* 'inventive', *dvor* 'yard', *huriavk* 'hubbub', *húžva* 'crumple', *hviezda* 'star', *javmi* 'phenomenon.pl.instr.', *jazva* 'scar', *Jevgenij* (proper name), *konvoj* 'convoy', *krivdiť* 'be unfair to', *kvet* 'flower', *larva* 'larva', *lichva* 'cattle', *návrat* 'return', *návšteva* 'visit', *navždy* 'forever', *názov* 'name', *nerv* 'nerv', *nevbehol* 'did not run in', *nevcháza* 'does not go in', *nevhodný* 'unsuitable', *obuv* 'shoe', *obrovský* 'huge', *obvod* 'district', *ovca* 'sheep', *ovplyvniť* 'bias', *Pavla* 'Paul.sing.gen.', *podošva* 'sole', *pravda* 'truth', *právd* 'truth.pl.gen.', *prehvízda* 'whistle', *prejav* 'manifestation', *prevzal* 'took over', *rovno* 'straight', *rozplávb* 'qualifying round in swimming.pl.gen.', *rvať* 'grapple', *sálv* 'salvo.pl.gen', *sekvencia* 'sequence', *správne* 'in the right way', *stavba* 'building', *stromov* 'tree.gen.pl', *svet* 'world', *švagor* 'brother-in-law', *telocvik*

‘PE’, *Topvar* (brand name), *tvoj* ‘your.sing.masc.’, *vchod* ‘entrance’, *vdova* ‘widow’, *vďaka* ‘thank’, *vetva* ‘branch’, *vgúliť* ‘roll in’, *vhodne* ‘properly’, *vjazd* ‘drive in’, *vklad* ‘deposit’, *vlak* ‘train’, *vmietnuť* ‘cast sg in sy’s face’, *vnada* ‘appeal’, *vňať* ‘stem’, *Vojvodina* (geogr. name), *vpád* ‘fall’, *vrabec* ‘sparrow’, *vsadiť* ‘plant in’, *vysvetliť* ‘explain’, *vzácny* ‘precious’, *vžiť sa* ‘accustom to’, *zátvorka* ‘parentheses’, *zvuk* ‘noise’, *žviakať* ‘ruminate’.

Notes

* We are grateful to Katalin Mády for her assistance with the acoustic measurements and statistics, and the reviewers and editors for their helpful comments on the issues tackled in this paper. We thank Péter Siptár and Klara Young for their comments, too. All remaining errors are ours. Our work was supported by the HNRF grants no. TO49327 and no. PD050018, as well as by the Bolyai Grant.
1. A note on notation. We will simply be using the orthographic form *v* to refer to what is usually and traditionally described as the “voiced labiodental fricative”. When the exact phonetic identity is at issue, we will use proper IPA symbols in square brackets (e.g., “[ʋ]”). Later on, the exact phonetic identity (and variants) of this sound will be made more explicit, and from then on the appropriate symbols will be used. The IPA transcriptions of the Hungarian letters the interpretation of which is non-obvious are as follows: *ty* = [c], *gy* = [ɟ], *sz* = [s], *s* = [ʃ], *zs* = [ʒ], *c* = [ʦ], *cs* = [ʧ], *dzs* = [ʤ], *ny* = [ɲ], *ly* = [j]; *a* = [ɒ], *á* = [aː], *e* = [ɛ], *é* = [eː], ö = [ø], ü = [y]. An acute accent over vowel letters signals length. In Slovak orthography, the letters *ť, ď, ľ* represent palatalized/(pre)palatal [t^j]/[c], [d^j]/[ɟ] and [l^j]/[ʎ], respectively; the wedge (or hachek) also signals palatal quality, thus: *š* = [ʃ], *ž* = [ʒ], *č* = [ʧ], *dž* = [ʤ], *ň* = [ɲ]. *c*, like in Hungarian, stands for [ʦ], while *dz* represents [ʣ]. *ch* signals the voiceless velar fricative [x], whereas *h* is realized as the *voiced* laryngeal fricative [ɦ]. *y* is used for [i] (to indicate that the preceding alveolar segment is not palatalized). *ä* is pronounced as [æ] or [ɛ], whereas *ô* stands for the diphthong [u̯ɔ]. In the orthography of this language, too, an acute accent over vowels (and syllabic consonants, such as *ĺ*) signals length.
2. The phonetically grounded analysis of the phonology of *v* to be presented here is primarily rooted in aerodynamic factors; however, we assume that the articulation-based realizations of *v*’s targets may also influence the salience of the *perceptual* cues of this sound in the various positions it occurs in. A perception-based account of the phonology of *v* is thus also plausible (similar to Steriade’s (1997, 1999) analysis of voicing and place contrasts of stops, and their neutralization); however, due to the lack of additional and detailed (experiment-based) data on the perceptual cues of voiced fricatives

(especially those of *v*), the perceptual effects can only be assumed here, and must be the object of future research. See, however, Balise and Diehl (1994), who concentrate on the typology of (primarily sibilant) voiced fricatives, and their apparent cross-linguistic markedness in consonant inventories. They claim that the presence of voicing interferes with the perception of place cues in fricatives. Voicing-based laryngeal contrasts in fricative inventories tend to be neutralized because it is relatively hard to recover their place cues. Balise and Diehl cite two pieces of evidence in support of this: the presence of voicing in a fricative reduces the amplitude of frication noise, which is an important cue for place contrast; furthermore, studies of consonant confusions indicate that across various harmonicity medians voiceless fricatives are identified correctly more often than their voiced counterparts.

3. See Padgett (2002) for examples from Norwegian, Iberian Spanish and French. Hungarian [j] also displays a similar conduct in neutralization-prone contexts, such as after an obstruent and before a pause: *lépj* [leːpç] ‘step.imp’.
4. Word stress in both Hungarian and Slovak falls on the initial syllable.
5. Offglides always depend on a neighbouring vowel both articulatorily and perceptually. So the vocalization route is not available in a #_C context, for instance, in Slovak.
6. The acoustic characteristics of sonorant voicing in Slovak are worth further research.
7. *ná-* is a prefix indeed, however *všteva* is not a word/root in Slovak. This type of variation is observed in very few lexical items, and individual and dialectal differences exist.
8. Some of these issues are tackled in Bárkányi and Kiss (2007, 2009).

References

Balise, Raymond R., and Randy L. Diehl
1994 Some distributional facts about fricatives and a perceptual explanation. *Phonetica* 51: 99–110.

Barkaï, Malachi, and Julia Horvath
1978 Voicing assimilation and the sonority hierarchy: Evidence from Russian, Hebrew and Hungarian. *Linguistics* 212: 77–88.

Bárkányi, Zsuzsanna, and Zoltán Kiss
2007 A phonetically-based approach to the phonology of *v*: A case-study from Hungarian and Slovak. Paper presented at the 4th Old World Conference on Phonology, January 18–21, 2007, Rhodes.

Bárkányi, Zsuzsanna, and Zoltán Kiss
2009 Hungarian *v*: Is it voiced? In *Approaches to Hungarian 11: Papers from the New York Conference*, Robert M. Vago, and Marcel den Dikken (eds.), 1–28. Amsterdam/Philadelphia: John Benjamins.

Boersma, Paul
1993 Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences (University of Amsterdam)* 17: 97–110.

Boersma, Paul, and David Weenink
2005 Praat: Doing phonetics by computer (version 4.3.19) [Computer program]. Retrieved July 20, 2005. (http://www.praat.org/)

Browman, Catherine, and Louis Goldstein
1992 Articulatory Phonology: An overview. *Phonetica* 49: 155–180.

Farnetani, Edda
1997 Coarticulation and connected speech processes. In *The Handbook of Phonetic Sciences*, William J. Hardcastle, and John Laver (eds.), 371–404. Oxford/Malden, MA: Blackwell.

Hamann, Silke, and Anke Sennema
2005 Acoustic differences between German and Dutch labiodentals. *ZAS Papers in Linguistics* 42: 33–41.

Harris, John
2009 Why final devoicing is weakening. In *Strength Relations in Phonology,* Kuniya Nasukawa, and Phillip Backley (eds.), 9–46. Berlin/New York: Mouton de Gruyter.

Hayes, Bruce, Robert Kirchner, and Donca Steriade (eds.)
2004 *Phonetically Based Phonology*. Cambridge: Cambridge University Press.

Hume, Elizabeth, and Keith Johnson (eds.)
2001 *The Role of Speech Perception in Phonology*. New York: Academic Press.

Jansen, Wouter
2004 Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch. Ph.D. diss., Rijksuniversiteit Groningen.

Jesus, Luis Miguel Teixeira de
2001 Acoustic phonetics of European Portuguese fricative consonants. Ph.D. diss., University of Southampton.

Johnson, Keith
2003 *Acoustic and Auditory Phonetics*. 2nd ed. Oxford: Blackwell.

Kenstowicz, Michael, Mahasen Abu-Mansour, and Miklós Törkenczy
2003 Two notes on laryngeal licensing. In *Living on the Edge: 27 Papers in Honour of Jonathan Kaye*, Stefan Ploch (ed.), 259–282. Berlin/New York: Mouton de Gruyter.

Kiss, Zoltán, and Zsuzsanna Bárkányi
2006 A phonetically-based approach to Hungarian [v]. *Acta Linguistica Hungarica* 53: 175–226.

Kráľ, Ábel
1974 Spoluhláska *v* a jej varianty [The consonant *v* and its variants]. *Kultúra slova* 8: 292–296, 338–343.

Krane, Michael H.
2005 Aeroacoustic production of low-frequency unvoiced speech sounds. *Journal of the Acoustical Society of America* 118: 410–427.

Ladefoged, Peter, and Ian Maddieson
1996 *The Sounds of the World's Languages*. Oxford: Blackwell.

Ohala, John J.
1983 The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, Peter F. MacNeilage (ed.), 189–216. New York: Springer.

Padgett, Jaye
2002 Russian voicing assimilation, final devoicing, and the problem of [v]. Ms. University of California, Santa Cruz. http://people.ucsc.edu/~padgett/locker/newvoice.pdf

Pauliny, Eugén
1979 *Slovenská Fonológia*. [Slovak phonology] Bratislava: Slovenské Pedagogické Nakladateľstvo.

Rubach, Jerzy
1993 *The Lexical Phonology of Slovak*. Oxford: Clarendon Press.

Shadle, Christine H.
1985 *The Acoustics of Fricative Consonants (RLE Technical Report 506)*. Cambridge, MA: MIT Press.

Siptár, Péter
1996 A Janus-faced Hungarian consonant. *The Even Yearbook* 2: 83–96.

Siptár, Péter, and Miklós Törkenczy
2000 *The Phonology of Hungarian*. Oxford: Oxford University Press.

Steriade, Donca
1997 Phonetics in phonology: The case of laryngeal neutralization. Ms. University of California, Los Angeles.

Steriade, Donca
1999 Alternatives to syllable-based accounts of consonantal phonotactics. In *Proceedings of the 1998 Linguistics and Phonetics Conference*, Osamu Fujimura, Brian Joseph, and Bohumil Palek (eds.), 205–242. Prague: The Karolinum Press.

Stevens, Kenneth N.
1998 *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Stevens, Kenneth N., Sheila Blumstein, Laura Glicksman, Martha Burton, and Kathleen Kurowski
1992 Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America* 91: 2979–3000.

Vago, Robert M.
1980 *The Sound Pattern of Hungarian*. Washington DC: Georgetown University Press.

The laryngeal characterization of Korean fricatives: Acoustic and aerodynamic data

Hyunsoon Kim, Shinji Maeda, Kiyoshi Honda and Stephane Hans

1. Introduction

The present study aims to provide acoustic and aerodynamic evidence for the laryngeal characterization of the Korean lenis /s/ and fortis /s'/ presented in a recent stroboscopic cine-MRI study of these fricatives (Kim, Maeda, and Honda submitted).

It has often been assumed based on Kagaya's (1974) fiberscopic and acoustic data on Korean obstruents that the non-fortis fricative is not lenis (/s/) but aspirated (/s^h/). According to Kagaya (1974), the maximum glottal opening of the fricative is word-initially as wide as that of the aspirated stop consonants /p^h, t^h, ts^h, k^h/, and aspiration occurs after the fricative when it is followed by the vowel /e/ or /a/ in word-initial position. In the same contexts, however, Kagaya (1974) noted in his acoustic data that such aspiration was not observed in the fortis fricative /s'/. Based on the phonetic study of Kagaya (1974), Iverson (1983) proposed that the non-fortis fricative is specified for the features [+spread glottis, -constricted glottis] like the aspirated stops, and the fortis /s'/ is specified for the features [-spread glottis, +constricted glottis] in terms of glottal opening, following Halle and Stevens' (1971) laryngeal features, as shown in (1a) (henceforth, [s.g.] and [c.g.], respectively). On the other hand, the sound patterning of the non-fortis fricative with lenis stops led Iverson to group this fricative and the lenis stops in terms of the feature specifications [-stiff vocal folds, -slack vocal folds] for glottal tension (henceforth, [stiff] and [slack], respectively). Thus, in (1), the fricative shares laryngeal features with both the aspirated and the lenis stops. The fortis fricative /s'/ is specified as [+stiff, -slack] like the fortis and aspirated stops.

However, in a recent stroboscopic cine-MRI study of the fricatives, Kim, Maeda, and Honda (submitted) have suggested that the non-fortis fricative is lenis (/s/). The articulatory study showed that at release onset position as well as during frication, the fricative is intermediate between /s'/ and /t^h, ts^h/ in glottal opening in the contexts /ma_a/ and /_a_a/ both word-initially and -medially.

(1) The laryngeal feature specification of Korean obstruents (Iverson 1983)

	a. fricatives		b. stops		
	lenis /s/	fortis /s'/	lenis	fortis	aspirated
[s.g.]	+	-	-	-	+
[c.g.]	-	+	-	+	-
[stiff]	-	+	-	+	+
[slack]	-	-	-	-	-

For example, the glottal width of the fricative is quite similar to that of the lenis stop consonants /t, ts/ in word-initial position. And when frame-to-frame variations of the glottal width and of the tongue-apex position were combined in the context /_a_a/, it was found that aspiration noise occurs during transitions between the frication of the two fricatives and the vowel following them. In addition, the MRI study showed that the duration of the narrowest oral constriction is longer with the apex being closer to the mouth roof in /s'/ than in /s/; the pharyngeal width is longer in /s'/ than in /s/; and the highest tongue blade and glottal height is sustained longer in /s'/ than in /s/.

The MRI data led Kim, Maeda, and Honda (submitted) to propose that the laryngeal characteristics of the fricatives can be captured in terms of glottal opening and concomitant tongue/larynx movements, as is done for the Korean coronal stops /t, t^h, t', ts, ts^h, ts'/ in Kim, Honda, and Maeda (2005).[1] According to Kim, Maeda, and Honda (submitted) as well as Kim (to appear), the laryngeal characteristics of the fricatives –a) glottal opening and b) concomitant tongue/larynx movements– are incorporated into the features [±s.g.] and [±tense], respectively, as with the Korean coronal stops (Kim 2003, 2005). That is, in terms of glottal opening, the lenis and fortis fricatives are specified for the feature [-s.g.] like their stop counterparts, and they are different in [tense]: the fortis /s'/ is specified as [+tense], like fortis and aspirated stops, and the lenis /s/ as [-tense], like lenis stops, as shown in (2) (see Kim, Maeda, and Honda 2009 for discussion in favor of the two features over [c.g.], [stiff] and [slack] in Korean obstruents; see also Kim (2005, to appear) for phonological arguments for [±s.g.] and [±tense] in Korean stops and fricatives and for the singleton analysis of Korean fortis consonants instead of the length-based proposal of Avery and Idsardi (2001), among others).

(2) The laryngeal feature specification of Korean obstruents (Kim, Maeda, and Honda submitted; Kim to appear)

	a. fricatives		b. stops		
	lenis /s/	fortis /s'/	lenis	fortis	aspirated
[s.g.]	-	-	-	-	+
[tense]	-	+	-	+	+

In order to verify the proposed laryngeal characterization of the Korean fricatives in terms of their acoustic and aerodynamic aspects, we conducted experiments. In an acoustic experiment we investigated whether aspiration is a phonetic property of the non-fortis fricative, as suggested in Kagaya (1974) among others, or whether it occurs during the transition, regardless of the type of fricative, as shown in Kim, Maeda, and Honda (submitted). In addition, we examined whether aspiration is affected by the quality of the vowels adjacent to the fricatives and by speakers. Aerodynamic data were also obtained in order to better understand how airflow and intraoral pressure are manifested during the production of the two types of fricatives. Considering that airflow and intraoral pressure data can provide substantive information to better understand the production of a speech sound, we believe that an aerodynamic experiment in conjunction with an acoustic experiment will validate the MRI study of Kim, Maeda, and Honda (submitted).

This paper is structured as follows. In the next two sections, the methods and results of our acoustic and aerodynamic experiments on the fricatives are presented, and in section 4 we discuss the results of the phonetic experiments in regard to how they support the laryngeal feature specification of the fricatives in (2a). In section 5 we briefly conclude the paper.

2. Acoustic data

2.1. Method

The two types of fricatives /s, s’/ were put into /_V_V/ contexts, where V is one of the eight Korean monophthongs /a, i, u, o, ɛ, æ, ʌ, ɨ/, as shown in (3).

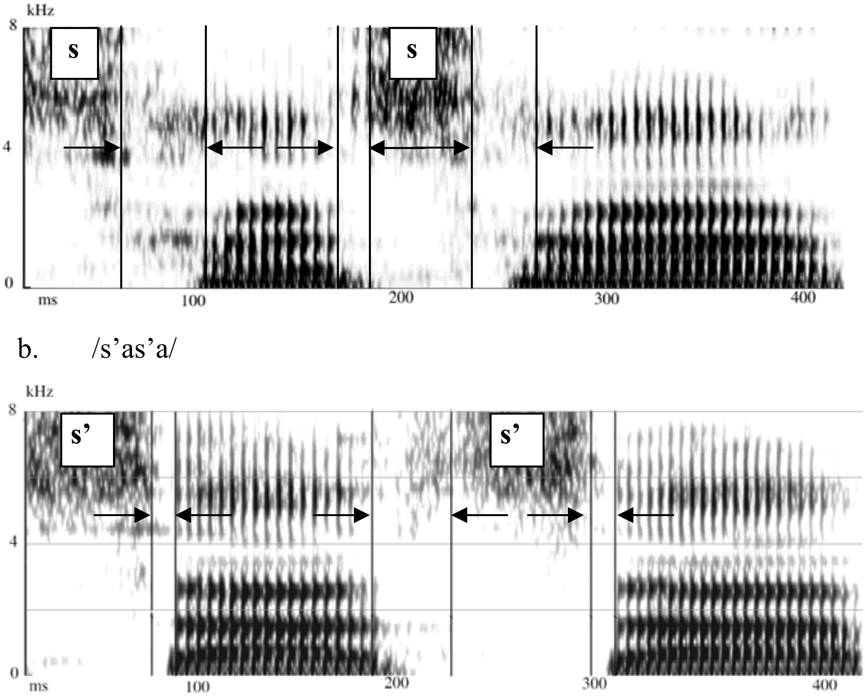
(3) /sasa/ /s’as’a/
/sisi/ /s’is’i/
/susu/ /s’us’u/
/soso/ /s’os’o/
/sɛsɛ/ /s’ɛs’ɛ/
/sæsæ/ /s’æs’æ/
/sʌsʌ/ /s’ʌs’ʌ/
/sɨsɨ/ /s’ɨs’ɨ/

The test words, which are all nonsense words, were embedded in the frame sentence /næka __ palɨmhapnita/ ‘I pronounce __’. On a single page, sentences with the test words written in Korean orthography were randomized with two filler sentences at the top and the bottom. The sentences were read six times at a normal speech rate by the same two (one male, one female) subjects who participated in the MRI study of Kim, Maeda, and Honda (submitted). Each subject familiarized him/herself with the test words by reading them a few times before recording and read them as naturally as possible during recording. A Shure SM57-LC microphone and a Sony TDC-D8 digital audio tape recorder were used in recording the subjects. The total of 192 tokens obtained in this way (16 test words x 2 subjects x 6 repetitions) were then analyzed.

2.2. Results

Figure 1 presents representative wide-band spectrograms of /sasa/ and /s’as’a/ as produced by the male subject. The frication phase of the fricatives, as the noise generated at oral constriction, is marked by an arrow with a solid line at the bottom of the spectrogram, and is identified by the major region of noise energy above 4 kHz as an alveolar fricative (see, e.g., Kent and Read 2002). Following the frication noise there is another type of noise that corresponds to aspiration; this is marked with a dashed line. The aspiration phase is identified by noise covering a broad range of frequencies with relatively weak energy. It is noticeable in Figure 1 that aspiration occurs during the transition from a fricative to a vowel and from a vowel to a fricative, regardless of the phonation types of the fricative.

a. /sasa/



b. /s’as’a/

Figure 1. Wide-band spectrograms of /sasa/ (a) and /s’as’a/ (b) as produced by the male subject, aspiration noise = interval between arrows.

Table 1 presents the average aspiration duration at the offset of each fricative in word-initial and word-medial positions for the female and male subjects.

Figure 2 shows the average aspiration duration at the offset of each fricative in word-medial position for the female subject. A comparison of the aspiration duration after the offset of the two fricatives reveals that the fricatives /s, s’/ differ in how long aspiration occurs.

Table 1. The average aspiration duration (ms) after the offset of the fricatives /s, s’/ word-initially (a) and word-medially (b) in the context /_V_V/, where V is one of the eight Korean monophthongs /a, i, u, o, ɛ, æ, ʌ, ɨ/, for the female and male subjects.

	a. initial				**b. medial**			
	female		**male**		**female**		**male**	
_V; V_V	**/s/**	**/s’/**	**/s/**	**/s’/**	**/s/**	**/s’/**	**/s/**	**/s’/**
/a/	47	9	43	13	15	7	21	13
/i/	19	7	18	10	13	12	14	10
/u/	18	9	31	17	14	8	19	13
/ɛ/	31	7	26	13	19	11	24	15
/æ/	33	9	26	13	22	11	22	13
/o/	21	8	18	12	11	7	14	9
/ʌ/	37	8	41	10	16	8	22	12
/ɨ/	14	8	20	16	12	8	17	12

A paired samples two-tailed t-test showed that the average aspiration duration after the offset of /s/ in word-initial position is significantly longer than in word-medial position (female subject: t(7) = 3.5, p < .01; male subject: t(7) = 3.1, p < .05). In contrast, the aspiration duration after the offset of /s’/ in word-initial position is not significantly different from that in word-medial position (female subject: t(7) < 1; male subject: t(7) < 1).

In addition, another paired samples two-tailed t-test showed that aspiration duration tends to be significantly longer after the offset of the lenis fricative /s/ than after the offset of the fortis /s’/ both word-initially and -medially for the two subjects. For example, when compared with that after the offset of /s’/ in the same vowel contexts, the aspiration duration after the offset of /s/ is significantly longer in the word-initial contexts /_a, _i, _ɛ, _æ, _o, _ʌ/ for the female subject (t(5) = 11.9, p < .001 for /sa/ vs. /s’a/ ; t(5) = 5, p < .005 for /si/ vs. /s’i/; t(5) = 13.2, p < .0001 for /sɛ/ vs. /s’ɛ/; t(5) = 14, p < .0001 for /sæ/ vs. /s’æ/; t(5) = 4.2, p < .01 for /so/ vs. /s’o/; t(5) = 10.1, p < .0005 for /sʌ/ vs. /s’ʌ/). Yet, in the two word-initial contexts /_u, _ɨ/, aspiration duration is not significant after the offset of the two fricatives in the subject (t(5) = 1.3, p = .244 for /su/ vs. /s’u/; t(5) = 2.2, p = .0785 for /sɨ/ vs. /s’ɨ/). Similarly, aspiration duration after the offset of the lenis /s/ is significantly longer in most of the word-medial contexts /_a, _u, _ɛ, _æ, _ʌ/ in the female subject (t(5) = 4.5, p < .01 for /sa/ vs. /s’a/ ; t(5) = 3, p < .05 for /su/ vs. /s’u/; t(5) = 4.5, p < .01 for /sɛ/ vs. /s’ɛ/; t(5) = 8, p < .0005 for /sæ/ vs. /s’æ/; t(5) = 7.5, p < .001 for /sʌ/ vs. /s’ʌ/). But in the word-medial contexts /_i, _o, _ɨ/, aspiration duration is not significant (t(5) = .5, p = .638 for /si/ vs. /s'i/ ; t(5) = 2.4, p = .0582 for /so/ vs. /s'o/; t(5) = 1.4, p = .2082 for /sɨ/ vs. /s'ɨ/).[4]

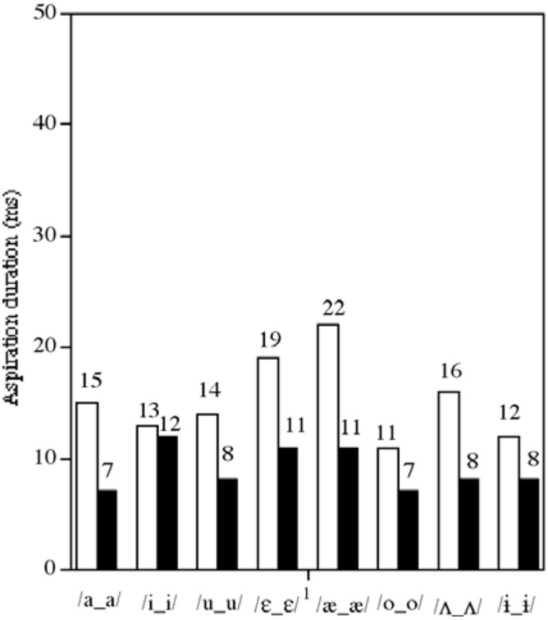


Figure 2. The average aspiration duration (ms) after the offset of the fricatives /s, s'/ word-medially in /_V_V/, where V is one of the eight Korean monophthongs /a, i, u, o, ɛ, æ, ʌ, ɨ/, for the female subject. Black bars: /s'/ and white bar: /s/.

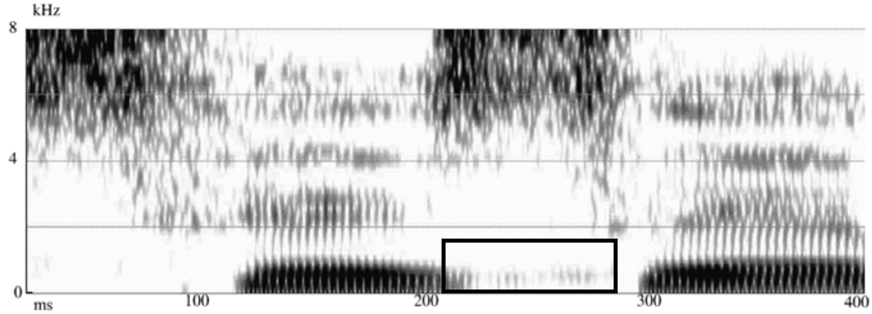
The male subject also showed that aspiration duration is likely to be significantly longer after the offset of /s/ than after the offset of /s'/ word-initially (p < .05 in the contexts /_a, _i, _ɛ, _æ, _ʌ/) and word-medially (p < .02 in the contexts /_a, _i, _u, _ɛ, _æ, _ʌ/). However, it is not significant after the offset of the two fricatives in the word-initial contexts /_u, _o, _ɨ/ (p > .05) and in the word-medial contexts /_o, _ɨ/ (p > .05).

Moreover, aspiration duration is affected by vowel contexts after the offset of /s/ but not of /s'/.[2] We ran multiple repeated measures ANOVAs with vowel context in (3) as the main factor and aspiration duration as the dependent variable. Vowel contexts in relation to aspiration duration were highly significant after the offset of /s/ (female subject: F(7, 47) = 12.5, p < .0001 for the word-initial /s/; F(7, 47) = 4.4, p < .005 for the word-medial /s/; male subject: F(7, 47) = 8.3, p < .0001 for the word-initial /s/; F(7, 47) = 4.1, p < .005 for the word-medial /s/), but not after the offset of /s’/ (female subject: F(7, 47) < 1 for the word-initial /s’/; F(7, 47) = 1.8, p < .05 for the word-medial /s’/; male subject: F(7, 47) = 2.2, p > .05 for the word-initial /s’/; F(7, 47) = 2.5, p > .1 for the word-medial /s’/).

In a Scheffe’s post hoc comparison, we also found that aspiration duration after the offset of the word-initial /s/ is significantly longer in /sasa/ than in /sisi, susu, soso, sɨsɨ/ (female subject: F(1, 47) = 27.3, p < .001 for /sasa/ vs. /sisi/, F(1, 47) = 28.3, p < .0001 for /sasa/ vs. /susu/, F(1, 47) = 26.3, p < .001 for /sasa/ vs. /soso/, F(1, 47) = 33.3, p < .0001 for /sasa/ vs. /sɨsɨ/ ; male subject, F(1, 47) = 25.5, p < .005 for /sasa/ vs. /sisi/, F(1, 47) = 25.5, p < .005 for /sasa/ vs. /soso/, F(1, 47) = 23.3, p < .05 for /sasa/ vs. /sɨsɨ/). In contrast, in word-medial position, no pair was found to be significant regarding aspiration duration in the two subjects, no matter whether the fricative is /s/ or /s’/.[3] This indicates that there is no vowel effect in aspiration duration in any pair of vowels after the offset of the two types of fricatives in word-medial position.

Furthermore, it is worth considering the occurrence of voicing in the fricatives. Cho, Jun, and Ladefoged (2002) suggest that the fricative /s/ may be lenis because about 47% of their tokens of the fricative were fully voiced in intervocalic position, like lenis stops. However, our acoustic data show that voicing can occur in both of the fricatives /s, s’/ although probably not as systematically as in the case of lenis stop consonants in intervocalic position. Of the total of 192 tokens obtained in our acoustic data (16 test words x 2 subjects x 6 repetitions), 38 tokens (19.8%) of /s/ and /s’/ were observed to be voiced fully (8 tokens) or partially (30 tokens) in intervocalic word-medial or word-initial position. For example, as shown in Figure 3, the word-medial lenis fricative in /sæsæ/ (a) has voice bars throughout oral constriction, and the word-medial fortis fricative in /s’is’i/ has voice bars at the beginning of oral constriction (b). The frequency of voicing of /s/ is 8.3% word-medially (16 tokens) and 3% word-initially (6 tokens), while that of /s’/ is 4.2% both word-medially and word-initially (8 tokens each). In contrast, the frequency of voicing of a lenis stop consonant in Jun’s (1994) acoustic study is much higher. For example, the frequency of voicing of a word-initial lenis stop consonant was 70% (196 tokens out of 281) when preceded by a vowel in a frame sentence, and of a word-medial lenis stop consonant 76% (90 tokens out of 120) with the word-initial consonant being a lenis stop.

a. word-medial /s/ in /sæsæ/



b. word-medial /s'/ in /s'is'i/

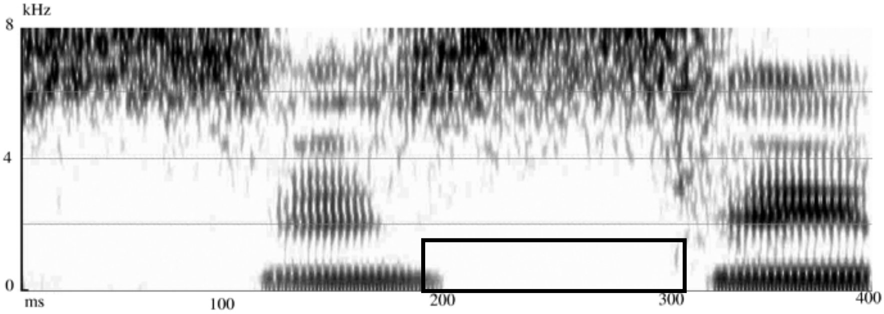


Figure 3. Wide-band spectrograms of /sæsæ/ (a) and /s'is'i/ (b) as produced by the female subject show full and partial voicing, respectively. Voicing during the fricative interval is marked with a black box

Thus, the much lower frequency of voicing of the fricatives makes it hard to equate voicing of the fricatives with that of a lenis stop consonant.

In short, the acoustic data presented here on the two types of fricatives have shown that aspiration occurs during transitions, regardless of the phonation type of the fricative, and that its duration tends to be longer after the offset of the lenis /s/ than after the offset of the fortis /s'/ both word-initially and word-medially. Moreover, aspiration duration during transitions in the word-initial lenis fricative /s/ can be affected by which vowels follow and also by speakers. However, there is no vowel effect in aspiration duration after the offset of the two types of fricatives in word-medial position or after the offset of the fortis /s'/ in word-initial position. This suggests that the absence or presence of aspiration is not relevant for the distinction of the fricatives.

3. Aerodynamic data

In conjunction with the acoustic experiment presented above, we conducted an aerodynamic experiment in order to better understand the production of the two Korean fricatives.

3.1. Method

Intraoral pressure and airflow were recorded with the Aerophone II at the European Georges Pompidou Hospital in Paris. Two speakers of Seoul Korean participated: one female, who participated in the acoustic experiment discussed above, and one male, who was living in Paris as a student. The two subjects held a face mask which covered the mouth and the nose to record airflow as well as a small tube with the open end inside their mouth between the hard palate and the tongue so as to record air pressure in the mouth. The test words we used for our aerodynamic data are listed in (4) for the fricatives and in (5) for the coronal stops.

(4) a. /masa/ /mas’a/
/sasa/ /s’as’a/

b. /sisi/ /s’is’i/
/susu/ /s’us’u/
/isi/ /is’i/
/usu/ /us’u/
/si/ /s’i/
/su/ /s’u/

(5) /mata/
/mat^{h}a/
/mat’a/
/matsa/
/mats^{h}a/
/mats’a/

The two types of fricatives were presented in the contexts /ma_a/ and /_a_a/ (4a), as in the stroboscopic cine-MRI study of Kim, Maeda and Honda (submitted), and also in /_V_V/, /V_V/ and /_V/ (V = i or u) (4b) in order to examine whether the quality of adjacent vowels plays a role in the intraoral pressure and airflow of the fricatives. The coronal stops in (5) were put in the context /ma_a/, as in the MRI study of Kim, Honda and Maeda (2005).

During the experiment, the two subjects were asked to repeat the test words ten times at a natural speech rate. Recorded airflow and intraoral pressure were digitized at a sampling rate of 1 kHz and analyzed. After the segmentation of the speech signal envelope, intraoral air pressure and airflow, the three curves were plotted for each test word on a single graph, as shown in Figure 4 below. From this plotted graph together with a relevant waveform, we set a cursor at the release of oral constriction of the fricatives, the point where airflow is the highest. Then the highest airflow and its corresponding intraoral air pressure at the same point were measured to obtain the values of airflow and intraoral pressure of the fricatives. As for the stops, the highest intraoral pressure during oral closure and the highest airflow during release were measured (see also Dart 1987 and Cho, Jun, and Ladefoged 2002 for aerodynamic data on Korean stops).

3.2. Results

Figure 4 illustrates representative aerodynamic records of the fricatives /s/ (a) and /s'/ (b) in /_a_a/ as spoken by the female subject. The two solid lines are aligned with the offset of the word-initial fricatives (i) and with that of the word-medial ones (ii). During the oral constriction of the fricatives /s/ and /s'/, regardless of their position, the intraoral pressure is high and the airflow is low, while airflow reaches its peak and intraoral pressure goes down at the offset of the fricatives. The peak value of the intraoral pressure (Pio in cmH_2O) and the peak value of the airflow (U in liters/s) were thus measured at two different time points.

As can be seen in Figure 4, the difference between the two fricatives lies in the magnitude of the airflow or intraoral pressure peak: the fortis fricative /s'/ is higher in intraoral pressure and much lower in airflow than the fricative /s/, not only in word-initial (i) but also in word-medial (ii) position. From the intraoral pressure and airflow of the two fricatives, we can calculate the airflow resistance, which is expected to be higher in the fortis /s'/ than in the lenis /s/, given the MRI data in Kim, Maeda, and Honda (submitted) that shows that the oral constriction is narrower and longer in the former than in the latter.

Table 2 shows the average values of the intraoral pressure peak and airflow peak of the two fricatives word-initially (a) and word-medially (b) in the test words in (4a) (repeated ten times each). From the average values of intraoral pressure (Pio in cmH_2O) and airflow (U in cm^3/s), we get the Pio/U ratio, that is, the airflow resistance (R in Ohm) in the table.

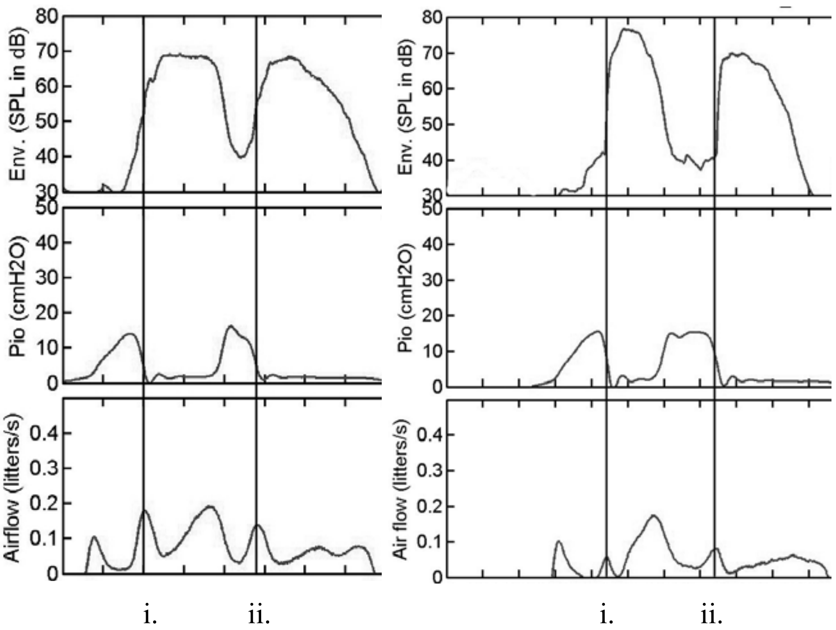


Figure 4. Aerodynamic data on /sasa/ (left) and /s’as’a/ (right) from the female subject accompanied by the signal envelope (upper panels); the solid lines are aligned with the offset of the word-initial fricatives (i) and with that of the word-medial ones (ii).

Intraoral pressure is higher in /s’/ than in /s/ for the two subjects, while airflow tends to be lower in /s’/ than in /s/ except for the word-initial fricative in /s’as’a/ for the male subject. Airflow resistance is greater in /s’/ than in /s/ in all the examined contexts in both subjects, providing a more solid criterion than intraoral pressure and airflow.

For example, a paired samples two-tailed t-test showed that the airflow resistance of /s’/ is significantly greater than that of /s/ in the word-medial position in the contexts /ma_a, _a_a, _i_i, _u_u/ (female subject: $p < .005$; male subject: $p < .0005$). This is also true of the fricatives in word-initial position. A paired samples two-tailed t-test showed that the airflow resistance of /s’/ is significantly greater than that of /s/ word-initially in the contexts /_a_a, _i_i, _u_u/ (female subject: $p < .001$; male subject: $p < .0001$).

Table 2. Airflow resistance as well as intraoral pressure peak and airflow peak values of initial fricatives (a) and medial fricatives (b) averaged over ten repetitions (N = 10) for the two subjects.

a. aerodynamic data on initial fricatives			
	air pressure (Pio = cmH_2O)	airflow (U = cm^3/s)	airflow resistance (R = Pio/U)
female subject			
/sasa/	11.95	0.031	385.52
/s’as’a/	11.05	0.026	431.47
male subject			
/sasa/	13.25	0.116	114.21
/s’as’a/	14.21	0.119	119.43
b. aerodynamic data on medial fricatives			
female subject			
/masa/	12.24	0.038	322.2
/mas’a/	12.99	0.036	363.98
/sasa/	14.08	0.041	343.29
/s’as’a/	14.96	0.016	935
male subject			
/masa/	14.99	0.144	104.13
/mas’a/	15.98	0.141	113.36
/sasa/	15.48	0.199	77.77
/s’as’a/	15.78	0.117	134.87

It is noteworthy that the fricatives are more similar to the lenis and fortis coronal stops than to the aspirated stops /t^h, ts^h/, in terms of airflow. As can be seen in Figure 5, the airflow of the aspirated stops is higher than that of not only the lenis and fortis coronal stops but also than that of the fricatives in the context /ma_a/. We ran multiple repeated measures ANOVAs with laryngeal setting (lenis, aspirated, fortis) as the main factor, and airflow as the dependent variable. Laryngeal setting was highly significant for the two subjects (/t, t^h, t’/: F(2, 29) = 334.5, $p < .0001$ (female); F(2, 29) = 299.3, $p < .0001$ (male); /ts, ts^h, ts’/: F(2, 29) = 269.1 $p < .0001$ (female); F(2, 29) = 27.4, $p < .0001$ (male)). We also compared the fricatives to the aspirated stops (/t^h/ and /ts^h/): airflow was found to be significantly higher for the stops (/s, s’, t^h/: F(2, 29) = 269.1, $p < .0001$ (female); F(2, 29) = 27.4, $p < .0001$ (male); /s, s’, ts^h/: F(2, 29) = 421.7, $p < .0001$ (female); F(2, 29) = 240.4, $p < .0001$ (male)).

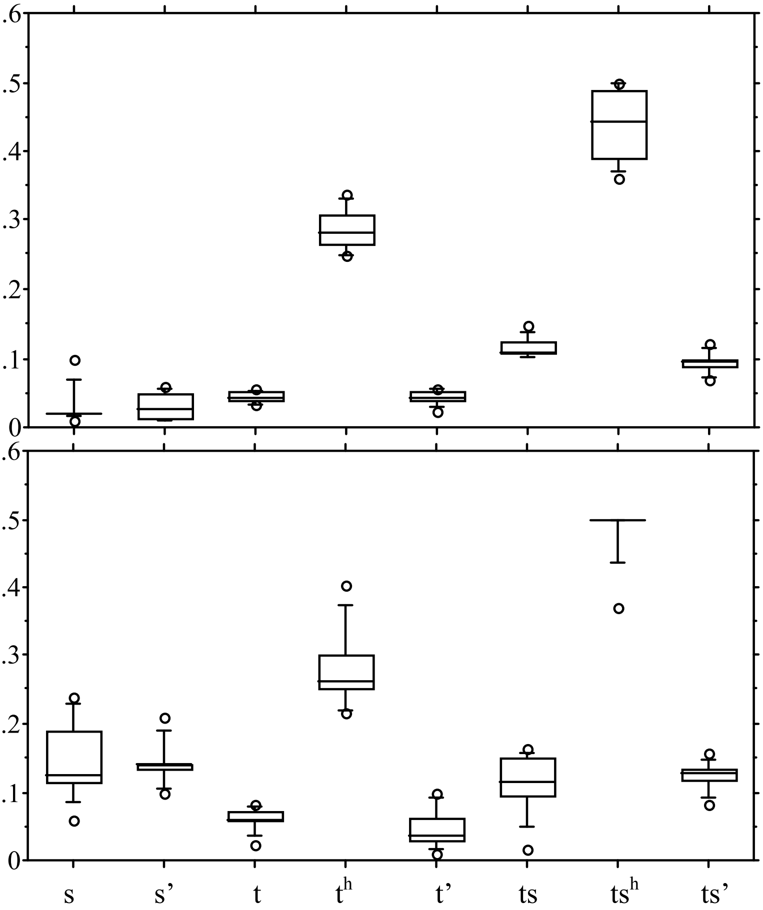


Figure 5. Comparison of the airflow (U in cm^3/s) of the fricatives /s, s’/ with that of the coronal stops /t, t^h, t’, ts, ts^h, ts’/ in /ma_a/ repeated ten times by the female (a) and male (b) subjects.

The average airflow resistance values of the fricatives in (4a) and the stops in (5) in the context /ma_a/ are given in Figure 6. They tend to pattern from high to low in the order fortis < lenis < aspirated stop consonants, with the two aspirated consonants /t^h, ts^h/ being less than 50 Ohms in the two subjects, as shown in Figure 6. We ran the same ANOVAs as above with airflow resistance as the dependent variable. Airflow resistance was significantly higher for the fortis consonants (/t, t^h, t’/: F(2, 29) = 39.4, $p < .0001$ (female); F(2, 29) = 7.7, $p < .005$ (male); /ts, ts^h, ts’/: F(2, 29) = 86.4, $p < .0001$ (female); F(2, 29) = 3.5, $p < .05$ (male)).

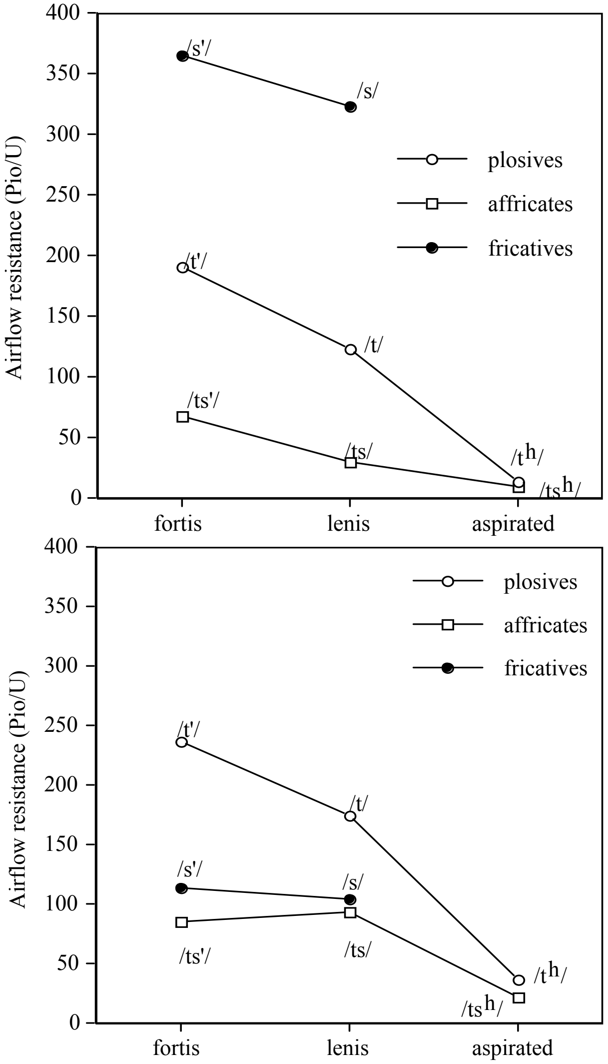


Figure 6. Airflow resistance of the fricatives /s, s'/ in comparison with that of the coronal stop consonants in /ma_a/ repeated ten times by the female (upper plot) and male (lower plot) subjects.

The airflow resistance of /s/ vs. /s'/ shows the same tendency as found with the lenis vs. fortis stops: it was higher for /s'/ than /s/, with airflow resistance values well above 50 Ohms for the two subjects. A paired samples two-tailed t-test showed that the airflow resistance of /s'/ is significantly greater than that of /s/ (female subject: $p < .0001$; male subject: $p < .0001$).

Given that the airflow resistance calculated from the intraoral air pressure and airflow is the oral-constriction resistance, we can say that the fricative /s/ is very similar to the lenis stops rather than to the aspirated ones in terms of oral-constriction resistance. Note that the aerodynamic variables U, Pio, and R in Table 2 are related to the frication and not to the aspiration part of the fricatives. Considering that airflow resistance is directly related to the oral constriction shape and that it is consistently higher in /s'/ than in /s/, we can suggest that the constriction during the frication of /s'/ is stronger than during that of /s/ in that the stronger the constriction is, the higher the resistance (see, e.g., Stevens 1998).

To sum up, we have noted that the airflow of the two fricatives /s/ and /s'/ is more similar to that of the lenis and fortis coronal stops than to that of the aspirated ones (Figure 5). We have also observed that the airflow resistance of the fricatives is much higher than that of the aspirated stops, with /s'/ being higher in airflow resistance than /s/ (Figure 6). The aerodynamic data suggests that the fricative /s/ is lenis, not aspirated, and that the constriction during the frication of /s'/ is stronger than during that of /s/.

4. Discussion

Given the results of the experiments presented above, we suggest that the acoustic and aerodynamic data support the laryngeal characterization of the fricatives in terms of the two binary features [±s.g.] and [±tense] in (2a).

First, the acoustic and aerodynamic data presented here provide evidence that /s/ and /s'/ are specified as [-s.g.], in contrast to the aspirated stops, which are specified as [+s.g.]. As shown in the acoustic data, aspiration is not solely a phonetic characteristic of the lenis fricative, although it does tend to be statistically longer than in the fortis fricative. Longer aspiration in /s/ than in /s'/ for the female subject (Figure 2) and the male subject can be attributed to a slower speed of transition due to the wider glottal opening in the lenis fricative than in the fortis fricative (Kim, Maeda, and Honda submitted). Thus, the adduction necessary for a following vowel takes a little longer with the wider glottal opening in the fricative /s/, resulting in longer aspiration than with the fortis fricative which has a narrower glottal opening (Kim, Maeda, and Honda submitted).

Even after the offset of the fricative, it is noticeable that the aspiration duration may depend on which vowel follows the fricative in word-initial position. As in the above Scheffe's post hoc comparisons, the significantly longer aspiration duration before the vowel /a/ in the two subjects can also be attributed to a longer transition or slower speed of transition from the word-initial fricative /s/ to the low vowel /a/ than to any other vowel.

In addition, the acoustic data showed that, regardless of the types of fricative, aspiration noise could occur during transitions from a fricative to a vowel and/or from a vowel to a fricative (Figure 1). This is in accordance with the MRI data of Kim, Maeda, and Honda (submitted) in that the presence of aspiration noise in the two fricatives can result from the relation of oral constriction size to glottal width at the transition from a fricative to a vowel and from a vowel to a fricative. If the oral constriction and glottal area are small, then airflow becomes so low that noticeable noise is not generated at the glottis and at the oral constriction. In this case, aspiration noise is hardly expected during transitions from a fricative to a vowel and/or from a vowel to a fricative. For example, in the tongue apex-glottis phasing in /sasa/ and /s'as'a/, Kim, Maeda, and Honda (submitted) have noted that aspiration can arise when the glottal width is less than the distance of the tongue apex from the mouth roof, regardless of the phonation types of the fricative. Thus, aspiration can arise not only word-medially but also word-initially during transitions between the frication of the two fricatives and a vowel following them, but its duration is likely to be shorter after /s'/, which has been confirmed in the present study.

Moreover, the aerodynamic experiment discussed above showed that the two fricatives are more similar to the lenis and fortis coronal stops than to the aspirated /t^h, ts^h/ in terms of airflow, as shown in Figure 5. When the fricatives and the aspirated stops were compared in terms of airflow, the aspirated stops were found to have significantly greater airflow than the lenis and fortis fricatives. This indicates that the glottis does not open as much during the fricative /s/ as it does during the aspirated stops. The airflow data is in accordance with the coronal data on the fricatives and the coronal stops in Kim, Maeda, and Honda (submitted), according to which the glottis opens less in the fricatives than in the aspirated stops. Consequently, the acoustic and aerodynamic data as well as the MRI study of the fricatives support the specification of the lenis and fortis fricatives for the feature value [-s.g.] in (2a).

Second, the aerodynamic data in this study agree with the MRI study of Kim, Maeda, and Honda (submitted), according to which the two fricatives are different in terms of the feature [±tense]: /s/ is specified as [-tense] like the lenis stops and /s’/ as [+tense] like the fortis and aspirated stops, as in (2a). The higher airflow resistance in /s’/ than in /s/ in the aerodynamic experiment (Table 2) corresponds to the higher apex position with a narrower and longer oral constriction in the /s’/ as discussed in Kim, Maeda, and Honda (submitted). That is, airflow resistance becomes higher as the cross-sectional area of the oral constriction becomes narrower and the length of the constriction becomes longer. Thus, the stronger constriction during /s’/ is correlated with a narrower and longer oral constriction. It is also noteworthy that the narrower and longer the oral constriction, the higher or longer the glottal raising in /s’/ (Kim, Maeda, and Honda submitted). In this respect, the aerodynamic data discussed above and the MRI data corroborate each other, though airflow resistance is a function of the three-dimensional oral constriction shape while the apex position measured by MRI indicates only the height dimension.

Given that stronger constriction together with higher or longer glottal raising and narrower and longer oral constriction in /s’/ than in /s/ is involves the tensing of both the primary articulator of the tongue blade and the vocal folds in the sense of Kim, Maeda, and Honda (submitted, 2009), we can say that the aerodynamic data provide evidence for the feature [tense] in the feature specification of the fricatives in (2a) (see C.-W. Kim (1965) for the traditional feature [tense] in Korean stop consonants; this feature is further elaborated by Kim, Maeda, and Honda (2009)).

As a result, the acoustic and aerodynamic data in the present study support the laryngeal feature specification of the fricatives in (2a). Note that Halle and Stevens’ (1971) laryngeal features [c.g.], [stiff] and [slack] in (1) are not used in the feature specification of the fricatives or of the stops in (2). For example, according to Kim, Maeda, and Honda (submitted), the fortis fricative /s’/ has a slightly wider glottal opening than the fortis stops /t’, ts’/ word-medially in /_a_a/ due to a continuous airflow during its oral constriction in the female and male subjects. Both of the fortis stop consonants have a slightly greater glottal opening than their lenis counterparts for the female subject in the same context (Kim, Honda, and Maeda 2005). Furthermore, the same is true of the fortis plosives /t’, p’/ word-medially in /_a_a/ (Kim, Maeda, and Honda 2009). What is invariant in glottal opening is that the glottis opens more in the aspirated stops /p^h, t^h, ts^h, k^h/ than in the other lenis and fortis consonants, including the fricatives /s, s’/. The systematic variation in glottal opening indicates that the feature [c.g.] for the presence/absence of the complete adduction of the vocal folds is not empirically supported for the fricatives or the stops in Korean.

In addition, the stiffening and slackening of the vocal folds for the features [stiff] and [slack], respectively, are considered as part of the articulatory bases of the proposed feature [tense] (Kim 2003, 2005, to appear; Kim, Maeda, and Honda submitted, 2009). According to Kim, Maeda, and Honda (submitted), the higher or longer glottal raising in the fortis fricative /s'/ than in the lenis /s/ correlates with higher glottal tension in the former, in that larynx raising is associated with glottal tension, which is in turn acoustically reflected by F0 rises (e.g., Stevens 1977, 1998; Honda 1995, 1999). Given that both vertical larynx movements associated with glottal tension and the supralaryngeal articulator (tongue blade) are proposed to be invariant articulatory bases of the /s/-/s'/ contrast (Kim, Maeda, and Honda submitted, 2009), the feature [tense] incorporates these characteristics, but not the separate features [stiff] and [slack]. See also Kim (to appear) for phonological arguments for the use of the features [tense] and [s.g.] in the fricatives as well as the stops in Korean.

One might raise the question of why there is a gap with the aspirated fricative (/s^h/), though stop consonants show a three-way laryngeal contrast. Recall that aspiration occurs during transitions between a fricative and a vowel and/ or a vowel and a fricative, regardless of the phonation types. Aspiration alone is not decisive in the perception of the fricatives either, when we consider Moon's (1997) and Park's (1999) perception experiments of the two fricatives. In the wide-band spectrograms of /sal/ 'flesh' and /s'al/ 'rice', Moon (1997) cut the aspiration part after the frication of /s/ and put it between the frication of /s'/ and the vowel /a/ in /s'al/. His three subjects all perceived this edited sound as /s'al/, regardless of the presence of the aspiration. In addition, the sequence of the frication of /s'/ and the vowel /a/ in /sal/ was perceived either as /sal/ or as /s'al/. When the vowel /a/ in /sal/ was replaced with that in /s'al/, thus the sequence of the frication of /s/, its aspiration in /sal/, the added vowel and /l/ were heard, all three subjects perceived the edited stimulus as /s'al/, though it sounded a little unnatural. Similar results come from Park's (1999) perception experiment. From the two tokens /sata/ 'to buy' and /s'ata/ 'to be cheap', he made the following four stimuli: (a) /sata/ with the frication of /s/ and its aspiration; (b) the frication of /s'/ in /s'ata/ was replaced with that of /s/ and its aspiration in /sata/; (c) the vowel /a/ after /s'/ in /s'ata/ was replaced with that after /s/ in /sata/; and (d) the vowel /a/ after /s/ in /sata/ was replaced with that after /s'/ in /s'ata/. His twelve subjects perceived the stimuli (a) and (c) as /sata/ and (b) and (d) as /s'ata/.

Therefore, Park's experiment suggests that aspiration does not play a role in the /s/ vs. /s'/ distinction. Given the perception experiments of the two fricatives as well as the acoustic and aerodynamic data presented here, we can say that if the aspirated fricative /s^h/ were present together with the lenis and fortis /s, s'/ in Korean, it would be difficult to auditorily distinguish /s^h/ from the fricatives /s, s'/ in terms of aspiration.

5. Conclusion

In the present study, we have examined whether the articulatory bases for the laryngeal feature specification of the Korean fricatives in Kim, Maeda, and Honda (submitted) are acoustically and aerodynamically supported. The results of the acoustic experiment have shown that aspiration occurs during transitions, regardless of the phonation types of the fricative, and that aspiration duration tends to be significantly longer after the offset of /s/ than after the offset of /s'/. In addition, aspiration duration during transitions in the word-initial lenis fricative /s/ is dependent on which vowel follows the fricative and also on the speaker. In the aerodynamic data, we have noticed that a higher airflow occurs in the aspirated stops /t^h, ts^h/ than in the fricatives or in the lenis and fortis coronal stops. Furthermore, we have also observed that airflow resistance (that is, oral-constriction resistance) is consistently lower in /s/ than in /s'/, showing a tendency to be more similar to that in the lenis stops /t, ts/, than to that in the aspirated ones /t^h, ts^h/. It has been noted that the stronger constriction during /s'/ correlates with narrower and longer oral constriction in the MRI study of Kim, Maeda, and Honda (submitted). Given the present data, it is concluded that the laryngeal feature specification of the fricatives in terms of the two binary features [±s.g.] and [±tense] in (2a) is acoustically and aerodynamically supported.

Notes

1. It has been postulated in the literature that the Korean affricates are post-alveolar (see the literature review in Kim 2004). However, throughout the text, we assume that the place of articulation of the affricates is alveolar in line with Skaličková (1960) and Kim (1999, 2001, 2004). For example, based on

stroboscopic cine-MRI data of affricates, Kim (2004) has shown that affricates are alveolars like the coronal plosives /t, t^h, t'/ and that the difference between the consonants lies in tongue body position: the affricates are all made laminally and the plosives apically or apico-laminally.

2. According to the acoustic study of Kim and Park (to appear), aspiration duration is affected by vowel contexts after the offset of the two fricatives in both word-initial and -medial positions. In Kim and Park (to appear), the fricatives /s, s'/ were put in the context /_V_V/, where V is one of the eight Korean monophthongs /a, i, u, o, ɛ, æ, ʌ, ɨ/. The test words, which are all nonsense words, were embedded in the frame sentence /næka __ palɨmhapnita/ 'I pronounce __' and randomized, as in the present study. Ten (5 male and 5 female) native speakers of Seoul Korean in their early 20s participated in the acoustic experiment and read the test words five times at a normal speech rate. The total number of tokens was 800 (16 test words x 5 repetitions x 10 subjects). We ran multiple repeated measures ANOVAs with Vowel context as the main factor and aspiration duration as the dependent variable. Vowel contexts in relation to aspiration duration were highly significant after the offset of /s/ (F(7, 280) = 20.5, p < .0001 for the word-initial /s/; F(7, 280) = 7.8, p < .0001 for the word-medial /s/) and also after the offset of /s'/ (F(7, 280) = 6.3, p < .0001 for the word-initial /s'/; F(7, 280) = 6.5, p < .0001 for the word-medial /s'/). Another repeated measure ANOVA with Vowel context (/a/ vs. one of the other vowels in (3)) as the main factor and aspiration duration after the offset of the fricatives as the dependent variable also revealed that aspiration duration is dependent on vowel contexts both word-initially and word-medially, regardless of the phonation types of the fricatives. That is, we found that aspiration is significantly longer after the offset of /s'/ in /_a/ than in the other vowel contexts and at the offset of /s/ in /_a/ than in /_u, _i, _ɨ/ in word-initial position. In word-medial position, aspiration duration is significantly greater after the offset of the fricatives in /a_a/ than in /u_u, i_i, ɨ_ɨ /.
3. See Kim and Park (to appear) for different results.
4. See Kim and Park (to appear) for similar results.

Acknowledgements

We would like to express our sincere thanks to two anonymous reviewers for their comments. We are also grateful to our subjects for participating in our acoustic and aerodynamic experiments and to Pierre Hallé for his help with statistics. Earlier versions of the present study were presented at the Conference on Turbulences at ZAS, Berlin and also at the Laboratory of Phonetics and Phonology, the Sorbonne-Nouvelle (University of Paris 3) in 2005. We appreciate the feedback from the audiences at our presentations, especially from Nick Clements and Jacqueline Vaissière as well as John Ohala, Martine Toda, Marzena Zygis and Susanne Fuchs. Any errors remain our own. This work was supported by the 2009 Hongik University Research Fund given to the first author.

References

Avery, Peter, and William J. Idsardi
2001 Laryngeal dimensions, completion and enhancement. In *Phonology and Phonetics*, Tracy Hall (ed.), 41–70. Berlin/New York: Mouton de Gruyter.

Cho, Taehong, Sun-Ah Jun, and Peter Ladefoged
2002 Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30: 193–228.

Dart, Sarah N.
1987 An aerodynamic study of Korean stop consonants: Measurements and modeling. *Journal of the Acoustical Society of America* 81: 138–147.

Halle, Morris, and Kenneth N. Stevens
1971 A note on laryngeal features. *Quarterly Progress Report* 101: 198–212. Cambridge, MA: Research Laboratory of Electronics, MIT.

Honda, Kiyoshi
1995 Laryngeal and extralaryngeal mechanisms of F0 control. In *Producing Speech: Contemporary Issues*, F. Bell-Berti and J. H. Abbs (eds.), 215–232. New York: American Institute of Physics.

Honda, Kiyoshi
1999 Interactions between vowel articulation and F0 control. In *Proceedings of LP'98*, O. Fujimura, B. D. Joseph, and B. Palek (eds.), 517–527. Charles University in Prague.

Iverson, Gregory K.
1983 Korean *s*. *Journal of Phonetics* 11: 191–200.

Jun, Sun-Ah
1994 The status of the lenis stop voicing rule. In *Theoretical Issues in Korean Linguistics*, Young-Ki Kim-Renaud (ed.), 101–114. CSLI: Stanford Linguistics Society.

Kagaya, Ryohei
1974 A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics* 2: 161–180.

Kent, Raymond D., and Charles Read
2002 *Acoustic Analysis of Speech*. 2nd ed. Australia, Canada, Mexico: Singular.

Kim, Chin-Wu
1965 On the autonomy of the tensity feature in stop classification. *Word* 21: 339–359.

Kim, Hyunsoon
1999 The place of articulation of the Korean affricates revisited. *Journal of East Asian Linguistics* 8: 313–347.

Kim, Hyunsoon
2001 The place of articulation of the Korean plain affricate in intervocalic position: An articulatory and acoustic study. *Journal of the International Phonetics Association* 31: 229–257.

Kim, Hyunsoon
2003 The feature [tense] revisited: the case of Korean consonants. *NELS* 34: 319–332.

Kim, Hyunsoon
2004 Stroboscopic cine-MRI data on Korean coronal plosives and affricates: Implications for their place of articulation as alveolar. *Phonetica* 61: 234–251.

Kim, Hyunsoon
2005 The representation of the three-way laryngeal contrast in Korean consonants. In *The Internal Organization of Phonological Segments*, Marc van Oostendorp, and Jeroen van de Weijer (eds.), 263–293. Berlin/New York: Mouton de Gruyter.

Kim, Hyunsoon
to appear What features underline the /s/ vs. /s'/ contrast in Korean? Phonetic and phonological evidence. To appear in *Where do features come from? The nature and sources of phonological primitives*, George N. Clements and Rachid Ridouane (eds.), Amsterdam/Philadelphia: John Benjamin Publishing Company.

Kim, Hyunsoon, Kiyoshi Honda, and Shinji Maeda
2005 Stroboscopic cine-MRI study on the phasing between the tongue and the larynx in Korean three-way phonation contrast *Journal of Phonetics* 33: 1–26.

Kim, Hyunsoon, Shinji Maeda, and Kiyoshi Honda
submitted The laryngeal characterization of Korean fricatives: Stroboscopic cine-MRI data.

Kim, Hyunsoon, Shinji Maeda, and Kiyoshi Honda
2009 Invariant articulatory bases of the features [tense] and [spread glottis] in Korean plosives: New stroboscopic cine-MRI data. In press. *Journal of Phonetics*.

Kim, Hyunsoon, and Chae-Lim Park
to appear An acoustic study of the Korean fricatives /s, s'/: Implications for the features [s.g.] and [tense]. To appear in *Tones and Features*, Beth Hume, John Goldsmith and Leo Wetzels (eds.). Berlin/New York: Mouton de Gruyter.

Moon, Seung-Jae
1997 An acoustical study of Korean /s/. *Malsori* [Journal of the Phonetic Studies of Korea] 33–34: 11–22. Seoul.

Park, Hansang
1999 The phonetic nature of the phonological contrast between the lenis and fortis fricatives in Korean. *Proceedings of the 14th International Congress on Phonetic Sciences*, 424–427. San Francisco.
Skaličková, A.
1960 *The Korean Consonants*. Prague, CSAV, Rozpray.
Stevens, Kenneth N.
1977 Physics of laryngeal behavior and larynx modes. *Phonetica* 34: 264–279.
Stevens, Kenneth N.
1998 *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Preaspiration as a correlate of word-final voice in Scottish English fricatives

Olga B. Gordeeva and James M. Scobbie

1. Introduction

This chapter investigates the acoustics of aspiration noise in the intersegmental transition between a vowel and a following fricative, and how Scottish English speakers use this turbulence to convey phonological-phonetic structure. ‘Preaspiration’ – the perceptually salient aspiration present in vowel-obstruent transitions – is usually associated with stops rather than fricatives, both at phonological and phonetic levels of description. This study describes the occurrence of phonetic (non-normative) preaspiration of voiceless fricatives in Scottish Standard English (SSE), spoken in the Central Belt of Scotland. This variety-specific optional characteristic is variably present in different SSE speakers, and results from a learnt dissociation of the lingual and supralaryngeal gestures required for voiceless fricatives.

The aims of this study are to explore the acoustic characteristics of preaspirated fricatives in SSE and the potential linguistic functioning of preaspiration as a correlate of the fricative /voice/ (see note 1 for the use of the notation) contrast. In doing so, we will contribute to the sparse acoustic literature on preaspirated fricatives; bridge the gap between possible functional and co-articulatory explanations of this phenomenon; and present a new analytical method to quantify the glottal aperiodic turbulence in the vowel and vowel-fricative transitions independently from the offset of periodic phonation.

1.1. Preaspiration in obstruents

Preaspiration has been described as a co-ordinatory relationship between a vowel and a following voiceless segment (Laver 1994). It involves an early offset of modal voicing in the vowel triggered by the anticipatory opening of the vocal folds associated with the voiceless segment (Ladefoged and

Maddieson 1996). Generally, the terms ‘breathiness’, ‘aspiration’, and ‘whisper’ relate to turbulent flow of air passing through the abducted glottis (Laver 1994: 189-190; Stevens 1998: 428). However, since the turbulent flow produced at the glottis is inseparable from the supraglottal constriction, we also follow in viewing aspiration as a source-filter composite varying in the precise combination of a glottal opening and a gradient supraglottal constriction shaping the turbulent flow (Kim 1970: 111; Ohala and Solé in this volume). The composite source-filter view of aspiration noise is helpful in explaining the notorious variety of phonetic labels ascribed to preaspiration in the literature (ranging from glottal [h ɦ] to palatal fricatives [ç], see e.g the overview in Silverman 2003), as well as in explaining the difficulty of assigning exact phonetic labels to preaspiration based on supraglottal characteristics.

Across languages and varieties, preaspiration can be ‘normative’ (a term introduced by Helgason 2002): i.e. a consistent characteristic, as in Icelandic (Thráinsson 1979), where it seems to be a major correlate of phonological contrasts exhibited by pairs like /viht/ (“wide”) vs. /vitt/ (“breadth”). Scottish Gaelic exhibits subtle dialectal variation in /ptk/ (Ó Murchú 1985; Bosch 2008), varying from no preaspiration in the Eastern and Southern periphery, to the presence of clear preaspiration [hp ht hk] in the North and [xp xt xk] in the South. The Western parts also have [hp ht], but [xk] appears as a special co-articulatory variant, and there are more subtle patterns found in other dialectal areas. Such structured dialect variation reveals both language-specificity and, probably, how the dissociated laryngeal and supralaryngeal gestures of /-voice/ stops may become more-or-less clusterlike.

Preaspiration can also be ‘non-normative’: i.e. variably present/absent in different speakers of a variety. This is the case in many Swedish and Norwegian dialects (van Dommelen 1998; Helgason 2002; Schaeffler 2005), where it variably complements phonological contrasts between short/long vowels and consonants.

An intermediate situation is when preaspiration is socially structured, as it is, for example, in Tyneside English, where some young working class females use preaspirated word-final stops while such variants are virtually absent from the 45-65 age group (Docherty and Foulkes 1999).

Only some languages are uncontroversially labeled as having normative preaspiration. For example, it only occurs in a limited number of languages, prompting claims that it is rare (Bladon 1986; Silverman 2003). Moreover, low intensity aspiration following full vowels is prone to masking effects from the human auditory system that negatively influence its perception (Bladon 1986). However, such considerations are not strong enough to mean that preaspiration cannot be a significant part of a language's high-level phonological system. Just on the perceptual side of things, speakers of the languages with /VhC/ sequences, such as Turkish and Arabic (where /h/ is a phoneme rather than obligatory intersegmental transition) are better attuned to the presence of [h] than speakers of English or French lacking segmental /VhC/ phonotactics (Mielke 2003). These considerations make close phonetic examination of comparable or more arguable cases of language varieties with unstable non-normative preaspiration a very worthwhile research strategy, especially if the phonetic underpinnings of this phenomenon and its functions are to be understood, as well as the influence of more general processes of co-articulation on synchronic and diachronic sound change (Ohala 1981, 1993; Ohala and Solé in this volume).

Most accounts of 'non-normative' preaspiration have so far considered preaspirated stops (e.g. van Dommelen 1998; Helgason 2002; Schaeffler 2005). While there are phonological contrasts with word-initial postaspirated voiceless fricatives reported in languages like e.g. Burmese (see discussion in Vaux 1998), there have been no reports of clearly normative preaspiration for fricatives. The possibility of preaspirated fricatives in non-normative forms has been acknowledged only recently in a very limited number of studies (Helgason 2002; Jones and Llamas 2003; Gordeeva 2007; Gordeeva and Scobbie 2007). However, its possible functions in speech remain unclear. The following sections clarify the latter point.

1.2. Preaspirated fricatives in Scottish Standard English

Scottish Standard English (SSE) spoken in the Lowland Central Belt has different preaspirating languages and varieties as geographical neighbours. The systematic preaspiration of stops in Scottish Gaelic was discussed above. There is also the socially structured non-normative preaspiration of stops in Tyneside English (Docherty and Foulkes 1999; Watt and Allen 2003). For the geographically neighbouring SSE, there have been no reports of wide and early glottal abduction before word-final stops that have lead to claims of preaspiration. (Although, cases of preaspiration before stops do occur, Stuart-Smith p.c.). On the contrary, SSE stops are often glottalized (Wells 1982; Chirrey 1999; Stuart-Smith 1999; Gordeeva

2008) or even produced with complete glottal closure as strong ejectives (Gordeeva and Scobbie 2006).

Despite the seeming lack of preaspiration before stops, SSE female speakers often produce word-final fricatives with preaspiration of substantial duration (Gordeeva and Scobbie 2004). That study looked into aspirated (whispery or weakly glottal) vowel-fricative (VF) transitions that were longer than a threshold of 30 ms. Although preaspiration was variable in frequency of occurrence, it was observed in all five female middle class (MC) speakers, and in 41% of all tokens (out of a total *n*=300). One speaker used them almost exclusively. In terms of duration, in the more open vowels, preaspirated transitions could be as long as the modal vowel itself. Phrase-final location of target words increased the frequency of occurrence and yielded longer duration of preaspiration.

Figure 1 shows an example of a preaspirated transition produced by a female MC speaker from Edinburgh in phrase-final “grass”.

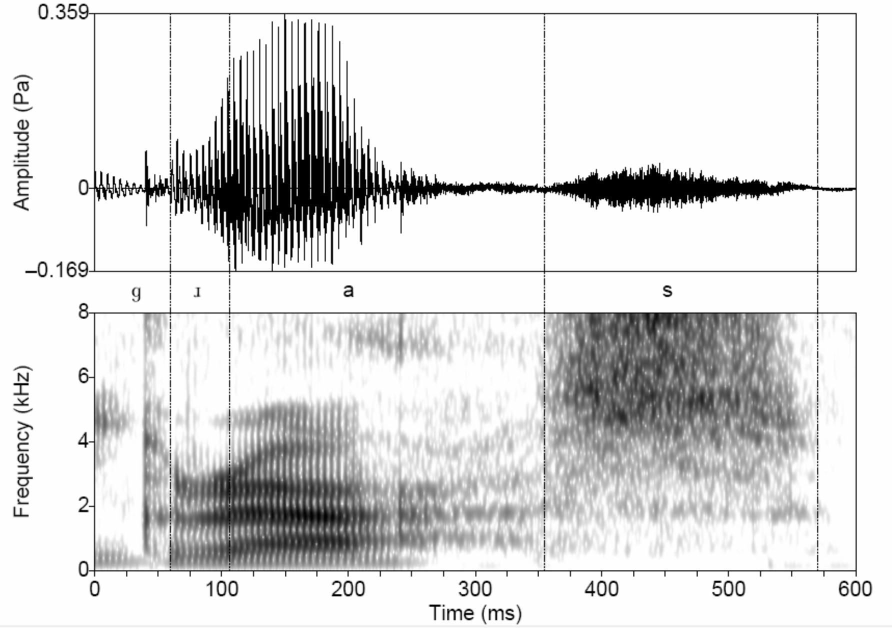


Figure 1. An example of whispery VF-transition in phrase-final “grass” token produced by a female MC speaker from Edinburgh (an example from a different dataset than this study).

The whispery to weakly aspirated VF-transition is 118 ms long, while the preceding full vowel is 129 ms. Such extensively long transitions after open (-mid) SSE vowels were often realized as glottal fricatives [h] or [ɦ] (see Figure 1). For more narrowly constricted (mid-)close vowels aspiration noise resulted in palatal [ç]. This phenomenon is also known from the literature on (post-) aspiration (Kim 1970; Stevens 1998: 445), i.e. phonemic /h/ before /i/ and /j/ is often [ç] in various languages, and is conform to our definition of preaspiration in section 1.1. Recall too the Gaelic preference for [x] before velars discussed above.

Although the extent of preaspiration shown in Figure 1 can be problematic for defining vowel offsets in acoustic analysis, previous phonetic studies that looked into vowel duration in SSE vowel-obstruent sequences (e.g. Agutter 1988; McKenna 1988) made no mention of extensively preaspirated fricatives in SSE speakers. McKenna’s discussion of segmentation criteria noted the problems arising from the partial devoicing of /z/ and voicing of /s/, but did not mention preaspiration. The lack of other specifically Scottish English reports on preaspirated fricatives raises a question (beyond the scope of this paper) as to whether we are observing an ongoing diachronical change in the gestural coordination of SSE vowel/voiceless fricative sequences. Turk, Nakai, and Sugahara (2006: 8) noted that post-aspirated /s/ in Japanese and preaspirated /ʃ/ in British English pose problems for segmental annotation, and remarked that preaspiration of voiceless fricatives is absent in American English. So, preaspiration may be a long-standing but previously unreported characteristic, or it may be a sound change in progress.

1.3. Glottal abduction in vowel/fricative sequences: automatic or learnt?

From a phonetic perspective, we could *a priori* expect to find some aspiration noise in vowel/voiceless fricative transitions. It is known that the laryngeal control mechanisms before voiceless fricatives seem to create better aerodynamic conditions for aspiration to occur compared to the contexts before voiceless stops. The glottal abduction is initiated earlier in fricatives (than in stops) relative to the formation of the supraglottal constriction (Löfqvist and McGarr 1987; Hoole 1999); and it involves a substantially greater amplitude of vocal fold abduction (Löfqvist and McGarr 1987). Similar laryngeal control mechanisms have also been shown to influence phonological /voice/ and postaspiration in word-initial stops (Lisker and Abramson 1964; Kim 1970).

It has been previously hypothesized that early glottal abduction (relative to supraglottal constriction) and a gradual breathy vowel offset might be an *automatic* (aerodynamically explained) feature of voiceless fricatives (Gobl and Ní Chasaide 1999). The study found an earlier onset of glottal abduction in voiceless (compared to voiced) fricatives irrespective of the languages considered (Swedish, Italian and German), whereas there were language-specific patterns for stops. These languages showed striking similarity in the vowel source characteristics before voiceless fricatives: i.e. gradually falling excitation strength, gradually rising dynamic leakage and increasingly symmetrical shape of the glottal pulse: i.e. all signs of early glottal abduction and increasing breathiness.

Despite these arguments, there is also evidence that the dissociation of laryngeal and supralaryngeal gestures in word-final voiceless fricatives is learnable in a language-specific way (similarly to the situation well-known in stops). Preaspirated fricatives have been noted in Central Standard Swedish (Helgason 2002) and in the Middlesbrough variety of British English (Jones and Llamas 2003). Jones and Llamas's study on the characteristics of word-final fricated and preaspirated stops in Middlesbrough was based on the data of three speakers. The materials included the word "mat" compared to control fricatives in words like "mass" and "mash". The authors concluded that the duration and auditory quality of non-modal breathy offsets in vowel-voiceless fricative transitions was substantial, and was comparable to preaspirated "mat" tokens.

In a more direct cross-linguistic study, Gordeeva (2007) showed that both Scottish Standard English (SSE) and Southern Standard British English (SSBE) speakers had a greater and earlier increase in the mid-frequency aspiration noise in vowel/voiceless fricative transitions, while in Russian (a language phonetically neutralizing /voice/), the increase was significantly smaller and later. The study suggested a tighter overlap of laryngeal-oral gestures in Russian compared to the more extreme dissociation found in the two British English varieties. In SSE and SSBE, the aspiration increase was greater for the Scottish speakers, confirming SSE-specific reports in Gordeeva and Scobbie (2004, 2007). This finding suggests that the more extreme dissociation between laryngeal and supralaryngeal gestures might be typical of other British English varieties than just Scottish or Middlesbrough English.

We assume, thus, that earlier in timing glottal abduction before voiceless (compared to voiced) fricatives can be an aerodynamically explained characteristic as previously shown by Gobl and Ní Chasaide (1999). In this limited sense, its existence can be viewed as "automatic". However, *the extent* of this gestural dissociation can be learnable in a variety or language-specific way, and can, thus, also be *actively controlled*. This assumption implies that creating aspirated turbulence in VF-transitions must be learnt either in terms of timing (Browman and Goldstein 1992) and/or in terms of the abduction amplitude (Kim 1970) in a way similar to aspirated stops. This also implies that language varieties like Scottish English that specify large laryngeal-oral dissociation in their grammars do so because this dissociation can then serve a significant non-biological function (Fuchs and Toda in this volume) such as the linguistic /voice/ contrast, or to encode sociolinguistic identity.

1.4. Transitional aspiration as a correlate of word-final fricative /voice/

There are reasons to believe that the transitional aspiration found in vowel/voiceless fricative sequences could be a facilitating cue to the phonological fricative /voice/ contrast for at least some speakers of SSE. There is, namely, an interesting parallel between the frequent lenition of voicing as a phonetic correlate of phonological /voice/ in British English varieties in phrase-final prosodic contexts (Haggard 1978; Docherty 1992) and the promotion of transitional aspiration in this context found in SSE (Gordeeva and Scobbie 2004).

The existence of preaspirated phrase-final /-voice/ fricatives may well be associated with phonetic devoicing of /+voice/ fricatives in that context. In SSE, unlike other Germanic languages, there is no phonological neutralisation of final /voice/. A priori this is because devoicing is incomplete, but perhaps the transitional aspiration is present as an optional facilitating cue, maintaining the contrast as in some Norwegian varieties (van Dommelen 1998). If the "same" contrast between /+voice/ and /-voice/ fricatives is expressed across a range of different areas of phonetic space, then the variation should be detectable in interspeaker variation, perhaps spanning traditional phonemic categories. Just such a pattern has been found in Shetlandic English in the word-initial stops /voice/ contrast (Scobbie 2006), challenging traditional accounts of the phonetics/-phonology interface (Scobbie 2007). A socially stratified sample of speakers indicated that a lower duration of VOT for /p/ meant a higher frequency of stop pre-voicing for /b/, where the speaker-specific VOT targets ranged from long lag (83ms) right down to short lag (22ms).

It is well known that depending on the phonetic context, phonological /voice/ of word-final fricatives in English can be controlled by a multitude of cues. Voiceless fricatives show much earlier cessation of voicing (Haggard 1978; Docherty 1992; Smith 1997), longer consonantal and shorter vowel duration (Smith 1997), higher voice source airflow during the consonant (Smith 1997), and subsequently higher frication noise amplitude (Balise and Diehl 1994). Most of these studies (apart from Smith 1997 that also looked at vowel duration) analyzed the acoustic correlates of /voice/ within the fricative's temporal scope and did not include the preceding vowel.

Aspiration in SSE VF-sequences is an anticipatory event with the glottal abduction gesture of the fricative starting quite early in the vowel. Therefore, in order to find out whether aspiration in the VF-transition can be promoted as a phonetic correlate of phonological /-voice/, we must subject the whole VF-rhyme to the aspiration/voicing analysis. In this study, we have developed a voicing offset ratio: i.e. a measure reflecting the extent of periodicity offset in the whole VF-sequence. This ratio also applies a normalisation procedure for the individual segments variable in duration. This voicing offset measure is applied in addition to analyses of aspiration handled in the following sections.

1.5. Acoustic measures of (pre)aspiration

With the conception of a /voice/ contrast (whether stable or not) being analysable in different regions of multidimensional phonetic space as well as on different locations (V, F, or VF), and aspiration potentially playing a role as a correlate, we need a broader range of phonetic measures in addition to the traditional measures of /voice/ discussed in the previous section.

One way to extend the range would be to use direct laryngeal techniques such as fiberoptic filming, transillumination or photoelectric glottography (e.g. Löfqvist and Yoshioka 1980; Ní Chasaide 1987) that have proven to be very useful in understanding the glottal abduction mechanisms behind (pre)aspiration. However, such experiments usually concern single case studies due to the procedural challenges imposed by the techniques. As opposed to that acoustic analysis does not require expensive specialist equipment and permits quantification of aspiration, while processing larger samples of subjects.

Generally, acoustic literature suggests that the most robust predictors aspiration and breathiness are the amount of noise present in spectral mid-frequencies (Klatt and Klatt 1990; Hillenbrand, Cleveland, and Ericson 1994) and the amplitude of the first (H1) relative to the second harmonic (H2) (Klatt and Klatt 1990; Hillenbrand, Cleveland, and Ericson 1994; Holmberg et al. 1995). Klatt and Klatt (1990) reported that increases of aspiration noise in the harmonics around the third formant (F3) in syllables with open vowels accounted for 60% of variance in the listeners' perception of long-term breathiness. Hillenbrand, Cleveland, and Erickson's (1994) study of production and perception of breathy vowel quality produced by male and female subjects confirmed the results of Klatt and Klatt (1990). The mid-frequency noise accounted for about 80% of breathiness ratings in perception, while the amplitude of the first harmonic was the second best predictor.

Despite these findings, the acoustic quantification of preaspiration (other than involving duration) has so far received very little attention in the phonetic literature (see e.g. Ní Chasaide and Gobl 1993; Bombien 2006). One reason for this is that spectral estimates of voice quality settings (e.g. modal vs. whispered) relying on formant levels (like spectral tilt in Hanson 1997) do not necessarily differentiate between contributions from periodic or aperiodic excitation crucial in aspiration analysis (see discussion in Ní Chasaide and Gobl 1993: 320). Another reason is the extreme phonetic variability in the supralaryngeal friction accompanying preaspiration ranging from more anterior (like [ç] or [x]) to more posterior place of articulation like in [h]; this on top of the differences in approximation from more close to more open settings for whispered vowels (Laver 1994; Silverman 2003). Finally, many acoustic measures of aspiration are periodicity-dependent: i.e. they are incomputable in non-periodic portions of aspirated transitions. Hillenbrand, Cleveland, and Erickson (1994) used automatic acoustic measures based on cepstral and pitch autocorrelation algorithms. Other periodicity-dependent techniques have been used (e.g. harmonics-to-noise ratio in Yumoto, Gould, and Baer 1982). As is well-known, (pre)aspiration can be periodic (Koenig, Mencl, and Lucero 2005), but often it is not, particularly around the offset of preaspirated parts (see examples in Figures 1, 2 and 4). For these reasons, preaspiration analysis raises some serious challenges for quantifying its amount in the signal. A *single* periodicity-independent measure should ideally be usable for either modal or whispered voice, or any states inbetween.

In order to explore the acoustic characteristics of preaspirated fricatives in SSE and their potential linguistic functioning as a correlate of the fricative /voice/ contrast, we developed a new periodicity-independent automatic measure of aspiration derived from the standard zero-crossing rate measure in the time domain. This measure was complemented by a set of more traditional (but periodicity-dependent) acoustic correlates of aspirated phonation: i.e. H1*-H2* and HTN (for details see sections 2.4.3. and 2.4.4.).

2. Method

2.1. Subjects and recordings

Data were gathered from five female (F1 – F5) and five male (M1 – M5) speakers of Scottish Standard English. The subjects were recruited amongst Queen Margaret University staff members. All subjects were informally of Middle Class background and were between 23 and 50 years old. All SSE speakers were long term residents of Edinburgh. Nine speakers were born in the Lowland Central Belt. One speaker (F1) was born in Aberdeen.

The recordings were made in a sound-treated recording studio using an omnidirectional condenser microphone. The recording volume settings and each subject's distance from the microphone were kept constant. The subjects read a set of sentences containing target words from the computer screen. No specific instructions were provided towards the pitch accent placement in the utterances. The subject's speech rate was controlled by the prompt sentences being made to appear at regular time intervals. The preaspiration materials were presented randomly and contained interspersed utterances from three additional experiments as distractors. The data was digitized at a sampling rate of 44100 Hz with 16-bit resolution, and was subsequently downsampled to 11025 Hz sufficient for all acoustic analyses in this study. The male data also included parallel laryngographic recording (Laryngograph Processor ™) that was not used in this study.

2.2. Materials

2.2.1. Female data

The original female data in which preaspiration was found was designed as control for a child language study (Gordeeva 2005). The materials varied phrasal accent locations and fricative /voice/, and contained three vowel heights: i.e. close, close-mid and open-mid. A subset of the complete data included five target words each repeated five times over four positions (two phrase-final and two non-final). The data are summarized in Table 1. The carrier sentences were of the form “A fish is a fish, and nothing but a fish” and “It’s a fish”. This yielded 100 (5x4x5) tokens per female speaker, and a total of 500 tokens for all speakers.

Table 1. Control conditions, materials and carrier sentences used for the female and male speakers. The uppercase words indicate phrasal accent (see note 2) in the carrier sentences typically produced by the speakers.

Female speakers:	**Phrasal contexts and carriers**
Preaspirated target	
/ʉ/ goose, /ɪ/ fish, /ʌ/ bus	*Final 1:* It’s a <TARGET>.
Voice contrast targets	*Non-final 1, Non-final 2 and Final 2:*
goo**se**/shoe**s**/choo**se**	A <TARGET> is a <TARGET>, and nothing but a <TARGET>.
Male speakers:	**Phrasal contexts and carriers**
Preaspirated target	
/ɩ/ fish, dish;	*Final 1:*
/e/ place, base;	That’s the word <TARGET>.
/ɛ/ best, Beth; ɔ/ boss;	*Non-final 1:*
/ʌ/ bus; /a/ bath	I can say <target> AGAIN.
Voice contrast targets	
bu**s**/bu**zz**	*Final 1:*
pla**c**e/play**s**	That’s the word <TARGET>.
ba**s**e/bay**s**	*Non-final 2* and *Final 2:*
	I say <TARGET>, and not <TARGET>.
	Non-final 1:
	I can say <target> AGAIN.

2.2.2. Male data

The male data were recorded at a later stage and contained additional vowel heights and three minimal voicing contrast pairs (see Table 1) to provide greater descriptive detail. The vowel height data included nine target words repeated over two phrasal positions, yielding 18 (9x2) tokens per speaker, and a total of 90 tokens for all speakers. The /voice/ contrast pairs contained six words recorded over four phrasal positions, yielding 24 (6x4) tokens per speaker, a total of 120 tokens for all speakers. These materials were of the form "That's the word bus" and "I can say bus again". For the targets involving the fricative /voice/ contrast we used the additional carriers such as "I say bus, and not buzz" containing minimal pairs (in both orders).

2.3. Phonetic analyses

Phonemic transcription was annotated along with the segment duration for each target word. All analyses were done in PRAAT 4.3 (Boersma and Weenink 2006) at a sampling rate of 11025 Hz with 16-bit resolution. All annotations were performed by the first author.

In defining segmental boundaries, we mainly followed the annotation criteria specified in van Zanten, Damen, and van Houten (1991). The criteria concern rapid changes in shape of the amplitude envelope of an acoustic waveform. CV boundaries were marked at the point of rapid change in the amplitude envelope. The VC boundary between the vowel and the following fricative was annotated at the beginning of a visually identifiable supralaryngeal friction in the higher frequency partials of the spectrogram.

Preaspirated parts were separately time-marked. We used a combination of acoustic landmarks (occurring prior to the onset of the following fricative noise) to determine the onset of whispery or glottal/supraglottal friction in the vowel offsets: (1) qualitative changes in the spectrum around F2-F4 in the excitation patterns from periodic to aperiodic along with visible weaking of the formant levels; (2) points of rapid (concave) decrease of amplitude envelope to consonantal levels; and (3) total offset of periodicity in the waveform amplitude if no other landmarks were present.

In Figure 2, the glottal fricative is 115 ms long, and the preaspiration onset is determined by the onset of non-[s] mid-frequency noise after the more modally voiced vowel. What matters here is the presence of glottal friction, not the loss of voicing, so both [h] and [ɦ], or any intermediate form, will count as preaspiration for the following lingual fricative. If the glottal fricative [ɦ] has uninterrupted voicing throughout the consonant (Koenig, Mencl, and Lucero 2005; Löfqvist, Koenig, and McGowan 1995), any measure of voicing offset will be unable to determine the acoustic boundaries of preaspiration. In Figure 2 the voicing continues a third of the way into the aspirated part.

Preaspiration was time-marked, however short, but if no preaspiration was found, the preaspiration marker was placed at the end of the vowel with zero duration. In voiced fricatives, the VF-boundary marker was determined by the onset of target lingual friction in spectrograms irrespective of the voicing.

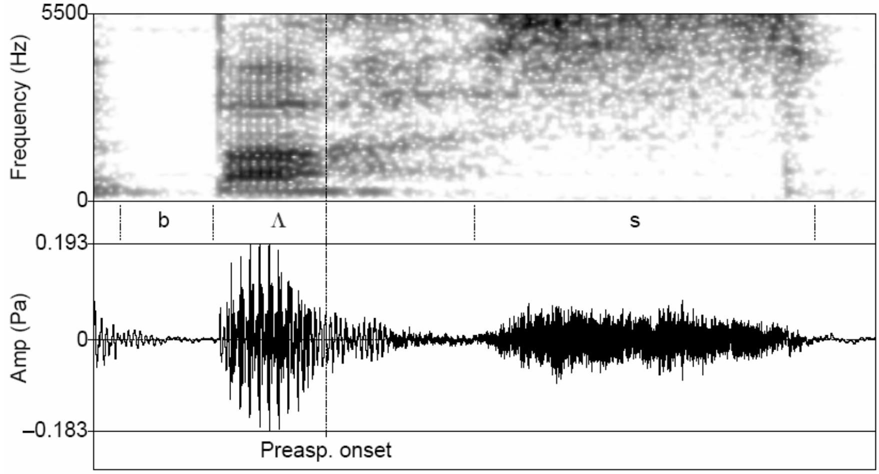


Figure 2. An example of annotation of preaspirated VF-transition in the word "bus" (speaker F1) following downsampling.

In any study of VF-transitions, we would expect an early onset of laryngeal abduction relative to the supraglottal constriction. One of the initial motivations for this study was that the voicing offset in the SSE female data observed by Gordeeva (2005) was so large to merit careful study. In order to obtain a categorical quantitative count of such preaspirated tokens, a threshold is required. Therefore, individual tokens were counted as 'preaspirated' if the time delay from the preaspiration onset to the onset of the following fricative was longer than 30 ms (independent of the duration of V or F). This seems to be a reasonable choice for a durational limen given a similar perceptual threshold for preaspiration as cue to voicing contrast reported by van Dommelen (1998).

For each token, we determined the presence/absence of phrasal accent and accordingly labeled them as "accented" or "de-accented" based on general impressions of prominence. Uncertain cases were consistently labeled as "de-accented". To avoid any prosodic ambiguities, only "accented" tokens were used during data analysis.

2.4. Acoustic analyses

In addition to a categorical binary analysis of preaspirated vs. non-preaspirated tokens, we performed continuous acoustic measures of preaspiration in VF-boundaries in order to better understand the underlying phonetic processes.

The continuous measures allowed us to establish the correlates of the observational categorisation into tokens containing preaspirated vs. non-preaspirated fricatives. Finally, we analysed the correlates of the conventional phonological categorisation into materials containing phonologically /+voice/ vs. /-voice/ fricatives. The acoustic measures used in this study are listed in Table 2. We assumed that these acoustic measures were potentially important correlates of both preaspiration and /voice/, but assigned them no *a priori* ranking. Instead their ranking was quantitatively defined in a bottom-up fashion using statistical analyses. Intensity measures were excluded from this set, because of the contextual heterogeneity of the recording materials. The parameters in Table 2 were automatically derived based on manual annotations of segment duration. All acoustic analyses were performed in PRAAT 4.3. The aspiration-related parameters were measured as averages in the middle and final parts of the vowel or as parameter change in the last part of the vowel relative to the middle vowel part (see Table 2).

Table 2. Overview of the acoustic measures used in this study. Further details are given in the text.

Measure	Description
Voicing:	
voicing_offset (%)	voicing offset ratio
Duration:	
V_dur (ms)	vowel duration (including preaspiration)
f_dur (ms)	duration of the coda fricative
Aspiration-related measures:	
ZCR mid (per sec)	zero-crossing rate in middle (third fifth) part of the vowel
ZCR final (per sec)	ZCR in the final (fifth) part of the vowel
ZCR change (per sec)	ZCR difference between the final and middle parts of the vowel
HTN mid (dB)	harmonics-to-noise ratio in middle part of the vowel
HTN final (dB)	HTN in the final (fourth fifth) of the vowel
HTN change (dB)	HTN difference between the final and middle parts of the vowel
H1*-H2* mid (dB)	H1*-H2* ratio in the middle (third fifth) part of the vowel
H1*-H2* final (dB)	H1*-H2* in the final (fourth fifth) part of the vowel
H1*-H2* change (dB)	H1*-H2* difference between the final and middle parts of the vowel

2.4.1. Voicing offset ratio

Voicing offset ratio (%) is a measure reflecting the timing of the offset of periodicity in the complete VF-rhyme. The offset of periodicity here refers to the complete offset of voicing rather than to the offset of modal voicing. The timing is expressed as percentage on the scale between 100 and -100%, whereby the V onset marks 100% and the f-offset marks -100%, both relative to the onset of final fricative (set at 0%) (see Figure 3). This new method has the advantage of indicating whether periodicity cessates in vowels or fricatives by means of the positive and negative scaling, and it quantifies the periodicity extent in percent relative to the 0% landmark (see example in note 3). At the same time the use of separate percentages for V and F parts normalizes for variable segmental durations of V or F independently from each other.

This method is preferable to voicing offset being expressed relative to the entire VF-part, as this cannot reveal the extent of periodicity in the separate segments (V or F). It also has the advantage of having a *single* measure for periodicity compared to two separate temporal measures for vowels or consonants that have been traditionally applied. This single measure is usable for any analysis of word-final /±voice/ obstruents: periodicity is likely to stop prior to the fricative onset when the fricative is /-voice/ (i.e. a partly devoiced vowel, giving a positive percentage offset), while in /+voice/ cases the offset is more likely to occur within the fricative portion (a negative offset).

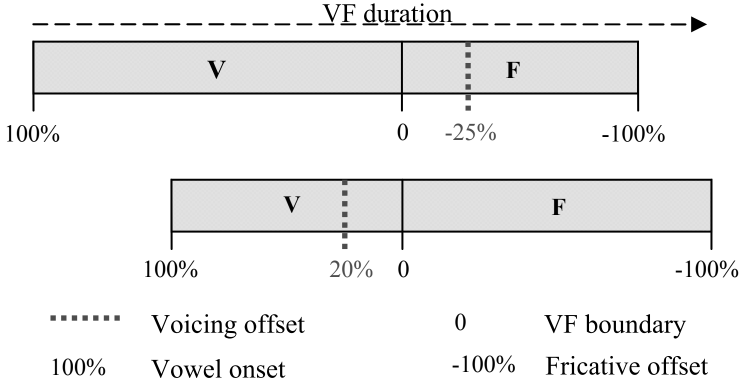


Figure 3. Representation of voicing offset ratio measure. The two gray horizontal bars represent absolute VF-duration of two different VF-sequences. The vertical 0 markers show the oral fricative onset time (so the two horizontal VF-bars are aligned at this point). The vertical dotted lines represent the relative timing of voicing offset in the VF-part. The percentages below are calculated between 0 and the fricative offset (-100%), or 0 and the vowel onset (100%), depending on whether voicing cessates in the fricative (as in the top bar representing a partially voiced fricative) or in the vowel (as in the bottom bar representing a preaspirated voiceless fricative).

Periodicity was measured from speech waveforms using the cross-correlation algorithm with the minimum of 75 Hz and maximum of 350 Hz for the male data, and 75 Hz and 400 Hz for the female data. The minima and maxima were based on vowel F0 ranges in each group. Prior to this, the speech waveforms were high-pass filtered at 50 Hz to get rid of the DC component.

2.4.2. Band-pass filtered zero-crossing rate

Zero-crossing rate (ZCR per sec) as implemented here reflects the amount of aspiration in the spectrum above the fundamental frequency. ZCR is a standard measure calculated in the time-domain of a waveform as the number of crossings by the wave of the time-axis (or zero-crossings) within a certain period of time, divided by the number of samples in this part (e.g. Rowden 1992: 45-46). ZCR tends to be highest for voiceless fricatives. Standard ZCR has recently been applied to study voiceless sonorants in Icelandic (Bombien 2006). However, the standard ZCR measure is heavily affected by the presence of low-frequency periodicity. In breathy speech, such low frequency (H1) components have the effect of quasi-sinusoidally displacing the wave away from the zero-line, reducing the number of zero-crossings caused by breathiness (Klatt and Klatt 1990; Stevens 1998) (see Pane A in Figure 4).

In this study, we propose an adaptation of the standard ZCR measure for periodicity-independence. The band-pass filtered ZCR reflects only mid-frequency ranges rather than the full spectrum. In order to achieve this, the waveforms were band-pass filtered with a flexible lower limit (defined at 1.5*maximum F0 for each vowel token) and an upper limit at 5.5 kHz (i.e. Nyquist frequency). The lower limit removes low frequency deviations away from the zero-line due to potential presence of periodic excitation. Without high-pass filtering the presence of this low-frequency harmonic renders a ZCR measure meaningless as an indicator of midrange aperiodic noise. The effect of band-pass filtering is illustrated in Figure 4 on Pane B, where the quasi-sinusoidal H1 domination in the time-domain is reduced (compared to Pane A), and the zero-crossings more accurately represent the midrange non-modal aspiration visible in the top spectrogram pane. The upper frequency limit reflects the main vocal tract resonance for vowels (whether phonetically voiceless or voiced). If the upper band is set higher than this, a side-effect of any friction generated locally to some supraglottal constriction is that it may inadvertently increase ZCR (e.g. if the VF-boundary is annotated somewhat later than the onset of the following fricative). Therefore, setting a too high upper limit should be avoided. Zero-crossings are counted in frames of 10 ms per second divided by the number of frames (Rowden 1992: 45-46). High values of ZCR reflect turbulent noise in the open vocal tract, while lower values reflect more modal excitation patterns.

Apart from periodicity independence, band-pass filtered ZCR should also be a better indicator of the midrange noise than more traditional measures such as spectral tilt (H1*-A3*, dB) (Hanson 1997), where the spectral level measures of the third formant can be boosted by aperiodic components (see discussion in Ní Chasaide and Gobl 1993: 320). Our measure of breathiness/aspiration has further advantage of being independent of the accuracy of pitch trackers, it does not require amplitude normalisation (as e.g. in Hillenbrand, Cleveland, and Erickson 1994) and is fully automatic and is easy to compute.

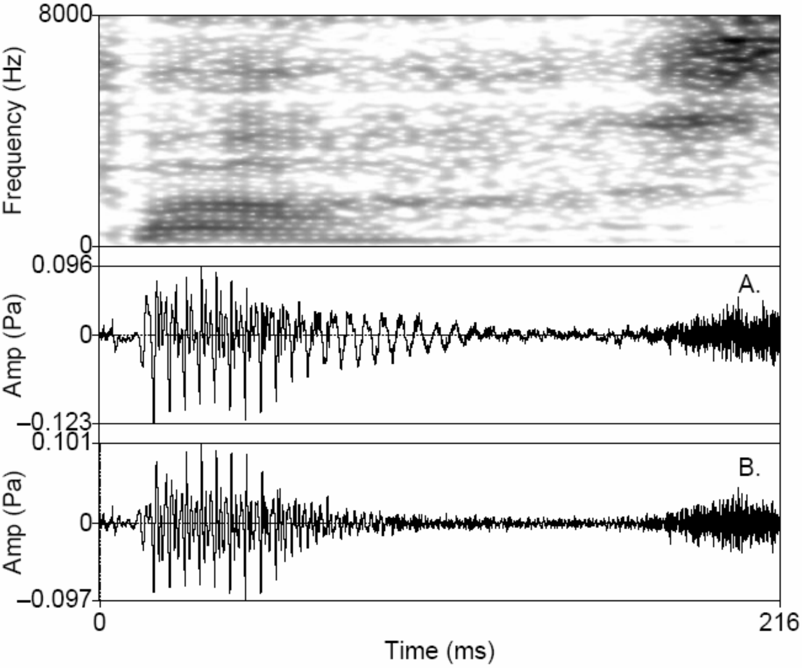


Figure 4. The effect of band-pass filtering on zero-crossing rate in a time-domain waveform with aspirated vowel /ʌ/ followed by fricative /s/: the unfiltered waveform with sinusoidal H1 domination in the VF-transition is represented in Pane A., the band-pass filtered form in Pane B, the top pane represents the spectrogram of the unfiltered speech.

2.4.3. Harmonics-to-noise ratio

Harmonics-to-noise ratio (HNR, dB) reflects a relative amount of aperiodicity in otherwise periodic portions of speech spectrum. This ratio does not show which aperiodic component (noise, jitter or shimmer) contributes to aperiodicity (Murphy 1999), and as such can be a correlate of hoarseness (Yumoto, Gould, and Baer 1982) as well as aspiration. An HNR ratio of 0 dB means that there is equal energy in the harmonics and noise, while 1% of noise equals 20 dB. HNR is calculated here using the Harmonicity autocorrelation method in PRAAT in frames of 10 ms and minimal pitch of 65 Hz. We only performed HNR comparisons in the first 80% of the vowel, since the final 20% often lacked periodicity due to voiceless preaspiration, in which case the ratio is incomputable.

2.4.4. H1-H2**

H1*-H2* (dB) is the difference between the levels of the first (H1) and second (H2) harmonics of the vowel. It serves as an approximation of the open quotient (OQ): i.e. the timing of open phase of the glottal cycle relative to the total time of the period. Due to the increase of the glottal area, larger OQ should lead to greater losses and more aspiration noise (Holmberg et al. 1995). The more breathy types of phonation are characterized by lower values of H2 relative to H1 (Hanson 1997; Löfqvist 1995). H1 and H2 levels were measured from the spectrum in a Hamming window with a window length covering two pitch periods. Raw H1 and H2 values were then corrected for the 'boost' effect resulting the proximity of the first formant following the procedure in Hanson (1997: 475). This resulted in boost-effect corrected H1*-H2* values.

2.5. Statistical analyses

All statistical analyses were carried out using SPSS 12 software. Multivariate Analysis of Variance (MANOVA, $\alpha = .05$) was used with the acoustic variables in Table 2 as dependent variables, and with the categories PREASPIRATION or VOICE as fixed factors to determine which of the variables have a significant effect on either of the fixed factors. The fixed factor VOICE was determined by purely phonological convention and PREASPIRATION was derived as discussed in section 2.3.

Subsequently, Linear Discriminant Analysis (LDA) was used to evaluate the relative ranking of each of the significant variables in predicting PREASPIRATION or VOICE. The 'stepwise' LDA was chosen, since it makes no assumptions about which predictor should have higher priority than others, and the order of predictor entry is determined by statistical criteria (Wilks' Lambda with an F-value of 3.84 for predictor entry and 2.71 for removal).

2.6. Reliability

In order to evaluate the consistency of the manual timing annotations, the first author re-measured the timing of the VF- and preaspiration boundaries eight months after the original analysis. The test was based on a random 10% of the data, both male and female. The RMS-error for the VF-boundary and onset of preaspiration was 8.3 ms and 8 ms respectively, and corresponded to only a small mean 1.6% and 1.5% of the total VF-duration. Based on these results, we considered the manual annotation of segment duration in the acoustic data to be reliable for the dataset.

3. Results

3.1. Acoustic correlates of observed preaspiration in /-voice/ fricatives

In this part, we aimed to establish the hierarchy of acoustic correlates of preaspiration found in SSE vowel - voiceless fricatives transitions. In order to find the candidate variables for the ranking, we ran MANOVA with the acoustic measures in Table 2 as independent variables and PREASPIRATION ("yes" or "no") as a fixed factor to determine which of the acoustic variables were significantly affected by preaspiration. The analysis was carried out on a selection of all instances carrying phrasal accent and ending with /-voice/ fricatives (n = 298).

Table 3 summarizes the acoustic variables that were significantly affected by preaspiration with means and standard deviations, as well as the significance levels (with F- and p-values) for those variables. The consonantal duration and H1*-H2* in the mid part of the vowel were not significantly affected.

In order to measure the ability of each of the significant acoustic correlates to predict PREASPIRATION and establish their ranking, we subjected the acoustic measures (n = 298) in Table 3 to stepwise LDA. PREASPIRATION ("yes" or "no") was used as the predicted variable, and the acoustic measures as independent variables (predictors).

The results indicate that 84.9% of all targets were correctly classified into our annotation-based categories of non- and pre-aspirated sounds. The relative ranking of the acoustic variables is summarized in Table 4, where they are ordered by their correlation size with standardized canonical discriminant functions. Amongst the significantly affected correlates of preaspiration, the best predictors of the presence or absence of PREASPIRATION are vowel duration, zero-crossing rate in the final vowel part and its change throughout the second part of the vowel, followed by the voicing offset ratio and harmonics-to-noise ratio in the second part of the vowel.

Table 3. Means, one standard deviation, and MANOVA results for the acoustic variables significantly affected by preaspiration pooled for all subjects.

Acoustic variables	Preaspirated? (Total *n*=298)				Significance	
	No (n=178)		Yes (n=120)		df=1	
	Mean	Stdev	Mean	Stdev	F	p
V_dur (ms)	123	30	170	46	115	0.000
Voicing offset ratio (%)	-9.2	18.8	2.7	14.5	34	0.000
ZCR mid (per sec)	1119	329	1488	580	49	0.000
ZCR final (per sec)	2271	1131	3444	1201	73.4	0.000
ZCR change (per sec)	1152	1050	1956	1119	39.9	0.000
HTN mid (dB)	15.2	5.2	12.7	4.2	18.7	0.000
HTN final (dB)	13.6	4.4	10.1	4.7	41.8	0.000
HTN change (dB)	1.6	9.6	3.9	7.4	5	0.028
H1*-H2* final (dB)	10.0	8.6	12.1	8.3	4	0.046
H1*-H2* change (dB)	1.6	9.6	3.9	7.4	4.9	0.028

Individual speaker means for preaspirated and non-preaspirated VF-transitions for the five selected LDA parameters and HTN final measure are plotted in Figure 5. To give an idea about the individual and sex differences for each of the parameters, the individual means of the female speakers are plotted in solid lines, while the male speakers are represented by dotted lines. Speaker M5 produced only preaspirated realisations under phrasal accent, therefore this subject's data lacks any non-preaspirated data points.

Figure 5 shows that preaspirated variants have longer vowel duration (where "vowel" includes any preaspiration); have substantially higher ZCR values (as an absolute value vowel-finally and as a relative change in the second part of the vowel) reflecting the increasing aspiration levels above the fundamental frequency; have less periodicity compared to noise (lower HTN values), and that the voicing offset is earlier relative to VF-boundaries. Although there are both individual and sex differences for most of the parameters, what matters is that the individual speakers are consistent in producing the same direction of the differences for the top four LDA parameters. This shows the coherency of the MANOVA and LDA results. There is more individual variation in HTN final measure (with the speaker F4 producing similar levels for non- and preaspirated variants), and H1*-H2* change measure.

Table 4. Pooled within-groups correlations (see note 4) between discriminating variables and standardized canonical discriminant functions for PREASPIRATION. Variables are ordered by absolute size of correlation within function. Variables marked with (a) were not selected by LDA as predictors.

Acoustic variable	
V_dur	0.627
ZCR final	0.501
ZCR mid (a)	0.409
ZCR change	0.369
Voicing offset ratio	0.341
HTN final (a)	-0.268
HTN mid (a)	-0.261
H1*-H2* change (a)	0.129
HTN change	0.129
H1*-H2* final (a)	0.012

Figure 5 also shows some sex-related differences between the individual speakers. Because of the heterogeneity of the recording materials between the males and females, any conclusions with respect to sex differences should be drawn with caution. For the two ZCR measures the female subjects produce higher ZCR values (more mid-frequency aspiration), which is consistent with the literature on breather voice quality in female speakers (Klatt and Klatt 1990; Fant, Kruckenberg, and Nord 1991; Hillenbrand, Cleveland, and Erickson 1994; Hanson 1997). The sex differences in vowel duration and voicing offset ratio may reflect the differences in materials (i.e. the female data contained less open vowels and more phrasal contexts). The higher HTN values (reflecting more periodicity than males) can be a result of creakier phonation type in male subjects, but cannot be interpreted with certainty because of the already discussed ambiguity of the HTN measure as to breathiness or creakiness (Yumoto, Gould, and Baer 1982).

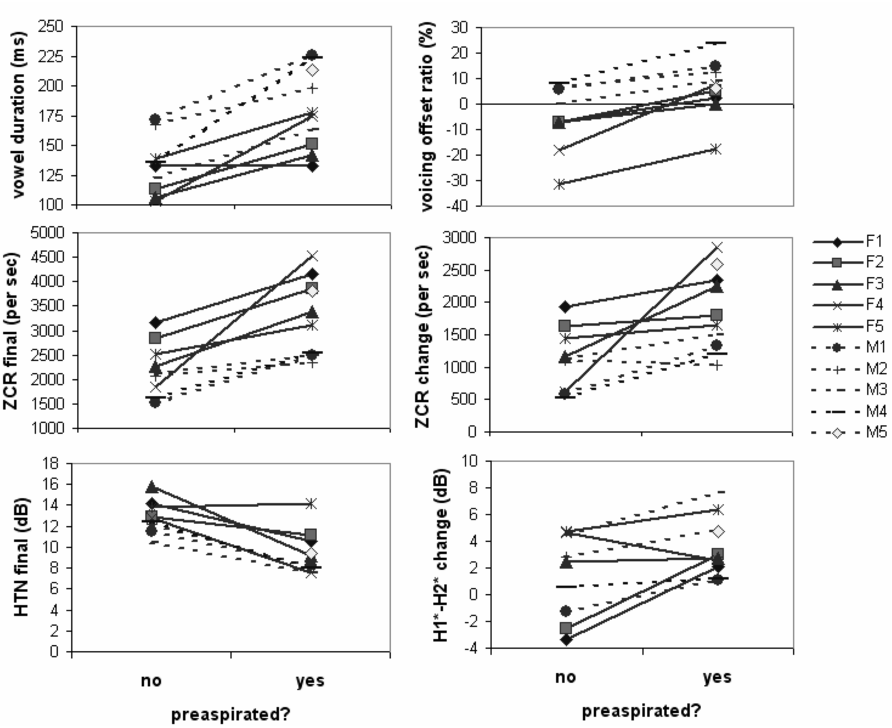


Figure 5. Individual speaker means of preaspirated and non-preaspirated voiceless fricatives for six top-ranked acoustic predictors of preaspiration. Solid lines indicate female speakers. Dotted lines indicate male speakers.

3.2. Acoustic correlates of phrase-final fricative /voice/

In the previous section, we established the hierarchy of the acoustic correlates of preaspiration in VF-transitions. In this section, we explored how the set of acoustic characteristics defining preaspiration relates to the production of the fricative /voice/ contrast. We evaluated the hypothesis that preaspiration enhances or preserves the phonological /voice/ contrast in phrase- and word-final fricatives prone to the partial phonetic neutralisation of voicing.

We ran MANOVA with the acoustic measures in Table 2 as independent variables, and VOICE ("yes" or "no") as a fixed factor to determine which of the acoustic variables were significantly affected by underlying fricative /voice/. The analysis was carried out on the targets in phrase-final positions realized with a phrasal accent and ending with voiced and voiceless fricatives (*n*=358). Table 5 summarizes the acoustic variables that were significantly affected by VOICE with means and standard deviations for the affected acoustic measures, as well as the significance levels (with F- and p-values) for those variables.

The measures ZCR mid, H1*-H2* mid and change and HTN change were not significantly affected by VOICE.

Table 5. Means, one standard deviation, and the MANOVA results for the acoustic variables significantly affected by VOICE pooled across all subjects.

	/Voice/ (Total *n*=358)				Significance	
	No (n = 147)		Yes (n =211)		df=1	
Acoustic variable	Mean	Stdev	Mean	Stdev	F	P
V_dur (ms)	154	51	200	57	62.8	0.000
F_dur (ms)	207	66	165	61	38.2	0.000
Voicing offset ratio (%)	-2.4	18.4	-28.4	21.4	142.4	0.000
ZCR final (per sec)	3064	1215	1545	953	175.1	0.000
ZCR change (per sec)	1727	1144	206	944	188.6	0.000
HTN mid (dB)	14.4	5.3	18.1	5.6	38.3	0.000
HTN final (dB)	11.8	5.0	17.7	5.8	101.9	0.000
H1*-H2* final (dB)	10.0	8.0	10.7	10.8	3.8	0.053

In order to measure the ability of each of the significant acoustic correlates to predict phonological /voice/ and establish their ranking, we subjected the significantly affected acoustic measures to stepwise LDA with the same targets as for MANOVA (*n*=358). VOICE ("yes" or "no") was used as predicted variable, and the acoustic measures in Table 5 as predictors.

The results indicate that a very high 92.2% of all phonologically voiced or voiceless fricatives were correctly classified. The ranking of the acoustic measures is listed in Table 6.

The results in Table 6 show that zero-crossing rate change through the second half of vowel (ZCR change) and its amount at the end of the vowel (ZCR final) are the most successful predictors of the phrase-final fricative /voice/, followed by the voicing offset ratio. The importance of voicing offset, and vowel and consonantal duration in cueing fricative /voice/ is well-known from the literature (Docherty 1992; Haggard 1978; Smith 1997), and is also corroborated in this study, since these parameters were significantly affected by VOICE in the MANOVA. However, the higher importance of the transitional aspiration parameter (ZCR) as a correlate of fricative /voice/ has so far not been attested.

The result supports our hypothesis that given the fact that phonetic devoicing of /-voice/ fricatives is likely to occur in phrase-final positions, it is also likely that transitional aspiration helps to maintain the /voice/ contrast in this prosodic context. This conclusion is also supported by the parallel significance and good LDA ranking of HTN ratio in the final part of the vowel. The durational correlates seem to score less successfully than both ZCR measures and voicing offset ratio in the % correct classification of fricative /voice/. This is even more surprising considering the fact that the SSE close vowels in the female data set are also affected by the Scottish vowel length rule with very big differences in vocalic duration (e.g. Scobbie, Hewlett, and Turk 1999).

Table 6. Pooled within-groups correlations (see note 4) between discriminating variables and standardized canonical discriminant functions for VOICE. Variables are ordered by absolute size of correlation within function. Variables marked with (a) were not selected by LDA as predictors.

Acoustic variable	Correlation
ZCR change	-0.518
ZCR final	-0.500
Voicing offset ratio	-0.451
HTN final	0.381
V_dur	0.299
HTN mid	0.234
F_dur (a)	-0.148
H1*-H2* final (a)	-0.070

Individual subjects' means for the top four LDA VOICE predictors and vowel and consonantal duration are plotted in Figure 6. The relative differences in the top four predictors of VOICE are consistent for all individual subjects (and irrespective of sex). The direction of the differences is similar to that for PREASPIRATION: i.e. higher ZCR values in the VF-transition and in the second half of the vowel, earlier voicing offset and lower HTN values (less periodicity). The sex-related differences are also similar to those presented in Figure 5.

Additionally, there are also remarkable individual differences observable in Figure 6. Male subject M5 (who produced exclusively preaspirated realisations in all tokens ending with voiceless fricatives shown in Figure 5) produces the highest difference in ZCR parameters, but unlike other subjects fail to produce the durational differences in VOICE for both vowels and consonants, suggesting that the /voice/ contrast is primarily encoded by aspiration-related parameters and less so by duration or offset of phonetic voicing.

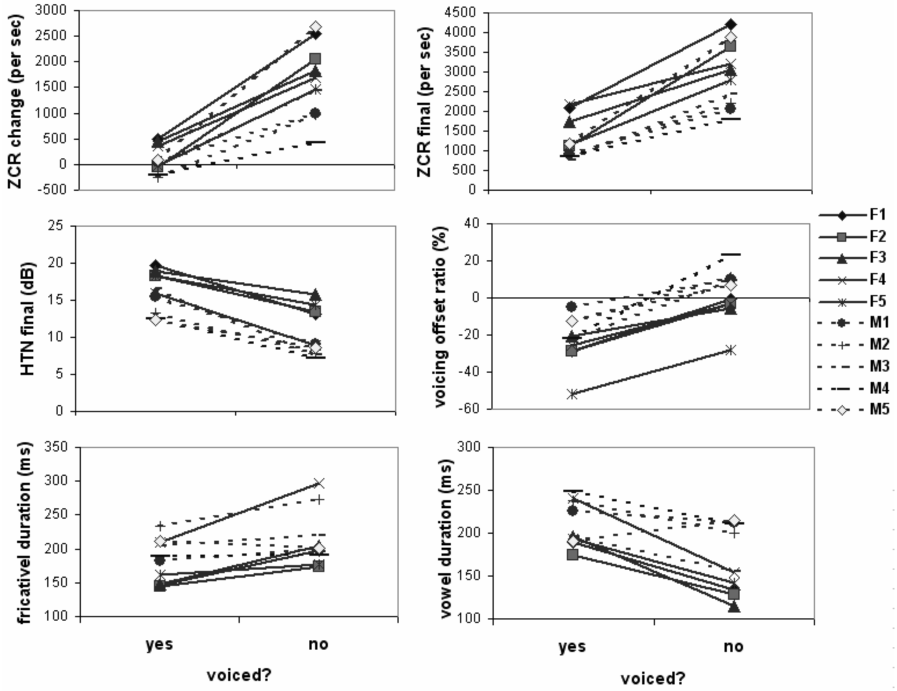


Figure 6. Individual subject means of VOICE ("yes" for /+voice/) plotted for the four highest LDA ranked acoustic parameters and additionally vowel and consonantal duration means. Solid lines indicate female speakers. Dotted lines indicate male speakers.

4. General discussion

The aim of this study was to establish the acoustic footprint of preaspiration in vowel-fricative transitions in Scottish Standard English and to link it to its potential functioning in the linguistic system as a correlate of word-final fricative /voice/ contrast.

4.1. The acoustics of transitional aspiration

With regard to the acoustic footprint of preaspiration, the main finding is that vocalic duration is the most successful predictor of preaspirated and non-preaspirated realisations. The fact that longer vowel duration is associated with the occurrence of preaspiration in VF-transitions is not surprising given the greater likelihood of preaspiration in phrase-final positions reported in Gordeeva and Scobbie (2004; 2007), a context which triggers longer segmental durations marking prosodic edges (Edwards, Beckman, and Fletcher 1991; Cho 2001). Therefore, the primacy of vowel duration as a correlate of preaspiration possibly reflects these phonetic/prosodic distributions. However, a direct comparison of phrase-final and non-final accents would be needed to assert this with any certainty.

Apart from that, preaspiration in the SSE VF-transitions is best predicted by mid-spectral aspiration parameters in the final part of the vowel (band-pass filtered ZCR). This finding confirms previous reports that in acoustic terms aspiration is best reflected in sustained vowels by the spectral noise components present in midfrequency ranges (Klatt and Klatt 1990; Hillenbrand, Cleveland, and Erickson 1994). In this study, we have shown that midfrequency noise is a successful predictor of aspirated voice quality applied to short-term transitional aspiration throughout the second half of the vowel. Aspiration correlates less strongly to the harmonics-to-noise and voicing offset ratio's. The fact that band-pass filtered ZCR substantially incraeases in the preaspirated cases towards the end of the vowel, while the HTN-ratio decreases, reflects a substantially smaller amount of periodicity compared to noise in the second half of the preaspirated vowel. The individual speakers in Figure 5 are consistent in realising the directions of differences irrespective of subject and sex.

Although the H1*-H2* final and change measures were significantly affected by preaspiration in having about 2 dB bigger differences between

H1 and H2 (suggesting breathier offsets due to somewhat bigger glottal area), the measures did not contribute sufficiently to the % correct classification and were amongst the lowest in ranking. This fact supports previous reports in the literature that, in so far as H1*-H2* reflects the degree of glottal opening, it seems to be a parameter independent from the ones reflecting the amount of glottal airflow (Klatt and Klatt 1990; Hanson 1997), like band-pass filtered zero-crossing rate in this study. This supports the idea that it is possible to adjust the timing of glottal opening relative to the supraglottal constriction without increasing the glottal airflow, as well as to adjust the amount of glottal flow without changing the timing of glottal opening (Hanson 1997).

A previous cross-linguistic study (Gordeeva 2007) showed that this large timing dissociation of linguo-laryngeal gestures in Scottish Standard English VF-sequences resulting in preaspiration is a language-specific learnt characteristic, while the dissociation is tighter in other languages like Russian. The high rate of mid-frequency aspiration prior to the consonantal constriction shows the importance of the amplitude of the glottal abduction gesture (Kim 1970).

4.2. Transitional aspiration as a correlate of /voice/

The possibility of linguistic function of large dissociated laryngeal and supralaryngeal gestures of /-voice/ fricatives has received little attention in the phonetic literature. In particular, the view typified by Gobl and Ní Chasaide (1999), in which the timing of glottal abduction in voiceless fricatives reflects automatic co-articulation resulting from aerodynamic demand must be tempered by awareness that languages vary in arbitrary ways, functionally exploiting the same phonetic characteristics irrelevant to other languages. We hypothesized that in a way similar to stops (discussed in the introduction) the dissociated laryngeal and supralaryngeal gestures of /-voice/ fricatives can convey sociolinguistic meaning, and, moreover, may be an important (albeit optional) exponent of phonological contrast. The latter hypothesis was tested and corroborated in this study.

In testing this hypothesis, we introduced a new unifying methodology to study aspiration and /voice/ simultaneously: both in terms of the acoustic measures used (including band-pass-filtered ZCR and voicing offset ratio) and in terms of broadening of the scope of analyses to the whole VF-rhyme.

An important finding is that SSE fricative /voice/ contrast in phrase-final singleton targets is primarily encoded by VF-transitional mid-frequency noise. Band-pass filtered zero-crossing rate – a periodicity-independent version of standard ZCR specifically designed in this study – suggests earlier timing of glottal abduction prior to the supralaryngeal constriction for the /-voice/ compared to /+voice/ fricatives. The importance of transitional aspiration as a correlate of phonological /voice/ mirrors the importance of this parameter as an acoustic correlate of preaspiration in voiceless fricatives shown in section 2.4.3., where VF-transitional aspiration was only surpassed in strength by vowel duration (while remaining a near-top parameter).

As aspiration-related transitional parameters have not been considered previously as a correlate of /voice/ in English word-final fricatives, this preaspiration may be variety-specific to SSE. However, it may play a similar role in other preaspirating British English varieties like Middlesbrough English (Jones and Llamas 2003) or any other English variety with this characteristic. It is possible that the extent of dissociation of laryngeal and oral stricture in other studied English varieties (Docherty 1992; Non SSE British English: Haggard 1978; American English: Smith 1997) was present but less substantial than in SSE, so was not noticed. The traditionally studied acoustic parameters of phrase-final fricative /voice/ in English, such as voicing offset, vowel and consonantal duration (Haggard 1978; Docherty 1992; Smith 1997) also play a role in SSE, and contribute to the massive 92.2% correct LDA classification of /voice/. However, the contribution of these traditional variables in SSE phrase-final fricatives is less important than the mid-frequency transitional aspiration reflected in band-pass filtered ZCR.

It is interesting to note that while the high % correct classification for /voice/ nearly approached a one-to-one mapping of these multidimensional phonetic correlates, the correlation strength of individual predictors in the ranking was relatively low (despite the significance of the correlates in MANOVA). This shows that underlying fricative /voice/ contrast is controlled by a multitude of facilitating phonetic correlates with no one-to-one correspondences, including the important ones located prior to the onset of supralaryngeal constriction already in the preceding vowel.

The extensive relocation of SSE fricative /voice/ into the preceding vowel could be classified as "contact dissimilation" in Ohala's (1981; 1993) terms, defined as "the loss or change of one or more features, including whole segments when the same feature is distinctive on another cite of the word" (1993: 249). However, Ohala treats the feature /voice/ as being non-susceptible to dissimilation, because "… the primary cue to a segment being voiced is the generally robust cue of periodic pulsation in the lower frequencies. This cue operates in a relatively short time window and does not manifest itself by colouration of adjacent segments…" (Ohala 1993: 254). By increasing the scope of analysis to the whole VF-rhyme, we have evidence from this study to concur this latter statement with a note that this study only accounts for word-final fricatives in phrase-final contexts. Although the possibility of such systematic relocation and functioning in SSE linguistic system is noteworthy and deserves a study of potential sound change in Scottish English.

This study, therefore, supports the abstractness and non-neutralising nature of phonological /voice/ in English in general, such that it may be reflected in a number of acoustic correlates with no one-to-one mappings. In this multidimensional acoustic space, the various correlates can dynamically adjust and change in importance depending on the phonetic structures and prosodic contexts involved, sometimes crossing the boundaries that would separate categories under traditional transcriptional analysis, and sometimes varying within-category.

We can only hypothesize at this point about the perceptual relevance of this preaspiration as a cue to /voice/. Native language learners' attention to VF-transitions may be mediated by the completeness of phonological neutralisation of /voice/ in word-final fricatives (in a way similar to perception of /VhC/ phonotactics in Turkish reported in Mielke 2003; or language-specific sensitivity to the acoustics of FV transitions reported in Wagner, Ernestus, and Cutler 2006). Some British English varieties have partial phonetic devoicing of pre-pausal word-final /+voice/ obstruents, which is phonetically gradual without neutralising the phonological contrast (Docherty 1992). If there is pre-pausal devoicing without neutralisation, the /voice/ contrast must be maintained by other phonetic correlates. Therefore, in a pre-pausal context where important acoustic cues such as voicing offset and duration are demoted, transitional cues like preaspiration (large timing dissociation between supralaryngeal and laryngeal gestures and/or wide laryngeal abduction) may become promoted as more important in that specific context. In languages like e.g. Russian, Dutch or German where phonological neutralisation of /voice/ in phrase-final contexts has diachronically become complete, this diachronic process did not occur or was not maintained. From a synchronic perspective, there is no (need for) such promotion of alternative transitional cues, and there are thus no reports of preaspiration for these languages. Similarly, in

English varieties where final /+voice/ obstruents are not strongly devoiced phonetically, preaspiration for /-voice/ obstruents is probably less likely.

Gestural dissociation as conditioned by prosodic context would, therefore, vary dialectally as a function of the phonological system, and the preservation of the /voice/ contrast in turn may be dependent on socially mediated spread of the functional dissociation of laryngeal and supralaryngeal information. The following section elaborates more on individual variation patterns apparent from this study.

4.3. Interspeaker differences and multidimensional /voice/ correlates

There seem to be different relationships in multidimensional phonetic correlates of fricative /voice/ amongst the speakers in this study. While the SSE speakers clearly share the same phonology in terms of maintaining /voice/ word-finally, different speakers rely on different correlates of /voice/ available in their inventory. For example, the /voice/ contrast of speaker M5 is mainly cued by the aspiration-related parameters. While most other speakers also employed voicing offset, and segmental duration on top of primary aspiration-related ZCR (see Figure 6), speaker M5 did not employ durational characteristics as a correlate of /voice/.

The relationship between the multidimensional phonetic characteristics in this speaker is exemplified in Figures 7 and 8. The token of “bus” and “buzz” are produced by M5 in the same prosodic context and at quite similar speech rate. The duration of the vowel [ʌ] in this instance of “bus” is longer than that in “buzz”. Unlike for most other speakers, this relationship is mirrored in Speaker M5’s durational means (for both vowels and fricatives in Figure 6). The consonantal duration differs by only 13 ms, which is not found to be a perceptually relevant limen for fricative sounds (Jongman 1989). While both /±voice/ fricatives are phonetically voiceless nearly throughout their durations, the voicing offset in time-derived domain of the electroglottographic (EGG: see note 5) signal (Pane C in Figures 7 and 8) differs only by 15 ms between the two instances. The only substantial difference between the instances in fricative /voice/ appears on Pane A in the sound spectrogram: i.e. the 64 ms long mid-frequency aspiration noise in the vowel-voiceless fricative transition on Pane A in Figure 7, while the phonetic voicing persists, and there is no such abrupt interruption by noise in the VF-transition of “buzz” on Pane C in Figure 8.

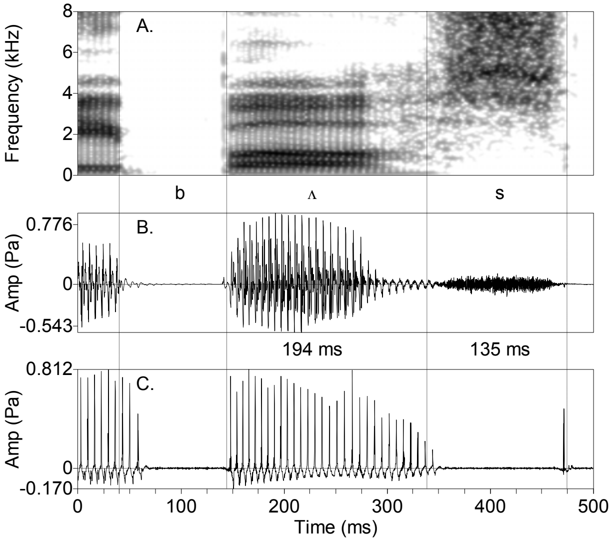


Figure 7. An example of /-voice/ fricative for the subject M5, producing "bus". Pane A represents the spectrogram; Pane B represents the acoustic waveform; Pane C represents the time derivative of EGG lx waveform.

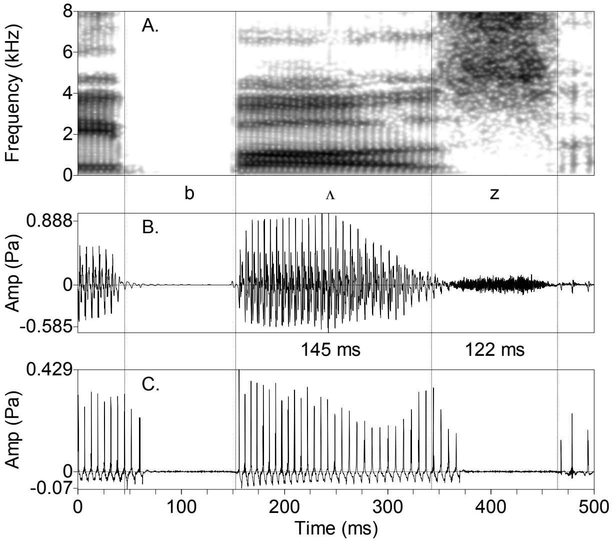


Figure 8. An example of /+voice/ fricative for the subject M5, producing "buzz". Pane A represents the spectrogram; Pane B represents acoustic wave; Pane C represents the time derivative of EGG lx waveform.

Figures 5 and 6 also show that for a number of the acoustic parameters, the range of phonetic space used by the speakers as a whole for /voice/ overlaps with the range used for preaspiration, so that a particular point on the range might encode either value of /voice/ in a way which is speaker-dependent. Thus the meaning of the feature /voice/ is in some regards arbitrary, gradient, and non-universal phonetically.

Nevertheless, the speakers are remarkably consistent in the ways in which they offset their own individual acoustic properties of /±voice/, as shown the parallel nature of the lines in Figure 6. Therefore, the systematic relationship between /s/ and /z/ remains relatively constant, and non-arbitrary, even if in transcriptional terms we would have to represent it as ranging from [ʰs] vs. [s] to [z] vs. [z̥].

Such patterns resemble those shown for postaspiration in Shetlandic English (Scobbie 2006), where the individual VOT patterns for /ptk/ were shown to function indexically ranging gradiently from short to long lag, and where there seemed to be a functional preservation of contrast between /p/ and /b/, such that speakers with increasingly shorter-lag /p/ tended to have more prevoiced /b/.

5. Conclusion

Our major descriptive finding is that glottal aperiodic energy systematically accompanies /-voice/ fricative single segment codas in SSE, appearing before the fricative during the vowel. We conclude that its extent is sufficiently large (at times as long as the fully voiced vowel) to merit the use of the phonetic label ‘preaspiration’. Similar conclusions regarding the theoretical status of such transitions have been drawn by Helgason (2002) for Central Standard Swedish and Jones and Llamas (2003) for Middlesbrough variety of British English. However, our conclusions are supported by a wider range of acoustic measures than previous studies, new measures which themselves have been evaluated statistically for their ability to capture the phonological contrast in /voice/ and our basic durational annotation of preaspiration. These parameters related to voicing offset, segmental duration and voice quality (aspiration). One of these voice quality parameters, a band-pass filtered zero-crossing rate (ZCR), looked at the timing differences in the relative onset of a major increase in zero-crossing rate in VF-transitions. This measure reflects aspiration noise present in mid/high spectral frequencies in the vowel, and is insensitive to the presence or absence of phonation. ZCR change was found to be a more consistent predictor of both phonetic and phonological analytic categories than the more traditional measures that are used to examine the phonetic correlates of phrase-final fricative /voice/.

From a typological point of view, it is noteworthy that the system of preaspiration in these SSE speakers is asymmetrical, in that there is no preaspiration before final /-voice/ stops, where glottalisation seems to be preferred (Gordeeva 2008). This shows that preaspiration of voiceless fricatives is not a characteristic that is coupled to preaspiration in stops in a variety. At a more abstract phonological level they may be both expressing 'voicelessness', if such a generalisation is relevant, but aspiration and glottalisation are very different mechanisms. Though it may be that aspiration accompanying fricatives and glottalisation accompanying stops is explicable due to coarticulation in airflow, in common with other phonetic explanations, there is clearly no *deterministic* low-level cause.

There is still sufficient speaker variation, and contextual variation that we would not wish to claim that preaspiration itself is a normative phenomenon, but preaspiration of voiceless fricatives is nevertheless important for understanding unequivocally normative aspects of the Scottish English sound system. It is unclear why preaspiration has not been reported previously in anything but a sporadic fashion.

The nature of the contrast between words like *bus* and *buzz* is normatively expressed via a *shifting set* of articulatory and acoustic parameters. Variation in phonetics and phonology is inescapable, and though there are speakers and contexts in which preaspiration does not figure, its appearance is apparently not random. Our assumption is that the presence of preaspiration is related to the preservation of contrast in a system with a tendency for final devoicing.

What is also normative is the relative *way* in which different speakers' /voice/ opposition is maintained in final fricatives. As we have shown, different speakers display comparable jumps in how the contrasting fricatives are distinguished along a number of acoustic parameters. Thus the function of preaspiration may be to maintain or enhance (depending on your diachronic point of view) a long-standing phonological contrast. Though not all speakers are using preaspiration, it is a clearly observable yet previously unreported aspect of the Scottish English sound system, a situation which raises the exciting possibility that other theoretically important phenomena are either evolving rapidly, or in hiding, waiting to be investigated.

Acknowledgements

We would like to thank the ESRC (UK) (under grants PTA-026-27-0368 and R000-22-2032) and Queen Margaret University (Edinburgh, Scotland) for funding this research. We express our special thanks to Gerard Docherty and the anonymous reviewers for suggestions and comments on the earlier drafts of this work, as well as to Steven Cowen for continuous technical support through the various stages of this study.

Notes

1. The continuing influence of notation used in “The Sound Pattern of English” (Chomsky and Halle 1968) has meant that distinctive features are traditionally notated with square brackets thus: [+/- voice]. This is because features were bundled into matrices, represented notationally as large vertical columns of text grouped by such brackets. In discussion of single features, in diagrams and in text, the notation was applied consistently. In segmental transcription, on the other hand, square brackets are reserved for phonetic representations and sloping ones for phonological representations. The SPE notation has the effect of making square brackets ambiguous, something that is deeply problematic. However, since the demise of feature bundles in favour of feature geometry (e.g. Sagey 1986), there has been no rationale to maintain square brackets as a notational convention. The solution of Docherty (1992) was to use capitalisation for single phonological features. We prefer to use sloping phoneme-style brackets, to keep the abstract phonological status of features clear and unambiguous. This easily interpretable usage means that /+voice/ is a phonological feature just as /b/ is a phonological segment. As is widely understood, the typographical label within such brackets does not imply anything definitive about articulation or acoustics, but is an abstraction: one of the main functions of phonological features in the first place.
2. The materials enable a full comparison of word-final voiced and voiceless fricatives, produced in a controlled range of prosodic contexts rather than a single one, but were not specifically designed as a homogenous whole to study prosodic conditioning of preaspiration across those contexts.
3. For example, -90% means that the whole vowel and the following 90% of the fricative is voiced, while 20% means that final 20% of the vowel and the following fricative completely lack periodicity.
4. The pooled within groups correlation matrix uses sums of squared Pearson correlations and cross-product deviations from the mean of *each* variable (group), summing these sums *across* groups.

5. Electroglottogrphic (EGG) data and analyses were available for the male subjects in this study.

References

Agutter, Alex

1988 The non-so-Scottish Vowel Length Rule. In *Edinburgh Studies in the English Language,* John M. Anderson and Norman MacLeod (eds.), 120-132. Edinburgh: John Donald Publishers.

Balise, Raymond R. and Randy L. Diehl

1994 Some distributional facts about fricatives and a perceptual explanation. *Phonetica* 51: 99–110.

Bladon, Anthony

1986 Phonetics for hearers. In *Language for Hearers,* G. McGregor (ed.) 1–24. Oxford, UK: Pergamon.

Boersma, Paul and David Weenink

2006 *PRAAT, a System for Doing Phonetics by Computer.* www.praat.org, version 4.3.04.

Bombien, Lasse

2006 Voicing alterations in Icelandic – a photoglottographic and acoustic investigation investigation. In *Arbeitsberichte des Instituts für Phonetik der Universität Kiel.* Jonathan M. Harrington, Christine Mooshammer, F. Kleber (eds.) 37: 63–82. Kiel: Universität Kiel.

Bosch, Anna

2008 Revisiting preaspiration: Evidence from the survey of the Gaelic dialects of Scotland. In *Harvard Celtic Colloquium*; Oral Paper.

Browman, Catherine P. and Louis Goldstein

1992 Articulatory Phonology: An overview. *Phonetica* 49: 155–180.

Chirrey, Deborah

1999 Edinburgh: descriptive material. In *Urban Voices: Accent Studies in the British Isles,* Paul Foulkes and Gerard J. Docherty (eds.), 223–229. London: Arnold.

Cho, Taehong

2001 Effects of prosody on articulation in English. *UCLA Dissertation in Linguistics* 22: 1–202.

Chomsky, Noam and Morris Halle

1968 *The Sound Pattern of English*. New York: Harper & Row.

Docherty, Gerard J.

1992 *The Timing of Voicing in British English Obstruents*. Berlin (Netherlands Phonetics Archives 9): Foris.

Docherty, Gerard J. and Paul Foulkes

1999 Derby and Newcastle: instrumental phonetics and variationist studies. In *Urban Voices: Accent Studies in the British Isles,* Paul Foulkes and Gerard Docherty (eds.), 47–71. London: Arnold.

Edwards, Jan, Mary E. Beckman and Janet Fletcher

1991 The articulatory kinematics lengthening. *Journal of the Acoustical Society of America* 89: 369–382.

Fant, Gunnar, Anita Kruckenberg and Lennart Nord

1991 Prosodic and segmental speaker variations. *Speech Communication* 10: 521–531.

Gobl, Christer and Ailbhe Ní Chasaide

1999 Voice source variation in the vowel as a function of consonantal context. In *Coarticulation: Theory, Data, Techniques,* William J. Hardcastle and Nigel Hewlett (eds.), 122–143. Cambridge: Cambridge University Press.

Gordeeva, Olga B.

2005 Language interaction in bilingual acquisition of sound structure: a longitudinal study of vowel quality, duration and vocal effort in pre-school children speaking Scottish English and Russian. Ph.D. diss., Edinburgh: Queen Margaret University College.

Gordeeva, Olga B.

2007 Learnability of laryngeal abduction gesture in voiceless fricatives: cross-linguistic evidence. In *Proceedings of the 16th International Congress of Phonetic Sciences*; Jürgen Trouvain and William J. Barry (eds.), Saarbrücken: Saarbrücken University.

Gordeeva, Olga B.

2008 The relative importance of laryngeal, supralaryngeal and temporal levels of speech production in the Scottish Vowel Length Rule. In *Colloquium of the British Association of Academic Phoneticians.* Oral Paper, Sheffield (UK): University of Sheffield.

Gordeeva, Olga B. and James M. Scobbie

2004 Non-normative preaspiration of voiceless fricatives in Scottish English. In *Colloquium of the British Association of Academic Phoneticians.* Oral Paper, Cambridge: University of Cambridge.

Gordeeva, Olga B. and James M. Scobbie

2006 Phonetic conditioning of word-final ejective stops in the speech of Scottish English pre-school children. In *Colloquium of the British Association of Academic Phoneticians.* Oral Paper, Edinburgh: Queen Margaret University College.

Gordeeva, Olga B. and James M. Scobbie
2007 Non-normative preaspirated voiceless fricatives in Scottish English: Phonetic and phonological characteristics. In *QMU Speech Science Research Centre Working Papers, WP*-12; James M. Scobbie, Ineke Mennen and Jocelynne Watson (eds.), 1–43, Edinburgh: Queen Margaret University.

Haggard, Mark
1978 The devoicing of voiced fricatives. *Journal of Phonetics* 6: 95–102.

Hanson, Helen M.
1997 Glottal characteristics of female speakers: acoustic correlates. *Journal of the Acoustical Society of America* 101: 466–481.

Helgason, Petúr
2002 *Preaspiration in the Nordic languages: synchronic and diachronic aspects*. Ph.D. Diss., Stockholm: Department of Linguistics, Stockholm University.

Hillenbrand, James, Ronald A. Cleveland and Robert L. Ericson
1994 Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37: 769–778.

Holmberg, Eva, Robert Hillman, Joseph Perkell, Peter Guiod and Susan Goldman
1995 Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research* 38: 1212–1223.

Hoole, Philip
1999 Techniques for investigating laryngeal articulation. Section A: Investigation of the devoicing gesture. In *Coarticulation: Theory, Data, Technique;* William J. Hardcastle and Nigel Hewlett (eds.), 294–300, Cambridge: Cambridge University Press.

Jones, Mark J. and Carmen Llamas
2003 Fricated pre-aspirated /t/ in Middlesbrough English: an acoustic study. In *Proceedings of the 15th International Congress of Phonetic Sciences*; Maria J. Solé, Daniel Recasens and J. Romero (eds.), 655–658. Barcelona.

Jongman, Allard
1989 Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America* 85(4): 1718–1725.

Kim, Chin-Wu
1970 A theory of aspiration. *Phonetica* 21: 107–116.

Klatt, Dennis H. and Laura C. Klatt
1990 Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87(2): 820–857.

Koenig, Laura L., W. Einar Mencl and Jorge C. Lucero

2005 Multidimensional analysis of voicing offsets and onsets in female subjects. *Journal of the Acoustical Society of America* 118(4): 2535–2550.

Ladefoged, Peter and Ian Maddieson

1996 *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

Laver, John

1994 *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lisker, Leigh and Arthur S. Abramson

1964 A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20(3): 384–422.

Löfqvist, Anders

1995 Laryngeal mechanisms and interarticulator timing in voiceless consonant production. In *Producing Speech: Contemporary Issues for Katherine Safford Harris*; F. Bell-Berti and L. J. Raphael (eds.), 99–116. Woodbury, NY: AIP Press.

Löfqvist, Anders, Laura L. Koenig, and Richard S. McGowan

1995 Vocal tract aerodynamics in /aCa/ utterances: measurements. *Speech Communication* 16: 49–66.

Löfqvist, Anders and Nancy S. McGarr

1987 Laryngeal dynamics in voiceless consonant production. In *Laryngeal Function in Phonation and Respiration*; Thomas Baer, C. Sasaki and Katherine S. Harris (eds.), 391–402, Boston: College Hill.

Löfqvist, Anders and Hirohide Yoshioka

1980 Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America* 68(3): 792–801.

McKenna, Gordon

1988 *Vowel duration in the standard English of Scotland*. M.Sc. Diss., Edinburgh: University of Edinburgh.

Mielke, Jeff

2003 The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica* 60(3): 208–229.

Murphy, Peter J.

1999 Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis. *Journal of the Acoustical Society of America* 105(5): 2866–2881.

Ní Chasaide, Ailbhe

1987 Glottal control of aspiration and of voicelessness. In *Proceedings of the 11th International Conference of Phonetic Science* 6: 28–31. Tallinn, Estonia.

Ní Chasaide, Ailbhe and Christer Gobl
1993 Contextual variation of the vowel voice source as a function of adjacent consonants. *Language and Speech* 36(2): 303–330.

Ó Murchú, Máirtín
1985 Devoicing and preaspiration in varieties of Scottish Gaelic. In *Ériu* 36: 195–198.

Ohala, John J.
1981 The listener as a source of sound change. In *Papers from the Parasession on Language and Behavior*; C. S. Masek, R. A. Hendrick and M. F. Miller (eds.), 178–203. Chicago: Chicago Ling. Soc..

Ohala, John J.
1993 The phonetics of sound change. In *Historical Linguistics: Problems and Perspectives*; Charles Jones (ed.), 237–278. London: Longman.

Ohala, John J. and Maria-Josep Solé
Turbulence and Phonology. This volume.

Rowden, Chris
1992 Analysis. In *Speech Processing*; Chris Rowden (ed.), 35–73. London: McGraw-Hill Book Company.

Sagey, Elisabeth
1986 The representation of features and relations in non-linear phonology. Ph.D. Diss., Massachusetts: MIT.

Schaeffler, Felix
2005 *Phonological quantity in Swedish dialects: typological aspects, phonetic variation and diachronic change*. Ph.D. Diss., Umeå University, Sweden.

Scobbie, James M.
2006 Flexibility in the face of incompatible English VOT systems. In *Papers in Laboratory Phonology 8: Varieties of Phonological Competence*; Louis M. Goldstein, Catherine Best and Douglas H. Whalen (eds.), 367–392. Berlin: Mouton de Gruyter.

Scobbie, James M.
2007 Interface and overlap in phonetics and phonology. In *The Oxford Handbook of Linguistic Interfaces*; Gillian Ramchand and C. Reiss (eds.), 17–52, Oxford: Oxford University Press.

Scobbie, James M., Nigel Hewlett and Alice Turk
1999 Standard English in Edinburgh and Glasgow: the Scottish Vowel Length Rule revealed. In *Urban Voices: Accent Studies in the British Isles*; Paul Foulkes and Gerard J. Docherty (eds.), 230–245. London: Arnold.

Silverman, Daniel
2003 On the rarity of pre-aspirated stops. *Journal of Linguistics* 39: 575–598.

Smith, Caroline L.
1997 The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics* 25: 471–500.

Stevens, Kenneth N.
1998 *Acoustic Phonetics*. Cambridge (Massachusetts): The MIT Press.

Stuart-Smith, Jane
1999 Voice Quality in Glaswegian. In *Proceedings of the XIVth International Congress of Phonetic Sciences*; John J. Ohala, Y. Hasegawa, Ohala M, D. Granville and A. Bailey (eds.), 2553–2556. San Francisco: University of California.

Thráinsson, Höskuldur
1979 *On Complementation in Icelandic*. New York: Garland Publishing Inc..

Turk, Alice, Satsuki Nakai and Mariko Sugahara,
2006 Acoustic Segment Durations in Prosodic Research: A Practical Guide. In *Methods in Empirical Prosody Research*; S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek and N. Richter (eds.), 1–27. Berlin: Walter de Gruyter.

van Dommelen, Wim A.
1998 Production and perception of preaspiration in Norwegian. In *Proceedings of FONETIK* 98; P. Branderud and Hartmut Traunmüller (eds.), 20–23. Stockholm: Stockholm University.

van Zanten, Erik, L. Damen and E. van Houten
1991 *The ASSP Speech Database*. Utrecht: Speech Technology Foundation.

Vaux, Bert
1998 The laryngeal specification of fricatives, *Linguistic Inquiry,* 29(3): 497–511.

Wagner, Anita, Mirjam Ernestus and Anne Cutler
2006 Formant transitions in fricative identification: the role of native fricative inventory, *Journal of the Acoustical Society of America,* 120(4): 2267–2277.

Watt, Dominic and William Allen
2003 Tyneside English, *Journal of the International Phonetic Association,* 33(2): 267–271.

Wells, John
1982 *Accents of English*. Cambridge: Cambridge University Press.

Yumoto, Eiji, Wilbur J. Gould and Thomas Baer
1982 Harmonics-to-noise ratio as an index of the degree of hoarseness, *Journal of the Acoustical Society of America,* 71(6): 1544–1550.

Phonetic characteristics of ejectives – samples from Caucasian languages

Sven Grawunder, Adrian Simpson
and Madzhid Khalilov

1. Introduction

The production of ejective consonants typically "involves a complete closure of the vocal folds followed by an upward movement" (cf. Kingston 2005; Ladefoged and Maddieson 1996; Lindau 1984; Maddieson 2005). Catford (1977a) calls such "bellows-like or piston-like movement of an organ or organ-group (an initiator), which generates positive or negative pressure in the part of the vocal tract adjacent to it", initiation. We will henceforth consider egressive glottalic initiation to involve a stricture of complete oral closure followed by elevation of the larynx with the glottis closed, ending with the release of the oral then the glottal closure. Ladefoged and Maddieson (1996: 79) point to the importance of the timing of these movements. Cross-linguistic differences arise on the one hand from the variation made by individual speakers, depending among other things on speech rate and context, as well as on interindividual variation within the same speech community. Fallon (2001) describes various processes which ejectives synchronically and diachronically undergo. Such processes are certainly driven by the various systemic contexts ejectives are embedded in, but are also due to particular differences which can result in perceptual confusion, which eventually results in different discrimination patterns (Ohala 1993). According to Lindau (1984) and Kingston (2005) ejectives could be classified, based on certain acoustic features, into "stiff" (e.g. Navajo, Athabaskan) and "slack" (e.g. Hausa, Chadic) ejectives (cf. Wright, Hargus and Davis 2002). Similarly, Maddieson (2005) differentiates the glottalized stops into glottalized egressive ejective stops vs. glottalized egressive non-ejective stops. The latter ones would "lack the raising [larynx] characteristic" (e.g. Yurok, Algic) or "are produced with a closer than usual vocal fold position, although they are not otherwise reminiscent of ejectives" (Maddieson 2005: 34).

In this chapter we will describe certain acoustic and articulatory aspects of ejective consonants in a selection of Caucasian languages (from the Nakh-Daghestanian and Kartvelian family). We do not provide a comprehensive overview, rather we aim to give an idea of the cross-linguistic variation within this diverse language group, which here is more in the form of inter-speaker variation. Nonetheless, we intend to focus on aspects which we think are relevant in the description of the phonetics of ejectives in general, but crucial for those occurring in the area of the Caucasus.

2. Typology of ejectives or (synchronic) phonology of ejectives

2.1. General typology

According to larger surveys on genealogical samples (Maddieson 1984; Maddieson 2005), ejectives are not rare. Within the WALS-sample “ejective-like” consonants occur in 16.3% (92 out of 566 languages). Whereas, according to WALS, ejectives do co-occur with implosives in 13 languages (2.3%) and with glottalized resonants in 19 languages (3.3%), ejectives do occur as ‘ejectives only’ in 57 languages (10%).

However, ejectives are fairly concentrated in certain areas of the world. According to their geographical distribution we could define 5 areal clusters of languages and language groups which have ejectives in their consonant system. Two of these so-called hotbeds can be found in Africa, one in southern Africa (including mainly Khoisan) and one in east-central Africa around the Nilo-Saharan phylum. For North America mainly Athabaskan and Salishan families on the northwest coast host languages with ejectives. For Central and South America there are the Southern Cordilleres. And finally, of course, the area of the Caucasus. In other geographical areas, like Europe or Australia, ejective consonants are either absent or may occur marginally, e.g. as allophonic variants of fortis and lenis plosives in some varieties of English (Catford 1977a: 70; Gordeeva and Scobbie 2006).

It is also worth noting that the occurrence of ejectives in a language seems to coincide with moderate to large consonant inventories.

2.2. Typological aspects of ejectives in Caucasian languages

Ejectives make up a core feature of Caucasian languages (Catford 1977b; Klimov 1994). The majority make extensive use of this feature within the systems of obstruents, others do so to a smaller degree, like Georgian, but there is no Caucasian language without ejectives. Caucasian languages show relatively small vowel inventories (especially in the Adyghe-Abkhaz group of the Northern Branch, e.g. Kabardian, Circassian) and moderate (Georgian, South-Kartvelian) to relatively large consonant inventories (Ubykh, Adyghe-Abkhaz).

Prototypically we find a three-way contrast of voiced, voiceless aspirated (pulmonic) and ejective, maintained at three (labial, coronal, velar/uvular) places of articulation. Exceptions within Daghestanian languages are e.g. Lezgian and Khinalug, where we have a four-way-contrast, after 'adding' a voiceless non-aspirated stop (see Table 1).

A classical typological rule would then be formulated: ej > pl; if there is an ejective, then there is also a(n aspirated) pulmonic stop. This rule can not be reversed and would imply a general markedness of ejectives regarding the occurrence in a given system. However, in some places ejectives do not adhere strictly to this rule, and the markedness may change regarding either frequency of occurrence or the default substitution of stops in loans. First, bilabial ejectives are not attested as fully contrastive in all Caucasian languages, and second, bilabial ejectives are absent in some languages. The lateral ejectives, or better ejective lateral affricates, behave similarly.

There are only four cases of deglottalization, i.e. cases of ejectives becoming pulmonic stops, attested for Caucasian languages (three dialects of Udi; Usukh-Čai Lezgian, Laz, and Eastern Circassian, cf. Fallon 2001: 102-103). And as Catford (1977b) notes, ejectives are also to be found in non-Caucasian languages of the Caucasus and Transcaucasus, e.g. Eastern Armenian, some Ossetic languages, and Northern Kurdish dialects. These languages would probably provide the examples for change in the opposite direction for this area, i.e. a pulmonic sound becoming glottalic.

In a quick survey, mainly based on Kibrik and Kodzasov (1990) and Alekseev (1999), we compiled the number of contrasts per place of articulation throughout the Daghestanian languages (plus Ingush and Georgian). Other 'secondary' articulations like gemination (see also 4.3.), labialization, pharyngealization (see also 4.4.) and palatalization enhance the contrasts subphonemically, or have phonematic value, increasing the

Table 1. Distribution of ejectives in Daghestanian languages (cf. Alekseev 1999; Kibrik and Kodzasov 1990) (if there was more than one dialect only one was taken, indicated by "a" after the reference number). c represents /ts/ and varies between dental and alveolar; č (/tʃ/) varies between alveolar and palatal and the laterals, ƛ (/tɬ/) represents the lateral affricate. -1 indicates a case of unclear classification, mostly in the sense that the feature often appears at a subphonemic (non-contrastive) level, e.g. for Udi the authors (K&K) place unaspirated stops at the ejective position.

no.	langoid	bilabial	dental	alveolar	alveolar-palatal	lateral	velar	uvular	no. of ejectives	geminate	pharyngealized	labialized	palatalized
		p	t	c	č	ƛ	k	q					
1	Avar	3/-1	3	2	2	2	3	2	7	0	0	1	0
2	Andi	2	3	2	3	2	3	3	6	1	0	1	1
3	Akhvakh	3	3	2	3	2	3	2	7	1	0	1	0
4a	Čamalal	2	3	1/-1	2	2	3	2	6	-1	0	1	0
5	Tindi	2	3	2	2	2	3	2	6	-1	0	1	1
6	Inkhokvari	3	3	2	2	2	3	2	7	-1	1	1	0
7	Tsez	3	3	2	2	2	3	2	7	-1	1	1	0
8	Hinukh	3	3	2	2	2	3	2	7	0	0	1	0
9	Bezhta	3	3	2	2	2	3	2	7	1	-1	-1	0
10	Hunzib	3	3	2	2	2	3	2	7	1	0	-1	0
11	Lak	3	3	2	2	0	3	2	7	1	-1	1	0
12	Dargwa	3	3	2	2	0	3	2	7	1	0	-1	-1
13	Arči	3	3	2	2	2	3	2	7	1	-1	1	0
14a	Tabasaran	3	3	3	3	0	3	3	6	1	0	1	0
15a	Agul	3	3	2	2	0	3	2	6	1	1	1	0
16	Lezgi	4	4	3	3	0	4	3	6	0	-1	1	0
17	Rutul	3	3	3	3	0	3	3	6	0	0	1	0
18	Tsakhur	3	3	2	3	0	3	3	6	1	1	1	1
19	Krys	3	3	3	3	0	3	3	6	0	0	-1	0
20	Budukh	3	3	2	3	0	3	3	6	0	0	-1	-1
21	Khinalug	4	4	3	4	0	4	4	6	1	0	0	0
22	Udi	3/-1	3/-1	3/-1	3/-1	0	3/-1	2/-1	(6)	0	1	0	0
31	Ingush	3	3	3	3	0	3	2	6	1	-1	1	-1
32	Georgian	3	3	3	3	0	3	1	6	0	0	-1	0

number of possibilities of contrast at some places of articulation. Hence, we would check whether the system employs such a feature, although we need to account for cases 'under current discussion of its phonemic status' with "-1" (see Table 1). On the other hand, there are, for example, several authors suggesting affricates instead of stops at the uvular position (e.g. Isakov/Khalilov for Hinukh and Magomedova for Chalamal in Job 2004; see also Catford 1977b). Here we were treating the categorization affricate vs. non-affricate more rigorously, so all uvular stops were coded as non-affricate. Fricatives with glottalic initiation are absent in our sample.

These occur in Caucasian languages mainly within the North West Caucasian branch (e.g. Adyghé and Kabardian; cf. Alekseev 1999).

3. Data

The languages we were able to include in this study are the following (cf. Koryakov 2006: 21 for genealogical classification and numbers of speakers):

1. Georgian (Kartvelian, South; 3.4 Mio speakers)
2. Avar (Nakh-Daghestanian, Andic; 800,000 speakers)
3. Ingush (Nakh-Daghestanian, Nakh, Veynakh; 425,000 speakers)
4. Tsez (Nakh-Daghestanian, Tsezic, West; 16,000 speakers)
5. Bezhta (Nakh-Daghestanian, Tsezic, East; 9,000 speakers).
6. Lezgi (Nakh-Daghestanian, Lezgic, Proper, 450,000 speakers)

In this study data has been investigated from 4(+4) speakers of Georgian, 3 speakers of Avar, 2 speakers of Ingush, 1 speaker of Tsez, and 1 of Bezhta, and 1 of Lezgi.

Apart from Georgian (South Caucasian), this is a sample of languages of the northeast branch, which are relatively closely related. The data consists of individual recordings of one or two speakers per language of word-list and sentence material. The elicitation materials were designed to investigate particular aspects in each of the languages, and were not originally intended for comparative purposes. However, due to similarities in data acquisition and the features captured in the recordings, we consider the data to be adequate and representative for the demonstration undertaken here. All recordings were made in the sound booth of the phonetics laboratory at MPI EvA, Leipzig. For each recording a microphone track plus an electroglottographic track was digitized at a sampling rate of 44.1 kHz with 16-bit amplitude resolution.

4. Phonetic analysis of ejectives

Aiming for a description of perceptually relevant cues we assume several relational acoustic measures to be useful and meaningful for the distinction of ejectives vs. other non-ejectives within contrastive situations of a given language (see for a discussion e.g. Wright 2004). And specifically, for the description of glottalic obstruents a number of (durational) measures based on quasi-stable acoustic landmarks (Figure 1) are examined: Along with the total duration (TD) of the stop and also the duration of the oral closure – closure time (CT) – (Gordon and Applebaum 2006; Lindau 1984; McDonough and Ladefoged 1993; Warner 1996; Wysocki 2004), the voice onset time (VOT) – release to voicing onset of a following vowel – (Catford 1977a; Hogan 1976; Ingram and Rigsby 1987; Lindau 1984; Wright, Hargus and Davis 2002; Wysocki 2004) and the lag between the end of the burst (i.e. frication noise after release) and the voicing onset of a following vowel (Vicenik 2008) have been considered relevant.

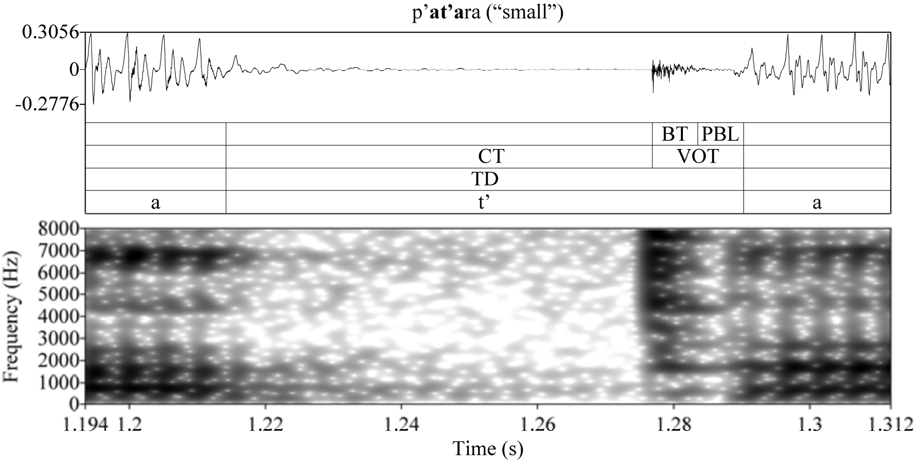


Figure 1. Signal labeling on the bases of landmarks of the Georgian word /p'at'ara/: TD = total duration, CT = closure time (closure duration), BT = burst time (burst duration), PBL = post burst lag (= burst voicing lag), VOT = voice onset time.

The instant of the oral release provides a very stable landmark in the acoustic signal as a sudden burst after a period of relative silence, which serves as an end point of the oral closure on the one hand and as a starting point of the release noise phase and of the voice onset time (VOT) on the other hand. The voice onset is usually defined by the first positive zero crossing of the first vocal period. However, the beginning of the oral closure is more ambiguous, since voicing often proceeds into the closure, and is therefore identified on the basis of abrupt spectral changes around the second formant of the preceding vowel (cf. Turk, Nakai and Sugahara 2006: 2) or an amplitude drop along with a change of the period structure in the wave form. Similarly, the end of the actual burst or pulse (cf. Stevens 1998: 347) in the stop's release and its adjacent frication, as initial parts of the release noise, are sometimes hard to distinguish from a following aspiration noise and can therefore only be estimated by means of the amplitude. A phase of relative silence or low energy following the burst and preceeding the voice onset of the adjacent vowel or preceeding the onset of another consonant is described very often as the post-burst lag (PBL) or burst-voicing lag (BVL) (Warner 1996; Wright, Hargus and Davis 2002).

Additionally, burst intensity (Vicenik 2008; Warner 1996), burst spectrum (McDonough and Wood 2008) and the slope to the amplitude maximum of burst frication (Vicenik 2008) can be considered to be useful, whereas here the term 'burst' comprises the actual burst or pulse and its following frication. Those measures serve especially to disambiguate components, such as additional lip rounding, or in cases where details of the end of the burst, the laryngealized post-burst phase and voice onset are unclear (Kong and Beckman 2006; Wright, Hargus and Davis 2002). Creaky voice and other forms of non-modality (aperiodicity) may be present in vocalic portions preceding and following an ejective consonant (Warner 1996). Especially the following creaky portion can be quite problematic for the interpretation of VOT and PBL, since the laryngealization starts early (in our material ca. 15 ms before the voice onset within the usual PBL) but it becomes quasi-periodic after only 2-3 pulses. In other words voice onset overlaps the glottalic closure phase. Thus we also explore assessments of voice quality in terms of H1–H2 and HNR (harmonics-to-noise ratio) measures.

Since these landmarks and their derived parameters are used repeatedly, the following sections concentrate primarily on particular features, that we consider to be important, when looking at ejectives in the Caucasus. First of all, there is the manner of articulation, or more specifically phonation type and initiation. Similar to, e.g. Athabaskan languages, there is in the majority of Caucasian languages a voiced series of stops (and sometimes also affricates). The actual phonetics of this contrast is so far not well documented. Whereas the findings of McDonough and Wood (2008) for Athabaskan reveal a laryngeal contrast in plain stops and affricates of unaspirated (“voiced”) and aspirated (“unvoiced”) non-ejectives, the Caucasiological literature suggests true voicing, so that we would find a 3-way distinction in terms of manner employing the features [+voiced, –checked (“closed glottis, glottalized”)] for the voiced, [–voiced, –checked] for the voiceless non-ejective and [–voiced, +checked] for the ejective stops. Here the row of voiceless and non-glottalic stops remains fairly unspecified and one needs to take into account that these stops can be observed as having aspiration. And what if the voiced stops become voiceless in certain positions? Could these devoiced ‘voiced’ and ejectives then better described as being both [–aspirated]? Would this hold true for all languages in the Caucasus area? The true phonetic nature of these contrasts has still to be described and explored in order to be finally tested.

Furthermore, there are additional articulatory features which supposedly enlarge the number of distinctions within a system of obstruents and which we would like to address here. One of these cross-linguistically more frequent secondary articulations is labialization, and in a few Caucasian languages also pharyngealization and palatalization. The latter one we will leave out here, since it occurs only in a small number of Northwest-Caucasian languages, and there is a clear need of data in order to say more about it.

4.1. Pulmonic vs. glottalic initiation as voicing contrast – a sample from Georgian

Georgian is perhaps the most well-known of the Caucasian languages. Although the consonant and vowel systems themselves are not particularly large in comparison with other Caucasian languages, the cumulation of morphemes preceding a verb stem can give rise to consonant sequences containing up to 8 elements. We will not address questions of consonantal clusters, as these have been analysed from both a phonetic and phonological perspective in other studies (Chitoran 2002; Ritter 2006; Wysocki 2004).

Together, pulmonic and glottalic initiation are used to maintain a three-term stop system at labial (/b, p, p’/), dento-alveolar (/d, t, t’, ʣ, ʦ, ʦ’, ʤ, ʧ, ʧ’/) and velar places (/g, k, k’/). At the uvular place (/q’, χ/), an ejective term is expounded by a range of phonetic variation from a dorso-uvular plosive ejective stop through to glottal closure without any oral stricture of complete closure or close approximation (Shosted and Chikovani 2006).

The data were elicited in a single recording session (four repetitions) from one 34-year-old male speaker of Georgian from Zestafoni. Observations are based on a small selection of di- and trisyllabic nouns, verbs and adjectives read as words in isolation as well as in the sentence frame glossed literally as "The child on the paper *ITEM* wrote."

Figure 2 contains spectrograms of tokens of the words /k'alati/ "basket" and /kalaki/ "town", representing examples of word-initial ejective and voiceless aspirated plosives, respectively. Symbols are centered approximately over the consonantal and vocalic portions. In both cases tokens have been extracted from the context of the sentential frame. Table 2 summarizes stop closure durations and the duration of the interval between plosive release and the onset of voicing (VOT) for a selection of items with glottalic and pulmonic stops in different structural positions. Each value is a mean of five tokens.

As can be seen from Table 2, VOT for both glottalic and pulmonic plosives is longer in isolated productions than in tokens embedded in the sentential frame. A comparison of VOT for pulmonic and glottalic plosives in Table 2, as well as visual inspection of the initial plosives in the tokens in Figure 2, show that the VOT for ejectives is significantly shorter than it is for aspirated plosives. VOT in the ejectives ranges from a mean of 25 ms for the dental ejective in /t'it'ini/ "baby talk" in the sentential context to a mean of 83 ms for the initial dorsal in /k'araki/ "butter" read from the word list. In other words, one speaker of Georgian producing plosives at two lingual places of articulation in different linguistic activities almost completely exhausts the durational range reported by Catford (1977a: 69) for a number of Caucasian languages. These values are also in line with those reported by Wysocki (2004) and Vicenik (2008).

By contrast, VOT – chiefly aspiration – for pulmonic plosives is consistently longer than for the ejectives. The aspiration phase ranges from a mean of 39 ms for the dental plosive in the word list tokens of /katami/ "chicken" to 119 ms following the release of the dorsal plosive in /kalaki/ "town". Not only is there a long aspiration phase, it is also a consistent feature across different places in the word, often being produced at the expense of vocalic voicing. The period of aspiration following the release of the dorsal plosive in the token of the word /kalaki/ "town" in Figure 2 is a representative example of this. The aspiration phase is more than 70 ms long. More striking than the duration of the aspiration phase itself, however, is the proportion of the syllable following stop release that it takes up. On average, the aspiration in syllables of this type (see Table 2) takes up two thirds of the duration of this stretch.

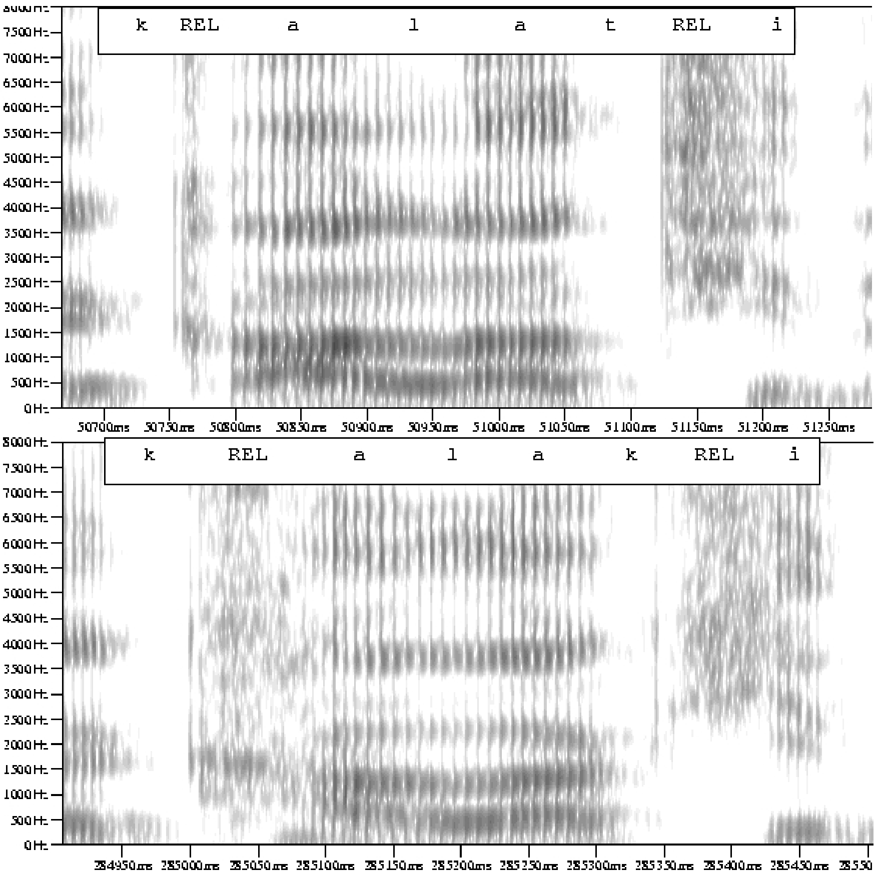


Figure 2. Tokens of the words /k’alati/ and /kalaki/ excised from the sentential frame. Symbols are approximately centered over the relevant portions. REL refers to the release phase prior to the onset of regular voicing.

The three-term plosive system found at different places of articulation, e.g. /g, k, k’/ suggests one term which is voiced and two lacking voice. The acoustic record suggests another picture. The plosives in the sentence frame sample analysed here are all in an intervocalic context. As is evident from the low frequency energy which can be seen in the spectrograms in Figure 2, voicing is present well into the closure of the each of the plosives. This is most striking for both the aspirated plosives in /kalaki/ shown in Figure 2. Here voicing would appear to cease just prior to plosive release. Furthermore, energy visible at frequencies above the fundamental in the initial portion of the stop stricture suggests incomplete oral closure, or possibly incomplete velic closure giving rise to nasality.

Table 2. Durations of stop closure, interval between release and voicing onset (VOT) and duration of vowel following plosive (Cn). Values from isolated word productions are in italic. Values are means of five repetitions.

	Closure C1	VOT C1	Duration V1	Closure C2	VOT C2	Duration V2
/kalaki/	53	75	94	50	76	41
		92	*121*	*67*	*119*	*142*
/katami/	53	59	90	54	39	74
		86	*115*	*71*	*42*	*143*
/k’alami/	58	47	101			
		70	*135*			
/k’alati/	48	45	99	58	67	43
		75	*135*	*81*	*91*	*144*
/k’amati/	53	40	106	55	60	45
		69	*139*	*76*	*99*	*142*
/k’araki/	49	45	122	53	73	40
		84	*153*	*62*	*116*	*132*
/t’it’ini/	78	34	69	70	25	84
		47	*106*	*84*	*28*	*114*
/arak’i/				67	37	53
				80	*45*	*174*
/ʃiraki/				57	77	38
				59	*117*	*118*

The voicing during these stop closures is remarkable, since it is too long to be a few remaining cycles from the preceding vowel. Rather it would seem that in the absence of an active glottal gesture either opening the glottis for voicelessness and aspiration or closing for glottalic initiation, the glottis remains configured for voice. This implies that the phonetic correlates distinguishing the different terms in the plosive system are concentrated on the release and post-release phase. It is worth speculating that the longer (in comparison with languages such as English or German) aspiration phase is not necessarily the result of a longer glottal opening gesture, but rather of an opening gesture synchronized later in relation to the stop closure. Examination of the spectrograms in Wysocki (2004) suggest that her speakers are producing similar patterns of voicing in stop closures, although the presence of echo in her recording prevents her from making any substantial claims about this in her own analysis (Wysocki 2004: 33).

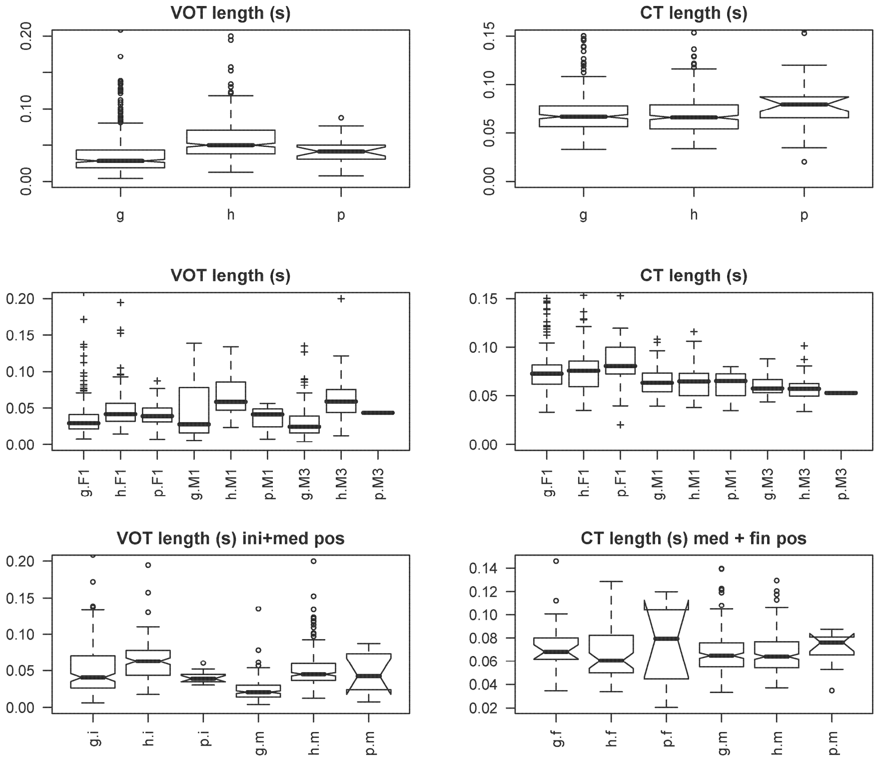


Figure 3. Voice onset times and closure time (duration) for initiation manners only (top row; g = (glottalic) ejective, h = aspirated, p = (plain) voiceless unaspirated), grouped by subjects (middle row; F1, M1, M3) and grouped by position (bottom row; i = initial, m = medial); N = 825.

By contrast, in a recent investigation of acoustic data from 6 female Georgian speakers, Vicenik (2008) focussed this phenomena again and found "a main effect of manner on closure voicing at each place of articulation" (bilabial, alveolar and velar). In phrase intial stop closures only the phonemically voiced stops with positive VOT were taken and showed on average 50-75% of voicing in the closure, whereas aspirated stops showed only 17% (cf. Vicenik 2008: 10).

Two further observations on our data seem to underline relative insignificance of the stop closure itself. First, there was no evidence to suggest any qualitative difference in the voicing found prior to or during the stop closure, i.e. there was no evidence of creaky phonation before or during the closure phase of an ejective. Secondly, there were no significant durational differences in the stop closure either for voiced, aspirated and ejective plosives at the same place of articulation or for plosives at different places of articulation. Wysocki's (2004) durational measurements of stop closures in different contexts (initial, intervocalic, etc.) for five speakers support this finding, and also Vicenik (2008: 9) concludes the same for his data.

Nonetheless we wanted to test our observations again and gathered recordings for an additional study, the previous Georgian speaker from Zestofani and three other speakers (two male, one female) from Tbilisi. The speakers (all aged between 35 and 50) produced items from an additional wordlist in the pattern of a word in isolation followed by the word within a phrase: “ITEM. me vt^{h}k^{h}vi sit’q’va ITEM ara-ert^{h}xel (“I said the word ITEM.”). For each speaker a sample of at least 10 tokens per position (initial, medial, final) and initiation manner was labeled. As found in the other studies, the phonemically voiced stops in Georgian are in the majority (>80%), realized as voiceless unaspirated pulmonic stops in initial position (cf. Vicenik 2008; Wysocki 2004). The truly voiced realizations with negative VOT were dropped for this analysis.

As expected, statistical analysis of closure duration reveals no significant differences (Figure 3). However, the voice onset time differs significantly between ejectives and aspirated stops ($t(380, 395) = -13.57$, $p < 0.001$) and between ejectives and voiceless unaspirated stops ($t(395, 50) = -3.77$, $p < 0.001$), but not between the two pulmonic stops.

Other aspects of differences between the aspirated and ejective plosives are of general phonetic interest. As might be expected, the phonetic correlates of plosives are temporally more extensive in isolated word productions than in a sentential context. Catford (1977a: 69) reports a large variation in the duration of the interval between plosive release and the onset of voicing for ejectives in Caucasian languages, ranging from 12 ms in Abkhaz to around 100 ms in Avar and the Bzhedukh dialect of Adyghe. In his study of Georgian stops, Vicenik (2008) addresses the relationship between duration and prosodic position position by comparing stops at the beginning of an intonational phase (IP), an accentual phrase (AP) and in word medial position. Such differences had also been described for VOT values by Wysocki (2004). In particular, ejectives showed a strong decrease in VOT from word initial to word medial position. In fact Vicenik (2008) found that certain higher prosodic positions (AP initial vs. word medial position) exhibited significantly higher CT values. But he only found significant differences of the same kind for VOT values.

Given these results one might also ask how sensitive these measures are to differences in speaking rate. Our recording of the female Georgian speaker involves three different speaking rates, going from careful slow to moderately fast in the relation 1.5:1.2:1 (slow:moderate:fast). In our sample there are significant negative correlations of VOT with increased speaking rate in all positions. However, it is probably more insightful to look at the different stop initiation types individually.

The ejectives in our sample show a correlation with the indicated speaking rate steps in terms of Spearman rank correlation coefficient (Spearman 1987) of ρ [rho] = 0.517 for CT, ρ = 0.179 for BT, ρ = 0.335 for PBL and ρ = 0.261 for VOT. Likewise, for the voiceless aspirated stops we find for CT ρ = 0.457 and ρ = 0.367 for VOT. The correlations suggest that CT as well as well PBL function as a kind of "buffer" phase. In fact, in cases of fast connected speech it is often hard to distinguish ejectives from voiceless pulmonic stops, since these usually show a low intensity burst, they lack a clear PBL and have a short VOT. This VOT overlap of ejective stops and voiceless (realized voiced) pulmonic stops can be observed especially in final and medial position. Phonetically (based on VOT) we have a four way split in the stops in Georgian, which at least suggests a need for other cues maintaining the 3-way contrast (Figure 4).

The phonetic correlates of uvular ejectives in Georgian range from a canonical uvular ejective plosive [q'] through an ejective fricative [χ'] to a glottal stop [ʔ]. For our speakers the initial position seems to be the most salient, since here only a small portion (approximately 10%) is 'deglottalized', whereas for the other positions this relation is reversed.

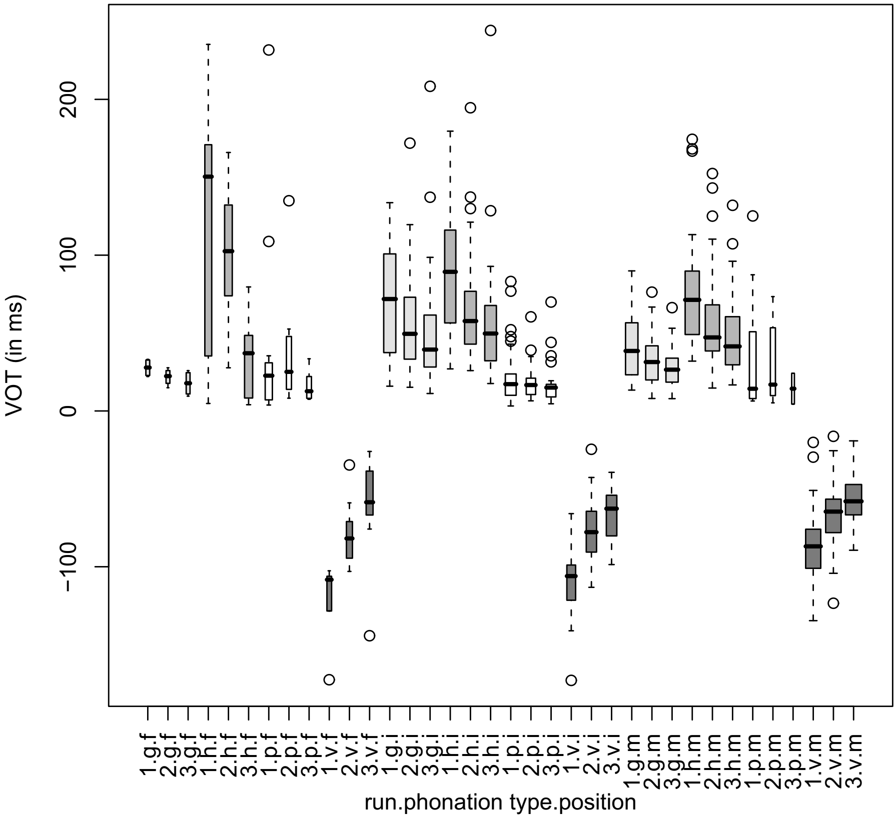


Figure 4. VOT measures (N = 1010) for three wordlist reading runs of increasing speaking rate produced by one female speaker of Georgian; all obstruents are grouped together; the labels in the x-axis indicated the run/rate (1-slow, 2-moderate, 3-fast), the phonation type (g – glottalic, h – aspirated, p – plain (as category for voiceless realized voiced stops), v – (true) voiced) and the position (f – final, i – initial, m – medial (intervocalic)).

4.2. The ejective vs. aspirated contrast in bursts – a sample from Ingush

Warner (1996) had already suggested that a possible difference in oral pressure of ejective and pulmonic voiceless stops should be found in burst intensity measures. In fact she found significantly smaller power values in ejectives for her data from Ingush.

Ingush is spoken mainly in the Republic of Ingushetia and also in parts of Daghestan. Our two speakers (one 52-year-old male, one 43-year-old female) are both from the capital Magas. A word list of 150 items was read in isolation and in a carrier phrase (ITEM *Az (xoga)* ITEM *al eandar. "ITEM. I said ITEM."*). For each speaker, 10–16 items per place of articulation (P, T, K, Q), including 5–8 per initiation manner (ejec, asp.) were analyzed with respect to VOT, zero-crossing rate, mean intensity in a window 20 ms after oral release targeting burst intensity, and the mean intensity of the last third of the VOT targeting the post-burst lag intensity. While the female speaker exhibits (particularly in the isolated context) the same PBL as in the other language samples, the male speaker has a rather noisy post-burst phase after short 'core' bursts of varying intensities with a subsequent (mostly) creaky voice onset of the vowel (Figure 5).

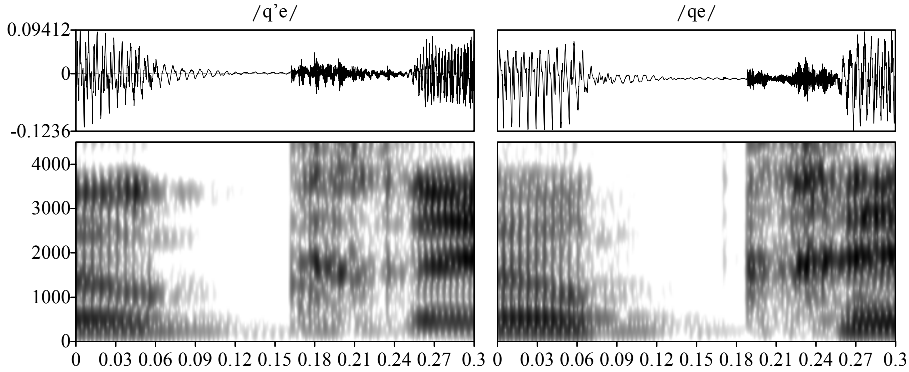


Figure 5. Ingush /q'e/ "impoverished" (VOT = 90 ms, rising pitch) and /qe/ "bean" (VOT = 70 ms, rising pitch); both sequences of 300 ms have the same left context /…œg / and the same vowel /e/ in the right context. Note that in both samples the preceeding voiced velar stop is unreleased.

CT and BT do not differ significantly between glottalic and pulmonic initiation, neither as a whole class, nor for individual places of articulation.

Based on pairwise comparisons using t-tests with pooled standard deviations (SD) and false discovery rate (FDR) p-value adjustment method for type I errors – recently described by Verhoeven, Simonsen and McIntyre (2005) and available in the R-package (http://www.R-project.org/) – we investigated the individual parameters as z-standardized pooled data per token. In Figure 6 the data is plotted pooled per token over subjects. We find only significant differences for mean intensity values of the last third of the VOT and here for /k'/ < /kh/ ($p = 0.03$), /t'/ < /th/ ($p = 0.05$) and /q'/ < /qh/ ($p < 0.001$). Only for bilabials does it seem to be reversed, but this difference is also non-significant, though the general tendency remains and supports the observation of a post-burst low energy phase in glottalic stops (Figure 6).

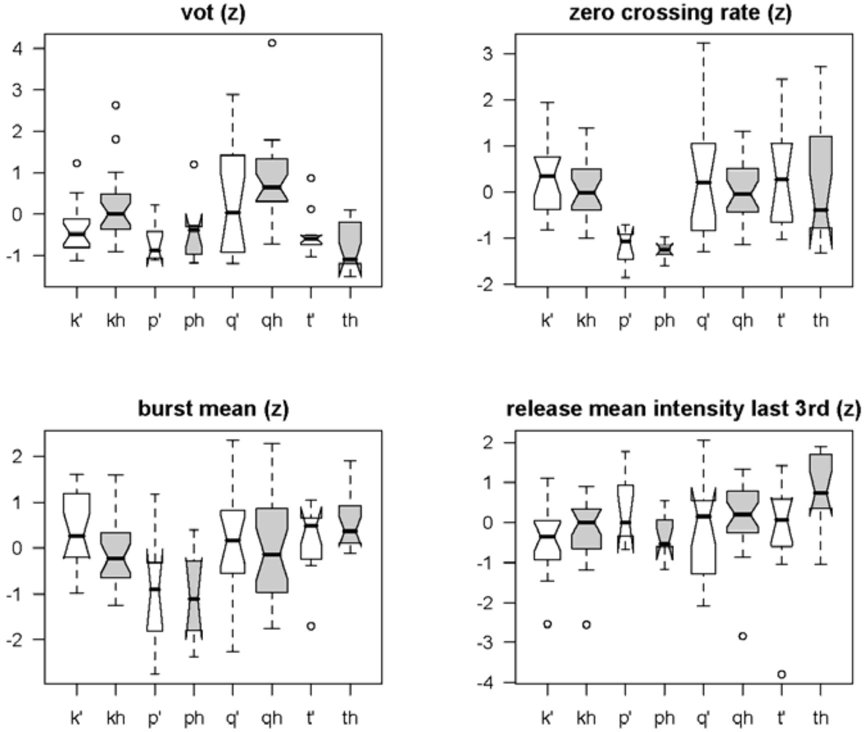


Figure 6. Burst and post-burst intensity measures (z-standardized) of two speakers of Ingush (5-8 tokens per type each); burst mean intensity (lower left) is measured 30 ms after release and post-burst lag intensity (lower right) is measured in the last third before voice onset; zero-crossing rate is calculated in the same window as burst mean.

This tendency in the order of burst intensities would also hold for our recordings of 3 Georgian speakers (see above), especially for initial and final position. We find significant differences between the initiation manners cumulated for all positions using the pairwise t-test comparison pooled SD and FDR p-value adjustment: ejec > asp (p < 0.001), ejec > unasp (p = 0.001) but not for asp > unasp (p = 0.112). Nonetheless, for all three positions the ‘burst intensity hierarchy’ is: ejec > asp > unasp. The burst of the (true) voiced stops had been excluded.

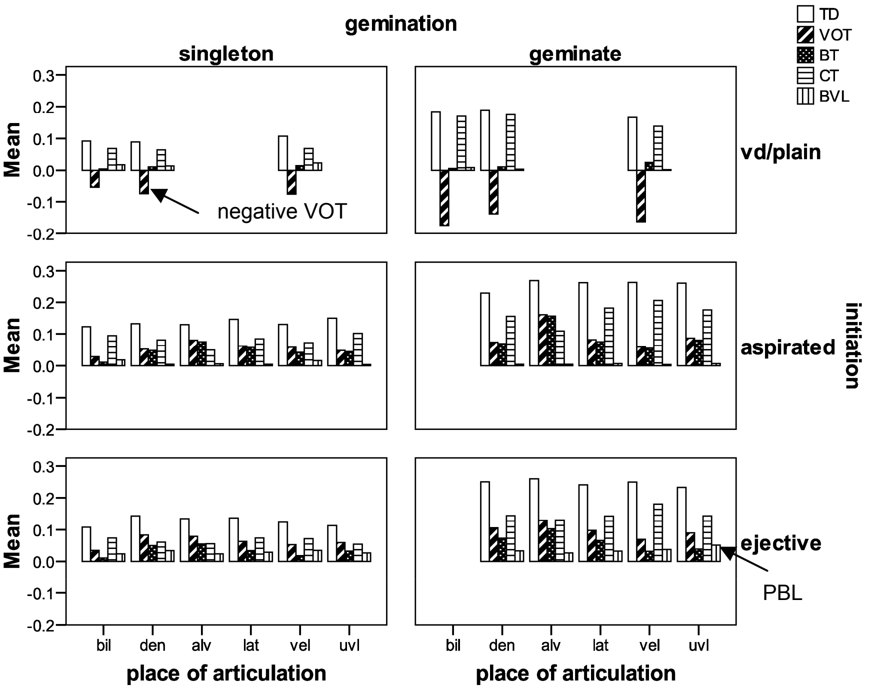


Figure 7. Durational measures (N = 516) of one speaker of Bezhta as bars representing the means; CT = closure time, VOT = voice onset time, BT = burst time (release burst+after release noise); BVL = burst-voice lag (= post-burst lag), TD = total duration; there are no geminate aspirated or glottalized bilabial stops, and no voiced alveolar, lateral or uvular obstruents in Bezhta.

4.3. Gemination – a sample from Bezhta

Geminated stops occur in a number of Dagestanian languages (see Table 1), especially in the Avar-Andi-Tsezic group. Bezhta (or Bezhtl’a, but actually [bɛʃk͡ʟ’a]) is one of them, and it also belongs to the less well described and smaller languages spoken in a few villages at the mountainous border to

Georgia in the west. Our speaker is a 56-year-old native Bezhta male from Makhachkala, who was born and raised in the village Bezhta and who still uses the language on a daily basis. Simultaneous audio and EGG recordings were made of a list of ca. 500 items in isolation as well as some 30 items in a carrier phrase (Figure 7).

Bezhta shows an extensive use of geminates at all places of articulation. The only exception to this is the absence of geminates of bilabial ejective and aspirated stops. There are, however, voiced labial geminates. In our data there is a significant difference between singletons and geminates in closure duration (F(386, 172) = 0.279, p < 0.001, t(215.892) = -28.33, p < 0.001), which also effects total duration (F(386, 172) = 0.80, p = 0.079, t(561) = -31.03, p < 0.001). The same can be stated for burst time (F(390, 173) = 167.4, p < 0.001, t(215.2) = 4.5e-14). For burst time the instant of release noise offset, corresponding to the beginning of the post-burst lag, was also labeled: duration of release noise. But post-burst lag and subsequently VOT do not differ significantly and seem only to be affected in alveolar affricate geminates [t͡ʃː t͡ʃʼː], which is reflected in the prolonged fricative portion. Geminate voiced stops, especially bilabial and velar stops, show a strong tendency for voicing to continue well into closure. This voicing remains clearly visible in the spectrogram, even in more frequent tokens like [βahag:ijo] "such" or [hugːijo] "this".

The Caucasiological tradition (Klimov 1994: 142) also refers to geminates as "strong" consonants because they have often been found to differ more in strength than in duration. This seems to be the case for Ingush (cf. Nichols 1994) (see also 4.2.) and for Avar (see 4.5.). In our Bezhta material the relation of singleton to geminate for total duration is approximately 1:2 (vd. 1: 2.2; asp. 1:2.0; ej. 1:1.9) and for closure duration ca. 1:2.3 (vd. 1:2.6; asp. 1:2.3, ej. 1:2.3).

4.4. Pharyngealization – a sample from Tsez

Pharyngealization is usually defined as a secondary articulation involving an additional constriction of the lower or upper hypo-pharynx. It is found in many Afro-Asiatic and American North-west coast languages but also within Northeast and Northwest Caucasian languages. One of them is Tsez (Dido), spoken in the Tsunta district of southern and western Daghestan. Our Tsez informant was a 34-year-old male speaker from Mokok. The recording consists of a read wordlist containing 110 items in isolation (2 repetitions).

With regard to Tsez, Maddieson et al. (1996: 102) have pointed to the long ongoing debate about the domain of pharyngealization in Caucasian languages and concluded from their analysis that "pharyngealization is a segmental feature, specifically a consonantal one".

Pharyngealized stops occur in Tsez with pulmonic and glottalic initiation, mainly as bilabial and as uvular stops in all three positions. Since for Semitic languages, especially Arabic, a merging of the features uvular and pharyngealized ('emphatic') is suggested, we should emphasize that in the majority of the Tsez lexicon the uvular place of articulation employs the pharyngealized/non-pharyngelized contrast with ejectives and pulmonics (see Figure 8 for illustration). Maddieson et al. (1996: 101) report for uvular stops a "significantly shorter closure duration than the other three places". The data from our speaker do not support this. However, regarding the distinction between glottalically and pulmonically initiated uvular stops we do find significant differences ($df = 54$, $p < 0.001$) in the means for closure time ($F = 0.03$, $t = 4.2$), burst time ($F = 2.6$, $t = 8$), and post-burst lag – using the non-parametric Mann-Whitney U test (Mann and Whitney 1947) – ($U = 17$, $p < 0.001$), but not for VOT ($t = 1.8$, $p = 0.075$). For velar stops ($df = 24$) there are significant differences for BT (including release and following noise; $F = 4.5$, $t = 6.8$, $p <0.001$), VOT ($F = 2.1$, $t = 2.2$, $p = 0.034$) and PBL ($F = 6$, $t = -2.9$, $p = 0.026$).

Although we also observe slight differences in timing for pharyngealized vs. non-pharyngealized stops, none of these differences is significant. Such differences in VOT have been considered characteristic for pharyngealized stops (cf. Laufer and Baer 1988). The only noticeable tendency is a small PBL (10 ms) in pharyngealized compared with a large PBL (80 ms) for bilabial stops in initial position, though PBL in the non-pharyngealized glottalic and pharyngealized pulmonic (aspirated) variants behave the same (60 ms). For uvular stops these differences are even smaller (Figure 8).

Regarding the burst spectra, we find a tendency for pharyngealized bursts to show raising of F1 and damping of F1 amplitude and a concentration between F2 and F3, which are unsurprisingly the characteristics of pharyngealized vowel portions (Figure 8). By comparison, for non-pharyngealized stops, energy is more evenly spread throughout the whole spectrum.

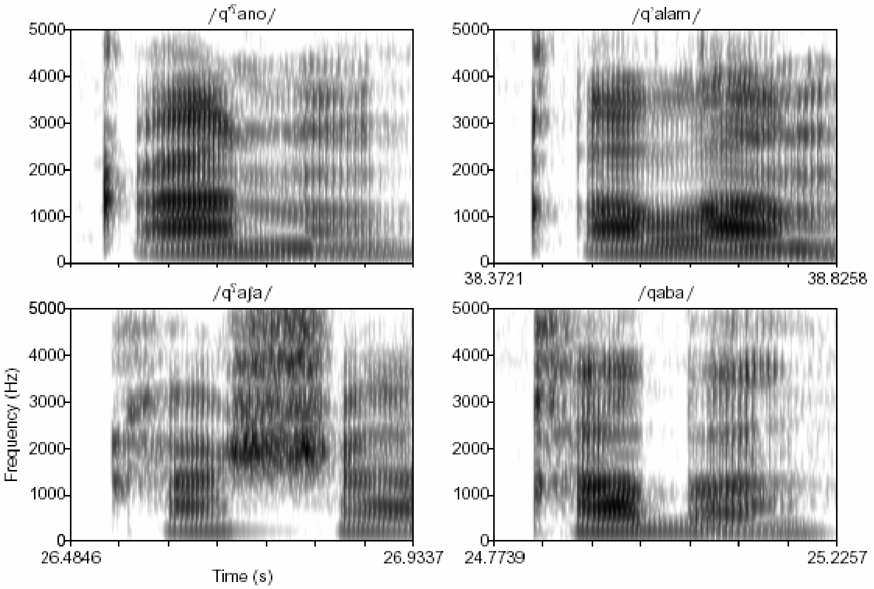


Figure 8. Pharyngealized vs. non-pharyngealized uvular stops (pulmonic vs. glottalic) in initial position in Tsez.

In order to assess possible distinctions on the basis of the acoustic characteristics of the bursts, the burst spectra have been investigated in terms of intensity measures (RMS, relative intensity) and energy distribution (COG within a range of 0-10 kHz, band energy difference, band energy density difference) and burst slope in terms of skewness and kurtosis. For the band differences a low band of 0 Hz to 500 Hz and a high band from 500 Hz to 4000 Hz have been selected in order to inspect the influence of the F1 damping. Although the burst time shows recognizable differences, e.g. in the form of shorter BT for velar and uvular ejectives, none of these measures showed significant effects.

4.5. Affricates and lateral releases – individual samples from Avar, Bezhta and Tsez

Avar, Bezhta and Tsez belong to the Avar-Ando-Tsezic group. Avar is one of the major languages of the area and serves as a lingua franca. So, speakers of Bezhta and Tsez are usually also fluent in Avar, which is taught at school as “their” native language. One feature that all 3 languages have in common is the use of voiceless laterals and lateral affricates, which also occur as ejectives. Whereas aspirated and ejective variants basically only differ in the post-burst lag, the place of the stop component ranges from alveolar to palato-velar. Note that this affricate is sometimes transcribed as [k͡ɬ] (Klimov 1994) instead of [t͡ɬ] (Maddieson, Rajabov and Sonnenschein 1996).

Analyses by other authors (Klimov 1994) or Kibrik and Kodzasov (1990: 320) who describe the lateral affricate as "apico-dorsal" suggest that this is not only an expression of interspeaker variation, but rather that this is a consistent feature of some languages, e.g. Avar and Bezhta. In fact the realisation of our (male) Bezhta speaker and one of the (female) Avar speakers should be characterized as a pre-velar affricate [kʟ] as also described for Archi (Lezgic, Nakh-Dagestanian) (cf. Ladefoged and Maddieson 1996: 206).

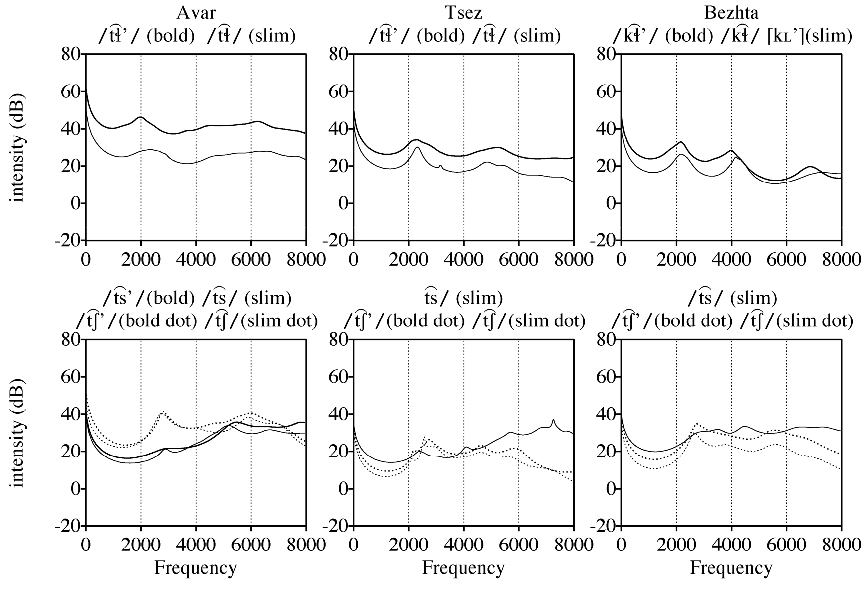


Figure 9. Averaged LPC-smoothed spectra of 10-15 bursts for lateral (top), dental and alveolar affricates (bottom), each ejective and non-ejective for individual speakers in Avar, Tsez and Bezhta; measured sequence taken 30ms from oral release.

Our Avar informant was a 50-year-old male speaker from Makhachkala, born in the village Katekh. Measurements of center of gravity (COG) in the burst spectra (0.03 sec measurement interval, 0-10 kHz) of lateral ejectives in our Avar data (N = 14) range between 3200 Hz and 3500 Hz, whereas in the Tsez and Bezhta (N = 10) data we find (lower) values between 1800 Hz and 2400 Hz, and between 1700 Hz and 2400 Hz, respectively. The COG of the lateral fricative (0.05 sec measurement interval, 0-10 kHz) ranges between 2900 Hz and 3400 Hz, which is apparently lower than values reported for Turkish Kabardian by Gordon and Applebaum (2006) which range from 4400 Hz to 4600 Hz. We assume that here also the two methods would bias the results, since we experienced COG to be very sensitive to the defined measurement onset, although no setting would have given us results similar to those found for Turkish Kabardian.

Regarding the specific sound quality, we can confirm for the lateral ejective affricates in Bezhta a scraping, often pulsing sound (cf. Gordon and Applebaum 2006), which seems to originate in the intermediate closure between the lateral part of the tongue body and base of the palatal arch behind the teeth, i.e. a secondary lateral passage outside the teeth. An investigation of further speakers will be necessary to confirm whether this is an idiosyncrasy of some speakers. However, we have observed similar patterns in (uncontrolled) recordings of other speakers of Bezhta or Avar.

Comparing the average burst spectra of all items in each of our samples of Avar, Tsez and Bezhta, we observe a positively skewed spectral shape with two peaks, one around 2000 Hz and one around 4000 Hz (Figure 9).

Since the voiceless lateral fricative is often realized as a prestopped fricative, these would potentially contrast with the lateral affricates. In fact, for initial position we observe only a more abrupt onset and longer burst phase (Figure 10). As with the lateral affricate, the geminated or strong variant of this lateral fricative does appear in initial and medial position, too. For geminate affricates we see stronger and longer bursts, and for geminate lateral fricatives longer frication phases with a strong tendency to be pre-stopped.

4.6. Labialization – a two speaker sample from Avar

In a sample of two speakers we investigated another secondary feature, labialization. As mentioned above this feature occurs throughout a number of languages of Nakh-Daghestanian and Adyghe-Abkhaz groups. It is either considered a secondary phonemic feature which provides another contrast or it is considered to have its own segmental status. In both cases it is usually reflected in the orthogaphy and would be represented as letter “в” following a consonant letter.

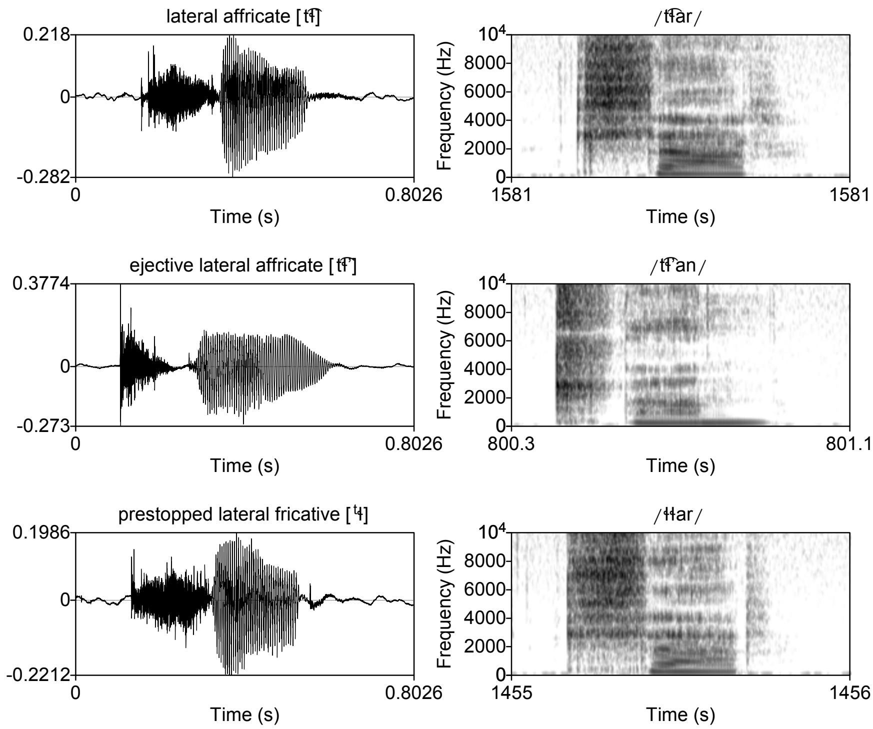


Figure 10. Triplet of close plain (aspirated) lateral affricate vs. ejective lateral affricate vs. prestopped lateral fricative.

Aside from the above discussion about the specific spectra of pharyngealized stops, we investigate here labialization as an additional factor that shapes the release spectrum alongside place of articulation and glottalization. In Avar labialization is also found with alveolar and uvular obstruents, but not with bilabial (ejective) stops as it is, for instance, in the Termirgoi variety of West Circassian (Adyghe-Abkhaz) (Hewitt 2004: 40). Words in isolation and in a carrier phrase (“Ditsa ITEM abuna.” “I said ITEM.”) were elicited and 25 items for each type (labialized/non-labialized) of lateral, velar and uvular pulmonic and glottalic stops were investigated.

Alternative approaches to the acoustic characterization of place of articulation in stop consonants have been the subject of recent discussion (Stevens, Manuel, and Matthies 2003; Suchato 2004a; 2004b; Suchato and Punyabukkana 2005; Flemming 2007).

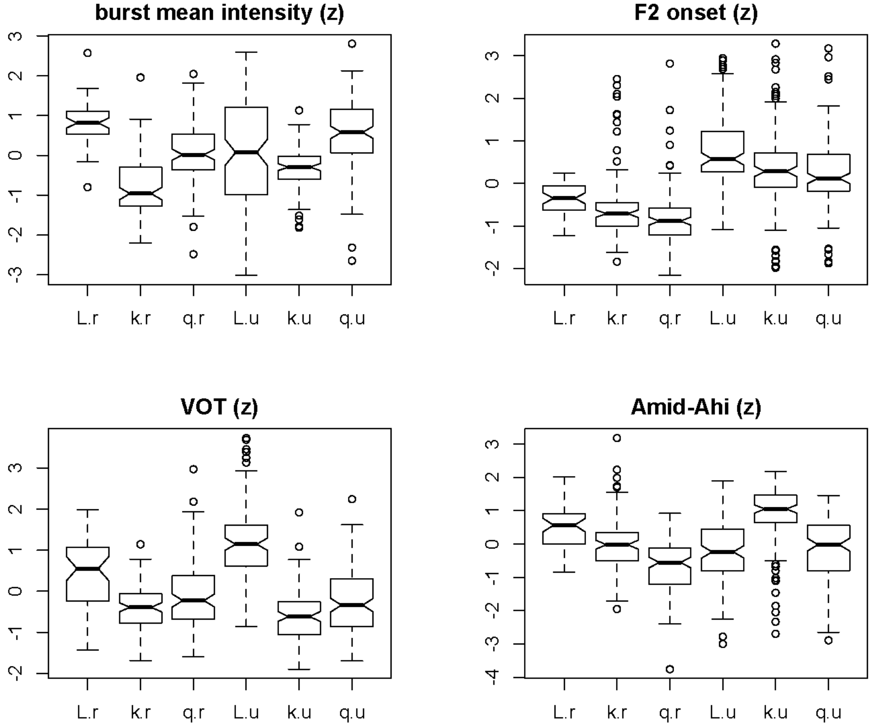


Figure 11. VOT and burst spectra measures of lateral (L= t͡ɬ), velar (k) and uvular (q) stops, including aspirated and ejective manner of initial position. Labialized (k’w, kw, t͡ɬ’w, t͡ɬ w, q’w, qw -> r) and non-labialized (k, k’, t͡ɬ, t͡ɬ’, q, q’ -> u) stops were z-standardized for two speakers (1 female/ 1 male) of Avar and cumulated over initiation at place of articulation.

We applied some of the suggested measures, such as burst peak frequency above 1 kHz and the amplitude difference (Amid-Ahi) of a mid-range frequency band (1.25–3 kHz) as well as a higher frequency band (3.5–8 kHz) (Figure 11).

Pairwise comparisons using t-tests with pooled SD and false discovery rate (FDR) as p-value adjustment method (Verhoeven, Simonsen and

McIntyre 2005) revealed that the velar and lateral, glottalic and pulmonic stops differ significantly – if we compare labialized vs. non-labialized variants – for burst duration (t͡ɬ’ > t͡ɬ’w: p < 0.001), for burst peak frequency above 1kHz (t͡ɬ’ > t͡ɬ’w: 0.006), for the Amid-Ahi measure (k’ > k’w: p < 0.001; t͡ɬ’ < t͡ɬ’w: p < 0.001) and (closely related) for PRAAT’s band energy difference (1.25–3 kHz vs. 3.5–8 kHz) (k < kw: p = 0.0089; k’ < k’w: p < 0.001; t͡ɬ’ < t͡ɬ’w: p = 0.0085). The difference reaches significance ‘more easily’ at the level of the individual and suggests that the contrast strategy by means of burst (or more precisely release noise spectra, intensity and duration) is speaker based.

Nonetheless, labialization itself is attested by means of F2-onset of the following vowel (cf. Suh 2007) such that we find significantly lower F2-onsets with labialized samples (asp.+glott. & lat.+vel.+uvul.: F = 30.181, p < 0.001, t(607) = -3.056, p = 0.0023). As expected (cf. Suh 2007), the spectral mean frequencies of the release burst noise do not allow us to distinguish the two types in a consistent way.

4.7. Contrast enhancement – Voice quality in adjacent vowels

Gordon (2001: 1) reports for ejectives in Athabaskan that the voice quality “in the vicinity of ejectives, especially in preceding vowels, becomes creaky”. Kingston (2005: 147) also notes a “constricted voice quality that preceded stem-final glottalic consonants…”

Within the Georgian material it was observed during face-to-face impressionistic recording sessions with one male informant that voice quality differences often characterized not only the syllable associated with an ejective consonant (Robins and Waterson 1952; Shosted and Chikovani 2006), but complete words. Impressionistically, a trisyllabic word with an initial ejective had a tighter (strangled, epiglottalized) voice quality than a word of an analogous structure with an initial pulmonically initiated plosive. While we did find large amplitude movements in electroglottographic recordings (Gx) commensurate with the vertical larynx movements visible in video recordings (see 0), no differences were observed in the shape or frequency of the vocal fold vibration itself (Lx). It would seem that the perceived voice quality differences are more likely to be related to the raised larynx extending over a number of syllables.

In order to study possible voice quality changes, we selected /a/-vowels following an initial stop in a sample of 20 tokens per phonation (initiation) type of the stop (glottalic, aspirated pulmonic, voiced pulmonic) per speaker. Harmonics-to noise ratio (HNR) and spectral slope in terms of H1-H2 was measured for the first third of the vowel and for total duration. The most stable measurement in our sample is that of HNR over the entire length (see Figure 12).

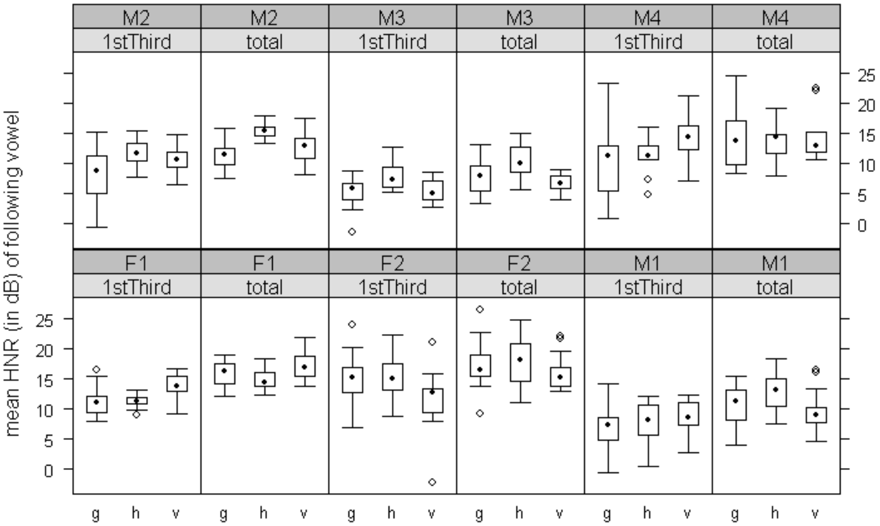


Figure 12. Mean harmonics-to-noise ratio values for /a/-vowels after glottalic (g), aspirated (h) and voiced (v) stops in four speakers (4 Male, 2 Female); for F1; M1-M3 N = 20 per type; panels start left in buttom row.

The three male speakers exhibited significantly (M1: *asp vs. glott* t = -2.149, p = 0.019; M2: *asp vs. voic* t = -7.407, p < 0.001; M3: glott vs. asp t = -3.495, p = 0.0012; df = 38) higher HNR values for vowels after aspirated stops. By contrast HNR values for vowel portions are in the same range as those for voiced stops. The pattern found for the female speakers is the reverse of that found for the male speakers. The HNR values in the vocalic portion after aspirated stops are lower (F1: *glott vs. asp* t(38) = 1.970, p = 0.028).

However, if we look at only the first third of the following vowels the picture changes quite distinctly and the dynamic nature of voice quality becomes visible. Whereas the HNR values for the voiced stops remain similar to those measures over the total duration of the following vowel, at least for the male speakers, the mean and median HNR values for the first third of the following vowel are lower for glottalic and aspirated stops (see Figure 12). Except for M3 the median HNR group, values for the first-third vowel portions after glottalic stops are always ranked lowest. The difference between vowels following glottalic and aspirated stops is significant, if averaged over all four speakers and after z-standardization ($t(38) = -3.430$, $p < 0.001$).

Finally, two additional samples (F2 & M4) had been added from the two recordings of Georgian speakers from the Phonetic Database of Victoria University (Esling 1994). Unfortunately here the samples lack homogeneity of those above in terms of balance per type (N(F2): g: 20 h: 16 v: 15; N(M4): g: 30 h: 10 v: 10) and consistency in vowel quality, and thus they have been omitted from the statistic analysis. HNR appears to be a very stable and robust measure for semi-automatic detection. Alternatively, one would choose H1-H2, but here pitch detection in the first third of the vowel becomes more complex (cf. Vicenik 2008).

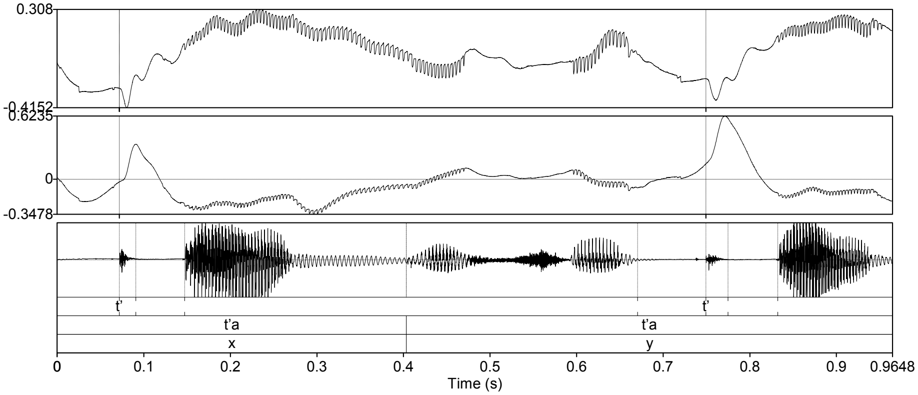


Figure 13. Double-channel EGG (top and middle signal) and synchronized audio signal (bottom) of two adjacent utterances of the word [t'a] by a female speaker of Avar; the first item in isolation (x) is immediately followed by the same token in carrier (y).

The measure of fundamental frequency (F0) in the initial portion of a following vowel was suggested by Warner (1996) and Wright, Hargus and Davis (2002). In our sample of Ingush (s. 4.2.), F0 of the following vowel does not seem to be a reliable indicator of glottalization, but there is, for example, a slight tendency for monosyllabic words to have a slightly higher F0 (1-2 semitones) after initial glottalic stops although we do not find a falling contour towards the center of the vowel, as Warner (1996) reports.

The F0 difference for our female Ingush speaker is significant (t(53.27) = -2.29, p = 0.0258) with F0 means per syllable of 177Hz for pulmonic vs. 203Hz for glottalic initial stops. However, if we look at our Georgian sample of four speakers (1 female / 3 male), F0 is higher in vowel portions (first 50ms) following an aspirated stop. Vowel onsets after glottalic stops (for two speakers) have significantly higher F0 than after voiced stops (M1: t(39) = 4.188, p < 0.001; M2: t(38) = 3.3037, p = 0.001). Thus, F0 after voiced stops is lower for only two of the speakers (M1, M2) to be clearly at the lowest rank. But the difference between voiced and glottalic is still significant (t(122.9) = 2.8043, p = 0.0058) in the z-standardized pool of the four Georgian speakers.

4.8. Elevated larynx – additional observations on larynx movements

As mentioned above, ejectives are usually described also by the articulatory characteristic of an upward movement of the larynx (see 1.).

Although we have no direct measures of larynx height, a sharp upward movement of the larynx was regularly observed during face-to-face impressionistic sessions with a Georgian informant. By contrast, in a word, containing only pulmonic consonants such as /oboba/ “spider”, only moderate displacement of the thyroid was observed. However, from external observations of larynx movements during impressionistic recording sessions, it is clear that upward larynx movement is not only present for the variants with an oral stricture of complete closure or close approximation, but also for the glottal stop itself.

Over the whole corpus we do observe fairly specific laryngographic Gx movements close to the release of glottalic stops. Although the low-frequency impedance signal offset (Gx) just reflects, as we assume, large tissue displacement (cf. Rothenberg 1992). Informed by our visual observations we are able to conclude that this Gx behavior (cf. Figure 13) reflects a short larynx elevation followed by a slower lowering. The elongation of the movement during pulmonic stops is less extensive.

Especially in the Bezhta ejectives (but also in those of Ingush) we repeatedly observed a peak in the EGG signal which occurred just prior to the burst in the acoustic record (see arrow in Figure 14). According to the vowel cycles, this is an abrupt decrease of impedance commensurate with a fast approximation of laryngeal tissue, or rather an increase of tissue contact. The peak is followed by a low amplitude oscillation, low relative to full voicing in the vowels. We find no other explanation than "passive" movement of the vocal folds or adjacent tissue during the burst.

Another very interesting observation on data from Lezgi, one of the major languages of Daghestan but a different branch from the languages examined thus far. Our speaker is a 55 year old male speaker from Makhachkala, born in the Kurakhanskij district. Here we find a voicing-like activity systematically preceding a majority of initial ejective stops (Figure 14). So far we see no indication that this is a pure idiosyncrasy of the speaker rather than perhaps an effect of hyperarticulation or even a characteristic feature of this variety of Lezgi, since (pre-)voiced ejectives are already known from Khoisan languages (cf. e.g. Ladefoged and Maddieson 1996: 80).

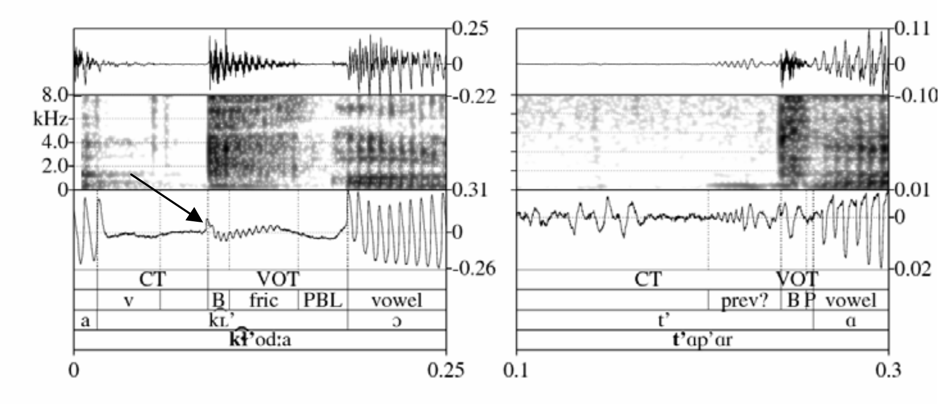


Figure 14. Bezhta [left] – positive peak in EGG (arrow) slightly before release instant and followed by low amplitude oscillation in the EGG-signal; Lezgi [right] – 'prevoicing'– like activity prior to the release; B – burst, fric – friction, prev-prevoicing, v – voicing.

5. Summary

Wright, Hargus, and Davis (2002), based on Lindau (1984), and Kingston (2005) make a general categorization of ejectives into the classical "stiff" ejectives and the "slack" ejectives. The authors proposed this 'fortis/lenis typology' based on a number of parameters, so that stiff vs. slack ejectives would be assumed to have long total duration, small closure duration/VOT, high burst intensity, high F0 of the following voice onset, modal or tense voice quality, and a fast rise to peak energy (energy slope). However, the Gitksan data presented by Ingram and Rigsby (1987) and the Ingush data by Warner (1996) already indicated that some languages would not follow this pattern. And Wright, Hargus, and Davis (2002) had to conclude the same for their Athabaskan material. Within our small sample of language material from Daghestanian languages we observe fairly homogeneous behavior, so that according to the two categories, all samples could possibly be classified as stiff ejectives. Nonetheless, the durational measures (e.g. VOT or CT) do not always produce significant results, and we must assume – similar to Wysocki (2004) and Vicenik (2008) – that compensation strategies need to be investigated on the level of each individual speaker (see our samples from Georgian and Ingush). Hence we would need to assume a heterogeneous quality of glottalization throughout the languages of the Caucasus, since variation within and between speakers would play a crucial role in the development of a cross-linguistic similarity of ejectives over a larger area. Factors, such as prosodic position and speech rate need to be considered, too.

The language material investigated for this study shows consistent use of a post-burst lag (after a relatively short burst) as the most likely distinctive and integral part of ejectives, at least at a moderate speech rate in word list and carrier phrase contexts. As a secondary and complementary accompaniment, an abrupt, aperiodic, often creaky voice onset or single glottal stop following the burst of ejectives possibly serves as an additional cue. Furthermore, voice quality differences (breathiness or tenseness) during the vocalic portion following a stop also seem to play a role, although a clear tendency can not be stated yet for all languages investigated here.

Additional illustrations of the data and audio-samples can be accessed via http://www.eva.mpg.de/~grawunde/cauc/fon_ejcauc01.html

Acknowledgements

We are very grateful to Martine Toda and Susanne Fuchs for their very helpful comments on earlier versions of this chapter, to Don Stilo for compiling the Georgian wordlist and to our informants and speakers Sultan Mereshkov, Lilija Teriva, Merab Geguchadze, Vano Nasidze, Manana Zurab, Arsen Abdulaev, Ibragim Dibirov, Zulaikhat Mallaeva, and Raisat Karimova.

References

Alekseev, Mikhail E., ed.
1999 *Jazyki mira - Kavkazskie jazyki* [The world's languages - Caucasian languages], Jazyki Mira / RAN. In-t. yazykoznania. Moscow: Academia.

Catford, John C.
1977a *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.

Catford, John C.
1977b Mountain of tongues: The languages of the Caucasus. *Annual Review of Anthropology* 6: 283–314.

Chitoran, Ioana
2002 Georgian harmonic clusters: Phonetic cues to phonological representation. *Phonology* 15(2): 121–141.

Esling, John
1994 *University of Victoria Phonetic Database*: Version 4.0.

Fallon, Paul D.
2001 *The Synchronic and Diachronic Phonology of Ejectives*. London/ New York: Routledge.

Flemming, Edward
2007 Stop place contrasts before liquids. *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany*: 233–236.

Gordeeva, Olga, and Jim Scobbie
2006 Phonetic conditioning of word-final ejective stops in the speech of Scottish English pre-school children. *Colloquium of the British Association of Academic Phoneticians. Queen Margaret University College.*

Gordon, Matthew
2001 Linguistic aspects of voice quality with special reference to Athabaskan. *Proceedings of the Athabaskan Languages Conference*: 163–178, retrieved from http://www.linguistics.ucsb.edu/faculty/gordon/athab-phonation.pdf (June/2007).

Gordon, Matthew, and Ayla Applebaum
2006 Turkish Kabardian. *Journal of the International Phonetic Association* 36(2): 255–264.

Hewitt, George
2004 *Introduction to the Study of the Languages of the Caucasus*. Munich: Lincom Europa.

Hogan, John T.
1976 An analysis of the temporal features of ejective consonants. *Phonetica* 33(4): 275–284.

Ingram, John, and Bruce Rigsby
1987 Glottalic stops in Gitksan: An acoustic analysis. *Proceedings of the of 11th International Congress of Phonetic Sciences, Tallin, Estonia* 2: 134–137.

Job, Micheal, ed.
2004 The Indigenous Languages of the Caucasus. vol. 3 The North East Caucasian Languages - Part 1. Ann Arbor: Caravan.

Kibrik, Alexander E., and Sandro V. Kodzasov
1990 *Sopostavitel'noe izucenie dagestanskix jazykov: Imja. Fonetika* [Comparative study of Dagestanian languages: Nomen, phonetics]: Moscow: MGU.

Kingston, John
2005 The phonetics of Athabaskan tonogenesis. In *Athabaskan Prosody (Current Issues in Linguistic Theory 269).* Sharon Hargus and Keren Rice, eds., 137–184. Amsterdam Studies in the Theory and History of Linguistic Science Series 4. Amsterdam/ Philadelphia: Benjamins.

Klimov, Georgij A.
1994 *Einführung in die Kaukasische Sprachwissenschaft.* Uebersetzt und bearbeitet von Jost Gippert. Hamburg: Helmut Buske.

Kong, Eun Jong, and Mary Beckman
2006 Fine-grained phonetics and developmental universals for glottal features, Poster presented at 10th LabPhon. Paris: retrieved from https://www.ling.ohio-state.edu/~edwards/ Labphon10VOT_June18.pdf (July 2007).

Koryakov, Yuri B.
2006 *Altas kavkazkikh yazykov* [Atlas of the Caucasian languages]. Moscow: RAN.

Ladefoged, Peter, and Ian Maddieson
1996 *The Sounds of the World's Languages*. Cambridge: Blackwell Publishing.

Laufer, Asher, and Thomas Baer
1988 The emphatic and pharyngeal sounds in Hebrew and in Arabic. *Haskins Laboratories Status Report on Speech Research* SR-95/96 New Haven: 51–71.

Lindau, Mona
1984 Phonetic differences in glottalic consonants. *Journal of Phonetics* 12: 147–155.

Maddieson, Ian
1984 *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian
2005 Glottalized consonants. In *The World Atlas of Linguistic Structures*. Martin Haspelmath, Matthew Dryer, David Gil, and Bernard Comrie, eds., 34–37. Oxford: Oxford University Press.

Maddieson, Ian, Ramazan Rajabov, and Aaron Sonnenschein
1996 The main features of Tsez phonetics. *UCLA Working Papers in Phonetics* 93: 94–110.

Mann, Henry Bertold, and Donald Ransom Whitney
1947 On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*: 50–60.

McDonough, Joyce, and Peter Ladefoged
1993 Navajo stops. *UCLA Working Papers in Phonetics* 84: 151–164.

McDonough, Joyce, and Valerie Wood
2008 The stop contrasts of the Athabaskan languages. *Journal of Phonetics* 36(3): 427–449.

Nichols, Johanna
1994 Ingush. In *The North East Caucasian languages, Part 2*. Rieks Smeets, ed., 77–145. (The indigenous languages of the Caucasus 4). Delmar, NY: Caravan Books.

Ohala, John J.
1993 Sound change as nature's speech perception experiment. *Speech Communication* 13(1-2): 155–161.

Ritter, Nancy
2006 Georgian consonant clusters: The complexity is in the structure, not the melody. *The Linguistic Review* 23: 429–464.

Robins, R. H., and Natalie Waterson
1952 Notes on the phonetics of the Georgian word. *Bulletin of the School of Oriental and African Studies, University of London* 14(1): 55–72.

Rothenberg, Martin
1992 A multichannel electroglottograph. *Journal of Voice* 6(1): 36–43.

Shosted, Ryan K., and Vakhtang Chikovani
2006 Standard Georgian. *Journal of the International Phonetic Association* 36(2): 255–264.

Spearman, Charles
1987 The proof and measurement of association between two things. *The American Journal of Psychology*: 441–471.

Stevens, Kenneth N.
1998 *Acoustic Phonetics*. Vol. 30, Current Studies in Linguistics. Cambridge, MA: The MIT Press.

Stevens, Kenneth N., Sharon Y. Manuel, and Melanie Matthies
2003 Revisiting place of articulation measures for stop consonants: Implications for models of consonant production. *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, USA*: 1117–1120.

Suchato, Atiwong
2004a Classification of stop consonant place of articulation, Ph.D. diss., MIT, retrieved 7th July 2008 from https://dspace.mit.edu/bitstream/1721.1/ 27197/1/Suchato_Atiwong_PhD_2004.pdf.

Suchato, Atiwong
2004b Classification of stop consonant place of articulation: combining acoustic attributes. *Proceedings of From Sound to Sense: 50+ Years of Discoveries in Speech Communication, MIT.*: 197–202, retrieved from http://www.rle.mit.edu/soundtosense/conference/pdfs/fulltext/Saturday%20Posters/SB-Suchato-STS.pdf.

Suchato, Atiwong, and Proadpran Punyabukkana
2005 Factors in classification of stop consonant place of articulation. In *Proceedings of Interspeech*, Lisbon, Portugal: 2969–2972.

Suh, Yunju
2007 Consonant-labiovelar glide combinations in Spanish and Korean. *Proc 16th International Congress of Phonetic Sciences, Saarbrücken, Germany*: 1773–1777.

Turk, Alice, Satsuki Nakai, and Mariko Sugahara
2006 Acoustic segment durations in prosodic research: a practical guide. In *Methods in Empirical Prosody Research*. Stefan Sudhoff, Denisa Lenertová, Robert Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter, and Johannes Schließer, eds., 1–28. Language, Context and Cognition, vol. 3. Berlin / New York: Walter de Gruyter.

Verhoeven, Koen J. F., Katy L. Simonsen, and Lauren M. McIntyre
2005 Implementing false discovery rate control: Increasing your power. *Oikos* 108(3): 643–647.

Vicenik, Chad
2008 Acoustic study of Georgian stop consonants. *UCLA Working Papers in Phonetics* 107(1): 1–30.

Warner, Natasha
1996 Acoustic characteristics of ejectives in Ingush. *Proceedings of the 4th International Conference on Spoken Language, ICSLP, vol. 3, Philadelphia, Pennsylvania*: 1525–1528.

Wright, Richard
2004 A review of perceptual cues and cue robustness. In *Phonetically Based Phonology*. Bruce Hayes, Robert Kirchner, and Donca Steriade, eds., 34–57. Cambridge: Cambridge University Press.

Wright, Richard, Sharon Hargus, and Katharine Davis
2002 On the categorization of ejectives: data from Witsuwit'en. *Journal of the International Phonetic Association* 32(1): 43–77.

Wysocki, Tamra M.
2004 Acoustic analysis of Georgian stop consonants and stop clusters. Ph.D. diss., Ann Arbor, Chicago University.

Tongue body and tongue root shape differences in N|uu clicks correlate with phonotactic patterns

Amanda L. Miller

1. Introduction

There is a phonological constraint, known as the Back Vowel Constraint (BVC), found in most Khoesan languages, which provides information as to the phonological patterning of clicks. BVC patterns found in N|uu, the last remaining member of the !Ui branch of the Tuu family spoken in South Africa, have never been described, as the language only had very preliminary documentation undertaken by Doke (1936) and Westphal (1953–1957). In this paper, I provide a description of the BVC in N|uu, based on lexico-statistical patterns found in a database that I developed. I also provide results of an ultrasound study designed to investigate posterior place of articulation differences among clicks.

Click consonants have two constrictions, one anterior, and one posterior. Thus, they have two places of articulation. Phoneticians since Doke (1923) and Beach (1938) have described the posterior place of articulation of plain clicks as velar, and the airstream involved in their production as velaric. Thus, the anterior place of articulation was thought to be the only phonetic property that differed among the various clicks. The ultrasound results reported here and in Miller, Brugman et al. (2009) show that there are differences in the posterior constrictions as well. Namely, tongue body and tongue root shape differences are found among clicks. I propose that differences in tongue body and tongue root shape may be the phonetic bases of the BVC.

The airstream involved in click production is described as *velaric airstream* by earlier researchers. The term *velaric airstream* is replaced by *lingual airstream* by Miller, Namaseb and Iskarous (2007) and Miller, Brugman et al. (2009). The majority of consonants found in the world's languages are produced using a *pulmonic egressive airstream*, meaning that sound is produced on the air pushed out of the lungs under the control of the respiratory muscles. Click sounds, on the other hand, are produced when air is rarefied between the two constrictions as the tongue dorsum moves backward and downward. The click burst occurs when the anterior constriction is released, allowing air to rush into the vacuum made by the tongue. The release of the posterior constriction is pulmonic egressive, because air is being pushed outward by the lungs. Due to the proximity of the releases of the anterior and posterior constrictions, the posterior release is inaudible in plain clicks, and there is often no visible pulmonic burst. However, I will provide data in this paper on the patterning of a class of clicks that have an audible pulmonic burst, which I refer to as linguo-pulmonic contour segments. The terms *complex segments* and *contour segments* refer to the distinction made by Sagey (1990). *Complex segments* are sounds that have two constrictions that are nearly simultaneous; and *contour segments* are single sounds that are sequences of articulations (Sagey 1990). In this paper, all clicks are referred to as *complex* following Sagey (1990) and Miller, Brugman et al. (2009), while affricates and linguo-pulmonic contour segments are referred to as *contour segments.*

I provide a model for click consonants that follows Zsiga (1997) and Fujimura (2000) in having both phonetic and phonological components. The phonetic component is based in Articulatory Phonology (Browman and Goldstein 1989). The mapping between the two components of the grammar may be viewed as an implementation of what Fowler (1980) refers to as coordinative structures.

1.1. The Back Vowel Constraint

Traill (1985) proposed a constraint that rules out the co-occurrence of plain clicks with front vowels to account for the lexical gap of words containing clicks and front vowels in !Xóõ, and stated it in terms of the feature [back]. Since plain clicks were assumed to all have velar posterior constrictions, they were all assumed to be marked for the feature [+back]. He called the constraint that rules out the co-occurrence of certain consonants with front vowels – the Back Vowel Constraint (BVC), and stated it in the form of the implication provided in (1):

(1) The Back Vowel Constraint

If	C_1	V_1
	<+back>	
then	C_1	V_1
	<+back>	<+back>

The existence of front vowels following dental and palatal clicks is captured by a rule, which Traill (1985) calls Dental Assimilation (DA). Sagey (1990) and Clements and Hume (1995) use the feature [+anterior] to classify the dental [ǀ] and palatal [ǂ] clicks separately from the central alveolar [!] and lateral alveolar [ǁ] clicks. The Dental Assimilation rule in (2), adopted from Sagey (1990), crucially requires both dental and palatal clicks to be [+anterior]. This is justified by the fact that palatal clicks have a long constriction, which covers a large area from the dental to the palatal region.

(2) Dental Assimilation (DA)
a → ə, ɨ / [+ant] ___ i, n

Miller-Ockhuizen (2000) showed that this so-called Dental Assimilation in Ju|'hoansi is not an assimilatory process, but rather a phonetic process of co-articulation, by showing that it does not change a back vowel to a front vowel categorically. Rather, co-articulation fronts a back vowel slightly following dental and palatal clicks, but this is largely inaudible. A separate process of height harmony raises the low vowel /a/ before the high vowels [i] and [u], which yields [ə], irrespective of the preceding consonant. Thus, DA cannot account for the presence of [i] following dental and palatal clicks in that language, as the co-articulatory process is not strong enough to change [ə] to [i] even between a dental click and a front vowel. Miller-Ockhuizen (2000, 2003) claims that there must be a phonological difference in the clicks themselves following Sands (1991) and Johnson (1993), and that the BVC must refer to that difference, targeting only central alveolar [!] and lateral alveolar [ǁ] clicks, along with pulmonic uvular consonants. Miller-Ockhuizen (2003) analyses the central alveolar [!] and lateral alveolar [ǁ] clicks as having a [pharyngeal] feature specified on the posterior constrictions as in (3), and captures the BVC as a co-occurrence constraint against pharyngeal consonants and front vowels as in (4).

(3) Specification of posterior constrictions in Ju|'hoansi clicks
(Miller-Ockhuizen 2003)

Dental Click	[ǀ]	unmarked for pharyngeal
Central Alveolar click	[!]	[pharyngeal]
Lateral Alveolar click	[ǁ]	[pharyngeal]
Palatal click	[ǂ]	unmarked for pharyngeal

(4) BVC (Miller-Ockhuizen 2003)
$*\{\ [\text{pharyngeal}]_{\text{Vplace}}\ \ [\text{coronal}]_{\text{Vplace}}\ \}_{\sigma}$
[pharyngeal] and [coronal] cannot be specified on the same or different v-place within a syllable.

Classification of dental and palatal clicks together, opposite the central and lateral alveolar clicks, in terms of the place of articulation of the anterior constriction is problematic, since alveolar clicks have an anterior constriction location in between the anterior dentals and further back palatals.

The unexplained patterning of clicks in terms of anterior place features caused Sands (1991) and Traill (1997) to classify clicks in terms of the acoustic feature [acute] vs. [grave] proposed by Jakobson, Fant and Halle (1952), which classifies sounds based on their spectral frequencies (for clicks and pulmonic stops, it is the frequencies of their bursts). [Acute] sounds are higher frequency than [grave] sounds. However, Miller-Ockhuizen (2000) showed that labial clicks and labial pulmonic consonants do not pattern together in !Xóõ, and thus [acute] vs. [grave] could not correctly classify clicks and pulmonic stops targeted by the BVC.

1.2. Phonetic differences among clicks

Miller, Namaseb and Iskarous (2007) and Miller, Brugman et al. (2009) have, by means of ultrasound, found that the palatal click [ǂ] involves tongue root raising, while the alveolar click [!] involves tongue root retraction, in Khoekhoe and N|uu respectively. Miller, Scott et al. (2009) have shown, using high frame rate ultrasound data, that posterior place of articulation differs among the four click types in Mangetti Dune !Xung. The palatal click displays the farthest back posterior constriction. The lateral and dental clicks display slightly more forward constrictions, and the posterior constriction of the alveolar click is the farthest forward. Contrary to traditional descriptions, none of the observed clicks has a velar posterior constriction location. Rather, the posterior constriction locations are all uvular. Thus, classification of clicks in terms of their BVC patterns does not match up with differences in place of articulation of the posterior constrictions.

Thomas-Vilakati (2009) shows using electropalatography and airflow data that IsiZulu clicks differ in terms of their rarefaction gestures. Some use tongue centre lowering and some tongue dorsum retraction. She suggests, based on indirect airflow measurements, that the palato-alveolar click [!] in IsiZulu must use mainly tongue centre lowering. Miller, Scott et al. (2009) use high frame rate ultrasound data to show that the dental and palatal clicks in Mangetti Dune !Xung display tongue centre lowering, while the central alveolar click displays tongue centre lowering, tongue tip retraction and tongue root retraction. The lateral alveolar click displays the widest region of tongue centre lowering, and involves formation of a low tongue centre plateau (as opposed to the narrow tongue well seen with the other clicks). Thus, the alveolar click, [!], which is subject to the BVC, involves tongue root retraction. Further investigation is needed to fully understand the dynamics of the lateral click, since only sagittal data have been analysed up to this point. These recent findings then suggest that the differences in the articulation of the posterior constrictions among clicks may help elucidate the phonetic bases of BVC patterns.

1.3. Clicks with airstream contours

I now turn to another class of clicks found in Khoesan languages, which Traill (1985, 1997), Bell and Collins (2001) and Nakagawa (2006) refer to as ‘uvular’ clicks, but Miller, Brugman et al. (2009) refer to as linguo-pulmonic stops, that is, clicks that have a contour in airstream. In this paper, I shall refer to these sounds as clicks with airstream contours.

Traill (1985), Ladefoged and Traill (1994) and Ladefoged and Maddieson (1996) claim that these ‘uvular’ clicks differ from ‘velar’ clicks (plain clicks) mainly in their posterior places of articulation, as seen in Table 1. Bell and Collins (2001) and Nakagawa (2006) have used the same symbols for ǂHoan and G|ui respectively. No phonological account of the claimed posterior place contrasts in Table 1 has been offered, and their co-occurrence patterns with front vowels, e.g. their BVC patterns, are unknown.

Miller, Brugman et al. (2009) have shown that clicks in N|uu that are phonetically similar to clicks transcribed with contrastive ‘uvular’ posterior place of articulation in !Xóõ, ǂHoan and G|ui, do not differ in terms of their posterior constriction locations from those termed ‘velar’ clicks.

Table 1. Claimed contrasts in posterior place of articulation (L&T refers to Ladefoged and Traill 1994; L&M refers to Ladefoged and Maddieson 1996; Miller refers to Miller, Brugman et al. 2009).

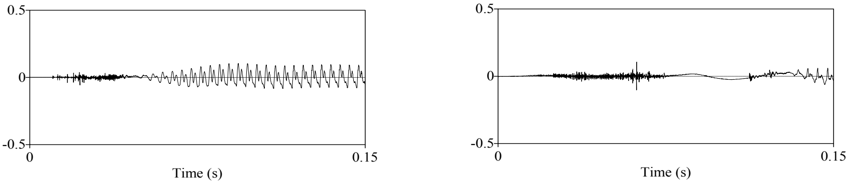
	L&T, L&M	L&T, L&M	Miller	Miller
Labial	[ʘk]	[ʘq]	[ʘ]	[ʘ͡q]
Dental	[ǀk]	[ǀq]	[ǀ]	[ǀ͡q]
(Central) Alveolar	[!k]	[!q]	[!]	[!͡q]
Lateral Alveolar	[ǁk]	[ǁq]	[ǁ]	[ǁ͡q]
Palatal	[ǂk]	[ǂq]	[ǂ]	[ǂ͡q]

Miller, Brugman et al. (2009) show that these clicks have an extended pulmonic airstream component involving audible posterior release bursts. Thus, they differ from so-called ‘velar’ clicks in terms of airstream, as they are single segments, which are produced with a loud lingual burst, followed by a second audible pulmonic burst that is the acoustic result of the posterior constriction release.

Ladefoged and Maddieson (1996) note that the posterior release in the so-called ‘uvular’ clicks is pulmonic, but they state that all clicks have a pulmonic posterior release. Miller, Brugman et al. (2009) show that while there are no posterior bursts in the N|uu clicks which were claimed to have a ‘velar’ pulmonic release (see Table 1), there is a shift from lingual airstream to pulmonic airstream. So-called ‘velar’ and ‘uvular’ clicks differ in the duration of the tongue dorsum lag phase, the phase that Thomas-Vilakati (1999) describes as the time that the tongue dorsum constriction stays in place after the release of the anterior constriction. In the so-called ‘velar’ clicks, the tongue dorsum and root are released nearly simultaneously with (in palatal clicks) or shortly after the release of the anterior constriction (in alveolar clicks), while in the so-called ‘uvular’ clicks, the posterior constriction involving the tongue dorsum and root is maintained for a long interval following the anterior release. Given the timing, Miller, Brugman et al. (2009) represent plain clicks as fully lingual complex stops, and so called ‘uvular clicks’ are represented as contour segments that are complex stops in the closure phase, and pulmonic simple stops in the release phase. I continue to use the symbol [q] to mark the release for these clicks as a matter of convenience following Miller, Brugman et al. (2009), although the posterior release location appears to be front uvular for [!͡q], but back uvular for [ǂ͡q], analogous to the posterior constriction locations found for [!] and [ǂ].

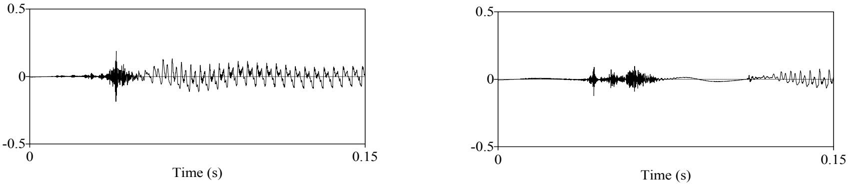
(a) Bilabial Click

(b) Bilabial Linguo-pulmonic



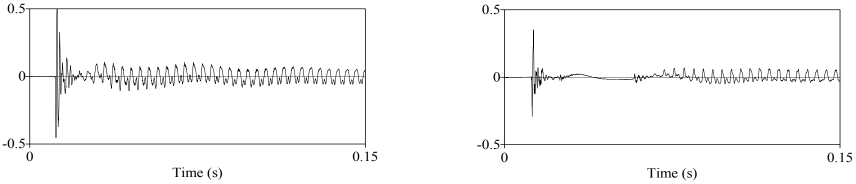
(c) Dental Click

(d) Dental Linguo-pulmonic



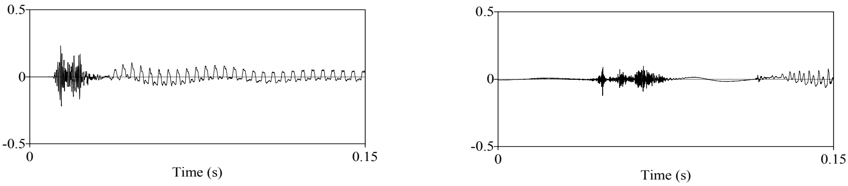
(e) Alveolar Click

(f) Alveolar Linguo-pulmonic



(g) Lateral Alveolar Click

(h) Lateral Linguo-pulmonic



(i) Palatal click

(j) Palatal Linguo-pulmonic

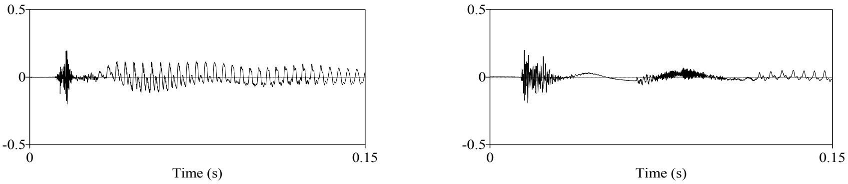


Figure 1. Waveforms of the 5 N|uu clicks and 5 linguo-pulmonic stops (clicks with airstream contours) in the words (a) [ʘuⁿuⁿ] ‘son’; (b) [ʘ͡qʰuiɑ] ‘sweat’; (c) [|uⁿuⁿ] ‘boil’; (d) [|͡quu] ‘tobacco’; (e) [!uu] ‘acacia’; (f) [!͡qui] ‘ashes’; (g) [ǁuu] ‘grasshopper’; (h) [ǁ͡quu] ‘urine’; (i) [ǂuuke] ‘fly’; and (j) [ǂ͡quu] ‘neck’(Speaker Katrina Esau).

Figure 1 provides waveforms showing the contrast between plain clicks, and clicks with airstream contours. As can be seen, the plain clicks have a single release burst formed by the release of the anterior constriction, and no acoustic signature of the posterior release. That is, the posterior release is inaudible. On the other hand, the clicks with airstream contours have both a clear click burst, which is the release burst of the anterior constriction that is made while the posterior constriction is held in place, and an audible second release burst resulting from the release of the posterior constriction. Since the anterior constriction has already released, this second burst is produced on a pulmonic airstream.

Miller, Brugman and Sands (2007) provide duration data for the four contrastive plain clicks, and the four clicks with airstream contours, in N|uu. The data show that the clicks with airstream contours have a second silent interval following the click burst with a mean of 40 ms, while there is no second silent interval in the plain clicks. Pulmonic bursts, which result from the release of the posterior constrictions in the clicks with airstream contours, are about 10 ms; while the plain clicks do not exhibit posterior bursts. The click bursts that result from the anterior releases range from 10-20 ms, and Voice Onset Time phases are about 20 ms. Each of these phases is similar in duration for the plain clicks and the clicks with airstream contours that have the same anterior places of articulation.

This study is similar in some aspects to Miller, Namaseb and Iskarous (2007) and Miller, Brugman et al. (2009). This paper differs from both of these earlier papers in that it provides a detailed lexical database study based on field recordings of the endangered Khoesan language N|uu recorded by the author and a team of linguists. Miller, Namaseb and Iskarous investigated Khoekhoe patterns. Miller, Brugman et al. (2009) focused on describing the inventory of N|uu clicks, and did not report on N|uu phonotactics. Though both report ultrasound data, this paper contains improved ultrasound traces that are plotted with the palate, and are discussed in more detail related to the phonotactic patterns. Miller (2010) provides an overview of known phonological patterns of clicks. The BVC patterns are only a small section of that paper. The linguistic analysis focuses on phonological features, rather than the phonetic model proposed here.

In this paper, I provide the results of two experiments. In section 2, I provide the N|uu consonant inventory. In section 3, I provide information about the methods, data collection and subjects used in this paper. In experiment 1, reported on in section 4, I provide lexico-statistical patterns from a database study in N|uu showing that there are two classes of clicks with respect to their patterning in the Back Vowel Constraint. I show that clicks with airstream contours in N|uu pattern the same as plain clicks with respect to the BVC. In experiment 2, reported on in section 5, I provide ultrasound traces from a single speaker of N|uu, illustrating that it is tongue body and tongue root shape differences that are the phonetic bases of the lexical patterns shown in experiment 1. In section 6, I provide a model for click articulation in terms of Browman and Goldstein's Articulatory Phonology, and in section 7, I conclude the paper.

2. N|uu consonant inventory

The N|uu consonant inventory described in Miller, Brugman et al. (2007, 2009), is provided in this section. Miller, Brugman et al. (2007, 2009) adopt a framework whereby airstream is used as a dimension to describe consonants, in addition to the standard *place of articulation* and *manner of articulation* dimensions. In the standard IPA consonant chart, consonants are separated into *pulmonic* and *non-pulmonic* consonants, and the full range of closure and release properties found on clicks are not included in the standard IPA consonant chart. This is much like aspiration, which is included as a diacritic in the standard IPA consonant chart (IPA 2006), but aspirated stops are included as a separate row in the consonant chart for Hindi where they serve as contrastive consonants (IPA 2006). In this paper, the N|uu stop inventory is presented in three tables based on the phonological categories of simple segments, complex segments and contour segments, with the complex vs. contour segment distinction used following Sagey (1990). Within each table, the rows represent manner of articulation, and the columns represent place of articulation as in the standard IPA chart. The airstream dimension is also used to group consonants within each table following Miller, Brugman et al. (2009). Glottalic airstream is the airstream used for ejectives. The full consonant inventory is provided in Miller, Brugman et al. (2009).

Table 2 provides the group of simple pulmonic stops. Table 3 provides the class of complex segments; that is clicks that are produced with two simultaneous constrictions, and a lingual airstream mechanism. These clicks are all those that are referred to earlier in this paper as plain clicks, and that were referred to in earlier descriptions as *velar clicks*. The term lingual airstream replaces velaric airstream mechanism, following Miller, Namaseb and Iskarous (2007) and Miller, Brugman et al. (2009), because the posterior constriction in clicks is not velar.

Table 4 provides the class of Nǀuu contour segments, segments that are sequences of articulations: affricates, that are stops in the closure phase with fricated release phases, and linguo-pulmonic and linguo-glottalic segments that have contours in airstream (e.g. clicks with an extended posterior constriction). These are the stops that were previously termed *uvular* clicks by Ladefoged and Traill (1994) and Ladefoged and Maddieson (1996). With acoustic data, the only way to identify the airstream of a stop is by looking at the stop bursts. Waveforms of the stop bursts for plain clicks, and clicks with a pulmonic release are seen above in Figure 1. Recall that contour segments in airstream are visible as such based on the presence of a typically higher amplitude lingual burst at the release of the first stop interval, and a typically lower amplitude pulmonic stop burst, which occurs at the end of the second silent interval formed by the extended posterior constriction.

Table 2. N|uu simple stops

	Pulmonic						
	Bilabial	Alveolar		Palatal	Velar	Uvular	Glottal
		Central	*Lateral*				
Stop	p b	(t) (d)		c cʰ ɟ c͡χ	k kʰ g	q	(ʔ)
Nasal	m	n		ɲ	ŋ		

Table 3. N|uu complex stops

	Lingual				
	Labial	Dental	Alveolar		Palatal
			Central	*Lateral*	
Stop	ʘ	ǀ ǀʰ ᶢǀ	ǃ ǃʰ ᶢǃ	ǁ ǁʰ ᶢǁ	ǂ ǂʰ ᶢǂ
Nasal	ᵑ̊ʘˀ ᵑʘ	ᵑ̊ǀh ᵑ̊ǀˀ ᵑǀ	ᵑ̊ǃh ᵑ̊ǃˀ ᵑǃ	ᵑ̊ǁh ᵑ̊ǁˀ ᵑǁ	ᵑ̊ǂh ᵑ̊ǂˀ ᵑǂ

Table 4. N|uu contour stops

Pulmonic					
	Labial	Dental	Palatal	Velar	Uvular
Affricate		t͡s	c͡χ		
Glottalic					
Affricate		t͡s’		k͡χ’	q͡χ’
Linguo-pulmonic					
	Labial	Dental	Alveolar		Palatal
			Central	*Lateral*	
Stop	ʘ͡q	ǀ͡q ǀ͡qʰ	!͡q !͡qʰ	ǁ͡q ǁ͡qʰ	ǂ͡q ǂ͡qʰ
Affricate	ʘ͡χ	ǀ͡χ	!͡χ	ǁ͡χ	ǂ͡χ
Linguo-glottalic					
Affricate		ǀ͡χ’	!͡χ’	ǁ͡χ’	ǂ͡χ’

N|uu has a simple five vowel inventory containing /u/, /i/, /o/, /e/, and /a/. However, it also has a large inventory of diphthongs.

3. Methods, data and subjects

I first provide the names of the consultants that I worked with to describe N|uu phonotactic patterns. I then describe the lexical database that I built in order to describe co-occurrence patterns found between consonants and vowels in experiment 1. In the third section, I describe the methodology used in collecting ultrasound data. The ultrasound data is used to describe the articulatory properties of clicks in N|uu, and to investigate the phonetic bases of the Back Vowel Constraint, in experiment 2.

3.1. Subjects

The data presented in this paper comes from fieldwork with speakers of N|uu, the last remaining member of the !Ui branch of the Tuu family, spoken in South Africa. There are less than 10 remaining speakers of this highly endangered language. I worked with a team of linguists: Johanna Brugman, Chris Collins, Levi Namaseb and Bonny Sands. We worked with the following N|uu speakers: Ouma Katrina Esau, Ouma Anna Kassie, Ouma Hanna Koper, Ouma |Una Rooi, Ouma Kheis Brou and Ouma Griet

Seekoei, who speak the Western dialect, and Ouma Hannie Koerant and Oupa Andries Olyn, who speak the Eastern dialect. All of these speakers are bilingual in Afrikaans and N|uu and are 65-75 years of age. None of the speakers currently resides in a household with other N|uu speakers, and Afrikaans is their dominant language.

3.2. Lexical Database

The lexical data in this paper comes from a dictionary of N|uu that is in progress, and is discussed in Sands, Miller and Brugman (2007). Transcriptions were agreed upon by all of the authors. A root database was developed by culling all of the Eastern dialect roots out of the dictionary. These roots were provided by the Eastern dialect speakers (HK and AO), and not known by the Western Dialect speakers. The resulting database contains 790 roots. This paper focuses on the Western dialect of N|uu, because that is the dialect of most of the remaining speakers. The majority of words have a $C_1V_1V_2$ or $C_1V_1C_2V_2$ word structure, though there are a few that have a $C_1V_1V_2C_2V_3$ word structure. Clicks, including clicks with airstream contours only occur in C1 position of roots, just as in Ju|'hoansi (Miller-Ockhuizen 2010), !Xóõ (Traill 1985) and Khoekhoe (Brugman 2009). Each root was coded for place, manner and airstream of the initial consonant (C_1), height and front/back distinctions on the two vowels in roots (V_1 and V_2), and place, manner and airstream of medial consonants (C_2) in bisyllabic roots. Loan-words that have not yet been assimilated to N|uu were marked as such in the dictionary and in the database, and they were not included in the lexico-statistical study reported in experiment 1 in Section 4.

3.3. Ultrasound study

Ultrasound investigations were undertaken with four of the N|uu speakers, and traces in this paper come from Ouma Katrina Esau. Data from other speakers show similar properties. Ultrasound videos were collected using a GE Logiqbook ultrasound machine with an 8C-RS 5-8 MHz pediatric transducer. Head and transducer stabilization were accomplished by using a microphone stand to hold the probe under the chin as in Gick, Bird and

Wilson (2005). The speakers sat on a bench with their heads against the wall as an aid to keep their heads stable.

The acoustic signal was simultaneously recorded with the ultrasound data, using a Shure SM10A head-mounted microphone, and the signal was channelled through a Shure FP23 pre-amp. All ultrasound recordings were made in the frame sentence [nɑ kɑ] _____ [nɑ kɑ qoˤɑˤiⁿ], meaning ‘I say _____, I say famished’. Tongue traces of clicks are plotted with and discussed relative to the place of articulation of [k] in the first [kɑ] token and/or the initial [q] in the word [qoˤɑˤiⁿ], as in Brugman (2005). Palates were traced from imaging a swallow following the method described in Epstein and Stone (2005). Note that all plots show the position of the tongue relative to the ultrasound probe, *not* the palate. For discussion of the methodological issues involved in getting from ‘probe space’ to ‘head space’ with ultrasound, see Stone (2005). We recorded 15 tokens of each word (5 repetitions, with 3 tokens per repetition), and the articulatory and acoustic signals were aligned. For each token, a frame was identified immediately before and after the click burst in the acoustic signal. The data presented here was recorded at 50 fps, meaning that we imaged the tongue every 20 ms. With the linguo-pulmonic stops (clicks with airstream contours), frames immediately before and after the pulmonic burst were also identified. The tongue edge was tracked for each of these frames using EdgeTrak software (Li, Khambamettu and Stone 2005). A complete description of the ultrasound setup used in this study, and the methodology used to align acoustic and articulatory data is provided in Miller, Brugman and Sands (2007).

The ultrasound traces provided here are similar to those found for all fifteen tokens produced by all three speakers in terms of the relative constriction locations and shapes, though due to the medium speed of the ultrasound imaging (50 fps) used in this experiment, and the high speed of the tongue in click production, there are significant aliasing effects in the data. The aliasing effects result in considerable variability in the position and shape of the tongue during the frames traced, making it problematic to average across tokens. Therefore, data is only plotted from one token produced by one speaker. However, the relative articulatory patterns found to differentiate the different segments reported here hold true for all of the data.

4. Experiment 1: Database study

In this experiment, I investigate N|uu consonant–vowel co-occurrence patterns. I hypothesize that Nǀuu plain clicks will pattern similarly to plain clicks found in Ju|'hoansi and !Xóõ. Namely, I hypothesize that the dental [ǀ] and palatal [ǂ] clicks will co-occur with both front and back vowels as they do in Ju|'hoansi and !Xóõ, while the central alveolar [!] and lateral alveolar [ǁ] clicks will not co-occur with front vowels, but instead will co-occur with a retracted and lowered [əi] allophone of /i/.

Clicks that exhibit airstream contours have never been accounted for in the statement of the BVC in any language. Thus, their phonological patterning is largely unknown. As noted above, Traill (1985) and Ladefoged and Traill (1994) have termed similar clicks in !Xóõ 'uvular' clicks, and claimed that these clicks contrast in the posterior place of articulation with so-called 'velar' clicks. However, the phonotactic patterning of 'uvular' clicks in !Xóõ does not comply with the predicted patterns given in this analysis. If uvular clicks all have posterior uvular releases, this predicts that all of these clicks should not occur with front vowels, similar to uvular pulmonic simple stop patterns. Phonological patterns involving such clicks in Traill's (1994) !Xóõ dictionary are difficult to interpret. We find words containing both clicks with airstream contours and following back vowels such as *!qāhi 'the hunt',* and words containing the retracted diphthong, such as ǂ*qái 'bird species'* and ǁ*qái 'nostril'*, which indeed seem to bear out the predictions of Traill's analysis. ('ai' is the orthographic form of the retracted diphthong [əi].) However, we also find words such as ǀ*qhái [ǀqhii] 'buffalo',* which do not bear out the prediction. The low frequency of clicks with airstream contours in !Xóõ make the interpretation even more difficult.

Based on preliminary investigations of ultrasound data showing that the palatal plain clicks and palatal clicks with airstream contours do not differ in anterior or posterior place of articulation, I hypothesize that N|uu clicks containing airstream contours will fall into two classes similar to those found with the plain clicks. Namely, I hypothesize that the dental and palatal clicks with airstream contours, [ǀq̑] and [ǂq̑], will occur freely with front vowels, while the central alveolar and lateral alveolar clicks with airstream contours, [!q̑] and [ǁq̑], will not occur with front vowels, but will instead co-occur with the retracted diphthong allophone of /i/. This hypothesis is based on the fact that Miller, Brugman et al. (2009) showed that the posterior place of articulation is the same in [ǂ] and [ǂ͡q], and [!] and [!͡q].

While this paper largely addresses the phonological patterning of clicks, pulmonic stop patterns provide further evidence as to the correct analysis of the BVC. As mentioned above, if labial pulmonic stops pattern differently from labial clicks, this rules out an analysis proposed by Traill (1997) in terms of the acoustic feature [acute] vs. [grave]. I hypothesize that the pulmonic stop patterns will be similar to those in Ju|'hoansi. Namely, I hypothesize that the labial and velar pulmonic consonants will occur with [i], while the uvular consonants will occur with [əi].

In Section 4.1., I describe the co-occurrence patterns found with consonants and monophthongal front and back vowels. In Section 4.2., I show that the retracted diphthong [əi] is in complimentary distribution with the vowel [i], suggesting that they are both allophones of /i/. Phonation contrasts shown in Tables 2-4, such as voicing, aspiration and glottalization, do not affect the patterning of stops, and neither does nasalization. That is, voiced, aspirated and glottalized stops pattern according to place of articulation, as do nasal stops. Therefore, voiceless unaspirated, voiced unaspirated, voiceless aspirated, voiceless nasal aspirated, voiced nasal, and nasalized glottalized consonants are all grouped together in the tables provided.

4.1. Results: The Back Vowel Constraint in N|uu

4.1.1. Pulmonic and click consonant phonotactic patterns

Figure 2 shows the co-occurrence patterns of front and back vowels with all of the root-initial simple pulmonic stop consonants found in the N|uu root database. Front vowels rarely occur in V_1 position of CV_1CV_2 roots. Therefore, only CVV and CVVCV roots are included in Figure 2 Back vowels are more frequent in the language overall, thus the lower frequency of front vowels across all segment types is reflective of the fact that 89% of roots contain an initial back vowel, while 11% of roots contain an initial front vowel. The alveolar, palatal, and velar initial pulmonic stops co-occur freely with both following front and back vowels, while the labial and uvular pulmonic stops occur only with back vowels.

The low lexical frequency of pulmonic stops in the language, and the particularly low frequency of labial segments, make it difficult to decide whether the lack of labial stop – front vowel sequences is due to a phonological constraint such as the Back Vowel Constraint, or whether this is just an accidental gap in the root patterns found in the database.

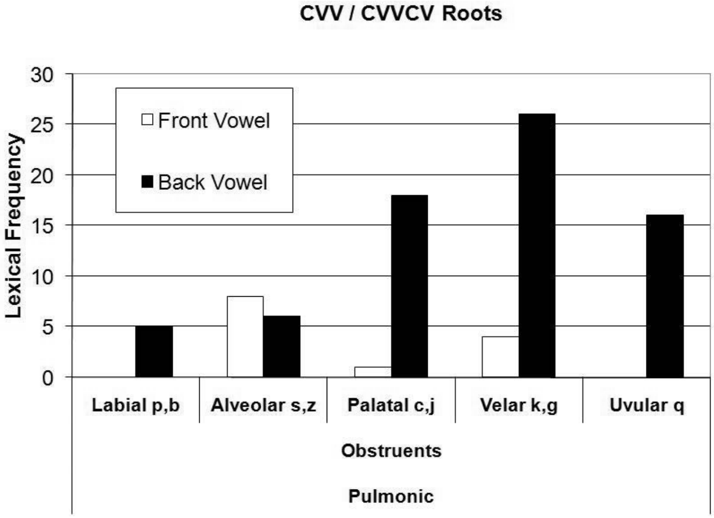


Figure 2. Co-occurrence of initial pulmonic consonants (simple consonants) with following front vs. back vowels in the 790 N|uu root database, CVV and CVVCV roots.

Figure 3 provides the co-occurrence patterns found between click consonants and front vs. back vowels. There are no front vowels in the database following central alveolar [!], lateral alveolar [ǁ], and labial [ʘ] clicks. However, note that labial clicks are low frequency, similar to labial pulmonic consonants, and thus the lack of front vowels following labial clicks could be either due to the Back Vowel Constraint, or be the result of an accidental gap of roots containing both low frequency labial clicks and low frequency initial front vowels.

Due to the ambiguity of patterns found with initial labials, I turn now to medial position, where labial consonants are quite frequent consonants, and high front vowels are also quite frequent. Medial consonant-vowel co-occurrence patterns in N|uu are shown in Figure 4. Crucially, we see that labial consonants occur freely with front vowels in this position. This differs from the lack of labial consonant-front vowel sequences found in CVV roots. Therefore, I attribute the gap of labial consonant – front vowel patterns in CVV roots to the low frequency of each of the sounds.

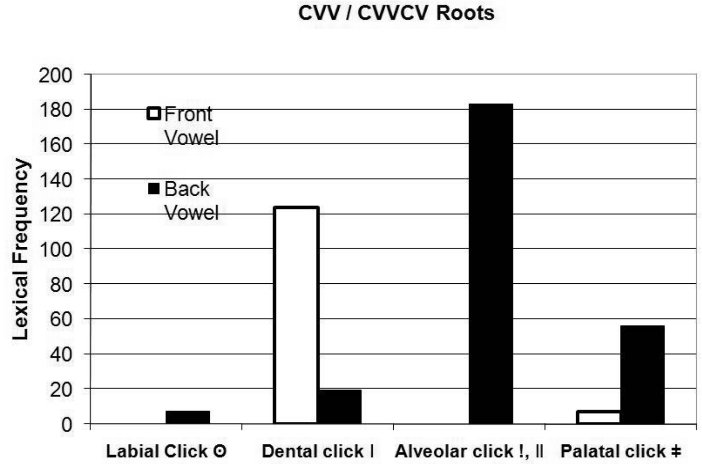


Figure 3. Co-occurrence of initial click consonants (complex consonants) with following front vs. back vowels in the 790 Nǀuu root database, CVV or CVVCV roots.

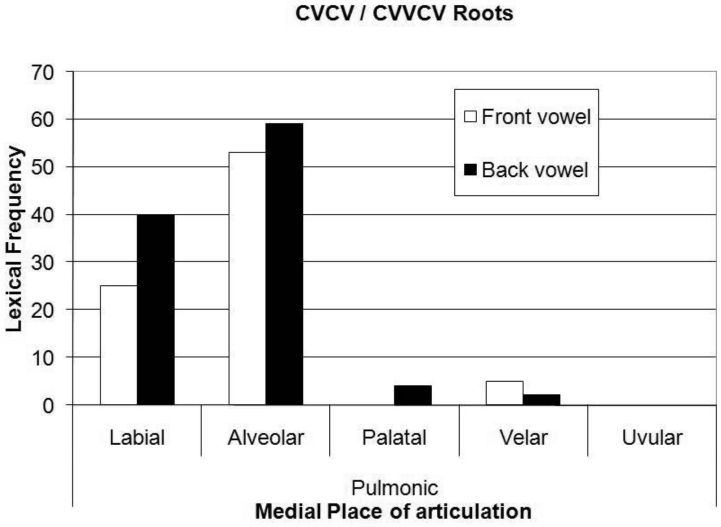


Figure 4. Co-occurrence of medial consonants and following front vs. back vowels in the 790 N|uu root database, CVCV and CVVCV roots.

Alveolar and velar consonants display the same distributional patterns as are found with velar consonants in initial position. There are no sequences of palatal consonants followed by front vowels in the second syllable of N|uu roots.

I attribute this gap to the fact that palatal consonants occur less frequently in medial position, and the fact that bisyllabic roots are less frequent than monosyllabic roots overall (23% of roots are bisyllabic in the database). Therefore, I suggest that it is the low frequency of palatals in C_2 position, and the low frequency of bisyllabic roots, which results in the gap of palatal consonant-[i] sequences in roots in the database.

4.1.2. Phonotactic patterns involving clicks with airstream contours

Figure 5 shows that linguo-pulmonic stops, which are phonetically analogous to those transcribed as 'uvular' clicks in !Xóõ, exhibit the same co-occurrence patterns to the complex stops (clicks).

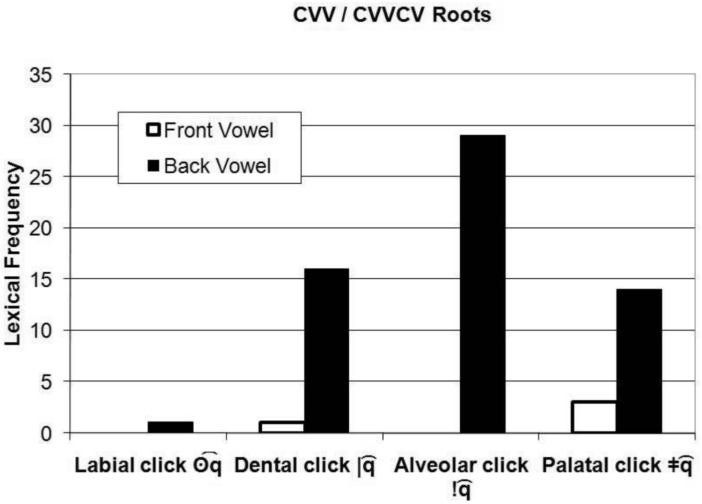


Figure 5. Co-occurrence of front vs. back vowels with root-initial clicks with airstream contours in N|uu in the 790 Root database, CVV and CVVCV roots.

That is, N|uu dental and palatal clicks with airstream contours, [ǀ͡q] and [ǂ͡q], co-occur freely with front vowels, while labial and alveolar clicks with airstream contours, [ʘ͡q], [!͡q], and [ǁ͡q], do not occur with front vowels, analogous to their plain click counterparts. I now turn to the investigation of co-occurrence patterns with [i] vs. [əi], which show that [əi] is an allophone of /i/.

4.2. Results: Allophonic patterns with the diphthong [əi]

The BVC patterns with respect to back vs. front vowels are striking, but the diphthong [əi] is even more constrained. It is in complementary distribution with the vowel [i]. That is, [əi] is an allophone of /i/ that occurs only after the same set of consonants that are limited in their co-occurrence with front vowels, namely [χ], [q], [ʘ], [!], [ǁ], [ʘ͡q], [!͡q], and [ǁ͡q]. Conversely, [i] occurs following labial, coronal and velar pulmonic consonants, as well as the clicks [ǀ] and [ǂ], and the clicks with airstream contours [ǀ͡q] and [ǂ͡q], as shown in Figure 6. Linguo-pulmonic affricate patterns are not provided here.

There is a maximality constraint in N|uu, which results in a diphthong never occurring in the second syllable of a bisyllabic root. Therefore, medial consonants are not relevant to this pattern.

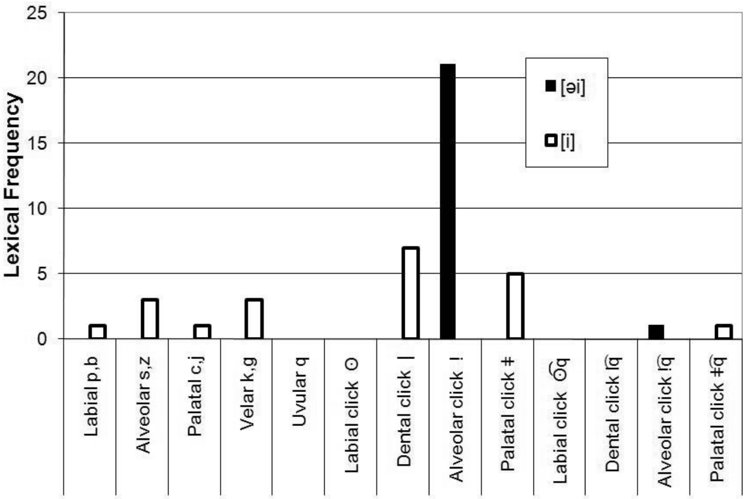


Figure 6. Co-occurrence of [i] vs. [əi] vowels with N|uu root-initial pulmonic stops, plain clicks and clicks with airstream contours.

4.3. Experiment 1: Discussion

I summarize my interpretation of the patterns seen with pulmonic stops, plain clicks, and clicks with airstream contours in Figures 2-5 in Table 5.

Table 5. Summary of C-V co-occurrence patterns in N|uu

	Occur with front & back V	**Occur with back V**
Pulmonic stops	Labial, Alveolar, Palatal, Velar	Uvular
Clicks	Dental clicks, Palatal clicks	Labial clicks, Central and lateral alveolar clicks
Clicks with airstream contours	Dental clicks, Palatal clicks	Labial clicks, Central and lateral alveolar clicks

These data show that N|uu has a Back Vowel Constraint, similar to that found in Ju|'hoansi (Miller-Ockhuizen 2003) and !Xóõ (Traill 1985). Given the BVC patterns found in N|uu, it is difficult to interpret The Back Vowel Constraint as being due to place of articulation of the anterior constriction. This is because alveolar clicks, which are articulated more forward in the mouth than palatal clicks, do not co-occur with front vowels, while palatal clicks, which have a farther back anterior constriction, do. The N|uu patterns provide further evidence that the acoustic feature [acute] vs. [grave] cannot account for BVC patterns. This is because labial pulmonic stops and labial clicks do not pattern together, and these are both classified as [grave] using Jakobson, Fant and Halle's acoustic feature. That is, their bursts both have lower frequency energy compared with the alveolar obstruents and palatal and dental clicks. The database results for the clicks with airstream contours show that the dental and palatal clicks of this type, [ǀ͡q], and [ǂ͡q], occur freely with front vowels, while the labial and alveolar clicks of this type, [ʘ͡q], [!͡q], and [ǁ͡q], occur only with back vowels and the retracted diphthong allophone of /i/. The different patterning of dental and palatal plain clicks and clicks with airstream contours, vs. the labial and alveolar plain clicks and clicks with airstream contours, leads me to hypothesize that these clicks may not have a posterior release that is the same across the board. I focus on the posterior constriction because of the patterning of uvular consonants, which are known to retract front vowels cross-linguistically. This hypothesis will be tested in a second experiment using lingual ultrasound imaging, described in section 5.

5. Experiment 2: Ultrasound study

5.1. Introduction

The phonotactic patterns found in experiment 1 with plain clicks and clicks with airstream contours lead me to hypothesize that there are two classes of clicks with respect to their articulatory properties. I hypothesize that the posterior constrictions of the central alveolar [!] and lateral alveolar [ǁ] clicks are similar in location to those found in the alveolar pulmonic consonants [q] and [χ]. Further, the phonotactic patterns seen with the clicks with airstream contours (stops) suggests that these fall into the same two classes, based on the anterior place of articulation. I hypothesize specifically that [!͡q] and [ǁ͡q] will have similar posterior constrictions to those found with [!] and [ǁ]. Moreover, [ǀ͡q] and [ǂ͡q] will have similar posterior constrictions to [ǀ] and [ǂ]. In this experiment, I investigate properties of the posterior constrictions of these four clicks using lingual ultrasound imaging.

5.2. Results

Figure 7 provides ultrasound traces of the tongue in N|uu palatal and alveolar clicks (lingual stops) and clicks with airstream contours (linguo-pulmonic stops). The tongue traces show that the alveolar click, [!], in Figure 7a, involves tongue dorsum and tongue root retraction, which result in a concave tongue body shape, and a convex tongue root shape. The posterior constriction in the alveolar click is at the same location as is found in the uvular pulmonic stop plotted with it. The cavity formed by the tongue body is fairly far forward in the oral cavity, and the anterior constriction is clearly apical. As has been noted by Traill (1985), Ladefoged and Traill (1994), and Thomas-Vilakati (2009), this configuration results in a large lingual cavity.

The production of the palatal click in Figure 7b, [ǂ], involves tongue root raising and a high flat tongue body shape. The posterior constriction of the palatal click is farther back than that of the alveolar click in Figure 7a, and the tongue tip shape is raised and flat. The broad anterior and posterior constrictions give rise to a narrow lingual cavity width and a shallower cavity depth, which results in a smaller overall cavity volume and a flatter tongue body shape. The tongue root proper does not retract, but rather it raises, as in the articulation of the [u] vowel in English described by Esling (2005). The rarefaction gesture involves gentle tongue centre lowering.

Ultrasound results of the N|uu alveolar and palatal clicks in Figures 7a and 7b show that these clicks exhibit a consistent difference in the posterior constriction locations. The posterior constriction location of the alveolar click, [!], in Figure 7a is in front of the posterior constriction location of the palatal click, [ǂ], seen in Figure 7b.

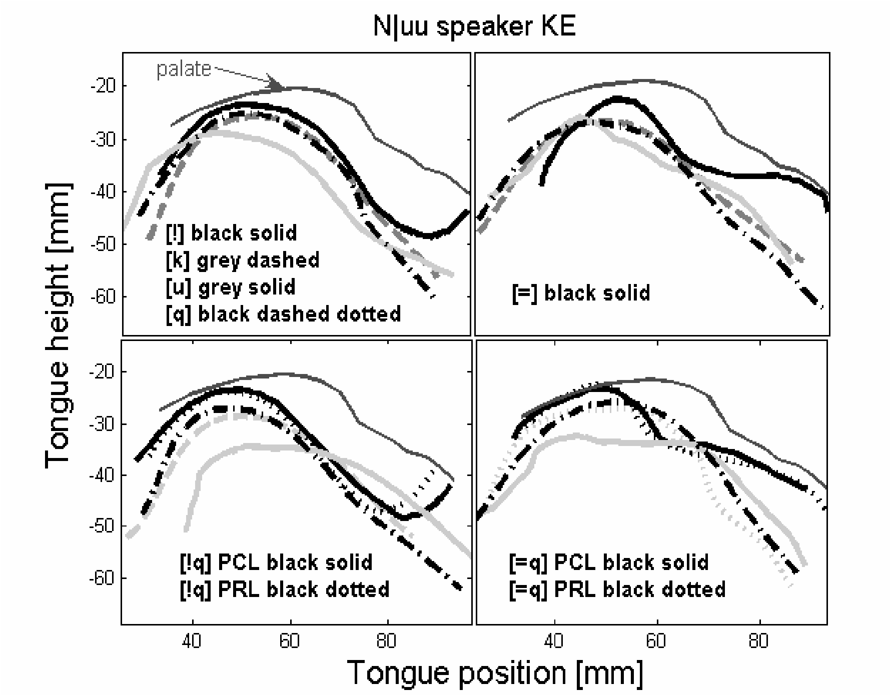


Figure 7. Ultrasound traces of the tongue in the click closure (solid black), click release (long dashed black), uvular stop (short dashed black), velar stop (dashed grey) and [u] (solid grey), and the palate (solid red) in the N|uu words !uu 'camelthorn' (7a: upper left), ǂuuke 'fly' (7b: upper right), !q͡ui 'ashes' (7c: lower left), and ǂq͡uu 'neck' (7d: lower right), (produced in the frame sentence Na ka _____, Na ka qoaqi., I say ___, I say famished. by Speaker Katrina Esau, PCL stands for Posterior Closure Location, and PRL stands for Posterior Release Location).

The posterior closures and releases in the words *!q̑ui ‘ashes’* and *ǂq̑uu ‘neck’* in Figures 7c and 7d do not differ from the posterior closures in *!uu camelthorn’* and *ǂuuke ‘fly’* seen in Figures 7a and 7b.

The alveolar and palatal clicks also differ in the length and breadth of the anterior and posterior constrictions. In the palatal clicks, both constrictions are long and broad. These contrast with the narrower anterior and posterior constriction shapes found in the alveolar click. The difference in the curvature of the tongue body in the two clicks is more pronounced earlier on, prior to the release of the posterior constriction. In the palatal click, [ǂ], the release of the anterior and posterior constrictions occur more simultaneously. Similar tongue tip, and tongue body shape differences are found among the clicks with airstream contours. The palatal linguo-pulmonic stop (click) in 7d is similar in shape to the plain palatal click in 7b, and the alveolar linguo-pulmonic stop (click) in 7c is similar in shape to the plain alveolar click in 7a.

5.3. Experiment 2: Discussion

Miller-Ockhuizen (2003) analysed the BVC in Ju|’hoansi, as involving a [pharyngeal] feature, given the assumed one to one mapping between pharyngeal articulations and tongue root retraction assumed by McCarthy (1994) and Rose (1996). The feature [pharyngeal] was proposed based on the phonotactic patterns seen in that language, with alveolar clicks behaving similarly to uvular consonants. Similar phonotactic patterns have been shown to exist for N|uu clicks in this paper.

The ultrasound data provided in Figure 7 of this paper for N|uu, in Miller, Namaseb and Iskarous (2007) for Khoekhoe, and Miller, Scott, et al. (2009) for Mangetti Dune !Xung, also show that tongue root retraction is not always a property of uvular constrictions, as proposed by McCarthy (1994). However, tongue root retraction and further forward uvular constrictions such as those seen for the alveolar click [!] in Figure 7a, may indeed go together.

Miller, Namaseb and Iskarous (2007) have claimed that the BVC is a phonological consequence of the difficulty of co-producing segments involving incompatible muscular systems, based on ultrasound results of alveolar and palatal clicks in Khoekhoe, and they propose that the tongue body shape differences among clicks account for the BVC patterns.

Thomas-Vilakati (2009) proposes rarefaction gestures for clicks. Further, she shows, via electropalatographic data with 6 speakers, that the rarefaction gestures involved in IsiZulu clicks differ for different clicks. For the IsiZulu dental click, the rarefaction gesture involves tongue centre lowering, and not dorsal retraction, while the IsiZulu palato-alveolar click [!] involves tongue dorsum retraction as well as a greater degree of tongue centre lowering. She notes that the dorsal release is uvular in nature. The lateral click in IsiZulu involves a further back dorsal position, and rarefaction involves mainly tongue centre lowering. Thomas' EPG data did not provide data on tongue shape, or on the dynamics of the tongue root.

The results of experiment 2 show that the N|uu alveolar click involves both tongue root retraction, and a concave tongue body shape, similar to that found in Khoekhoe. The palatal click, on the other hand, exhibits tongue root raising similar to the vowel [u] in English described by Esling (2005) and the vowel [u] in N|uu seen in Figure 8, and a high flat tongue body shape. These results support my hypothesis that there are articulatory differences in the posterior constrictions of the central alveolar [!] and palatal [ǂ] clicks in N|uu. The concave tongue body shape and tongue root retraction which leads to a convex tongue root shape found with [!], are incompatible with the high flat tongue body shape found with the vowel [i].

Browman and Goldstein's Articulatory Phonology theory (1989) originally propose that gestures can be produced by one of three relatively independent vocal tract subsystems: oral, velic and laryngeal. Within the oral tract, they propose three relatively independent sets of articulations: lips, tongue/blade, and tongue body. They recognize that tongue root gestures may be eventually needed. Clicks are one such case where the tongue root acts as an independent tongue segment. Further, the data suggest that tongue root shape is important in understanding the articulation of the alveolar click [!]. Thus, I suggest that just as tongue tip is specified for shape, both tongue body and tongue root must also be specified for shape.

The phonotactic patterns seen in experiment 1 led me to hypothesize that the articulation of [!͡q] would be similar to the articulation of [!], and that the articulation of [ǂ] would be similar to the articulation of [ǂ͡q], in terms of constriction locations and shapes. Results of experiment 2 show that this is indeed the case. The [!͡q] click involves tongue root retraction and a concave tongue body shape similar to [!], while the articulation of [ǂ͡q] is more similar to [ǂ] in terms of posterior constriction location and tongue body and tongue root shapes. The results refute earlier characterizations of clicks with airstream contours as involving a uvular posterior release that contrasts with a velar release in the plain clicks. The results show that, rather, both plain and contour clicks have uvular posterior constrictions as shown by Miller, Brugman et al. (2009), and that the clicks differ in terms of their tongue body and tongue root shapes.

6. Phonological model for N|uu clicks

I propose gestural scores for plain alveolar clicks and alveolar clicks exhibiting airstream contours using Browman and Goldstein's (1989) theory of Articulatory Phonology. The model requires the addition of the tongue root articulator, as well as tongue body shape and tongue root shape, that were not included in the original theory. Distinguishing consonants in Khoesan languages involves describing clicks involving a high flat tongue body shape, [ǂ], and a raised tongue root, as distinct from clicks involving a concave tongue body shape and a convex tongue root shape, such as [!].

I propose two levels of pressure to account for airstream, intra-oral pressure and pharyngeal pressure. This conforms to Mattingly's (1990) appeal that the basic units of speech should be described in terms of articulatory goals, and mirrors the types of aerodynamic components added to the task dynamics model by McGowan and Saltzman (1995).

I assume, following Zsiga (1997) and Fujimura (2000), that there are distinct phonetic and phonological components of grammar. Thus, I also provide the major phonological features that I propose are specified on the clicks described here. I follow Ladefoged (1982, 1997, 2007) in having an airstream feature. In order to capture the inventory of N|uu airstream contrasts, three airstreams are necessary: pulmonic, lingual and glottalic. All of these airstream contrasts occur as simple segments, and linguo-pulmonic and linguo-glottalic contour segments also exist (Miller et al. 2009). The use of an airstream feature allows me to distinguish between plain clicks that have a shift in airstream at the edge of the consonant, from clicks that I analyse as airstream contour segments, which have a shift at the centre of the segment. I assume that cavity volume is related to tongue shape, and thus does not need to be represented separately.

Figure 8 provides a gestural score of the alveolar click within Articulatory Phonology. The three tongue segments are divided and mapped to prosodic structure: moraic, syllabic and foot structure for the vowels, and syllable and foot position only for the consonants. The tongue shape specifications are mapped from the articulatory parametric representations. Time points are marked with reference to Thomas-Vilakati's (1999) phases of click production, as well as acoustic landmarks, which aid the reader in seeing the relationship between the articulatory and acoustic properties.

Thomas-Vilakati (1999) describes click articulation with three phases that parallel the phases of pulmonic stops: (A) the tongue dorsum lead phase, where both anterior and posterior constrictions are made in order to form a cavity (this parallels the shutting phase of pulmonic plosives); (B) the overlap phase, where air is rarefied in order to increase the volume of the velaric (lingual) cavity (this parallels the closure phase of pulmonic plosives); and (C) the tongue dorsum lag phase, which includes both the release of the anterior constriction and the release of the posterior constriction (this parallels the release phase of pulmonic plosives). In addition, the Anterior Release of the click is marked with "AR".

For the alveolar click, the tongue tip raises, forming the constriction at the alveolar ridge, leading to a convex tongue tip shape. The tongue body, which includes the tongue body and dorsum, has a concave shape, with the centre of the tongue body being the lowest point. The tongue root exhibits a convex shape, capturing the fact that the tongue root proper is protruded into the pharynx in the production of the alveolar click.

The first mora of the vowel in the word [!əi] obtains its tongue shape from the preceding consonant via co-production, and thus there is no tongue root retraction gesture associated with it. The vowel [i] has its own tongue shape, which is high and flat. The tongue root shape is in the neutral position. At the phonological level, these map to place of articulation specifications, in terms of [coronal] and [dorsal] specifications, as well as the feature [RTR], which I would classify as a tongue shape feature.

Airstream is specified at the phonological level, and can be either [pulmonic] or [lingual]. I assume that the [pulmonic] specification is the default specification.

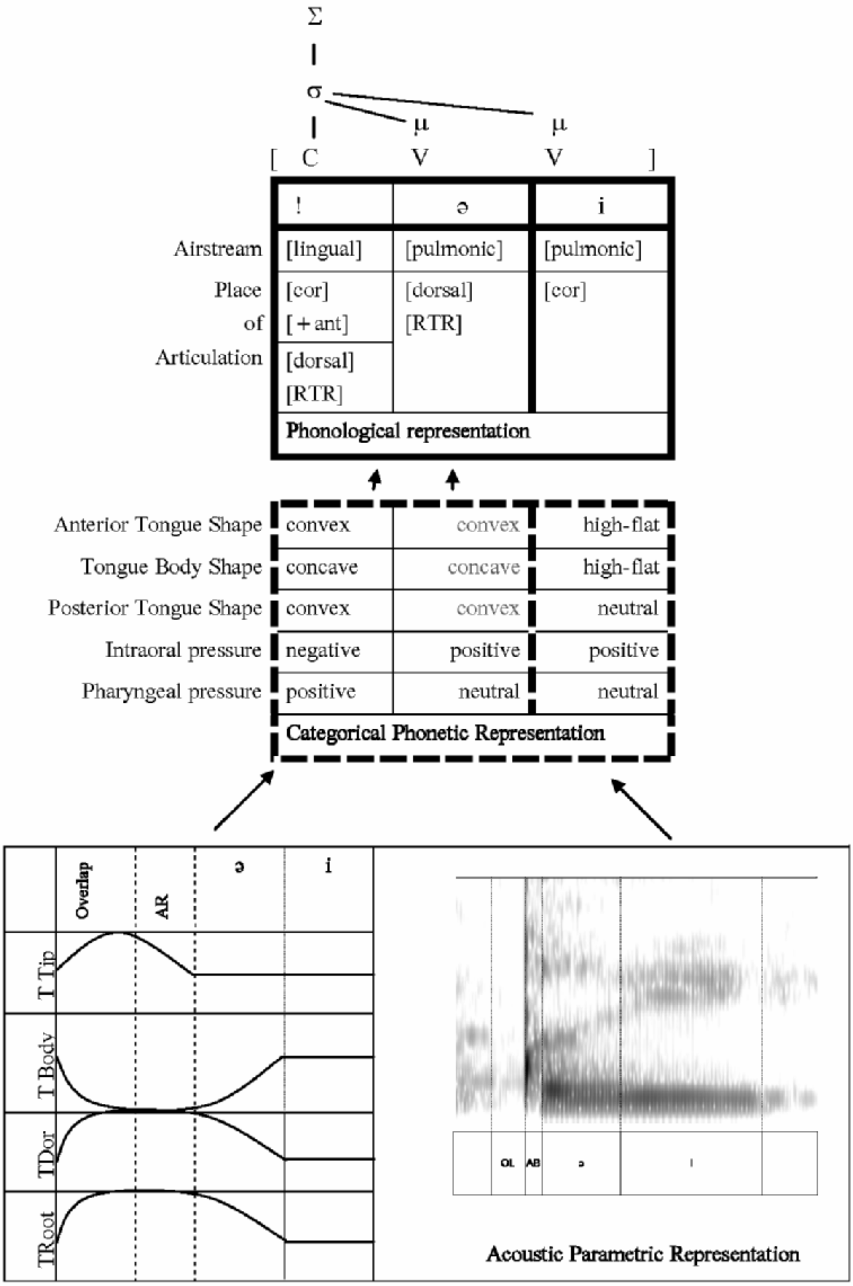
In the palatal click, the tongue body shape is high and flat, and the tongue root shape is neutral, just as in the high front vowel [i]. Thus, there is only a slight co-articulatory effect on a following high front vowel. As noted by Miller, Namaseb and Iskarous (2007) and Miller, Brugman et al. (2009), the muscles found in the articulation of [ǂ] and [i] are compatible, unlike those of [!] and [i]. That is, in a word like ǂii 'don't', there is a gentle lowering effect on the front vowel /i/, which causes it to be realized as [i̞].

There is not a strong backing effect as is found in the production of [!]. The backward movement of the upper tongue root lowers the tongue body. The tongue root proper is not retracted.

Figure 9 provides a gestural score for the alveolar linguo-pulmonic stop within this model. The tongue tip raises up to make an alveolar constriction just as in the fully lingual alveolar stop during the overlap phase (marked "OL"), but the tongue tip returns to neutral position earlier within the segment at the point marked "AR" for anterior release, which corresponds to the click burst in the spectrogram. We can see that the Tongue Dorsum Lag Phase ("DL") is much longer in this click than in the fully lingual alveolar stop (capturing the timing differences seen in the waveforms in Figure 1), and there is also a posterior release in this click that is not found in the plain alveolar click. The posterior release is marked "PR". The tongue root continues returning to neutral position during the vowel following the posterior release, and is responsible for the schwa articulation found in the first mora of the vowel. The lowered F2 and raised F3, as well as the slightly raised F1 seen at the beginning of the vowel in the spectrogram is the result of the lag seen in the tongue root gesture.

This segment is an airstream contour segment, and thus has two timing slots. The first slot is marked for [lingual] airstream, and the second is marked for [pulmonic] airstream. The negative vs. positive intra-oral pressure is marked in the middle panel of Figure 9. At the level of gestures, there is no representation of airflow (though this may need to be captured eventually in something akin to proposals made by McGowan and Saltzman 1995). At this point, I leave it so that the airflow is derived from the particular timing of the individual articulators.

Experiment 2 results showed that the posterior constriction locations in clicks are different for the two classes of clicks presented in Figure 7. Plain alveolar clicks, [!], and alveolar clicks with airstream contours, [!͡q], both involve further forward uvular constrictions, while plain palatal clicks, [ǂ], and palatal clicks with airstream contours, [ǂ͡q], exhibit farther back uvular constrictions. Note that the consonants that were shown to co-occur freely with front vowels in experiment 1 are farther back than those that are blocked from their occurrence with front vowels, and retract the high front vowel /i/ to [əi]. This mirrors the situation with the anterior constriction locations, since the palatal anterior constriction location is farther back than the alveolar one. The results of experiment 2, therefore, show that neither the anterior constriction locations nor the posterior constriction locations can be the phonetic bases of the Back Vowel Constraint in N|uu.



	!	ə	i
Airstream	[lingual]	[pulmonic]	[pulmonic]
Place of Articulation	[cor] [+ant]	[dorsal] [RTR]	[cor]
	[dorsal] [RTR]		
Phonological representation			

Anterior Tongue Shape	convex	convex	high-flat
Tongue Body Shape	concave	concave	high-flat
Posterior Tongue Shape	convex	convex	neutral
Intraoral pressure	negative	positive	positive
Pharyngeal pressure	positive	neutral	neutral
Categorical Phonetic Representation			

Figure 8. Phonetic and phonological representation of the word !əi ‘belch’.

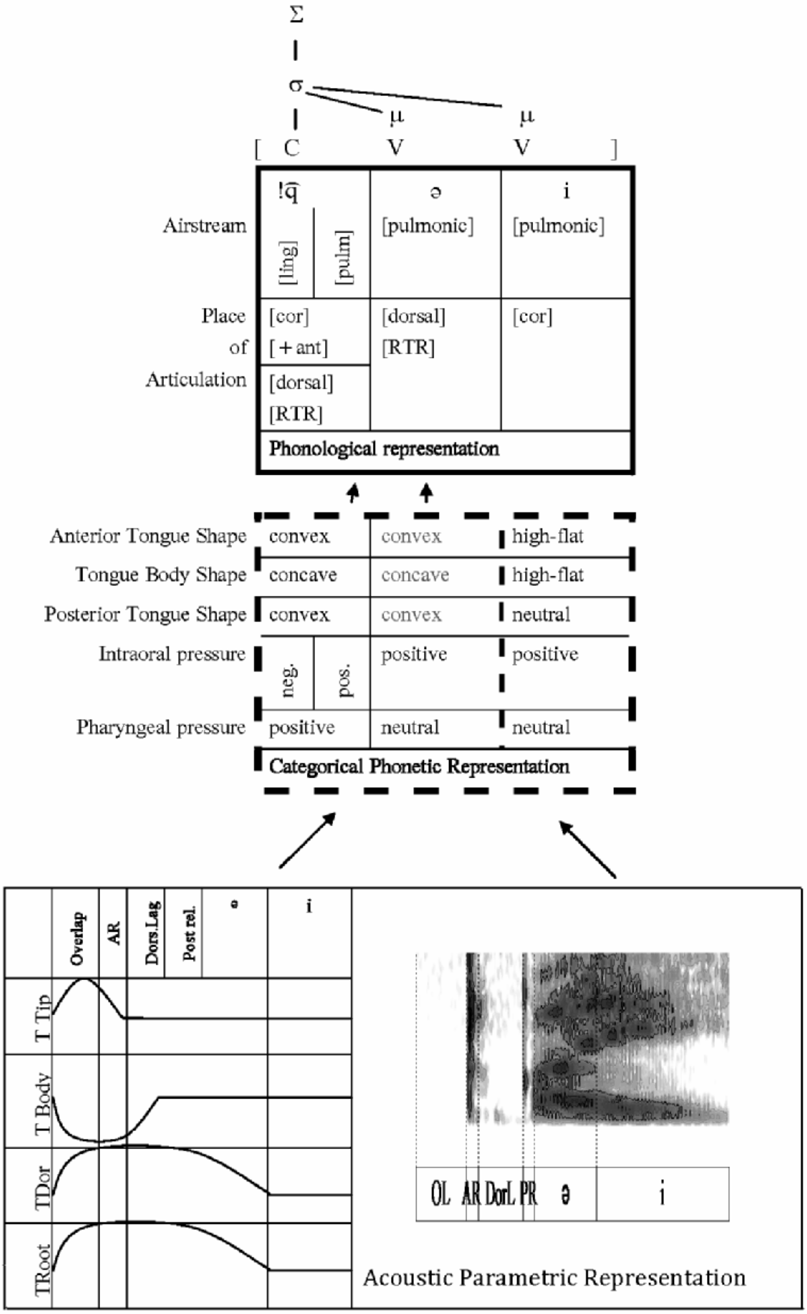


Figure 9. Linguistic representation of the word !q͡əi ‘be behind’.

Experiment 2 results have shown a contrast in the tongue tip, tongue body and tongue root shapes that are used in the production of the alveolar and palatal clicks, as well as the length of both the anterior and posterior constrictions. Thus, the results support Miller, Namaseb and Iskarous' (2007) claim that tongue body shape is the phonetic bases of the BVC. They also suggest that the shape of the tongue root may be, in part, responsible for the patterns seen. Tongue root shape is related to the presence of tongue root retraction in the alveolar [!] clicks and the tongue root raising in the palatal [ǂ] clicks.

Posterior place differences are not in themselves contrastive as they are tied to the anterior constriction differences seen in the clicks. However, the place of articulation of the anterior constrictions does not correctly predict the co-occurrence patterns seen. Thus, N|uu BVC patterns show that predictable phonetic differences (e.g. differences in posterior constriction locations in clicks) are phonologically relevant. Since there are two kinds of clicks that have the same posterior constrictions, anterior place differences are also contrastive. Therefore, redundant articulatory properties are relevant to the phonological patterns that these sounds exhibit.

7. Conclusion

I have provided data from co-occurrence of front and back vowels with simple stops (initial pulmonic stops), complex stops (clicks), and contour segments in terms of airstream (linguo-pulmonic stops), as well as medial consonants in the endangered language N|uu. Lexical frequency was calculated over a 790 root database compiled from the N|uu dictionary based on my field-work with a team of linguists. Disparate pulmonic stop and click (lingual stop) patterns show that anterior place of articulation in clicks is not responsible for the co-occurrence restrictions seen between a class of N|uu consonants and front vowels. I have provided ultrasound traces, which show that the posterior constriction locations also do not predict the patterns seen. It is the tongue tip, tongue body and tongue root shapes, which differ among the alveolar and palatal clicks, that act as the phonetic bases of the Back Vowel Constraint. Although most of the articulatory differences found in the tongue body, dorsum and root are predictable from the anterior constriction differences, the tongue dorsum and root differences found among the clicks are phonologically relevant.

That is, I propose that they are the phonetic bases of the Back Vowel Constraint in N|uu, and possibly other Khoesan languages.

Different places of articulation of the linguo-pulmonic stops do not exhibit differences between the posterior constriction closures and releases that are predicted by Traill's (1985) and Ladefoged and Maddieson's (1996) transcription of them. The contrastive element of these clicks is one of timing. They differ in the duration of the tongue dorsum lag phase. As shown by Miller, Brugman and Sands (2007), the release phase of the alveolar click has a duration of about 20 ms, while the release phase of the alveolar click with an airstream contour has a duration of approximately 70 ms. Previously transcribed 'velar' clicks are articulated with the lingual airstream, while previously transcribed 'uvular' clicks are contour segments, with a lingual closure phase, and a pulmonic release phase.

Acknowledgements

This research is supported by NSF # BCS-0726200 (Miller PI) and BCS-0726198 (Sands PI): "Collaborative Research: Phonetic and Phonological Structures of Post-velar Constrictions in Clicks and Laterals", and NSF # BCS-0236735 (Miller & Collins, co-PIs) and (BCS-0236795, Sands, PI). "Collaborative Research: Descriptive and Theoretical Studies of N|uu". Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

I would like to acknowledge Abigail Scott and the editors for comments on earlier versions of this paper. I would like to thank Johanna Brugman and Jonathan Howell, who assisted with the collection of the ultrasound data that is plotted here, and Peggy Renwick who assisted with the data processing.

References

Beach, Douglas M.
1938 *The Phonetics of the Hottentot Language*. Cambridge: Heffer & Sons.
Bell, Arthur and Christopher Collins
2001 ǂHoan and the typology of click accompaniments in Khoisan. *Cornell Working Papers in Linguistics* 18: 126–153.
Browman, Catherine and Louis Goldstein
1989 Articulatory gestures as phonological units. *Phonology* 6: 201–251.

Brugman, Johanna
2005 Pharyngeal constrictions in the vowels of N|uu. Paper presented at Ultrafest III, Tucson, Arizona.

Brugman, Johanna
2009 Segments, tones, and their distribution in Khoekhoe prosody. Ph.D. diss., Cornell University.

Clements, George N. and Elizabeth Hume
1995 The internal organization of speech sounds. In: Goldsmith, John. (ed.) *Handbook of Phonological Theory,* 245–306, Oxford: Blackwell.

Doke, Clement Martin
1923 Notes on a problem in the mechanism of the Zulu clicks. *Bantu Studies* 2(1): 43–45.

Doke, Clement Martin
1936 An outline of ǂKhomani Bushman phonetics. *Bantu Studies* 10(4): 433–461.

Epstein, Melissa and Maureen Stone
2005 The tongue stops here: Ultrasound imaging of the palate. *Journal of the Acoustical Society of America* 118(4): 2128–2131.

Esling, John H.
2005 There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics* 50(1/2/3/4): 13–44.

Fowler, Carol
1980 Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8: 113–133.

Fujimura, Osamu
2000 The C/D model and prosodic control of articulatory behavior. *Phonetica* 57: 128–138.

IPA
2006 The International Phonetic Alphabet (revised to 2005) [chart]. *Journal of the International Phonetic Association* 36(1): 135.

Gick, Bryan, Sonya Bird, and Ian Wilson
2005 Techniques for field application of lingual ultrasound imaging, *Clinical Linguistics and Phonetics* 19(6/7): 503–514.

Jakobson, Roman, Gunnar Fant, and Morris Halle
1952 *Preliminaries to Speech Analysis (Technical Report 13).* Acoustics Laboratory, MIT.

Johnson, Keith
1993 Acoustic and auditory analyses of Xhosa clicks and pulmonics. *UCLA WPP* 83: 33–45.

Ladefoged, Peter
1982 *A Course in Phonetics*, 2nd Edition, New York: Harcourt Brace.

Ladefoged Peter

1997 Linguistic phonetic descriptions, In: Hardcastle, William and John Laver (eds.), *Handbook of Phonetic Sciences*, 589–618. Cambridge: Blackwell.

Ladefoged , Peter

2007 Articulatory features for describing lexical distinctions. *Language* 83(1): 161–180.

Ladefoged, Peter and Anthony Traill

1994 Clicks and their accompaniments. *Journal of Phonetics* 22: 33–64.

Ladefoged, Peter and Ian Maddieson

1996 *Sounds of the World's Languages*, Cambridge, MA: Blackwell.

Li, Min, Chandra Khambamettu, and Maureen Stone

2005 Automatic contour tracking in ultrasound images. *International Journal of Clinical Linguistics and Phonetics* 19(6/7): 545–554.

Mattingly, Ignatius

1990 The global character of phonetic gestures. *Journal of Phonetics,* 18: 445–452.

McCarthy, John

1994 The phonetics and phonology of Semitic pharyngeals. In P. Keating (ed.) *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form.* Cambridge: Cambridge University Press, 191–233.

McGowan, Richard and Elliott Saltzman

1995 Incorporating aerodynamic and laryngeal components into task dynamics. *Journal of Phonetics* 23: 225–269.

Miller, Amanda

2010 The representation of clicks. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Rice, Keren (eds). *The Blackwell Companion to Phonology.* Oxford: Blackwell Publishing Ltd. (in press)

Miller, Amanda, Johanna Brugman, and Bonny Sands

2007 Acoustic and Auditory analyses of N|uu lingual and pulmonic stop bursts*, Proceedings of the 16th International Congress of Phonetic Sciences* (Trouvain, J. and Barry, W., Eds.), Pirrot GmbH, Dudweiler, Germany, 769–772.
(Available online at http://www.icphs2007.de/)

Miller, Amanda, Johanna Brugman, Bonny Sands, Levi Namaseb, Mats Exter, and Christopher Collins

2007 The Sounds of N|uu: place and airstream contrasts, In: Lee, J. and Pittyaporn, P., (eds.), *Working Papers of the Cornell Phonetics Laboratory* 16: 101–160.

Miller, Amanda, Johanna Brugman, Bonny Sands, Levi Namaseb, Mats Exter, and Christopher Collins

2009 Differences in airstream and posterior place of articulation among N|uu lingual stops. *Journal of the International Phonetic Association* 39(2): 129–161.

Miller, Amanda, Levi Namaseb, and Khalil Iskarous

2007 Posterior tongue body constriction locations in clicks. In: Cole, Jennifer and Hualde, Jose (eds.), *Laboratory Phonology 9*, (Phonology and Phonetics Series), 643–656, Berlin: Mouton de Gruyter.

Miller, Amanda, Abigail Scott, Bonny Sands, and Sheena Shah

2009 Rarefaction gestures and coarticulation in Mangetti Dune !Xung clicks, *In* M. Uther, R. Moore & S. Cox, (eds.), *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009),* Causal Productions: Brighton, U.K.: 2279–2282.

Miller-Ockhuizen, Amanda

2000 C-V coarticulation and complex consonants: Evidence for ordering in click place gestures, In: Fujimura, Osamu, Joseph, Brian and Palek, B., (eds.), *Proceedings of LP'98,* Prague: Charles University Press: 301–330.

Miller-Ockhuizen, Amanda

2003 The Phonetics and Phonology of Gutturals: A Case Study from Ju|'hoansi. *Outstanding Dissertations in Linguistics*. New York: Routledge.

Miller-Ockhuizen, Amanda

2010 A prosodic account of Ju|'hoansi consonant distribution patterns, *Proceedings of the 2nd International Symposium on Khoisan linguistics* (Brenzinger, M. and Konig, C., Eds.), Koeln: Ruediger Koeppe Verlag, 40–73.

Nakagawa, Hiroshi

2006 Aspects of the phonetic and phonological structure of the G|ui language. Ph.D. diss., University of the Witwatersrand, Johannesburg.

Rose, Sharon

1996 Variable laryngeals and vowel lowering, *Phonology* 13: 73–117.

Sagey, Elizabeth

1990 The Representation of Features in Non-linear Phonology: The Articulator Node Hierarchy. Hankamer, Jorge, (ed.) *Outstanding Dissertations in Linguistics*, New York: Garland.

Sands, Bonny

1991 Evidence for click features: acoustic characteristics of Xhosa clicks. *UCLA Working Papers in Phonetics* 80: 6–37.

Sands, Bonny, Amanda Miller, and Johanna Brugman

2007 The lexicon in language attrition: the case of N|uu. In: Payne, Doris and Peña, Jamie, (eds.), *Selected Proceedings of the 37th Annual Conference on African Linguistics*, Somerville, MA: Cascadilla Proceedings Project: 55–65.
(Available online at http://www. lingref.com/cpp/acal/37/index.html)

Stone, Maureen

2005 A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics* 19(6-7): 455–502.

Thomas-Vilakati, Kimberley

1999 Coproduction and coarticulation in IsiZulu clicks. Ph.D. diss., University of California at Los Angeles.

Thomas-Vilakati, Kimberley

2009 Coproduction and coarticulation in IsiZulu clicks. To appear in UC Publications in Linguistics.

Traill, Anthony

1985 *Phonetic and Phonological Studies of !Xóõ Bushman* (Quellen zur Khoisan-Forschung 1). Hamburg: Helmut Buske Verlag.

Traill, Anthony

1994 *A !Xóõ Dictionary* (Quellen zur Khoisan-Forschung 9). Köln: Rüdiger Köppe Verlag.

Traill, Anthony

1997 Linguistic phonetic features for clicks: articulatory, acoustic and perceptual evidence. In: Herbert, Robert K. (ed.), *African Linguistics at the Crossroads: Papers from Kwaluseni*, 99–117. Köln: Rüdiger Köppe.

Westphal, Ernst

1953–1957 *Ng|huki*. BC 1143 (C12), E O J Westphal Papers. Unpublished notes housed at the Rare Documents and Manuscripts Department, University of Cape Town.

Zsiga, Elizabeth

1997 Features, gestures, and Igbo vowels: an approach to the phonology-phonetics interface. *Language* 73(2): 227–274.

Do differences in male versus female /s/ reflect biological or sociophonetic factors?

Susanne Fuchs and Martine Toda

1. Theoretical background

For many decades researchers have been preoccupied with the search for invariance at the acoustic, articulatory and motor control levels which could potentially correspond to phonological entities. This undertaking has more recently been redirected towards the discussion of the *extent of variability* and towards the question of *what causes variability* (Perkell et al. 2004, Newman, Clouse and Burnham 2001). Explanations in the literature are manifold: variability can be a consequence of (a) phonemic inventory; (b) phonemic context; (c) stress, speech rate or loudness; (d) can be driven by perceptual requirements; (e) caused by the position of the segment in the syllable, word and other constituents; (f) can be a result of the dialectal origin of the speaker, the speaker's gender, the speaker's vocal tract, age or mood, among many other possibilities.

For the purposes of the current study we will differentiate between *biological* and *non-biological* explanations. In our view, biological factors are those that cause variability due to the bio-physical properties of the speech production and perception systems. These properties change during ontogenesis (Fuchs, Pompino-Marshall and Perrier 2007). For example, differences in fundamental frequency between young children and male adults should to a great extent be attributed to differences in the size of the vocal folds and their respective mass and tension, i.e., properties that are anatomically given.

In contrast, non-biological factors arise from speech and language as a collective behaviour which is learnt during speech acquisition and is both adaptable and robust during ontogenesis (Pierrehumbert 2006). So far empirical evidence from socially conditioned variability comes mainly from the areas of sociolinguistics, sociology, and psychology. However, as phonetics has become more and more interdisciplinary, researchers have also devoted increasing attention to this topic, for example in a special issue of the Journal of Phonetics (edited by Jannedy and Hay 2006) on socially conditioned variability in speech production and perception. Pierrehumbert and Clopper (forthcoming) even call sociophonetic issues the next challenge for Laboratory Phonology.

Our present work focuses on the potential biological versus sociophonetic origin of phonetic differences in male versus female /s/ production, an issue which has previously been raised on the basis of acoustic and perceptual data. We adapt the terminology used by Stuart-Smith (2007) and many others and call male-female differences *sex* differences when they are based on biological factors. The term g*ender* will be used instead whenever differences are grounded in sociophonetic issues. In other words:

> "Gender is described as what we 'do', or what we 'perform': gender doesn't just exist, but is continually produced, reproduced, and indeed changed though people's performance of gendered acts, as they project their own claimed gender identities, ratify or challenge others' identities, and in various ways support or challenge systems of gender relations and privilege "(Stuart-Smith 2007, p. 66 citing Eckert and McConnell-Ginet).

1.1. Acoustic and perceptual evidence for differences in male and female /s/ production

At this point we would like to provide some examples from the literature of differences in male versus female /s/ realization. The earliest work we are aware of comes from the perceptual domain. Schwartz (1968) studied listeners' ability to identify speakers' sex from a variety of isolated voiceless fricatives. Results of his study showed that listeners were able to identify speakers' sex for the sibilants, but not for the more front articulated /f, θ/. The former result was attributed to the higher frequencies in females' realizations in comparison to males' realizations of the sibilants. Schwartz supposed that these distinctions are based on vocal tract differences. Similarly, Johnson (1991) discussed his perceptual findings on speaker and vowel variability in a /s/ to /ʃ/ continuum with respect to biological factors, i.e., to anatomical differences between males and females.

A detailed literature review of studies of /s/ is provided in Flipsen et al. (1999). In their own acoustic investigation with 26 American English speakers between 9 and 15 years of age they found significant sex-related spectral differences, in particular in the frequency mean and skewness parameter at the fricative midpoint.

Gordon, Barthmaier and Sands (2002) investigated fricative inventories in seven genetically unrelated languages. They used the acoustic centre of gravity parameter and found male-female differences in Chikasaw. The authors reported no differences in the other languages, but this may reflect small sample sizes and that the authors grouped several fricatives together.

Additional evidence comes from an acoustic study by Heffernan (2004) which included 12 Canadian English speakers and 10 Japanese speakers. He found significant male-female differences for the acoustic centre of gravity parameter. The male-female distinction was generally more pronounced in the Canadian English speakers than in the Japanese speakers. He interpreted these results with respect to a social component to explain male-female differences, although a potential mixture with physiological factors was discussed too.

1.2. How can male-female differences in /s/ be explained?

The question which motivated our own work is: Are the acoustic differences in males' and females' realization of /s/ a consequence of differences in vocal tract size, or are they a consequence of a sociophonetic process in which females actively produce /s/ with different articulatory strategies than males as an index of their gender? The former explanation relates to the biological nature of speech underlying interspeaker variation whereas the latter is grounded in the social function of speech. A combination of the two explanations may be possible as well.

Support for the biological explanation includes the fact that acoustic male-female differences are found in various genetically unrelated languages (e.g., English, Japanese, and Chickasaw) and they always go in the same direction, with females realizing /s/ with higher frequencies than males. Stevens (1998, p. 398) also notes that the length of the cavity downstream of the oral constriction is somewhat smaller for females than for males. Since small differences in the size of the front cavity, especially in its length, are crucial for the spectral properties of /s/, sex differences can be expected. Another argument in favor of the biological explanation would be that acoustic differences are often found in sibilants where vocal tract differences could matter, but not in labiodentals, where vocal tract differences do not matter.

Support for the gender explanation comes from sociophonetically grounded work (Stuart-Smith 2007, Strand 1999, Heffernan 2004, Munson et al. 2006). One of the main arguments of these studies is that morphological differences between males and females are the most distinct in the back part of the vocal tract, i.e., in the pharynx (see Fitch and Giedd 1999 for a very comprehensive study), but are less influential (if they ever occur) at the place where alveolar fricatives are realized. Another piece of evidence comes from experiments including males and females of different ages and social classes. Stuart-Smith (2007), for instance, reports for Glaswegian English that although there may be vocal tract differences between younger (13–14 years) and older (40–60 years) males, the measured acoustic parameters were relatively homogeneous. By contrast, the young working class girls behaved similarly to all males, but differed from all other females (older working class woman, middle class women, and middle class girls).

Another recent study from Munson et al. (2006) supports the importance of sociophonetic factors in /s/ production. They investigated the acoustic and perceptual bases of women's and men's sexual orientation by means of homosexuals', bisexuals' and heterosexuals' speech. Their acoustic findings provide evidence that homosexual and heterosexual men's speech differs significantly in the spectral skewness parameter of /s/. (Results for lesbian and heterosexual females, however, did not reveal such patterns.) In a second experiment Munson et al. were able to show that listeners' judgements of perceived sexual orientation were related to the acoustic findings.

Stuart-Smith's and Munson et al.'s findings suggest a primary role of gender identity in the realization of /s/ and refute a primary role of biological causes.

However, none of the studies we are aware of obtained direct morphological or articulatory data from their subjects. The current study is a first attempt to close this gap by means of electropalatographic (EPG) data simultaneously recorded with acoustics. Additionally, we gathered morphological data for the size of the speaker's palate from the electrodes located in the EPG palate. The main goal of this study is to determine whether male versus female /s/ realization is best accounted for by biological or by sociophonetic factors. In the next section we will outline how this question was addressed.

2. Methodology

The corpus we used was originally constructed to study the influence of the palate shape on token-to-token variability (Brunner, Fuchs and Perrier 2009). Here we will only consider word-medial /s/ in the target words /sasa/ for the English speakers and /zasa/ for the Germans. Target words were embedded in the carrier phrase ‘Say __ please.’ for the English subjects and ‘Habe ___ gesagt.’(Have …. said.) for the German subjects. The target sibilants occurred word medially in an ambisyllabic post-stressed position.

Table 1. Participants of the study with their language and gender abbreviation (E = English, G = German, M = male, F = female) and number of repetitions (N).

Speaker	Language (regional origin)	N
EM1	English (Australian English)	30
EM2	English (American English)	30
EM3	English (Scottish English)	30
EM4	English (RP)	29
EM5	English (RP)	28
EM6	English (Scottish English)	30
EF1	English (RP)	30
EF2	English (Scottish English)	29
EF3	English (RP)	30
EF4	English (RP)	24
EF5	English (RP)	29
EF6	English (RP)	30
GM1	German (Bavarian)	30
GM2	German (Alemannic)	30
GM3	German (Northern German)	30
GM4	German (Northern German)	30
GM5	German (Bavarian)	30
GM6	German (Northern German)	25
GF1	German (Bavarian)	23
GF2	German (Northern German)	30
GF3	German (Northern German)	31
GF4	German (Northern German)	30
GF5	German (Saxonian)	30
GF6	German (Northern German)	31
Σ 24	English: 12 speakers German: 12 speakers	699

All sentences were presented in a randomized order and repeated thirty times (with some exceptions cf. table 1). Recordings of the German-speaking subjects took place in a professional soundproofed room at Zentrum für Allgemeine Sprachwissenschaft (ZAS) in Berlin using the EPG 3 Reading system, and the English-speaking subjects were recorded at Queen Margaret University College (QMUC) Edinburgh using the Win EPG system. All recordings, however, involved the same kind of electrode arrangement on the artificial palate of the EPG 3 system. All EPG data have a sampling frequency of 100 Hz. Acoustic data were recorded with a sampling frequency of 44 kHz for the English speaking subjects.

For the German speaking subjects (except GM2, GM4, GF3, and GF6) and one English speaker (EM4) data were recorded with 48 kHz, but they were further downsampled to 16 kHz since the corpus was originally not recorded to study sibilants. The acoustic data with the original sampling rate are no longer available. Altogether 24 speakers were recorded (4 more speakers than the original corpus), 6 male and 6 female native speakers each of German and English. Since our experimental set-up restricted us to those subjects who had an EPG palate available, our sample is not equally well balanced with respect to age (ranging from the early 20s to mid 50s), speaker's origin (see table 1), and as can be seen at a later point, to palate shape. All speakers are academics. Table 1 provides an overview of the participating subjects.

2.1. Measuring palatal morphology

The artificial palate of each subject was used to measure several morphological parameters. The Reading EPG system allows morphological data to be obtained since the electrodes in the artificial palate are placed with respect to anatomical landmarks (for more details see, Hardcastle, Gibbon, and Jones 1991). To do this, the palate was fitted to its dental cast and a high quality 1:1 photocopy was made for each speaker. The copy served as a reference to measure the horizontal (x) and vertical (y) coordinates of each of the 62 electrodes by means of a caliper.

The anterior palatal width (Width_ant) was defined as the horizontal distance between the two most peripheral electrodes in the first (anterior) row. The posterior palatal width (Width_post) was defined as the horizontal distance between the two most peripheral electrodes in the last row. The length of the palate (Length) was calculated as the vertical distance between the two most peripheral electrodes on the left side of the palate plus the vertical distance between those at the right side divided by two.

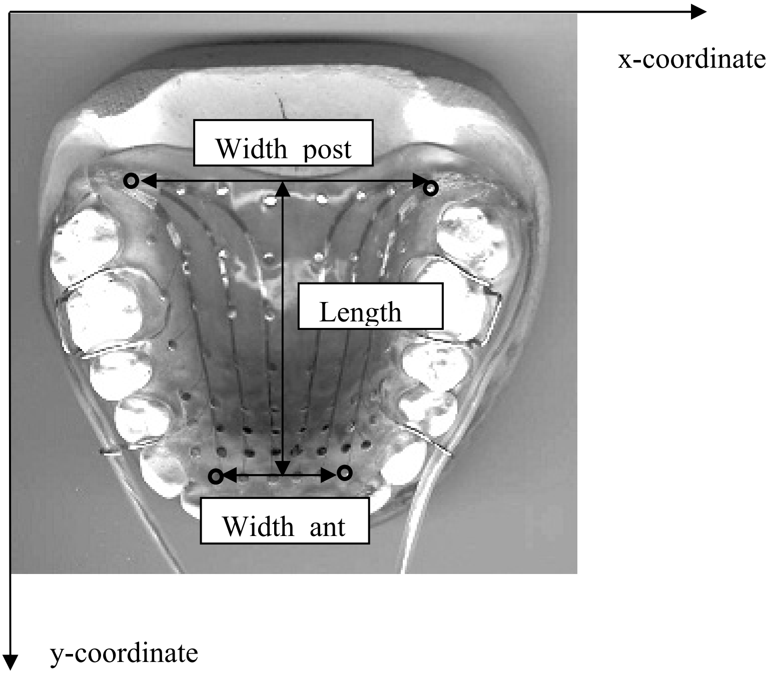


Figure 1. Copy of a dental cast with an EPG palate involving 62 electrodes. Width_post corresponds to posterior palatal width, Width_ant to anterior palatal width, and Length to the distance between the two most extreme points in the vertical dimension.

2.2. Acoustic and articulatory parameters

Based on the acoustic data we defined the beginning and end of the high frequency noise interval of intervocalic /s/ and calculated the acoustic fricative midpoint. Then we plotted the articulatory data across repetitions for each speaker at the fricative midpoint as contact frequency plots. In such a plot the percentage of tongue-palate contact of each electrode corresponds to a colour. White fields correspond to 0–25 percent contact, fields in light grey to 26–50 percent, fields in dark grey to 51–75 percent, and fields in black to 76–100 percent contact. Contact frequency plots for all speakers are shown in figures 4 and 5 in the results section.

The contact frequency plots and the morphological measures of the electrodes' x- and y-coordinates served as the basis for calculating the constriction width (Constr) and provided an equivalent parameter for the length of the front cavity (FrontCav). For measuring Constr, the electrode row which showed the narrowest air channel and the highest percentage of tongue-palate contact on both sides of the channel in the contact frequency plot was selected. For example, in EM3's data (see figure 2, left) the row with the narrowest air channel is the first row, since only one electrode shows no contact (marked with arrows). In the second row a wider channel with two electrodes can be found. In cases where two rows had similar properties we chose the most anterior one. For instance in GM1's data the third and fourth rows (see figure 2, right) both show the narrowest constriction, but the third row was taken for further calculations. We then calculated the distance between the electrodes on both sides of the smallest constriction. A crucial factor for the reliability of this measure is the density of the electrode placement in the artificial palate. As can be seen in figure 1, the more anterior the electrodes, the closer they are to each other (on average for all speakers 2.89 mm in the first two rows, 3.13 mm in the third and fourth rows), and the more posterior the electrodes, the greater the distance between them (more than 4 mm in the posterior rows). A larger electrode distance increases the potential error since the tongue may be in contact with the palate where no electrode is available. Hence the more back the articulation, the rougher the Constr parameter.

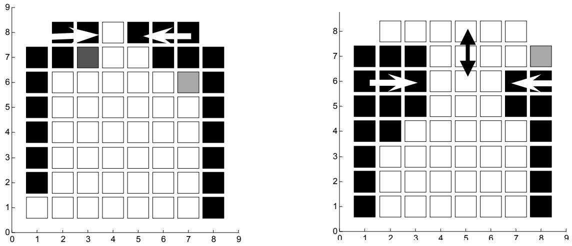


Figure 2. EPG frequency plots for EM3 (left) and GM1 (right). The white arrows mark the row chosen to define the Constr parameter. The distance between the electrodes on the right and left sides of the constriction was used to calculate the constriction width. The black vertical arrow on the right (GM1) connects the electrodes which were used for calculating the FrontCav parameter.

A similar issue arises for the calculated parameter FrontCav, where we subtracted the vertical placement (in mm) of the electrode in the channel showing minimal constriction from the corresponding electrode in the first row. If the minimal constriction occurred in the first row, the FrontCav was zero. The front cavity length is not necessarily zero in reality, since the first electrode row is approximately 2–3 mm away from the necks of the teeth. Hence the FrontCav parameter is again a rough estimate and is shorter than the real front cavity length.

We also calculated the articulatory Centre of Gravity index (COG_ar) for the acoustically defined fricative midpoint. The COG_ar is a weighted index in the front-back dimension with higher values corresponding to a more front place of articulation and lower values to a more back articulation (Hardcastle, Gibbon, and Nicolaidis 1991). The index is described by the formula (where R = row):

$$\text{COG\_ar} = \frac{0.5\ \text{R8}+1.5\ \text{R7}+2.5\ \text{R6}+3.5\ \text{R5}+4.5\ \text{R4}+5.5\ \text{R3}+6.5\ \text{R2}+7.5\ \text{R1}}{\text{total number of contacts}}$$

As can be seen, the more front the articulation, the more weight is added, e.g., in the first row R1 it is 7.5 while in the third row R3 it is only 5.5.

The three articulatory parameters COG_ar, FrontCav, and Constr were selected since we predict that they affect the acoustic properties of sibilants and may differ between males and females. As reported earlier, females often realize higher frequencies with a larger intensity than males (Flipsen et al. 1999). For the females we assume an underlying articulatory strategy with a more front articulation, a shorter front cavity length and/or a narrower minimal constriction while for males we suppose a more posterior place of articulation with a longer front cavity, and/or a wider constriction. The parameters COG_ar and FrontCav were both used since COG_ar is a normalized value and does not take into account individual palate length while FrontCav is an absolute value.

Choosing among the potential acoustic parameters for describing inter-speaker variation of /s/ was a challenging task, since there is no generally accepted parameter in the literature, and some of the parameters studied so far are partially redundant. Although partially redundant, we decided to use the main spectral peak (Peak), the acoustic Centre of Gravity (COG_ac), and Skewness. The Peak parameter was selected because Hughes and Halle (1956) suppose an inverse relation between the length of the front cavity and the frequency of the most prominent peak. The Peak was defined at the fricative midpoint in the LPC-smoothed spectrum of the frication noise with 3 LPC coefficients for the data with a sampling rate of 16 kHz and 4 coefficients for the data with a sampling rate of 22 kHz. The peak detection was limited to the range of 2–10 kHz for the English speakers and the German speakers GM2, GM4, GF3 and GF6 and to 2–8 kHz for the remaining German speakers. The lower limit is intended to avoid detection of voicing, if any should occur. This limit should not interfere with the detection of peaks related to place of articulation, since they are generally located above 4 kHz for alveolars.

Second, although it is similar to the Peak, we calculated the acoustic Centre of Gravity (COG_ac) as the first spectral moment or mean by the formula

$$\text{COG\_ac} = \sum (F_i * I_i) / \sum (I_i), \quad \text{for } i = 2 \text{ to } 8 \text{ kHz}$$

where I is the amplitude and F the frequency of the spectrum. The choice of the COG_ac parameter was inspired by Tabain (2001), who reports a high correlation between the articulatory and the acoustic COG (see also Newman et al. 1991). Again, the range below 2 kHz was excluded in order to avoid detection of low frequency prominence due to voicing and the upper boundary was set to 8 kHz in order to compare English and German data, since the limits of the spectral range influence the calculation of COG_ac values.

Third, skewness was calculated following Forrest et al. (1988) since Flipsen et al. (1999), Stuart-Smith (2007) and Munson et al. (2006) report gender and sex differences with respect to this parameter. The frequency range from 700 Hz to the main spectral peak of each token per speaker was considered for the analysis. A single DFT for each window (window length 12.5 ms) was calculated at the fricative midpoint. We did not choose a larger fricative portion in order to compare articulation and acoustics at the same time point.

For statistical analyses we used SPSS (version 15.0). The details of the statistical procedures are given in the results section.

3. Results

3.1. Are there differences in male and female palatal dimensions?

To answer this question a MANOVA (Wilks-Lambda) was carried out in SPSS with all palatal parameters as dependent variables and language and sex as independent factors. The results provide evidence for a significant difference between palatal parameters of English and German speakers ($F = 7.05$, $p = 0.002$), but neither an effect for sex ($F = 1.26$, $p = 0.318$) nor an interaction between language and sex was found. English speakers differ significantly from Germans in anterior palatal width and palatal length (Width_ant: $F = 11.54$, $p = 0.003$, Length: $F = 4.48$, $p = 0.047$); they have on average a longer, but narrower palate. A summary of the findings is given in table 2 below.

Table 2. Descriptive statistics for the morphological parameters of the palate: Min = Minimum, Max = Maximum, Std. dev. = Standard deviation.

Group	**Parameter**	**Min**	**Max**	**Mean**	**Std. dev.**
English males	Width_ant	11.80	15.50	13.56	1.34
	Width_post	27.90	39.80	33.70	4.32
	Length	38.80	49.45	45.21	4.87
English females	Width_ant	13.30	15.20	14.08	0.78
	Width_post	30.70	39.10	33.13	3.02
	Length	37.15	42.80	40.39	2.09
German males	Width_ant	14.10	19.70	16.17	2.19
	Width_post	31.50	50.00	39.00	6.73
	Length	36.60	47.45	40.02	3.81
German females	Width_ant	14.00	17.10	15.53	1.16
	Width_post	30.50	39.00	33.96	3.25
	Length	34.10	43.65	39.03	3.87

The fact that we did not find significant sex differences may lead to the conclusion that if any further acoustic or articulatory male-female differences in /s/ realization are found, they are grounded in sociophonetic factors. However, although not significant, we still found a trend for males to have longer palates than females for the English speakers only. The exceptions are two males who have a short palate. This finding is particularly interesting, since we assume that speakers should compensate for their palatal morphology in order to reach the required acoustic goal.

Thus, if /s/ realization is a marker of gender identity, males with a short palate may compensate for it by means of a more back articulation to increase the front cavity length or by means of a wider constriction to have lower frequencies in comparison to females. In contrast, females with a relatively long palate may compensate for it with a more front articulation to decrease the length of the front cavity or by using a narrower constriction in order to realize higher frequencies in comparison to males.

If only the length of the hard palate mattered (biological factor), no compensation would be expected in the productions of males with short palates and of females with long palates. In this case, palatal length should correlate with place of articulation and front cavity length – as can be seen for the English speakers in our sample.

3.2. Evidence from articulation

Figure 3 displays scatterplots for the English and German speakers with the length of the palate on the y-axis and COG_ar (weighted articulatory index in the front-back dimension) on the x-axis. For the English speakers a strong negative correlation was found (R = -0.78, $r^2 = 0.58$, F = 13.99, p = 0.004). Palatal length predicts about 60 percent of the variance in place of articulation. This influence even increases when we consider not only the COG_ar as a speaker-normalized index, but also the absolute value of the front cavity length, FrontCav (R = 0.79, $r^2 = 0.63$, F = 16.77, p = 0.002).

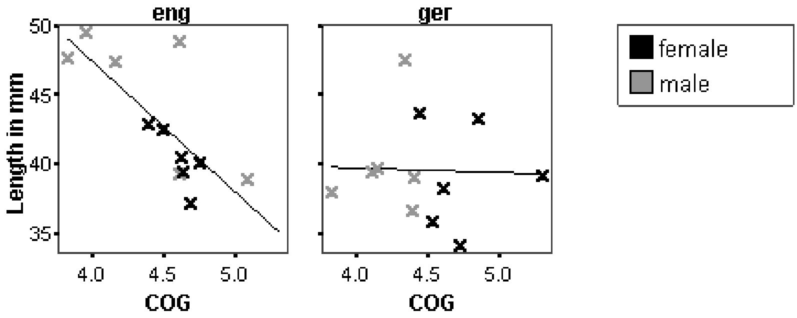


Figure 3. Scatterplots for palatal length and place of articulation (larger COG_ar values = more anterior articulation); left: English speakers, right: German speakers; grey x's = males, black x's = females, linear regression line for the English speakers.

The two males who have a short palate behave articulatorily like the corresponding females and do not compensate for their morphology, which supports a biological explanation of the question under investigation.

For the German speakers no relation between palatal length and COG_ar was found. Independent of palate length, females consistently realize a more anterior articulation in comparison to males. In at least three cases, German males (GM1, GM4, GM6) with a short palate length similar to that of females produce either a relatively posterior articulation or a wider constriction, suggesting a compensatory process.

Contact frequency plots for all the speakers were calculated and are displayed in figures 4 and 5. After a first evaluation two general findings become evident: First, females produce /s/ more anteriorly than males independent of their language. Second, German speakers realize a wider constriction than English speakers. For instance, most female speakers produce the minimal constriction in the first row of the palate; the plots for German females, however, more often show two electrodes with no contact corresponding to the air channel, whereas for the English females it is often only one electrode. Table 3 displays the descriptive statistics.

Table 3. Descriptive statistics for the articulatory parameters: Min = Minimum, Max = Maximum, Std. dev. = Standard deviation.

Group	**Parameter**	**Min**	**Max**	**Mean**	**Std. dev.**
English males	COG_ar	3.83	5.08	4.38	0.48
	Constr	6.00	10.60	7.20	1.76
	FrontCav	0.00	6.50	4.10	2.47
English females	COG_ar	4.39	4.75	4.60	0.13
	Constr	4.70	8.40	6.05	1.28
	FrontCav	0.00	3.00	0.50	1.22
German males	COG_ar	3.83	4.40	4.21	0.22
	Constr	8.20	15.70	11.13	3.09
	FrontCav	0.00	5.80	2.68	2.32
German females	COG_ar	4.44	5.30	4.74	0.31
	Constr	6.10	12.10	8.79	2.35
	FrontCav	0.00	3.70	0.62	1.51

To provide further evidence several ANOVAs were carried out with one of the articulatory parameters as the dependent variable and language and sex as independent factors. COG_ar and FrontCav clearly showed an influence of sex (COG_ar: $F = 8.87$, $p = 0.007$, FrontCav: $F = 12.58$, $p = 0.002$), and

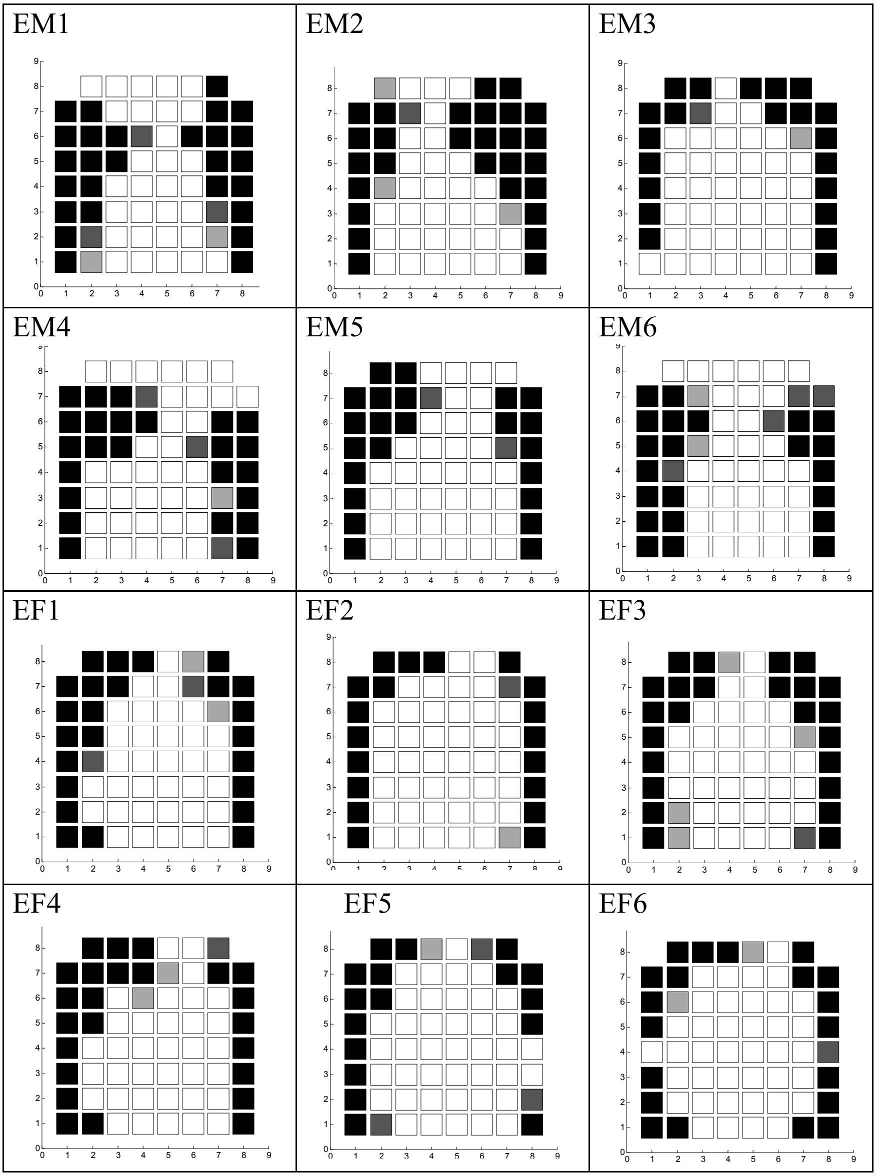


Figure 4. Frequency plots for all English speakers with the first two rows for males and the last two rows for females. Black markers correspond to 76–100% tongue-palate contact with respect to all the subject's repetitions, dark grey markers correspond to 51–75%, light grey to 26–50%, and white markers to 0–25%. Upper incisors are located above the first row.

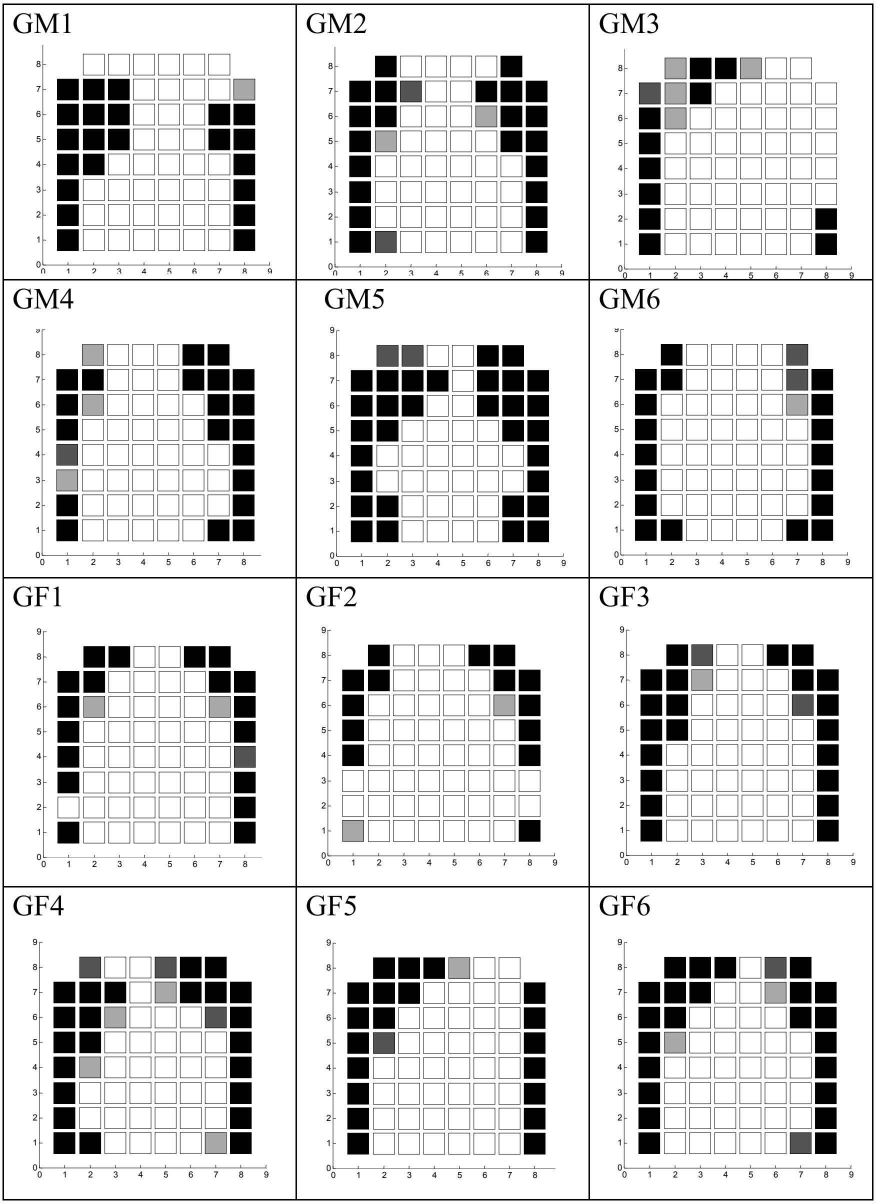


Figure 5. Same as figure 4, but for German subjects.

Constr differed significantly between the two languages (F = 13.47, p = 0.002). All results are significant below the p < 0.016 level after Bonferroni correction.

Hence, there is evidence for articulatory differences regarding place of articulation for males and females. Since we also noted a potential covariation of COG_ar and FrontCav with palatal length (see figure 3 for English speakers), it is uncertain whether the male-female differences are biological or sociophonetic in nature. To rule out possible biological factors, we ran the ANOVAs again, but included each time one of the palatal parameters as a covariate. If findings are still significant, the articulatory differences between males and females cannot be explained by morphological differences. Including palatal length as a covariate resulted in a decrease in the distinctive power, but findings are clearly significant with respect to sex and language (sex effect for COG_ar: F = 4.9, p = 0.039, FrontCav: F = 7.36, p = 0.014; language effect for Constr: F = 10.86, p = 0.004). Including Width_ant had only a marginal influence. The most pronounced influence was found for the Width_back covariate since it caused sex-specific differences not only in COG_ar and FrontCav, but also in constriction width (Constr: F = 4.48, p = 0.048).

To summarize, male-female differences in the articulation of /s/ are still persistent even if morphological factors are ruled out.

3.3. Results for acoustic parameters

An overview of the acoustic results is given in table 4. Results for the English females consistently show an extraordinarily high main spectral peak, on average above 8 kHz. This is quite different from the data for English speaking males and from all the German data, which do not show a male-female distinction in the Peak parameter. We do not believe that these differences between English and German females are due to the lower sampling frequency of the acoustic data of the Germans for the following reasons: (a) There were two German females for whom we had the same sampling frequency as for the English speakers, but peak values were still lower in comparison to the English females; (b) EPG data provide evidence for a wider constriction (Constr parameter) in the German data, which should affect the main spectral peak in the direction found in the acoustics, i.e., lowering the main spectral peak (Shadle 1991); (c) Peak values were not included in the statistics for those cases where the front slope was still rising or did not decrease after the peak. For GF4, 10 repetitions were discarded from the recordings and for GF1 seven, but mainly because their spectrum was relatively flat. For all other German females main peaks were consistently found.

Table 4. Descriptive statistics for the acoustic parameters: Min = Minimum, Max = Maximum, Std. dev. = Standard deviation.

Group	**Parameter**	**Min**	**Max**	**Mean**	**Std. dev.**
English males	Peak	4713	8140	6284	1229
	COG_ac	4757	6167	5632	644
	Skewness	-0.47	0.06	-0.25	0.20
English females	Peak	7041	9017	8282	722
	COG_ac	5722	6856	6412	428
	Skewness	-1.23	-0.47	-0.86	0.32
German males	Peak	4848	6785	5721	814
	COG_ac	4006	6345	5463	903
	Skewness	-0.89	-0.02	-0.44	0.31
German females	Peak	4778	6556	5841	716
	COG_ac	5358	6366	5859	328
	Skewness	-1.38	-0.20	-0.57	0.45

Several ANOVAs were also carried out for the acoustic data. Findings for the Peak parameter showed an interaction of the two main effects language*sex ($F = 6.6$, $p = 0.018$) with differences between male and female English speakers, but none for the German speakers. However, it failed to reach significance after Bonferroni correction ($p < 0.016$). For the Skewness parameter an effect of sex was found (Skewness: $F = 7.24$, $p = 0.014$), again with more negatively skewed data for the female spectra. COG_ac did not reach significance after Bonferroni correction.

In summary, only the acoustic Skewness parameter showed significant differences between males' and females' /s/ realization. All other parameters showed a strong trend in the expected direction, but did not reach significance after correction for multiple tests.

3.4. Linking articulation and acoustics

In a next step the relation between articulatory and acoustic parameters was investigated. Spearman Rho correlations showed that for the English speakers the most important articulatory correlate corresponding to the acoustics is the FrontCav (correlation with Peak: R = -0.755, p = 0.004, with COG_ac: R = -0.839, p = 0.001, with Skewness: R = 0.767, p = 0.004). It is followed by the articulatory COG_ar correlating with the acoustic COG_ac (R = 0.630, p = 0.028). This finding clearly supports earlier work from Hughes and Halle (1956), who suggest an inverse relation between front cavity length and the spectral peak. It is also in line with modeling work from Fant (1960), who mentions the importance of the length of the front cavity for the spectral characteristics of /s/.

For the German speakers no correlation at all was found, i.e., there is at least no simple linear relation between articulatory and acoustic parameters. But German speakers also differ significantly from the English speakers in palatal length, anterior palate width, and with respect to the produced constriction width.

4. Discussion and conclusion

In order to answer the question as to whether differences that have been observed for male versus female /s/ realizations are biological or sociophonetic in nature, articulatory, acoustic and morphological data were gathered from 12 English and 12 German speakers with 6 males and 6 females for each language.

In a first step we compared the palatal size parameters for males and females, assuming that this is the relevant part of the vocal tract where /s/ is realized and where males and females may potentially differ. Significant results for morphological parameters were not found that correlated with the sex of the speaker, but significant results were found which correlated with the language of the speaker. We are not certain whether these morphological differences are due to our small sample or whether they are representative for the differences between the German and English population. English speakers had on average a longer palate and narrower anterior palate width than Germans. Although not significant, we found a trend for English males to have a longer palate than English females, with two exceptions (two males with relatively short palates). This was particularly intriguing since we supposed that if /s/ distinctions are biological in nature, males with a short palate should not compensate for their palatal length. In contrast, if there are gender differences, males should compensate for their palate length, for instance by means of a back articulation or a wide constriction in order to decrease the high spectral energy. Pooling all data from males and females together yielded a particularly high correlation between palatal length and the length of the front cavity as well as the length of the front cavity and the acoustics for the English speakers. These results support the biological explanation of male-female differences in /s/ realization and speak for the need to obtain speakers' morphological data instead of simply splitting data into male and female results.

However, we not only found support for the biological explanation of /s/ differences, but also for the sociophonetic explanation, i.e., there was a mixture of effects. Potential biological influences were factored out by means of calculating several ANOVAs, where each time one morphological parameter was included as a covariate. Even if the morphological parameters reduced the power of the distinction, males and females still showed significant differences with respect to the articulatory COG_ar and FrontCav. Consequently, females actively produce a more front place of articulation and a shorter front cavity than males. These articulatory differences had an impact on the measured acoustic parameters for the English speakers. We found that the length of the front cavity was the most important parameter correlating with the main spectral peak, with the COG_ac, and with the spectral skewness.

Such a correlation was, however, not found for the German speakers, who differed from the English speakers with respect to palate size and constriction width: Although German females in most cases realized a similar anterior articulation to the English females, their constriction width was significantly wider. This also held true for the males. We assume that the wider constriction yielded the generally lower acoustic frequencies found in the German data.

An additional factor, which could be responsible for the language-specific differences, may be that the English phoneme inventory includes the neighboring /θ/ to the /s/, but the German inventory does not. Jongman, Wayland and Wong (2000) report lower spectral means for English /θ/ (averaged over speakers and vowel contexts ca. 5100 Hz) in comparison to /s/, and Narayanan, Alwan and Haker (1995) found a greater constriction area for /θ/ than for /s/. The way in which German /s/ differs in our data from English /s/ is in the direction of exactly these characteristics of /θ/. We therefore suggest that English /s/ may be more constrained than German /s/ in order to avoid perceptual confusion with /θ/.

Taken together, the findings of this study provide evidence for a mixture of effects contributing to the male-female differences in /s/ production.

Palatal size parameters did not differ with respect to sex, but a trend was found for the English males to have a longer palate than the English females. Hence, differences in palatal size may generally be negligible, but in some cases dependent on the recorded sample. Another important biological factor which was not investigated here, but may also contribute to the often reported male-female distinction might be the length of the incisors. Longer versus shorter incisors may cause differences in the front cavity length and thus, in the acoustics.

Our study provides evidence that if palatal size parameters are ruled out, male-female differences still remain in the articulatory production and acoustic realization of /s/.

Acknowledgements

We would like to thank Jana Brunner for making part of the data available to us, Jim Scobbie, Alan Wrench, and Mariam Hartinger for their help in doing the recordings at QMUC Edinburgh, and Jörg Dreyer for his technical expertise during the recordings at ZAS Berlin and for measuring the placement of the electrodes in the artificial palates. Thanks to Olessia Panzyga and Anke Busler for analyzing acoustic data, to Jonathan Harrington and Stefanie Jannedy for providing background information on sociophonetics, to Ralf Winkler for scripting, to Caterina Petrone for her expertise on statistics, to Laura Koenig and Bernd Pompino-Marschall for proofreading, and to Keith Johnson and Marija Tabain for their comments on an earlier version of this paper. This work was supported by a grant from the BMBF and a grant from the DFH to the PILIOS project. It is dedicated to Dieter Fuchs.

References

Brunner, Jana, Susanne Fuchs, and Pascal Perrier
2009 On the relationship between palate shape and articulatory behaviour. *Journal of the Acoustical Society of America* 125 (6): 3936–3949.
Fant, Gunnar
1960 *Acoustic Theory of Speech Production*. The Hague: Mouton.
Fitch, Tecumseh W., and Jay Giedd
1999 Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America* 106(3): 1511–1522.

Flipsen, Peter Jr., Lawrence Shriberg, Gary Weismer, Heather Karlsson, and Jane McSweeny
1999 Acoustic characteristics of /s/ in adolescents. *Journal of Speech, Language, and Hearing Research* 42(3): 663–677.

Forrest, Karen, Gary Weismer, Paul Milenkovic, and Ronald N. Dougall
1988 Statistical analysis of word initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America* 84(1): 115–123.

Fuchs, Susanne, Bernd Pompino-Marschall, and Pascal Perrier
2007 Is there a biological grounding of phonology? Determining factors, optimisation, and communicative usage. *In* Trouvain Jürgen, and William J. Barry (eds.)*, Proceedings of the 16th International Congress on Phonetic Sciences.* Dudweiler: Pirrot GmbH: 219–224.

Gordon, Matthew, Paul Barthmaier, and Kathy Sands
2002 A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32: 141–174.

Hardcastle, William J., Fiona Gibbon, and W. Jones
1991 EPG in assessment and remediation of speech disorders. *British Journal of Disorders in Communication* 26: 41–74.

Hardcastle, William J., Fiona Gibbon, and Katharina Nicolaidis
1991 EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics* 19: 251–266.

Heffernan, Kevin
2004 Evidence from HNR that /s/ is a social marker of gender. *Toronto Working Papers in Linguistics* 23(2): 71–84.

Hughes, George W., and Morris Halle
1956 Spectral properties of fricative consonants. *Journal of the Acoustical Society of America* 28(2): 303–310.

Jannedy, Stefanie, and Jennifer Hay (eds.)
2006 Modelling sociophonetic variation *Special Issue of Journal of Phonetics.*

Johnson, Keith
1991 Differential effects of speaker and vowel variability on fricative perception. *Language and Speech* 34(3): 265–279.

Jongman, Allard, Ratree Wayland, and Serena Wong
2000 Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108(3): 1252–1263.

Munson, Benjamin, Elizabeth C. McDonald, Nancy L. DeBoe, and Aubrey R. White
2006 The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics* 43(2): 202–240.

Narayanan, Shrikanth, Abeer Alwan, and Katherine Haker
1995 An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America* 98(3): 1325–1347.

Newman, Rochelle S., Sheryl A. Clouse, and Jessica L. Burnham
2001 The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America* 109(3): 1181–1196.

Perkell, Joe S., Melanie L. Matthies, Mark Tiede, Harlan Lane, Majid Zandipour, Nicole Marrone, Ellen Stockmann, and Frank H. Guenther
2004 The distinctiveness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech Language and Hearing Research* 47: 1259–1269.

Pierrehumbert, Janet B.
2006 The next toolkit. *Journal of Phonetics* 34(6): 516–530.

Pierrehumbert, Janet B., and Cynthia Clopper
forthcoming What is Labphon? And where is it going? In C. Fougeron, B. Kühnert, M. d'Imperio and N. Vallée (eds.) : *Laboratory Phonology, 10: Variation, Phonetic Detail and Phonological Modeling.* Berlin/New-York: Mouton de Gruyter.

Schwartz, Martin F.
1968 Identification of speakers' sex from voiceless, isolated fricatives. *Journal of the Acoustical Society of America* 43(5): 1178–1179.

Shadle, Christine H.
1991 The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics* 19: 409–424.

Stevens, Kenneth N.
1998 *Acoustic Phonetics*. Cambridge, Cambridge, MA.: MIT Press.

Strand, Elizabeth A.
1999 Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1): 86–99.

Stuart-Smith, Jane
2007 Empirical evidence for gendered speech production: /s/ in Glaswegian. In *Change in Phonology. Laboratory Phonology 9,* Jennifer Cole, and José I. Hualde (eds.), 65–86. Berlin: Mouton de Gruyter.

Tabain, Marija
2001 Variability in fricative production and spectra: Implications for the Hyper- and Hypo- and Quantal Theories of speech production. *Language and Speech* 44(1): 57–94.

Producing turbulent speech sounds in the context of cleft palate

Fiona E. Gibbon and Alice Lee

1. Introduction

Over the past 40 years, a combination of improved surgical technique, multidisciplinary working, and speech therapy has resulted in measurable progress in treating individuals with cleft palate (Kawano et al. 1997). Despite these gains, speech disorders in this group are a serious problem – it is estimated that between 50–90% of individuals with cleft palate require episodes of speech therapy (Albery and Grunwell 1993; Peterson-Falzone 1990a; Witzel 1991). If a speech disorder persists past the preschool years, it can adversely affect individuals' social communication and emotional wellbeing as well as their educational and ultimately their employment prospects.

This chapter focuses on abnormal articulations that can develop when individuals have structural (anatomical) abnormalities that interfere with their ability to create normal air turbulence in the oral cavity. Abnormal articulations that have a constriction located in the oral region are illustrated using data recorded from the instrumental technique of electropalatography (EPG). Studies that have used other instruments to investigate abnormal articulations located in the pharyngeal and laryngeal regions are also discussed. In preparing the ground for these discussions, there follows some introductory remarks for readers who may be unfamiliar with cleft palate and to explain the pivotal role of speech turbulence in the development of abnormal articulations.

1.1. Speech turbulence

Moin and Kim (1997: 62) define turbulence as "composed of eddies: patches of zigzagging, often swirling fluid, moving randomly around and about the overall direction of motion". They go on to say that technically, the chaotic state of fluid motion arises when the speed of the fluid exceeds a specific threshold. Thus, in order to create a turbulent flow of air for speech, under normal circumstances speakers must constrict the articulators to a greater or lesser extent and build up sufficient air pressure behind the constriction to create plosion or friction. Sounds produced this way, namely plosives, fricatives, and affricates, are known collectively as obstruent sounds and all involve creating turbulent airflow. The acoustic characteristics of turbulence in speech sounds are complex and manifold, and are determined by a combination of factors, such as the intraoral air pressure, oral airflow rate, the location and shape of the constriction through which the air passes and so on (e.g., Hixon 1966; Stevens 1998).

1.2. Cleft palate

Cleft palate is a congenital malformation of the hard or soft palate, or both, which under normal circumstances is repaired surgically during the first 12–18 months of life. Most developed countries have well organized and sophisticated health care regimes for the management of individuals born with cleft palate. A cleft palate is usually identified around the time of birth, although these days it can be identified in utero using ultrasound technology. Specialist multidisciplinary teams are responsible for developing and coordinating individualized treatment plans to ensure that each person’s complex medical, psychological, and social needs are met from birth to adulthood. The team consists of a surgeon, an orthodontist, and a speech and language therapist, with additional specialist input from otolaryngology, audiology, dentistry, paediatrics, genetics, psychology, and social work. Following primary surgery as infants, many undergo further surgery during childhood and adolescence in order to provide the best possible vocal tract structures for normal speech, hearing, and eating.

1.3. Early speech development in infants with cleft palate

The timing of the primary palatal repair at around 12–18 months means that babies born with cleft palate are, at least during the pre-speech period, “obligated to engage in vocal practice without the normal division between the oral and nasal cavities provided by the hard and soft palate and, in many cases, the absence of normal articulatory contacts in the anterior portion of the hard palate” (Peterson-Falzone et al. 2006: 5). These anatomical constraints, together with frequently occurring middle ear infection and accompanying conductive hearing loss, can have detrimental effects on babies' vocalization development. Studies have shown that early pre-surgery vocalizations in babies with cleft palate are restricted in the types of sounds they produce. Most noticeably they produce a smaller number of consonant types with fewer oral plosives, and fricatives than typically developing babies (Chapman 1991; Chapman et al. 2001; Grunwell and Russell 1988; Hutters, Bau, and Brøndsted 2001; O'Gara and Logemann 1988). In terms of place of articulation, they avoid the hard palate as a place of articulation with the result that they tend not to produce sounds with alveolar or postalveolar placement (Peterson-Falzone et al. 2006). Not surprisingly, babies with cleft palate produce more sounds that do not require velopharyngeal closure (e.g. nasals, glottal stops, glottal fricatives) during babbling than babies with intact velopharyngeal mechanisms.

Patterns of sound usage during babbling are relevant because during this period, children are developing their phonological systems, and previous studies on normal speech acquisition have shown that the sounds used frequently in babbling are the ones that appear in early words (Stoel-Gammon 1985; Vihman et al. 1985). As a result, a limited babbling inventory can adversely affect early phonological development and result in an equally reduced phonetic inventory during the first word stage at the 12–18 month period, even if primary palatal surgery has been successful.

1.4. Velopharyngeal dysfunction and oronasal fistulae

Structural defects, such as velopharyngeal dysfunction (VPD) and oronasal fistulae, can persist after primary surgery. These defects are potential hazards to an individual's ability to create and control turbulent airflow for speech sounds made in the oral cavity. A properly functioning velopharyngeal mechanism is a prerequisite in achieving normal oral-nasal resonance and adequate intraoral pressure for speech. VPD, on the other hand, results in an inability of the velopharyngeal structures to separate the nasal cavity adequately from the oral cavity during speech. The perceptual results of VPD are hypernasality and/or nasal emission due to excessive amounts of airflow escaping into the nasal cavity. An oronasal fistula is an opening between the oral and nasal cavity and, like VPD, can result in air escaping into the nose. The incidence of VPD after primary cleft palate repair has been found to range from 25% to 44%, depending on the cleft type and surgical procedures (Krause, Tharp, and Morris 1976; see also

Phua and de Chalain 2008, for a detailed review). The incidence of oronasal fistulae has been found to vary even more widely, from 0% to 34% (Cohen et al. 1991; Phua and de Chalain 2008) depending on a number of factors such as the extent of clefting and the type of repair. Due to air leakage into the nose, these structural abnormalities can make it difficult, or even impossible, for some speakers to build up the necessary oral pressure to produce obstruent sounds.

1.5. Compensatory errors affecting obstruent sounds

An appreciation of the concept of compensatory errors assists in understanding the nature of articulation errors, particularly errors affecting obstruent sounds that are produced by individuals with cleft palate. Compensatory errors are speech behaviours that are interpreted as being due to abnormal learning. Speakers are thought to adopt compensatory articulations in order to reduce the perceptual or acoustic consequences of a structural deficit, such as an oronasal fistula or VPD. Warren (1986), however, views compensatory articulations as arising more from a need to maintain a stable aerodynamic environment for speech production rather than achieving a perceptual or acoustic goal (see Netsell 1990). Regardless of how they arise, once learned, these abnormal learned patterns can persist due to habituation (Peterson-Falzone, Hardin-Jones, and Karnell 2001) even after the structural abnormality has been corrected (e.g. following surgery).

As described in the previous section, structural abnormalities such as VPD and oronasal fistulae can make producing obstruents in the oral cavity problematic due to air leakage and a difficulty building up sufficient pressure to produce turbulence in this part of the vocal tract. Nevertheless, it is still possible for individuals to compensate for this difficulty by producing obstruents at different locations in the vocal tract instead. By making an articulatory constriction “upstream” of the structural defect, speakers can circumvent air leakage and its consequences for obstruent sound production. In other words, individuals with VPD can still produce these sounds below the level of the velopharyngeal structures, for example, in the pharyngeal and laryngeal regions of the vocal tract. Similarly, an individual with a large oronasal fistula in the hard palate can circumvent air leakage by producing obstruents in the velar and uvular regions (provided there is no VPD) as well as in the pharyngeal and laryngeal regions. The shift of place of articulation upstream of the structural defect just described results in abnormally retracted place of articulation. A final possibility is that the speaker uses the tongue body to aid velopharyngeal closure, a phenomenon referred to as "lingual assistance" (Brooks, Shelton, and Youngstrom 1965; Trost 1981). In this remarkable maneuver, the tongue body moves upwards and backwards to assist velopharyngeal closure with the result that placement is also retracted. Not surprisingly, retracted place of articulation for obstruents is one of the most characteristic and pervasive features of cleft palate speech reported in the literature (Morley 1970; Trost 1981; Trost-Cardamone 1997; Wyatt et al. 1996).

Although speech disorders are frequent in individuals with cleft palate, studies conducted in the US have shown relatively low rates of compensatory errors. Peterson-Falzone (1990a) investigated a group of 240 children with cleft palate aged 4–11 years and found that 22% had compensatory errors. Dalston (1992) found a somewhat higher prevalence rate of 28%, and the latest study by Hardin-Jones and Jones (2005) reported a prevalence of 25% in 212 preschoolers with cleft palate. A recent study of 42 school children and adolescents with repaired cleft palate in Greece showed a prevalence of 28.5% for compensatory articulations (Paliobei, Psifidis, and Anagnostopoulos 2005). This study on Greek-speaking children seems to support the assumption that lower prevalence rates are associated with younger age of palatal surgery.

Peterson-Falzone (1989) used perceptual analysis to investigate the frequency of different types of compensatory articulations in a group of 112 individuals with repaired cleft palate. Consistent with the view that compensatory errors involve predominantly retracted placement, her study found that frequency varied, and reported the following (from most to least frequent): glottal stops; middorsum palatal stops; pharyngeal fricatives; velar plosives; pharyngeal affricates; palatal fricatives; pharyngeal stops; and velar fricatives (Peterson-Falzone 1989). Hardin-Jones and Jones (2005) reported a similar trend, with glottal articulations as the most prevalent errors, followed by middorsum palatal stops and pharyngeal productions.

1.6. Dental and occlusion abnormalities

Dental and occlusal abnormalities, which are common in individuals born with cleft palate, affect dentition and the way in which the upper and lower teeth meet when the jaws bite together. Dental and occlusal abnormalities do not cause a problem of air leakage into the nasal cavity in the same way as VPD or oronasal fistulae, but they can cause sounds made in the oral cavity to be distorted. Dental and occlusal abnormalities can therefore have a direct effect on articulation and the acoustic characteristics of obstruent sounds. For example, speech problems are likely to occur where a dental or occlusal problem significantly reduces intraoral area and consequently the space within which the tongue can move. Such a restriction can occur when individuals have small/narrow hard palates or when they have a Class III malocclusion (Peterson-Falzone 1990b), which occurs when the maxillary molars are posterior to mandibular molars when the jaws bite together. Avoiding placement in the alveolar region may be exacerbated by diminished sensation in this area due to scarring following surgery. Similarly, deviated anterior teeth that are rotated or ectopic may also adversely affect the tongue's ability to make alveolar placement or to form the anterior groove necessary for normal sibilant sound production.

In terms of effects on speech, Albery and Grunwell (1993) found that malocclusions can cause difficulties in forming a seal between the sides of the tongue and the hard palate. A seal is necessary to produce a normal central airstream and to prevent loss of air laterally into the buccal cavities for sounds such as /s/, /z/, /ʃ/, /ʒ/, /tʃ/, and /dʒ/. Lateral escape of air in these circumstances can lead to distorted, usually lateralized, productions of alveolar obstruent sounds. Class III malocclusions may also cause the tongue tip to protrude between the upper and lower teeth leading to interdental productions of alveolar plosives and a "lisp" produced for sibilant sounds (Vallino and Tompson 1993). Another distortion that may arise in speakers with severe malocclusion is a labio-dental substitution, which may replace bilabial targets because the malocclusion prevents the lips approximating in the normal way (Peterson-Falzone, Hardin-Jones, and Karnell 2001).

2. Abnormal articulations located in the oral region

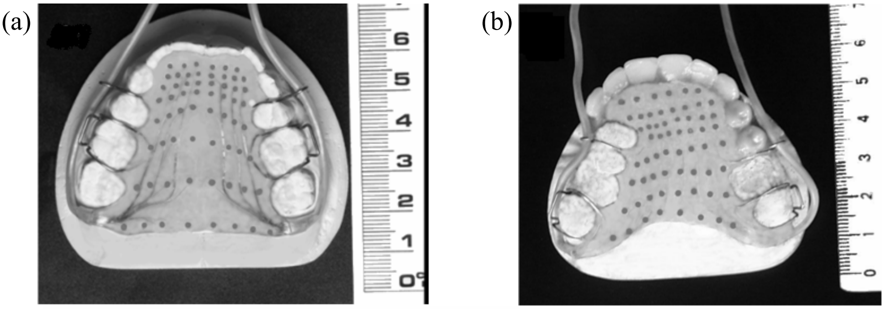
The sections that follow give examples of abnormal articulations produced with constriction in the oral region of the vocal tract. The examples are from speakers with repaired cleft palate and they illustrate errors described in the previous sections. As already mentioned, these errors primarily affect obstruent sounds located in the oral region, which makes electropalatography (EPG) an ideal instrument for recording the dynamic details of the tongue-palate contact associated with these abnormal articulations.

2.1. Electropalatography (EPG)

EPG is an instrumental procedure that records details of the location and timing of the tongue's contact with the hard palate during speech (Hardcastle and Gibbon 1997; Hardcastle, Gibbon, and Jones 1991). Two commercially available versions have dominated EPG research in the study of cleft palate speech: a British system – the EPG3 system developed at the University of Reading – has been used in the majority of cleft palate studies conducted by researchers in the UK and Hong Kong. A new Windows® version of the Reading EPG has recently been developed at Queen Margaret University, Edinburgh, UK (Scobbie, Wood, and Wrench 2004). The Rion EPG (Fujimura, Tatsumi, and Kagaya 1973; Hiki and Itoh 1986) is used in studies reporting Japanese cleft palate speech.

A component of all EPG systems is a custom-made artificial plate moulded to fit the speaker's hard palate. Figure 1 shows a Reading plate for a normal speaker and a similarly aged adult with a cleft palate. Embedded in the artificial plate are electrodes exposed to the lingual surface. The Reading and Rion plates are made from a relatively rigid acrylic, and are held in place by metal clasps that fit over the upper teeth. The Reading plates have 62 electrodes placed according to identifiable anatomical landmarks (Hardcastle, Gibbon, and Jones 1991). The electrodes are arranged in eight horizontal rows, with eight electrodes in every row except the most anterior, which has six. The most posterior row of electrodes on the Reading plates is located on the junction between the hard and soft palates. Figure 1c shows a single EPG frame, which is a typical contact pattern of alveolar stops /t/, /d/, and /n/,[1] divided into different phonetic regions (alveolar, postalveolar, palatal, and velar), and information on the part of the tongue that is assumed to make contact with these regions.

Of relevance to this chapter is that EPG systems that use Reading plates register characteristic patterns in normal speakers for all English lingual obstruents /t/, /d/, /k/, /g/, /s/, /z/, /ʃ/, /ʒ/, /tʃ/, and /dʒ/. Articulations that have their primary constriction either further forward than the most anterior row of electrodes (e.g. dentals or labials) or further back than the most posterior row of electrodes (e.g. velars in the context of open vowels, uvular, pharyngeal and glottal sounds) have minimal EPG contact patterns. Some EPG contact may be present during these articulations, however, due to the influence of surrounding vowels (Gibbon, Lee, and Yuen 2007). There have been recent advances in the design of EPG plate – the Articulate EPG plate, which has a similar design to the Reading plate and is compatible with the Reading EPG systems, has the first row of the electrodes advanced by 1 mm for capturing linguo-dental articulation and the last row placed straight across the soft palate, with the midsagittal electrodes about 7-12 mm behind the border of the hard and soft palate (Wrench 2007).



(c)

Normal	Row		
0 0 0 0 0 0	1	Alveolar	Anterior
0 0 0 . . 0 0 0	2		(tongue
0 . . . . . 0 0	3	Post-alveolar	tip/blade)
0 0	4		
0 0	5		
0 0	6	Palatal	Posterior
0 0	7		(tongue
0 0	8	Velar	dorsum)

Figure 1. A Reading plate of (a) a normal speaker, (b) a similarly aged adult with a cleft palate, and (c) a single EPG frame, showing a typical contact pattern of alveolar stops /t/, d/, and /n/; with EPG frame row numbers 1 through 8 indicated, as are the phonetic regions of the palate (alveolar, postalveolar, palatal, and velar), and the part of the tongue assumed to make contact with these regions (adapted with permission from Gibbon and Crampin 2001).

2.2. Middorsum palatal stops

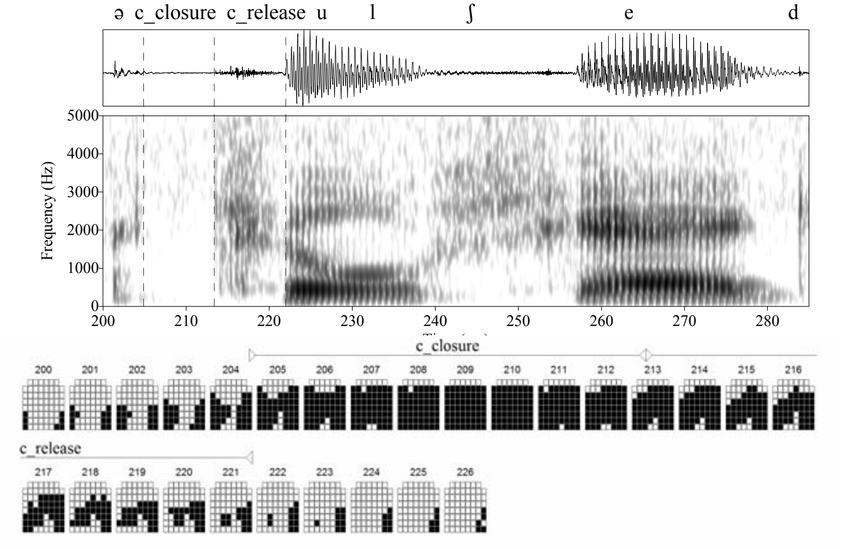
The middorsum palatal stop is one of the most frequently occurring types of compensatory articulations. Trost (1981) originally described [c, ɟ] as substitutions used to replace /t/, /d/, /k/, and /g/. Note that Trost-Cardamone and colleagues use a different set of symbols for transcribing compensatory articulations (see for example, Peterson-Falzone et al. 2006). Middorsum palatal stops are made in the approximate place of the glide /j/ with midpalatal lingual contact produced with the tongue dorsum raised and the tongue tip down. Trost (1981: 196) states that "perceptually, the phoneme boundaries between /t/ and /k/ or between /d/ and /g/ are lost". Although in middorsum palatal stops /t/ and /d/ targets are retracted from alveolar to palatal placement, /k/ and /g/ targets show the opposite trend as they are fronted from velar to palatal placement.

Palatal stops have been described using EPG in English, Japanese and Cantonese speakers with cleft palate (Gibbon and Crampin 2001; Whitehill et al. 1995; Yamashita et al. 1992). Yamashita et al. (1992) found that about three quarters of 53 Japanese speakers with cleft palate aged 4–20 years produced these abnormal articulations. The EPG patterns for these palatal misarticulations had either contact across the whole surface of the palate or contact that is confined to the posterior region of the palate.

Gibbon and Crampin (2001) used EPG data recorded from an adult with cleft palate who produced middorsum palatal stops, which were typical in the sense that listeners could not distinguish between alveolar and velar targets, with the result that the phoneme boundary between these classes of sounds was lost (Trost 1981). However, this speaker articulated alveolar targets in a subtly different way from velar plosives, a finding that was not predicted from a transcription-based analysis. Figure 2 illustrates this finding and shows EPG patterns for this speaker's production of a /t/ and a /k/, which were both transcribed as [c], with the same targets produced by a normal speaker for comparison in Figure 3.

The EPG patterns for the speaker in Figure 2 show that at the onset of closure, tongue placement approximates the normal pattern. Following onset, the EPG patterns during closure are distorted, with a high amount of contact compared to a normal speaker. Although both /t/ and /k/ involved increased contact, nevertheless /t/ has more contact compared to /k/ indicating an articulatory difference between these sound classes.

(a) /t/ → [c] in “a toolshed” produced by an adult male with cleft palate.



(b) /k/ → [c] in “a kettle” produced by an adult male with cleft palate.

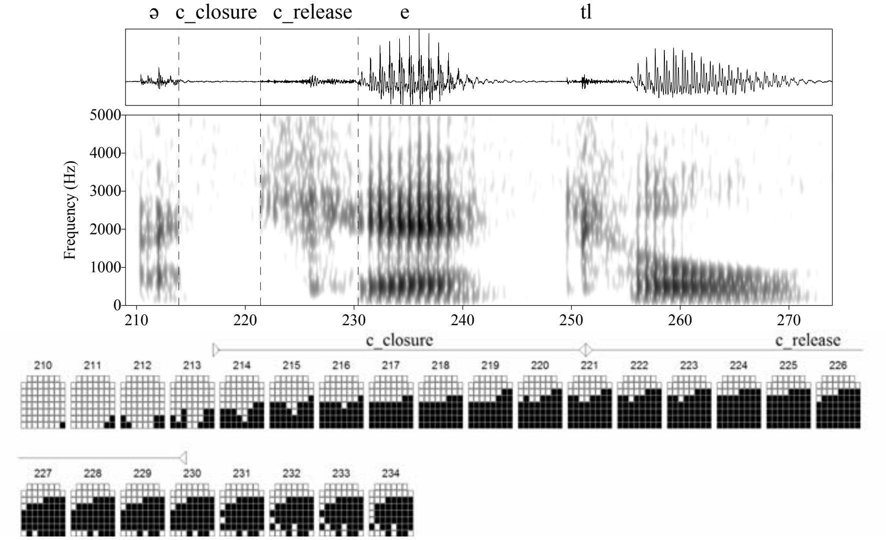
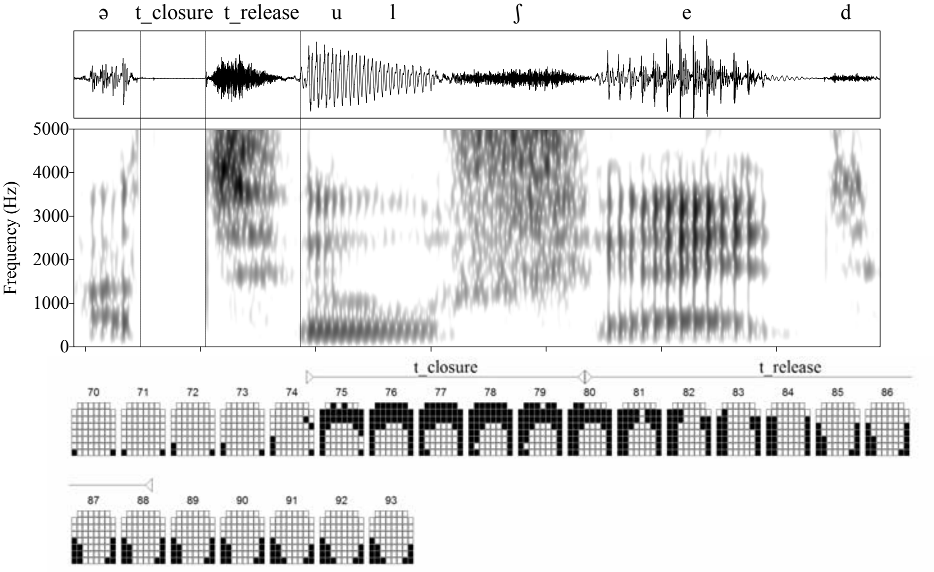


Figure 2. EPG data illustrating middorsum palatal stops used for /t/ and /k/ targets. Simultaneous acoustic and EPG data are shown of two phrases (a) “a toolshed” and (b) “a kettle”, which were realized as middorsum palatal stops and recorded from a 36-year-old man with repaired cleft palate.

(a) /t/ → [t] in “a toolshed” produced by adult male with normal speech.



(b) /k/ → [k] in “a kettle” produced by adult male with normal speech.

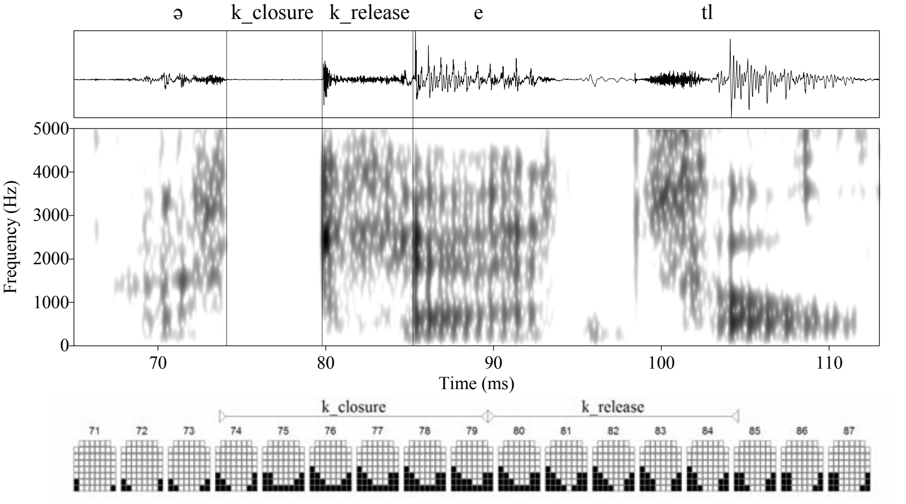


Figure 3. EPG printouts of normal patterns for (a) /t/ in “a toolshed” and (b) /k/ in “a kettle”.

So, although perceptual analysis showed that alveolar and velar targets were middorsum palatal stops, the EPG data revealed that this speaker produced consistent articulatory differences between these targets. (See Gibbon and Crampin 2001, for details of this case and the clinical implications of producing subtle articulatory differences between perceptually neutralized sound contrasts.)

2.3. Palatal/velar fricatives and affricates

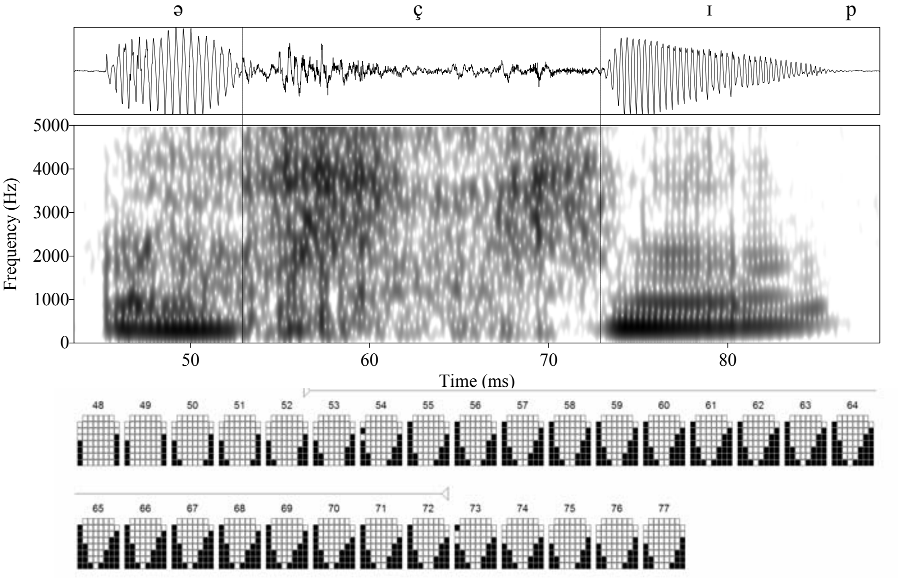
Perceptual and EPG studies have shown that palatal and velar fricatives and affricates usually occur as substitutions for alveolar and postalveolar fricatives and affricates (Howard 1998; Howard and Pickstone 1995; Michi et al. 1990; Trost 1981; Yamashita et al. 1992). Palatal and velar fricatives are produced by the tongue dorsum forming a constriction in the posterior region of the hard palate or on the soft palate. Yamashita et al. (1992) found in an EPG study that 25% of Japanese alveolar and palatal fricatives involved retracted placement. Howard and Pickstone (1995) described contact patterns produced by a 6-year-old girl with a repaired cleft of the hard and soft palate. EPG contact patterns for targets /s/, /z/, /ʃ/, and /ʒ/ produced by this child were transcribed as palatal fricatives, and typically involved retracted placement, a fairly broad central groove, and a wide band of side contact from postalveolar to the front of the velar region.

Figure 4a shows an example of a palatal fricative produced by a 9-year-old girl with cleft palate for the alveolar fricative /s/, and below is the same target produced by a normal speaker. The patterns produced by the child with cleft palate show an “inverted” pattern compared to the normal, with a narrow, central groove located in the posterior rather than the normal alveolar region of the palate. Unlike the speaker described by Howard and Pickstone (1995), the patterns in Figure 4 show that this speaker produced a narrow, rather than a broad, posterior central groove configuration.

2.4. Velar plosives

Velar plosives may occur as substitutions for alveolar plosives, and sometimes even for bilabial plosives. Velar plosives involve the tongue dorsum making closure in the region of the junction between the hard and soft palates. Retraction to velar placement can be seen clearly on EPG

(a) /s/ → [ç] in “a sip” produced by a girl with cleft palate.



(b) /s/ → [s] in “a sip” produced by a normal speaker.

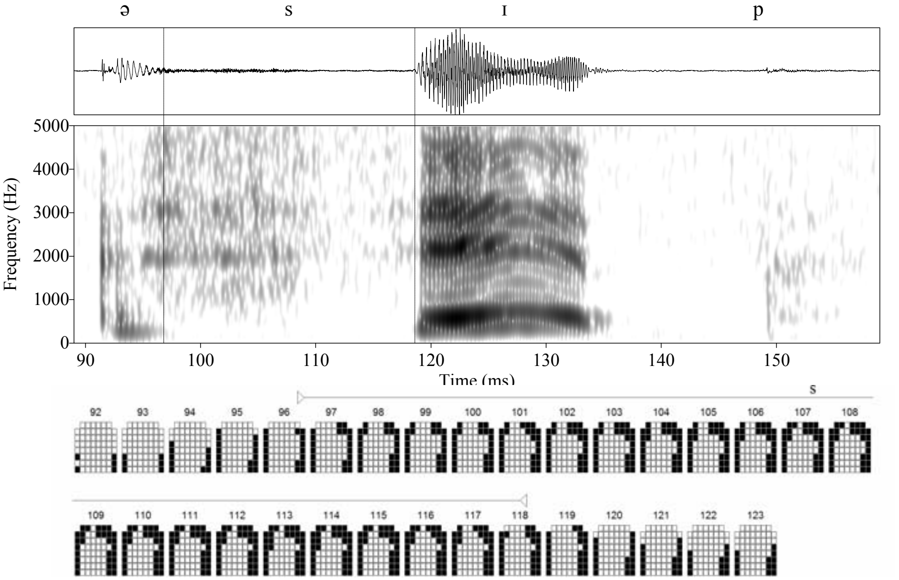
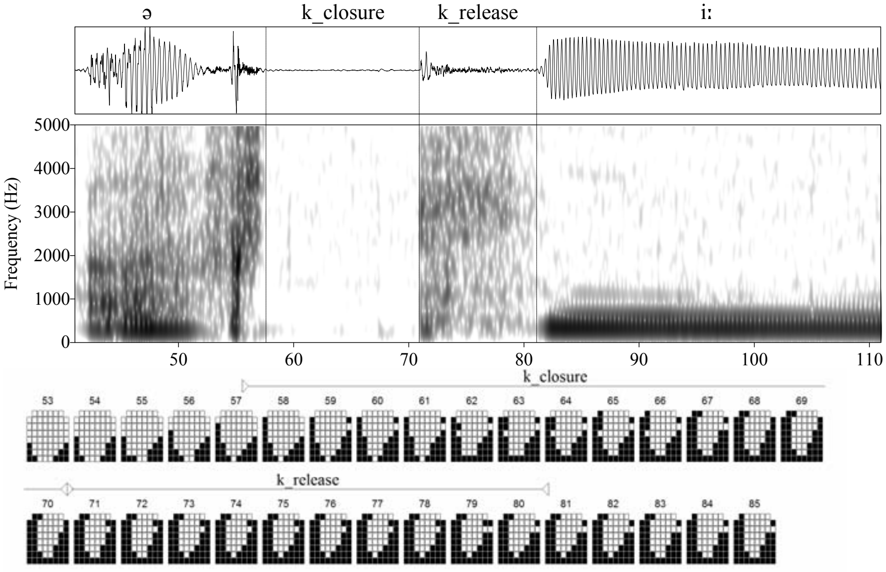


Figure 4. EPG data illustrating a palatal fricative used as a substitution for /s/. The EPG printouts in (a) show a palatal fricative for the /s/ in the phrase “a sip” produced by a 9-year-old girl with repaired cleft palate. A normal pattern for /s/ in the same phrase is shown in (b).

(a) /t/ → [k] in “a team” produced by a girl with cleft palate.



(b) /t/ → [t] in “a team” produced by a normal speaker.

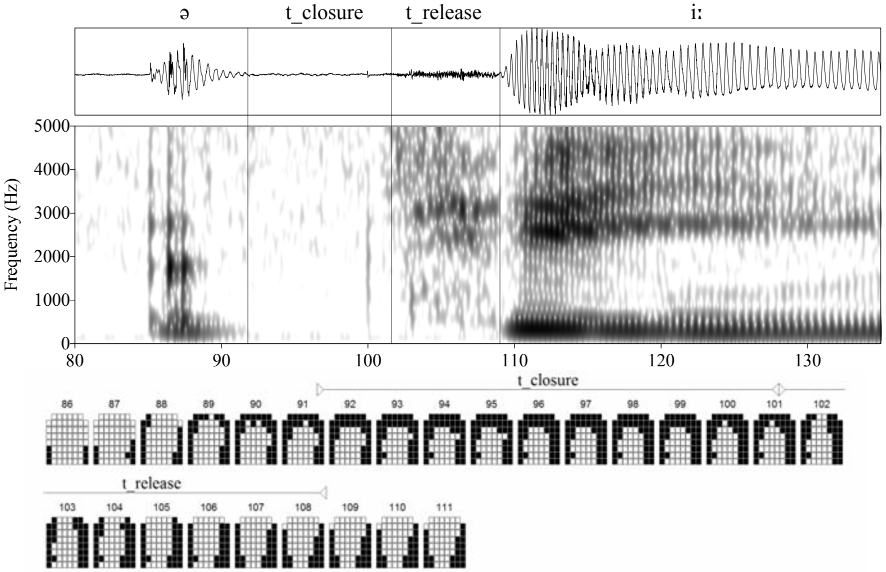


Figure 5. EPG data illustrating a velar plosive produced as a substitution for /t/. The EPG printouts in (a) show a velar plosive for the /t/ in the phrase “a team” produced by a 9-year-old girl with repaired cleft palate. A normal pattern for /t/ in the same phrase is shown in (b).

traces, and has been reported in previous studies of English (Gibbon and Crampin 2001; Gibbon and Hardcastle 1989; Hardcastle, Morgan Barry, and Nunn 1989), Japanese (Yamashita et al. 1992), and Cantonese (Whitehill et al. 1995; Whitehill, Stokes, and Man 1996) speakers with cleft palate. Figure 5a shows a retracted placement from alveolar region to velar region for the target /t/, observed in a girl with cleft palate, which contrasts to the normal pattern illustrated in Figure 5b.

2.5. Lateral fricatives and affricates

Lateral fricatives and affricates usually occur as substitutions for sibilant targets /s/, /z/, /ʃ/, /ʒ/, /tʃ/, /dʒ/ (Gibbon 2004). EPG patterns for lateral fricatives often have increased contact, but do not have an anterior central groove that normal speakers have for these sounds. Suzuki et al. (1981) defined lateral misarticulation as involving the tongue making complete contact across the palate (i.e. there is no evidence of tongue grooving), and lateral release of air (i.e. air directed out of the occluded dental arch posterior to the molar teeth). Likewise, lateral affricates usually involve complete articulatory closure that extends throughout the stop and fricative phases of the affricate. Where complete closure affects sibilant targets there is evidence from the acoustic signal and from perceptual analysis of friction, but the EPG patterns show complete constriction. With complete constriction across the palate, the possibility of a normal central flow of air through an anterior groove is reduced, and the likely escape of air is around the lateral margins of the tongue producing lateral friction.

Figure 6 shows EPG patterns from a girl with a cleft palate producing a target /s/ in the phrase “a sip”, transcribed as a lateral fricative [ɬ] (see Figure 4b for a normal speaker’s production of /s/ in “a sip”). Unlike a normal speaker’s production of /s/, this speaker does not produce an anterior groove configuration combined with lateral contact. Instead, there is complete contact across the palate in the alveolar/postalveolar region, and incomplete lateral contact on the left side. It is likely that the air is escaping out of the left side into the buccal cavity due to incomplete lateral seal on this side, but this information is inferred from the EPG patterns, and further diagnostic tests would be needed to confirm this.

The EPG configurations involved in speech sound distortions referred to as lateral fricatives, vary between speakers. A study by Yamashita et al. (1992) showed that only a minority of lateral misarticulations involved alveolar contact, such as illustrated in Figure 6. Instead, they found a significant proportion of lateral misarticulations involved contact in the posterior region (i.e. retracted placement) but that the overwhelming majority (68%) involved contact across the whole length of the palate. Lateral fricatives are almost always associated with complete contact across the palate, however.

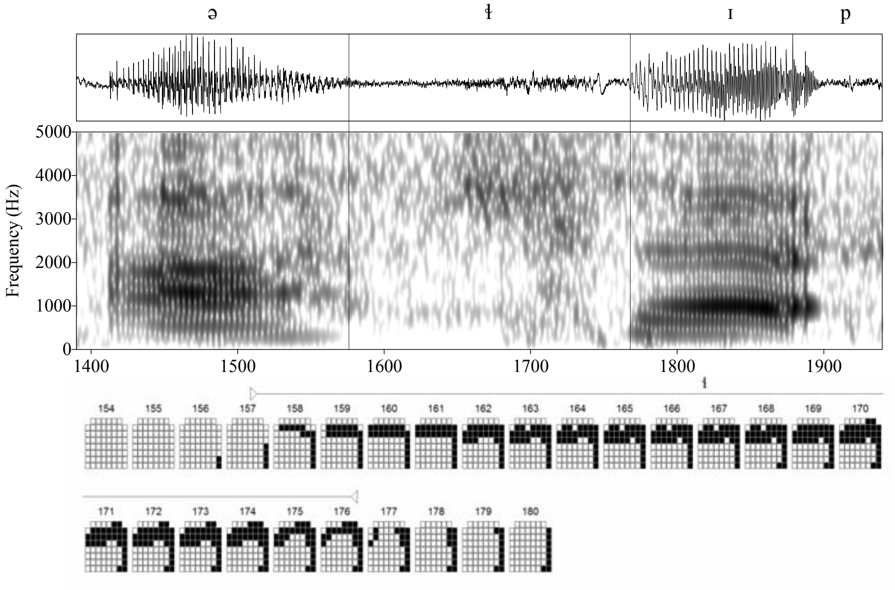


Figure 6. EPG data illustrating a lateral fricative used as a substitution for /s/ in the phrase “a sip” produced by a 12-year-old girl with cleft palate.

2.6. Lateral release of lingual plosives

Although lateral fricatives produced as substitutions have been described frequently in the literature, a less studied characteristic of cleft palate speech is lateral release as a secondary articulatory feature of alveolar and velar plosives (Albery 1991). The lack of research is surprising because lateral release degrades the acoustic cues that listeners use to identify primary place of articulation. Lateral release is therefore likely to have a detrimental effect on speech intelligibility because it leads to abnormal transitional placement cues. Of relevance here is that normal speakers can have lateral release of /t/ in words such as “bottle”, where a lateral /l/ follows a stop (Ball 1993). However, abnormal lateral release occurs in speakers who have this as a general feature of their speech, regardless of the phonetic context.

Gibbon and Crampin’s (2001) study of an adult male who produced middorsum palatal stops showed that he had complete tongue-palate contact occurring simultaneously with lateral release (see Figure 2a). Gibbon and Crampin showed that articulatory release of the tongue constriction in the central region of the palate for /t/ and /k/ targets was not closely timed with the acoustic plosive burst, as occurs in normal speakers. On average the first third of the aspiration period for their speaker’s production of /t/ and /k/ involved lateral escape of air, while the remaining period of aspiration involved a central airstream (Gibbon and Crampin 2001). The timing of articulatory release for /t/ and /k/ was variable, which is consistent with other studies (e.g. Wada et al. 1970) that have reported increased variability in articulatory release in cleft speakers with articulation disorders.

2.7. Double articulations

Abnormal double articulations have been identified in many studies of cleft palate speech (Grunwell 1993; McWilliams, Morris, and Shelton 1990; Morley 1970; Sell, Harding, and Grunwell 1994; Stengelhofen 1989; Trost 1981). Double articulations, which Trost (1981: 200) termed coarticulations or coproductions, involve “one manner of production with simultaneous valving at two places of production”. In many standard accounts of cleft palate speech, double articulations are described as involving glottal or pharyngeal articulations, which are usually combined with a second constriction in the oral region. Trost-Cardamone (1990: 233) goes so far as to state that “only the glottal stop and pharyngeal fricative occur as coarticulations”. To explain this statement further, Trost-Cardamone (1990: 233) says that during glottal and pharyngeal constriction, “the tongue is more free to make simultaneous and (more) anterior articulatory contacts”.

Evidence from EPG studies illustrated in the next sections show that a variety of other types of double articulations can occur. For example, it is possible to have tongue and lip closure occurring simultaneously, resulting in bilabial-lingual double articulations. It is also possible to have tongue body and tongue tip/blade double articulations resulting in alveolar-velar double articulations. In other words, the lips, tongue tip/blade, and tongue body are free to make simultaneous closures with each other, with the possibility that a wider variety of double articulations can occur than was previously thought.

2.7.1. Bilabial-velar double articulations

Bilabial-velar double articulation [p͡k] and [b͡g] are double articulations that involve bilabial closure occurring simultaneously with the tongue body making contact against the palate in the velar region. Several case studies have used EPG to reveal these types of double articulations in English speaking children with cleft palate. Gibbon and Hardcastle (1989) first described bilabial-velar double articulations in a case study of a 13-year-old boy with cleft palate. His EPG patterns showed consistent posterior lingual contact occurring throughout the closure period for consonants /p/, /b/, and /m/. Dent, Gibbon and Hardcastle (1992) described a similar case of a 9-year-old boy whose EPG data showed complete contact across the palate in the velar region for labial targets /p/ and /b/. These double articulations are relatively rare, however, with a study finding that just one out of 27 speakers with speech disorders associated with cleft palate produced them (Gibbon and Crampin 2002).

Figure 7a is an example of a bilabial-velar double articulation produced by a 9-year-old boy with a cleft palate. The target /b/ is in word-final position in the word “web”. Figure 7b shows the same word produced by a typically developing child of a similar age (see also Gibbon, Lee and Yuen, 2007). The typically developing child shows some contact in the posterior lateral region of the palate, which is normal in the context of bilabial following an /ɛ/ vowel (as occurs in “web”). In contrast, the child with a cleft palate shows complete closure across the palate in the velar region, which occurred simultaneously with bilabial closure (although it is not possible to record bilabial closure from the EPG patterns, the investigators observed that bilabial closure occurred). The occurrence of simultaneous velar closure during bilabials were not detected perceptually during single word productions, which were almost always heard by listeners as normal bilabial productions, although listeners sometimes detected velar substitutions for bilabial targets during connected speech.

2.7.2. Alveolar-velar double articulations

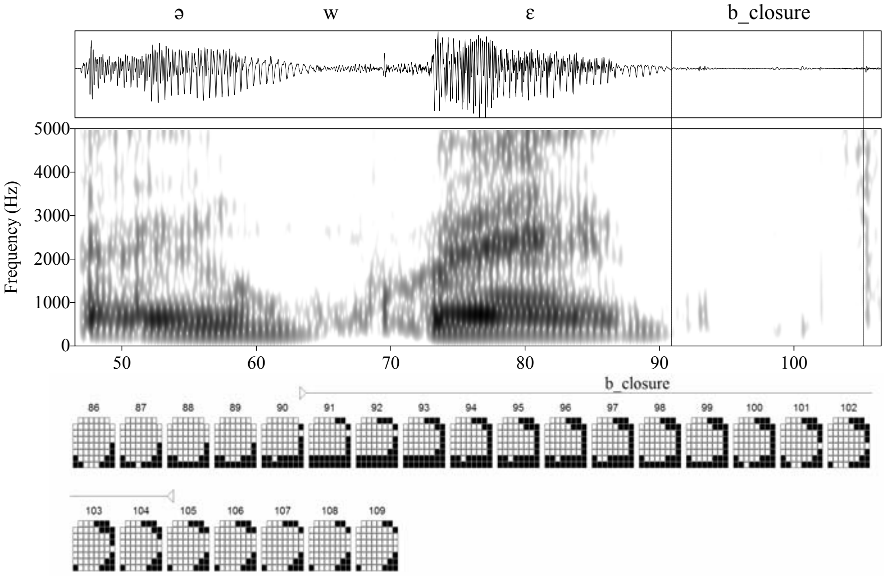
Alveolar-velar double articulations [t͡k] and [d͡g] have been reported more frequently than bilabial-velar double articulations in children with cleft palate. For example, alveolar-velar double articulations have been identified as frequent errors in a longitudinal transcription-based study of children with cleft palate (Harding and Grunwell 1993). They found that half of the 26 children produced these double articulations at some stage in their phonetic development towards correct production of /t/ targets. In terms of EPG studies, Gibbon, Ellis and Crampin (2004) also found that alveolar-velar double articulations were frequently occurring errors, with 10 out of the 15 children they studied producing this type of double articulation. They used an EPG classification scheme to identify alveolar-velar double articulations in 15 children with cleft palate. Their results showed that alveolar targets /t/ and /d/ were more likely to be produced as alveolar-velar double articulations than velar targets /k/ and /g/. For example, 28% of alveolars and 12% of velars were produced in this way.

Although there are relatively few EPG studies of cleft palate speech in languages other than English, there is evidence that double articulations also occur in other languages, such as Cantonese (Whitehill et al. 1995). Figure 8 shows an example of alveolar-velar double articulation in Kalantanese (a Malay dialect) produced for a /d/ target and recorded from a 10-year-old boy with repaired cleft palate. In the articulation shown in Figure 8, which was typical of this boy's articulations for all alveolar targets, the onset of closure is in the alveolar region at frame 221, which is followed by simultaneous alveolar and velar closure from between frame 224–232, followed by the release of the alveolar closure at frame 233 and then velar release at frame 235.

2.8. Clicks

Clicks are similar to the double articulations just described in that both involve velar closure occurring simultaneously with a second closure further forward in the oral cavity, either in the alveolar region or at the lips. They differ of course in the precise details of the timing and most importantly clicks involve a nonpulmonic, as opposed to a pulmonic, airstream mechanism. Unlike other types of compensatory errors already described, clicks are rarely reported in the cleft palate literature (see

(a) /b/ → [b͡g] in a 9-year-old with cleft palate.



(b) /b/ → [b] in a 12-year-old normal speaker.

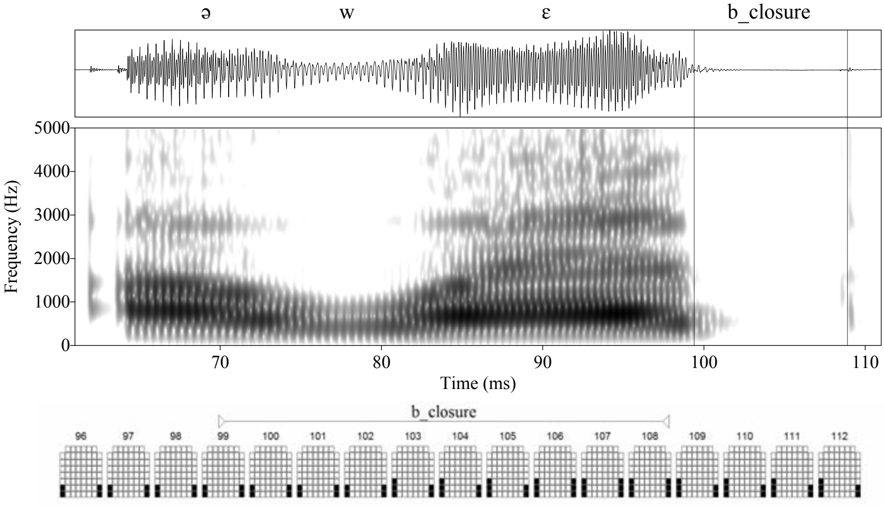
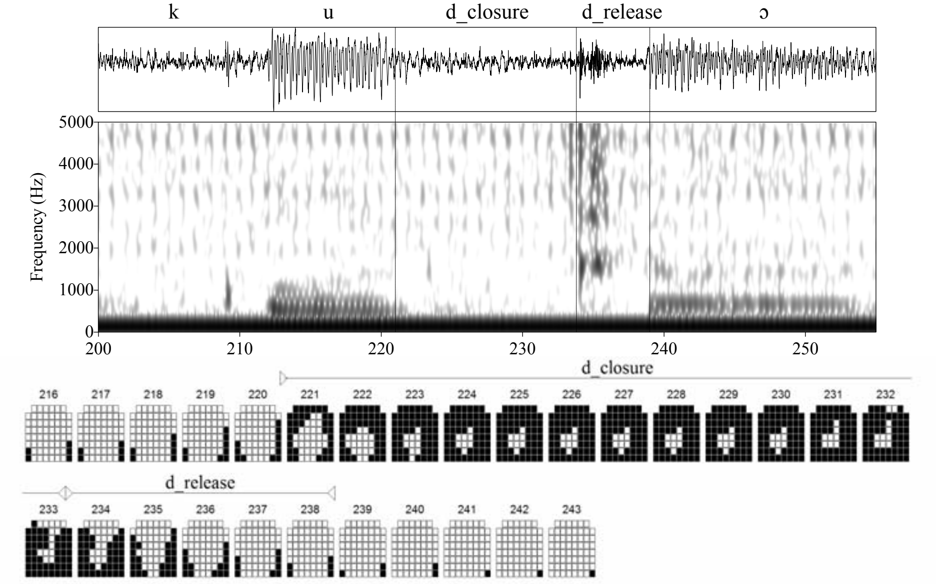


Figure 7. EPG data illustrating bilabial-velar double articulation. The EPG printouts in (a) show a bilabial-velar double articulation used for the /b/ in the phrase "a web" produced by a 9-year-old boy with cleft palate. A normal pattern for /b/ in the same phrase is shown in (b).

(a) /d/ produced as [d͡g] by a 10-year-old Kalantanese boy with cleft palate.



(b) /d/ produced as [d] by a normal Kalantanese speaker.

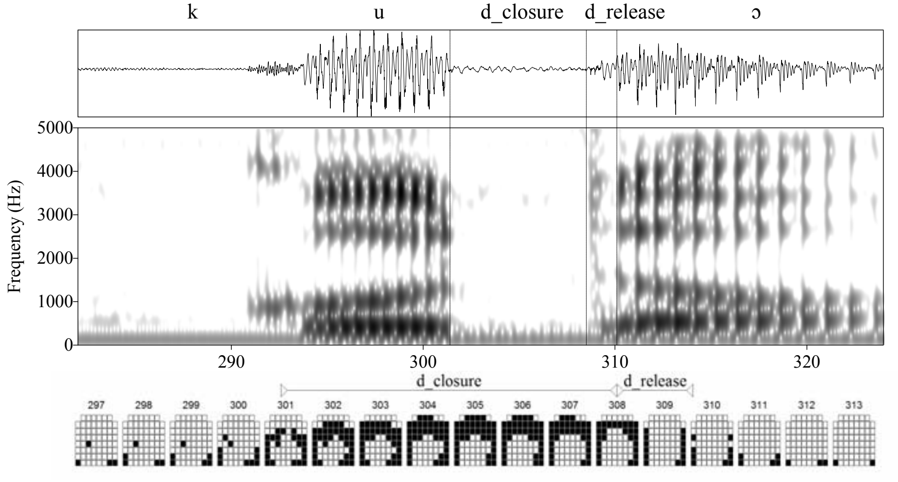


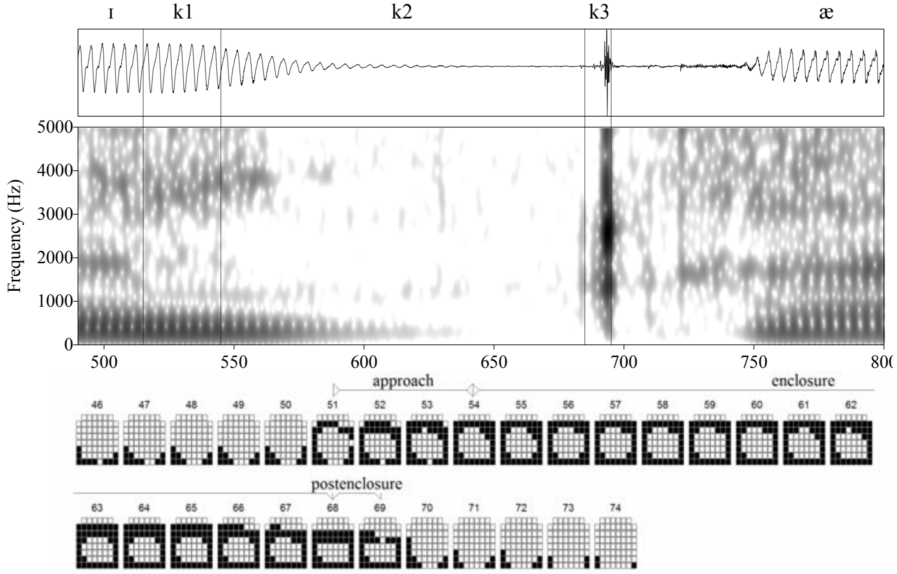
Figure 8. EPG data illustrating alveolar-velar double articulation. The EPG printouts in (a) show an alveolar-velar double articulation used for /d/ in the word “kuda” produced by a 10-year-old Kalantanese (a Malay dialect) boy with cleft palate. The word was produced in the Kalantanese sentence “daun muda makan ko kuda”, which in English means “the young leaves are eaten by the horse”. A normal pattern for /d/ produced by a Kalantanese speaker for the same phrase is shown in (b).

Miller's chapter on clicks in current volume). Howard (1993) reported a 6-year-old child with cleft palate who used bilabial clicks [ʘ] for /p/ targets. Peterson-Falzone, Hardin-Jones and Karnell (2001: 170) allude to the occurrence of clicks in some children with cleft palate, stating that they have "on rare occasion observed click substitutions of stop consonants in children with velopharyngeal inadequacy". A recent study by Mills, Gosling and Sell (2006) of 76 children with 22q11 deletion syndrome aged 3–10 years confirmed that clicks occur infrequently, with only a minority (4%) of this group producing these abnormal articulations.

Clicks are highly complex speech sounds that under normal circumstances occur almost exclusively in the languages of Southern Africa. Clicks are described as stops in which the essential component is "the rarefaction of air enclosed between two articulatory closures formed in the oral activity" (Ladefoged and Maddieson 1996: 246). Gibbon et al. (2008) reported perceptual and EPG data on clicks produced by two adolescents with VPD. The following EPG data are from one of the adolescents – a 14-year-old girl with ongoing VPD and associated hypernasality (Figure 9a). This girl was able to produce /d/, /k/, and /g/ targets as alveolar nasal clicks [!], and was able to produce them fluently in connected speech in all syllable and word positions. Clicks produced by this girl are probably abnormal compensatory articulations that are used in order to produce plosive sounds in the oral region of the vocal tract. This view is generally congruent with that of Warren and colleagues that compensatory articulations represent the speaker's attempt to achieve adequate pressure-valving for speech (Warren 1986; Warren, Dalston, and Dalston 1990).

Figure 9 shows that the click has an identifiable enclosure phase with simultaneous anterior and posterior closures (frames 54–68). At frame 69, enclosure terminates with the release of the anterior closure and the production of an audible click sound. A second feature is that clicks have an abruptness of the release of the anterior closure, and this can be seen clearly on the acoustic signal. The release goes from complete alveolar constriction to no contact without an intermediate phase showing fricative-like constriction. A third feature is evidence of the tongue blade moving backwards during enclosure (see frames 62–68). These features are consistent with previous studies of clicks in normal speakers of South African languages (Thomas 1997; Traill 1995). Thomas (1997: 382) showed in her EPG data a backward movement of the tongue blade during the enclosure phase of clicks, and suggests that "the fast movement of the

(a) /k/ produced as [ǃ] by a 14-year-old girl with velocardiofacial syndrome.



(b) /k/ produced as [k] by a normal speaker.

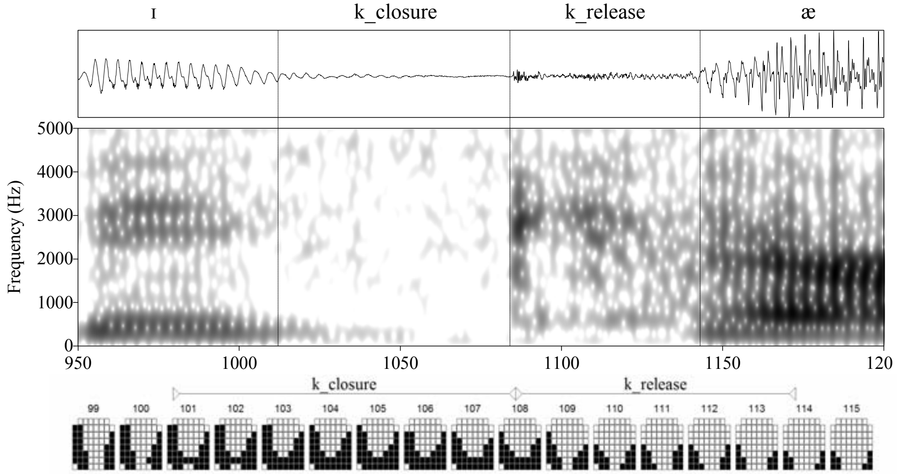


Figure 9. EPG data illustrating a nasal click produced as a substitution for /k/. The EPG printouts and acoustic trace in (a) show a click for the /k/ in the phrase "Happy Karen" produced by a 14-year-old girl with VPD. On the waveform, "k1" marks the approach phase (frames 51–54); "k2" marks the enclosure phase (frames 54–68); and "k3" marks the post-enclosure phase (frames 68–69). A normal pattern for /k/ in the same phrase is shown in (b).

tongue body in rarefaction, with its resulting large negative pressure, pulls the tongue tip backwards along the palate". Thomas suggested that the backward tongue blade movement is a strategy to facilitate tongue body lowering. In other words, downward movement of the tongue body, facilitated by tongue blade retraction, pulls the tongue blade back and abruptly away from the palate to produce the distinctive click sound.

2.9. Posterior nasal fricatives

Posterior nasal fricatives have been described radiographically by Trost (1981) as a substitution for alveolar and postalveolar fricatives and involving the velum approximating the posterior pharyngeal wall to create nasal turbulence or what is sometimes called "rustle". The movement of the velum can be seen using radiography as a blurring due to velar flutter. Although nasal turbulence can result in severely distorted speech, it is associated with a relatively small size of velopharyngeal gap compared to speakers with hypernasality (Kummer et al. 1992). These authors suggest that the nasal turbulence is generated by "friction of the air being forced through a small velopharyngeal gap" (Kummer et al. 1992: 155), and that this friction creates a more audible sound than through a larger opening when there is a larger gap size. The positive implication for the small velopharyngeal gap size is that speech therapy may be effective and should be attempted before embarking on surgical intervention.

Nasal misarticulations, similar to Trost's (1981) posterior nasal fricatives, have been described as occurring in Japanese speakers with cleft palate (Abe 1987). Yamashita et al. (1992) used EPG to describe three Japanese individuals with cleft palate whose speech contained what they called "nasopharyngeal misarticulations". The EPG data showed that all the speakers produced these abnormal articulations with complete closure across the palate. Dent, Gibbon and Hardcastle (1992) found a similar pattern in a 9-year-old English speaking child with a cleft palate who produced complete contact in the velar region during /s/ and /z/ targets, which were heard by listeners as posterior nasal fricatives.

Figure 10a shows an example of EPG data with velar closure during a posterior nasal fricative. In this example, the EPG records show evidence of complete closure in the posterior region of the palate during the fricative noise. The presence of complete closure here suggests that air is escaping into the nasal cavity, and this is confirmed by the perceptual analysis of a posterior nasal fricative.

3. Abnormal articulations located below the velopharyngeal structures

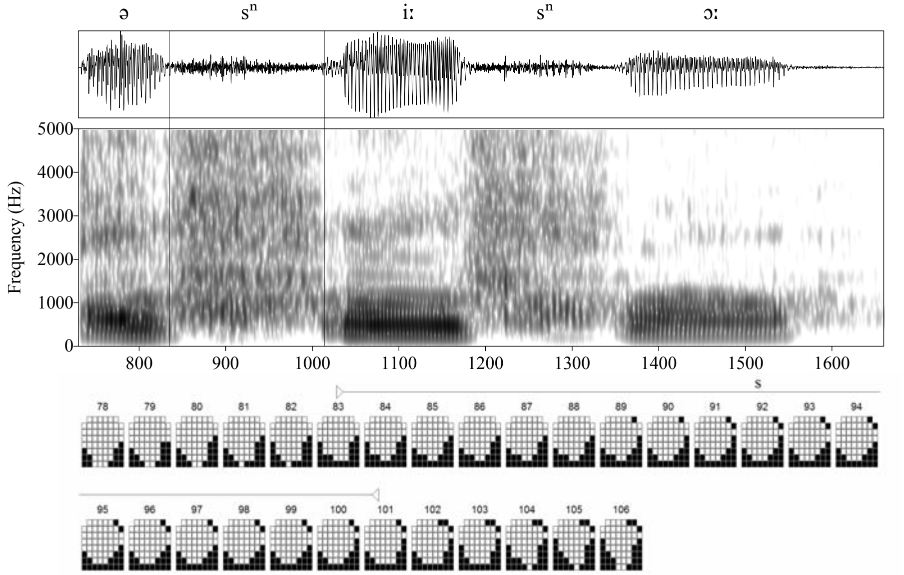
The previous sections described abnormal articulations where at least one major constriction was located in the oral region. However, many compensatory articulation errors have their major constriction below, or upstream of, the velopharyngeal structures. Glottal and pharyngeal articulations are examples, but they involve minimal or no EPG contact (Gibbon 2004), and as a result, EPG data recorded during these productions are often not illuminating. The following sections therefore discuss studies that have used techniques other than EPG to investigate errors with primary constrictions below the velopharyngeal structures.

3.1. Glottal stops

Glottal stops are the most frequently reported type of compensatory articulations in individuals with cleft palate (Peterson-Falzone 1989; Peterson-Falzone, Hardin-Jones, and Karnell 2001; Trost-Cardamone 1990). Glottal stops are usually produced as substitutions for oral plosive consonants, and also less frequently for fricatives and affricates (Peterson-Falzone, Hardin-Jones, and Karnell 2001). A study of children with severe articulation disorders reported their use of glottal stops for liquids and glides (Bzoch 1965). Like pharyngeal stops, glottal stops can be “coarticulated” with other consonants, such as bilabial plosives and alveolar plosives, resulting in double articulations.

Glottal stops are stop consonants produced at the level of the glottis, which begins with a forceful adduction of the vocal folds and build up of sub-glottal pressure, followed by the sudden opening of the vocal folds to release the pressure (Kummer 2001). Abnormal laryngeal constriction was reported in 15 out of 26 individuals with cleft palate or VPD who produced glottal stops (Kawano et al. 1997). They showed constriction of the whole larynx comprising the adducting elevation of arytenoids, approximation or contact of arytenoids with the epiglottis, and medial movement of the bilateral aryepiglottic folds. These authors found that constriction of the larynx with the adducting arytenoids elevated towards the epiglottis appears

(a) /s/ produced as [s^n] by a 9-year-old with cleft palate.



(b) /s/ produced as [s] by a normal speaker.

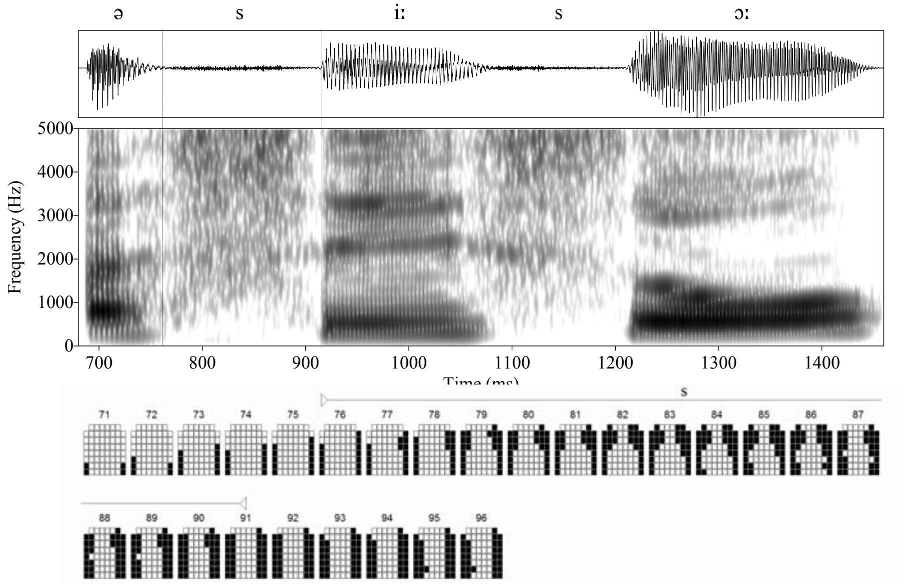


Figure 10. EPG data illustrating a nasopharyngeal fricative.The EPG printouts in (a) show a nasopharyngeal fricative used as a substitution for /s/ in the phrase “a seesaw” produced by a 9-year-old boy with cleft palate. A normal pattern for /s/ for the same phrase is shown in (b).

to enhance plosive production.

Using cineradiography, Henningsson and Isberg (1986, 1991) showed that limited or no velopharyngeal movement may be associated with glottal stop substitutions in speakers with repaired cleft palate. They found good to moderate velopharyngeal movement during non-glottal stop production; moderate, insufficient, or poor movement when producing weak pressure consonants; whereas poor to no velopharyngeal movement during production of glottal stops (see Figure 11). The findings supported the notion that glottal stop substitutions are compensatory articulations because the air stream was stopped at the glottis which is upstream from the velopharyngeal structures; velopharyngeal closure thus becoming unnecessary (Henningsson and Isberg 1991).

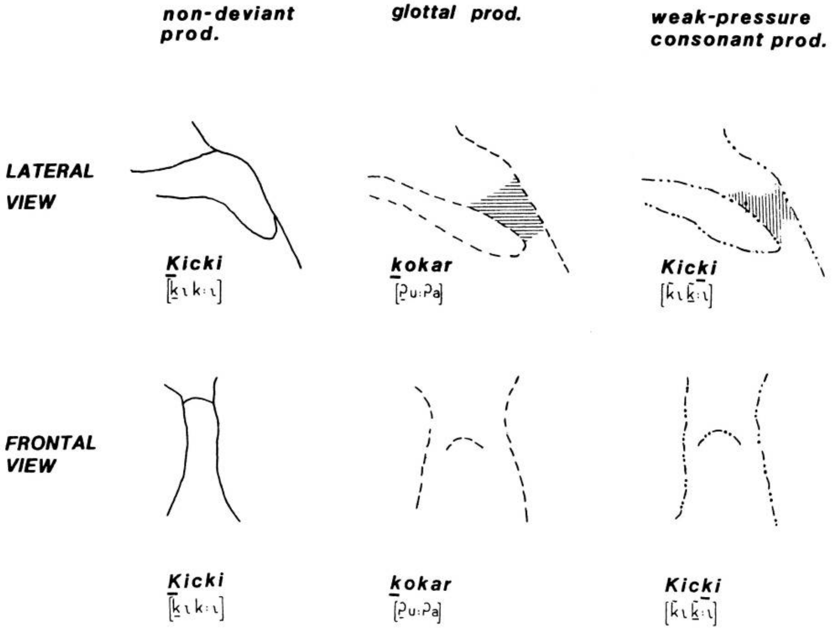


Figure 11. Tracings of velopharyngeal movements of a single speaker from cineradiographic frames for non-deviant, glottal, and weak-pressure consonant production, respectively. Shadowed areas indicate the position of pharyngeal flap. Good velopharyngeal movements were observed during non-glottal stop production (solid line) (adapted with permission from Henningsson and Isberg 1991).

It should be noted that glottal stops (referred to as preglottalization and glottalling) are part of normal English speech in certain phonetic contexts (Crystal 2005; Docherty et al. 1997; Ladefoged 2005; Wells 1982). It is important to distinguish between glottal stops that are produced as a normal part of everyday speech from those that are abnormal due to, for example, compensatory articulations in cleft palate speech. In terms of normal speech, Wells defined preglottalization as the insertion of a glottal stop /ʔ/ before voiceless plosives /p/, /t/, /k/, and affricate /tʃ/ when these consonants are in syllable-final position, preceded by a vowel, a liquid, or a nasal (e.g. “equal” is produced as /ˈiːʔkwəl/; “teacher” is produced as /ˈtiːʔtʃə/) (Wells 1982).

Glottalling, on the other hand, is the use of glottal stop to replace voiceless plosive /t/, and sometimes also /p/ and /k/. Glottalling is observed in many English accents (Wells 1982). For example, normal speakers can use the glottal stop for intervocalic /t/, so “butter” is realized as /ˈbʌʔə/, “bottle” is realized as /ˈbɒʔo/ (Ladefoged 2005; Wells 1982). In addition, nowadays some, particularly younger, speakers produce glottal stops at the ends of words such as “cap”, “cat”, and “back” (Ladefoged 2005). In American English, glottal stops are often produced in words such as “button” and “bitten”. Glottal stops are therefore an integral part of normal speech where they occur in specific phonetic contexts and accents, and should not be judged as abnormal except where they occur in other situations. This view contrasts directly with Larsen (2003: 557), who in our view stated erroneously that the glottal stop “represents substandard articulation, and may be heard in the speech of both the general population and the cleft population. Its use should be discouraged in both populations”.

3.2. Pharyngeal stops

Pharyngeal stops can be produced as substitutions for plosive targets, particularly /k/. Trost (1981) described how the pharyngeal stop can be produced at various locations in the pharynx, from high (oropharynx) to low (epiglottis). Kawano and colleagues’ videofluoroscopic data confirmed Trost’s radiographic findings, and they found some interesting additional features of the pharyngeal stop (Kawano et al. 1997). First, Kawano et al. (1997) found that a higher place of articulation (with tongue base contacting at oropharynx) was more frequently observed in older children and adults, and a lower site (at oro- and laryngopharynx) was more frequently seen in younger children. Second, several studies have noted that the pharyngeal stop is greatly influenced by preceding and following sounds (Brooks, Shelton, and Youngstrom 1965; Honjow and Isshiki 1971; Kawano et al. 1997; Trost 1981). These studies showed that pharyngeal stops occur primarily in the context of low/back vowels, such as /a/, /o/, /u/ but not high/front vowels such as /i/, /e/ (/k/ in these contexts was often produced correctly or as other types of errors). Their explanation of the vowel effect was that the production of pharyngeal stops involves the base of the tongue, and the pharyngeal location is too distal anatomically from that of the high vowels.

3.3. Pharyngeal/glottal fricatives and affricates

Pharyngeal fricatives are produced by a constriction between the tongue dorsum and the posterior pharyngeal wall (Morley 1970; Morris 1972; Trost 1981), and are most frequently seen as substitutions for oral fricatives. Kawano et al. (1997) used fiberscopy and videofluoroscopic analysis of a large group of over 250 individuals with cleft palate in order to investigate abnormal articulations that were judged perceptually as pharyngeal fricatives and affricates. They found that of 20 individuals (7.5 % of whole group) whose /ʃ/ and /s/ were judged as pharyngeal, the overwhelming majority (19) articulated the fricatives at the level of the larynx, with just one individual articulating at the pharynx. Similarly, of 20 individuals whose /tʃ/ and /ts/ were judged as pharyngeal, 16 were found to produce the sound at the level of the larynx. More specifically, these fricatives and affricates were produced in a narrow space between the arytenoids and epiglottis, or between the epiglottis and posterior pharyngeal wall or at both locations. Like the glottal stops, Kawano et al. (1997) found that velopharyngeal closure did not occur at the time the laryngeal sounds were produced, irrespective of whether the speaker had velopharyngeal competence. These authors concluded that in the overwhelming majority of cases, articulation errors that are heard as pharyngeal fricatives and affricates are much more likely to be in reality laryngeal fricatives and affricates. Kawano et al.'s study showed that distinguishing pharyngeal from laryngeal placement is difficult based on perceptual analysis alone, and that instrumental procedures greatly assist in the diagnostic process.

4. Summary

This chapter has illustrated how individuals with cleft palate produce a variety of distortions and substitutions in their efforts to produce turbulent speech sounds. Many atypical articulations are “compensatory errors”. The concept of compensatory errors is important in understanding the nature and treatment of articulation errors associated with cleft palate speech. In essence, compensatory errors are thought to arise as a result of abnormal learning, and not as a direct consequence of the structural abnormality. Consequently, these errors often persist in speech even when vocal tract function has improved through surgical or orthodontic intervention.

The presence of compensatory articulatory errors at any point in an individual’s development illustrates the complex relationship between vocal tract structure, function and quality of speech. This relationship means that the diagnosis of articulation difficulties in individuals with cleft palate requires close cooperation between a highly specialist multidisciplinary team. Team work is especially important to define the precise nature and extent of ongoing structural abnormalities, and the likelihood of improvement with surgical, medical, orthodontic or speech and language therapy intervention. These are highly complex tasks. For example, it has been shown that it is difficult to assess velopharyngeal function for speech when a speaker uses predominantly glottal articulations. Perceptual analysis combined with instrumental investigations of speech, such as those illustrated in this chapter, are important in making an accurate speech diagnosis.

Another factor that can give rise to individual differences in speech production is that speakers find different “solutions” when faced with similar structural anomalies. For instance, there are a number of possible ways to produce plosive sounds in the context of VPD. A speaker could produce a click or a glottal stop; both would achieve the same goal of plosion but using different articulatory mechanisms. Another possibility is that the speaker uses the tongue body to aid velopharyngeal closure, so-called “lingual assistance” (Brooks, Shelton Jr., and Youngstrom 1965; Trost 1981). Finally, a speaker may not make any attempt to compensate for the presence of VPD, with the result that obstruent targets are produced as weak pressure consonants, consonants with nasal emission or nasal consonants. Thus, individuals with similar anatomical abnormalities may present with quite different patterns of speech errors.

Accurate diagnosis is an essential prerequisite to effective intervention, underscoring the importance of a multidisciplinary approach. Many of the compensatory errors described in this chapter are amenable to positive behavioural change with speech therapy. Visual feedback using EPG has been found to be beneficial in establishing correct placement for obstruent targets located on the hard palate in children and adults with cleft palate (Dent, Gibbon, and Hardcastle 1992; Gibbon and Hardcastle 1989; Whitehill, Stokes, and Man 1996). Other studies have shown that apparently faulty velopharyngeal function can improve spontaneously by correcting abnormal compensatory articulations (Kawano et al. 1997). Other interventions are equally important because they can help prevent abnormal articulations arising in the first place. For example, speech therapy can help very young children expand their sound inventories using naturalistic speech therapy methods. Scherer (1999) has found that indirect therapy to develop vocabulary in young children with cleft palate was effective in increasing consonant inventories and reducing the production of glottal stops. With appropriate early management such as this, it is realistic to expect that most individuals born with cleft palate will attain speech that is indistinguishable from their typically developing peers within the preschool years.

Acknowledgements

We thank Rosmawati Aman for the Kalantanese EPG data. The research reported in this chapter was supported by various project grants including: BBC Children in Need; London Law; and Henry Smith Charity. We thank Alan Wrench who designed the EPG system.

Notes

1. Gibbon et al. (2007) compared the tongue-palate contact patterns between oral alveolar stops /t/ and /d/ and nasal alveolar stop /n/. They found that all stops showed similar spatial patterns, however, the oral alveolar stops had more contact and were more likely to have bilateral constriction than the nasal counterpart.

References

Abe, Masako

1987 Pathophysiology of nasopharyngeal articulation: Sound analysis and observation of articulatory movement. *The Japan Journal of Logopedics and Phoniatrics (Japanese)* 28: 239–250.

Albery, Elizabeth

1991 Consonant articulation in the different types of cleft lip and palate. M. Phil. thesis, Leicester Polytechnic.

Albery, Elizabeth, and Pamela Grunwell

1993 Consonant articulation in different types of cleft lip and palate. In *Analysing Cleft Palate Speech,* Pamela Grunwell (ed.), 83–111. London: Whurr Publishers.

Ball, Martin J. (ed.)

1993 *Phonetics for Speech Pathology*. 2nd ed. London: Whurr Publishers.

Brooks, Alta R., Ralph L. Shelton Jr., and Karl A. Youngstrom

1965 Compensatory tongue-palate-posterior pharyngeal wall relationships in cleft palate. *Journal of Speech and Hearing Disorders* 30: 166–173.

Bzoch, Kenneth R.

1965 Articulatory proficiency and error patterns of preschool cleft palate and normal children. *Cleft Palate Journal* 2: 340–349.

Chapman, Kathy L.

1991 Vocalizations of toddlers with cleft lip and palate. *Cleft Palate-Craniofacial Journal* 28: 172–178.

Chapman, Kathy. L., Mary Hardin-Jones, Julie Schulte, and Kelli Ann Halter

2001 Vocal development of 9-month-old babies with cleft palate. *Journal of Speech, Language, and Hearing Research* 44: 1268–1283.

Cohen, Steven R., Joseph Kalinowski, Don LaRossa, and Peter Randall

1991 Cleft palate fistulas: A multivariate statistical analysis of prevalence, etiology, and surgical management. *Plastic and Reconstructive Surgery* 87: 1041–1047.

Crystal, David

2005 *The Stories of English*. London: Penguin.

Dalston, Rodger M.

1992 Timing of cleft palate repair: A speech pathologist's viewpoint. *Problems in Plastic and Reconstructive Surgery* 2: 30–38.

Dent, Hilary, Fiona Gibbon, and William Hardcastle

1992 Inhibiting an abnormal lingual pattern in a cleft palate child using Electropalatography. In *Interdisciplinary Perspectives in Speech and Language Pathology,* Margaret M. Leahy and Jeffrey L. Kallen (eds.), 211–221. Dublin: School of Clinical Speech and Language Studies.

Docherty, Gerard J., Paul Foulkes, James Milroy, Lesley Milroy, and David Walshaw
1997 Descriptive adequacy in phonology: A variationist perspective. *Journal of Linguistics* 33: 275–310.

Fujimura, Osamu, I. F. Tatsumi, , and R. Kagaya
1973 Computational processing of palatographic patterns. *Journal of Phonetics* 1: 47–54.

Gibbon, Fiona E.
2004 Abnormal patterns of tongue-palate contact in the speech of individuals with cleft palate. *Clinical Linguistics and Phonetics* 18: 285–311.

Gibbon, Fiona E., and Lisa Crampin
2001 An electropalatographic investigation of middorsum palatal stops in an adult with repaired cleft palate. *Cleft Palate-Craniofacial Journal* 38: 96–105.

Gibbon, Fiona E., and Lisa Crampin
2002 Labial-lingual double articulations in speakers with cleft palate. *Cleft Palate-Craniofacial Journal* 39: 40–49.

Gibbon, Fiona E., Lucy Ellis, and Lisa Crampin
2004 Articulatory placement for /t/, /d/, /k/ and /g/ targets in school age children with speech disorders associated with cleft palate. *Clinical Linguistics and Phonetics* 18: 391–404.

Gibbon, Fiona E., and William Hardcastle
1989 Deviant articulation in a cleft palate child following late repair of the hard palate: A description and remediation procedure using electropalatography. *Clinical Linguistics and Phonetics* 3: 93–110.

Gibbon, Fiona E., Alice Lee, and Ivan Yuen
2007 Tongue-palate contact during bilabials in normal speech. *Cleft Palate-Craniofacial Journal* 44: 87–91.

Gibbon, Fiona E., Alice Lee, Ivan Yuen, and Lisa Crampin
2008 Clicks produced as compensatory articulations in two adolescents with velocardiofacial syndrome. *Cleft Palate-Craniofacial Journal* 45: 381–392.

Gibbon, Fiona E., Ivan Yuen, Alice Lee, and Lynne Adams
2007 Normal adult speakers' tongue palate contact patterns for alveolar oral and nasal stops. *Advances in Speech–Language Pathology* 9: 82–89.

Grunwell, Pamela
1993 *Analysing Cleft Palate Speech.* London: Whurr Publishers.

Grunwell, Pamela, and Jane Russell
1988 Phonological development in children with cleft lip and palate. *Clinical Linguistics and Phonetics* 2: 75–95.

Hardcastle, William J., and Fiona E. Gibbon
1997 Electropalatography and its clinical applications. In *Instrumental Clinical Phonetics*, Martin J. Ball and Chris Code (eds.), 149–193. London: Whurr Publishers.

Hardcastle, William J., Fiona E. Gibbon, and Wilf Jones
1991 Visual display of tongue-palate contact: electropalatography in the assessment and remediation of speech disorders. *British Journal of Disorders of Communication* 26: 41–74.

Hardcastle, William J., Rosemarie Morgan Barry, and Marion Nunn
1989 Instrumental articulatory phonetics in assessment and remediation: case studies with the electropalatograph. In *Cleft Palate: The Nature and Remediation of Communicative Problems* Jackie Stengelhofen (ed.), 136–164. Edinburgh: Churchill Livingstone.

Hardin-Jones, Mary A., and David L. Jones
2005 Speech production of preschoolers with cleft palate. *Cleft Palate-Craniofacial Journal* 42: 7–13.

Harding, Anne, and Pamela Grunwell
1993 Relationship between speech and timing of hard palate repair. In *Analysing Cleft Palate Speech,* Pamela Grunwell (ed.), 48–82. London: Whurr Publishers.

Henningsson, Gunilla, and Annika Isberg
1986 Velopharyngeal movement patterns in patients alternating between oral and glottal articulation: A clinical and cineradiographical study. *Cleft Palate Journal* 23: 1–9.

Henningsson, Gunilla, and Annika Isberg
1991 A cineradiographic study of velopharyngeal movements for deviant versus nondeviant articulation. *Cleft Palate-Craniofacial Journal* 28: 115–118.

Hiki, Shizuo, and Hidemi Itoh
1986 Influence of palate shape on lingual articulation. *Speech Communication* 5: 141–158.

Hixon, Thomas
1966 Turbulent noise sources for speech. *Folia Phoniatrica* 18: 168–182.

Honjow, Iwao, and Nobuhiko Isshiki
1971 Pharyngeal stop in cleft palate speech. *Folia Phoniatrica* 23: 347–354.

Howard, Sara J.
1993 Articulatory constraints on a phonological system: A case study of cleft palate speech. *Clinical Linguistics and Phonetics* 7: 299–317.

Howard, Sara J.
1998 A perceptual and electropalatographic case study of Pierre Robin Sequence. In *Clinical Phonetics and Linguistics,* Wolfram Ziegler and Karin Deger (eds.), 157–164. London: Whurr Publishers.

Howard, Sara J., and Caroline Pickstone
1995 Cleft palate: Perceptual and instrumental analyses of a phonological system. In *Case Studies in Clinical Linguistics,* Michael Perkins and Sara Howard (eds.), 65–90. London: Whurr Publishers.

Hutters, Birgit, Anja Bau, and Kirsten Brøndsted
2001 A longitudinal group study of speech development in Danish children born with and without cleft lip and palate. *International Journal of Language and Communication Disorders* 36: 447–470.

Kawano, Michio, Nobuhiko Isshiki, Iwao Honjo, Hisayoshi Kojima, Kyosuke Kurata, Fumiko Tanokuchi, Naohiro Kido, and M. Isobe
1997 Recent progress in treating patients with cleft palate. *Folia Phoniatrica et Logopaedica* 49: 117–138.

Krause, Charles J., Rosemary F. Tharp, and Hughlett L. Morris
1976 A comparative study of results of the Von Langenbeck and the V-Y pushback palatoplasties. *Cleft Palate-Craniofacial Journal* 13: 11–19.

Kummer, Ann W.
2001 *Cleft Palate and Craniofacial Anomalies: The Effects of Speech and Resonance*. San Diego: Singular.

Kummer, Ann W., Cindy Curtis, Melissa Wiggs, Linda Lee, and Janet L. Strife
1992 Comparison of velopharyngeal gap size in patients with hypernasality, hypernasality and nasal emission, or nasal turbulence (rustle) as the primary speech characteristic. *Cleft Palate-Craniofacial Journal* 29: 152–156.

Ladefoged, Peter
2005 *Vowels and Consonants.* 2nd ed. Oxford: Blackwell.

Ladefoged, Peter, and Ian Maddieson
1996 *Sounds of the World's Languages*. Oxford: Blackwell.

Larsen, Robert E.
2003 "Stops" in speech. *Cleft Palate-Craniofacial Journal* 40: 557.

McWilliams, Betty Jane, Hughlett L. Morris, and Ralph L. Shelton
1990 *Cleft Palate Speech.* 2nd ed. Philadelphia: B. C. Decker.

Michi, Ken-ichi, Yukari Yamashita, Satoko Imai, and Kohsuke Ohno
1990 Results of treatment of speech disorders in cleft palate patients: Patients obtaining adequate velopharyngeal function. In *Craniofacial Abnormalities and Clefts of the Lip, Alveolus and Palate,* Gerhard Pfeifer (ed.), 419–423. Stuttgart: Thieme.

Mills, Louise, Anne Gosling, and Debbie Sell
2006 Extending the communication phenotype associated with 22q11.2 Microdeletion Syndrome. *Advances in Speech-Language Pathology* 8: 17–27.

Moin, Parviz, and John Kim

1997 Tackling turbulence with supercomputers. *Scientific American* 276: 62–68.

Morley, Muriel E.

1970 *Cleft Palate and Speech.* 7th ed. Edinburgh: Livingstone.

Morris, Hughlett L.

1972 Cleft palate. In *Communicative Disorders: An Appraisal,* Alan J. Weston (ed.), 128–160. Springfield, IL: Charles C Thomas.

Netsell, Ronald

1990 Commentary. *Cleft Palate Journal* 27: 58–60.

O'Gara, Mary M., and Jerilyn A. Logemann

1988 Phonetic analyses of the speech development of babies with cleft palate. *Cleft Palate Journal* 25: 122–134.

Paliobei, Vassiliki, Anestis Psifidis, and Dimitrios Anagnostopoulos

2005 Hearing and speech assessment of cleft palate patients after palatal closure: Long-term results. *International Journal of Pediatric Otorhinolaryngology* 69: 1373–1381.

Peterson-Falzone, Sally J.

1989 Compensatory articulations in cleft palate speakers: Relative incidence by type. *Proceedings of the International Congress on Cleft Palate and Related Craniofacial Anomalies*, Jerusalem, Israel.

Peterson-Falzone, Sally J.

1990a A cross-sectional analysis of speech results following palatal closure. In *Multidisciplinary Management of Cleft Lip and Palate,* Janusz Bardach and Hughlett L. Morris (eds.), 750–756. Philadelphia: W. B. Saunders.

Peterson-Falzone, Sally J.

1990b Speech disorders associated with dentofacial anomalies. In *Pediatric Oral and Maxillofacial Surgery*, Leonard B. Kaban (ed.), 63–70. Philadelphia: W. B. Saunders.

Peterson-Falzone, Sally J., Mary A. Hardin-Jones, and Michael P. Karnell

2001 *Cleft Palate Speech.* 3rd ed. St. Louis: Mosby.

Peterson-Falzone, Sally J., Judith E. Trost-Cardamone, Michael P. Karnell, and Mary A. Hardin-Jones

2006 *The Clinician's Guide to Treating Cleft Palate Speech*. St. Louis: Mosby.

Phua, Yun Shan, and Tristan de Chalain

2008 Incidence of oronasal fistulae and velopharyngeal insufficiency after cleft palate repair: An audit of 211 children born between 1990 and 2004. *Cleft Palate-Craniofacial Journal* 45: 172–178.

Scobbie, James M., Sara E. Wood, and Alan A. Wrench

2004 Advances in EPG for treatment and research: An illustrative case study. *Clinical Linguistics and Phonetics* 18: 373–389.

Scherer, Nancy J.
1999 The speech and language status of toddlers with cleft lip and/or palate following early vocabulary intervention. *American Journal of Speech Language Pathology* 8: 81–93.

Sell, Debbie, Anne Harding, and Pamela Grunwell
1994 A screening assessment of cleft palate speech Great Ormond Street Speech Assessment. *European Journal of Disorders of Communication* 29: 1–15.

Stengelhofen, Jackie
1989 The nature and causes of communication problems in cleft palate. In *Cleft Palate: The Nature and Remediation of Communication Problems,* Jackie Stengelhofen (ed.), 1–30. Edinburgh: Churchill Livingstone.

Stevens, Kenneth N.
1998 *Acoustic Phonetics.* Cambridge, MA: The MIT Press.

Stoel-Gammon, Carol
1985 Phonetic inventories, 15–24 months: A longitudinal study. *Journal of Speech and Hearing Research* 28: 505–512.

Suzuki, Noriko, Yukari Yamashita, Ken-ichi Michi, and Tadashi-Ueno
1981 Changes of the palatolingual contact during articulation therapy in the cleft palate patients: Observation by use of dynamic palatography. Paper presented at the 4th International Congress of Cleft Palate and related Craniofacial Anomalies, Acapulco, Mexico.

Thomas, Kimberly
1997 EPG and aerodynamic evidence for the coproduction and coarticulation of clicks in IsiZulu. In *EUROSPEECH-1997,* George Kokkinakis, Nikos Fakotakis, and Evangelos Dermatas (eds.), 379–382. Retrieved July 4, 2006, from ISCA Archive: http://www.isca-speech.org/archive/eurospeech_1997

Traill, Anthony
1995 Place of articulation features for clicks: anomalies for universals. In *Studies in General and English Phonetics in Honour of Professor J. D. O'Connor,* Jack W. Lewis (ed.), 121–129. London: Routledge.

Trost, Judith E.
1981 Articulatory additions to the classical descriptions of the speech of persons with cleft palate. *Cleft Palate Journal* 18: 193–203.

Trost-Cardamone, Judith E.
1990 The development of speech: assessing cleft palate misarticulation. In *Cleft Lip and Palate: A System of Management,* Desmond A. Kernahan and Sheldon W. Rosenstein (eds.), 227–235. Baltimore: Williams and Wilkins.

Trost-Cardamone, Judith E.

1997 Diagnosis of specific cleft palate speech error patterns for planning therapy or physical management needs. In *Communicative disorders related to cleft lip and palate,* Kenneth. R. Bzoch (ed.), 313–330. Austin: Pro-ed.

Vallino, Linda D., and Bryan Tompson

1993 Perceptual characteristics of consonant errors associated with malocclusion. *Journal of Oral and Maxillofacial Surgery* 51: 850–856.

Vihman, Marilyn M., Marlys A. Macken, Ruth Miller, Hazel Simmons, and Jim Miller

1985 From babbling to speech: A re-assessment of the continuity issue. *Language* 61: 397–444.

Wada, Takuro, Motoyasu Yasumoto, Noriyuki Ikeoka, Yoshishige Fujiki, and Ryuichi Yoshinaga

1970 An approach for the cineflurographic study of articulatory movements. *Cleft Palate Journal* 7: 506–522.

Warren, Donald W.

1986 Compensatory speech behaviors in individuals with cleft palate: A regulation/control phenomenon? *Cleft Palate-Craniofacial Journal* 23: 251–260.

Warren, Donald W., Rodger M. Dalston, and Eileen T. Dalston

1990 Maintaining speech pressures in the presence of velopharyngeal impairment. *Cleft Palate Journal* 27: 53–58.

Wells, John C.

1982 *Accents of English I: An Introduction*. Cambridge: Cambridge University Press.

Whitehill, Tara, Stephanie Stokes, William J. Hardcastle, and Fiona Gibbon

1995 Electropalatographic and perceptual analysis of the speech of Cantonese children with cleft palate. *European Journal of Disorders of Communication* 30: 193–202.

Whitehill, Tara L., Stephanie F. Stokes, and Yonnie H. Y. Man

1996 Electropalatography treatment in an adult with late repair of cleft palate. *Cleft Palate-Craniofacial Journal* 33: 160–168.

Witzel, Mary A.

1991 Speech evaluation and treatment. *Atlas of the Oral and Maxillofacial Surgery Clinics of North America* 3: 501–516.

Wrench, Alan A.

2007 Advances in EPG palate design. *Advances in Speech-Language Pathology* 9: 3–12.

Wyatt, Rosemary, Debbie Sell, Jane Russell, Anne Harding, Kim Harland, and Elizabeth Albery

1996 Cleft palate speech dissected: A review of current knowledge and analysis**.** *British Journal of Plastic Surgery* 49: 143–149.

Yamashita, Yukari, Ken-ichi Michi, Satoko Imai, Noriko Suzuki, and Hiroshi Yoshida

1992 Electropalatographic investigation of abnormal lingual-palatal contact patterns in cleft palate patients. *Clinical Linguistics and Phonetics* 6: 201–217.

Formant-cavity affiliation in sibilant fricatives

Martine Toda, Shinji Maeda and Kiyoshi Honda

1. Introduction

Sibilant fricatives form a subgroup of fricatives that produce high intensity noise. This high intensity reflects the mechanism of noise generation, where the airflow is directed against an obstacle such as the upper or lower incisors (Shadle 1985, 1991). This mechanism is also supported by the high jaw position of sibilants (Lee, Beckman and Jackson 1994; Mooshammer, Hoole and Geumann 2007), implying that the position of the lower incisors with respect to the airstream is relevant for noise generation. The intense frication noise is known to carry perceptual cues for the place of articulation of sibilants (e.g. English /s/ vs. /ʃ/, Harris 1958). Contrary to stops, nasals and non-sibilant fricatives, sibilants are thus a rare kind of consonant whose place as well as manner cues are signaled primarily by the spectral structure of the segment itself, in a way similar to what is found in vowels.

The acoustic characteristics of fricatives, however, differ in many respects from those of vowels. Firstly, the spectral envelope of the noise source can maintain a high intensity level, up to about 15 kHz for some utterances, whereas the voiced source of vowels damps out rapidly above 5 kHz. Secondly, since the generation of fricatives' source requires a narrow constriction, the back cavity located behind the constriction tends to be acoustically inactive (e.g. Fant 1970 [1960]: 182-185). Therefore, the formant structure of sibilant fricatives is primarily determined by the front cavity resonances. Thirdly, fricatives generate both resonances and antiresonances, because the noise source is located midway in the vocal tract, in particular within the front cavity, near the teeth (Shadle 1991 for [s] and [ʃ][1]). The antiresonances, like resonances, contribute in shaping the overall spectrum. To summarize, while a source-filter independence can be assumed in obstacle-type fricatives (Shadle 1985: 178), sibilants' spectrum differs from that of vowels not only in the nature of excitation sources (noise source opposed to the periodic glottal source) but also in its specific spectral structure.

The noise spectra of fricatives are often described by global acoustic parameters such as the center of gravity, peak frequency, spectral tilt, etc. However, such descriptions are not adequate when the spectral structure is to be interpreted in articulatory terms.

This chapter attempts an analytical description of sibilants' spectral structure, in order to clarify the articulatory-to-acoustic mapping in sibilant fricatives. After a short summary of the literature about the noise source, the general acoustic (resonance) properties of simple fricative-like models will be discussed and illustrated through acoustic simulations. Then, the three-dimensional vocal tract shape of Polish sibilants will be examined through teeth-inserted high-resolution magnetic resonance imaging (MRI) data, in order to provide a detailed description of a variety of sibilant fricatives. Finally, we will attempt to clarify the formant-cavity affiliation in Polish /ç/ and /ṣ/[2] by running acoustic simulations with realistic and modified models derived from the MRI data. The results will be discussed in relation to their natural spectra.

2. Fricatives' noise source

Several types of noise sources are assumed to exist in fricatives. When an air jet forms at the constriction and flows into a wider cavity, such as the front cavity, the laminar flow turns into turbulences, at a certain threshold formulated by Reynolds' number. These turbulences give rise to monopole and quadrupole noise sources (Pastel 1987: 21). The effect of quadrupole sources is however considered to be minor in the speech signal.

In addition to the just-described sources resulting from a free jet, *obstacle* sources arise in sibilant fricatives such as [s] and [ʃ], where the air jet exiting from the constriction meets an obstacle. The upper or lower incisors constitute the obstacle, depending on the model (Shadle 1991). An obstacle configuration gives rise to dipole sources. The noise intensity is amplified and spreads over higher frequencies when an obstacle source is involved (Stevens, 1998: 107).

Dipole sources can be understood as two monopole sources pulsating in opposite phase, one releasing, the other absorbing [air] mass (Pastel 1987: 21). Those sources are oriented perpendicular to the obstacle (Curle 1955). It has been shown that dipole sources couple most efficiently to the principal acoustic mode along the length of the vocal tract, when the obstacle is oriented perpendicular to it (Pastel 1987).

In [ç] and [x], Shadle (1985: 178) has shown that the noise source can be well approximated by dipole sources distributed along a surface downstream from the constriction. In these articulatory configurations, the air jet exiting from the constriction impinges on the palatal roof, but it is not perpendicularly oriented with respect to the obstacle. Therefore, the resulting noise level might be lower than in the case where the obstacle is oriented perpendicularly (Stevens 1998: 102). Nevertheless, it can be suggested that such a source generation mechanism plays an important role in general in fricatives that are characterized by a long constriction formed by a domed tongue. Fant (1970) notes that the best fit between acoustic modeling and actual sound spectrum in 'š' ([ʃ]) is obtained when the source is placed within the palato-alveolar constriction. Narayanan and Alwan (2000) also obtained an optimal fit of the simulated spectrum to the actual noise spectrum of English /ʃ/ and /ʒ/ by using different source types: (1) a *wall* dipole source in the vicinity of constriction, in addition to (2) another dipole source located at the teeth, and (3) a constant monopole source located at constriction exit, as well as (4) a voiced source for /ʒ/. The authors note that these dipole sources greatly influenced the spectral shape over most of the frequency range, i.e. up to 10 kHz. Therefore, it can be supposed that sibilants with a similarly long constriction also involve, at least, the wall-obstacle source (near the constriction) in addition to the teeth-obstacle source.

Dipole source is considered to have a relatively flat long-term spectrum with a single broad peak (see Stevens 1998: 103), and its central frequency as well as the left and right slopes seemingly vary according to the subject and to the identity of the fricative (Narayanan and Alwan's 2000 three parameter dipole source models, see Figure 1). Although the spectral shape of fricatives is the product of both the source envelope and the vocal tract transfer function, the latter largely determines the individual spectral characteristics of the sibilant fricatives.

3. Theoretical considerations: Acoustic structure of fricatives

Like vowels, consonants possess an underlying formant pattern, the F-pattern defined by Fant (1970: 25), which explains the formant transitions at consonant-vowel boundaries. However, the F-pattern is not entirely visible in the acoustic realization of consonants. In fricatives, the front cavity resonances excited by a supraglottal source are thought to determine the overall spectral shape, because of the existence of a narrow constriction.

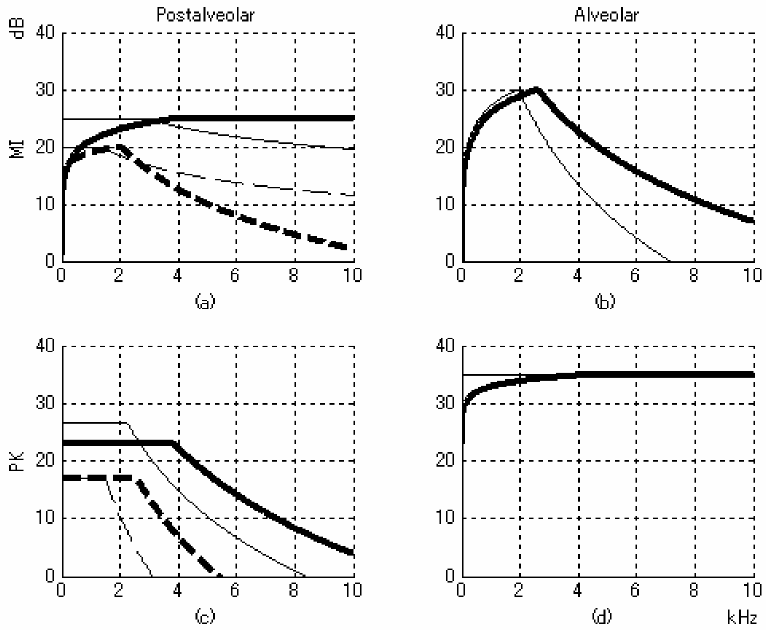


Figure 1. Spectra of three parameter dipole source models formulated by Narayanan and Alwan (2000) for English fricatives. The upper graphs are for subject MI (a and b); the bottom graphs for subject PK (c and d). On the left are graphs for postalveolar /ʃ/ and /ʒ/ (a and c), and on the right, those for aveolar /s/ and /z/ (b and d). In each graph, the curves corresponding to the voiceless fricatives are drawn with a thick line and those for the voiced fricatives with a thin line. Plain lines correspond to the teeth dipole source, broken lines to the wall dipole source. The monopole and voiced sources are not represented.

In addition, the transfer function of fricatives is characterized by antiformants of various origins, as will be discussed below. In passing, we note in this paper that only the visible formants will be referred to as fricative formants F1, F2, etc., regardless of their relation to the adjacent vowel formants.

3.1. Method

The simplest fricative model, similar to Flanagan (1972 [1965]: 74), is a straight-tubed tract that contains a constriction. In an [s]-like configuration, a narrow tube, which represents the tongue constriction and the small front cavity, as well as the lip opening, would be put at the frontmost portion of the model, as shown in Figure 2.

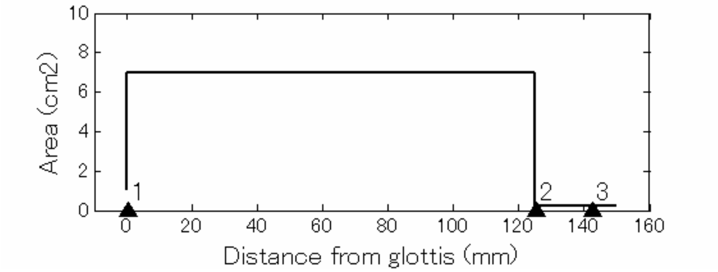


Figure 2. A simple [s]-like fricative model. Numbered black triangles indicate alternative source locations. Back cavity length = 12.5 cm; back cavity cross-sectional area = 7 cm^2; front cavity length = 2.5 cm; front cavity cross-sectional area = 0.2 cm^2.

It is possible to calculate the transfer function (i.e., the transfer ratio of the radiated sound pressure over the source level) of such a vocal tract model by using acoustic simulation. The program VTF_fric, a frequency domain version of VTCalcs (Maeda 1982), is used in this study. This program takes an area function as input, as schematized in Figure 2, and the parameters that are listed in Table 1 for default values.

Table 1. Default parameter values used with the acoustic simulation program VTF_fric.

wall properties	*radiation load (at lips)*	*subglottal system*	*source type*
Yielding	RL circuit	OFF	pressure
output type	*glottal area*	*length of sections*	*frequency samples*
radiated pressure	1 cm^2	1 mm	every 10 Hz from 10 to 24000 Hz

As in the classical lumped T-type transmission line representation of an acoustic tube, each section of the vocal tract model consists of 2 series elements representing the acoustic mass L/2, and a parallel element representing the compliance C. The mono-pole source is a flow source (a current-source in the electrical analog) connected in parallel with C, whereas the dipole source is a pressure source (a voltage source) connected in series with L/2. In the original form, the dipole source consists of a flow source and sink placed in parallel within a short distance of each other, which is an equivalent to a pressure source in the longitudinal direction of the vocal tract (Shadle 1985).

RL circuit means in the simulation that the acoustic load at the lip opening is approximated by a lumped resistance (R) and an acoustic mass (L) connected in parallel.

The subglottal system, comprising the tracheal tube and the lung volume, is neglected in the present calculations, and the glottal end is directly connected to a constant air-pressure source, 8 cmH_2O (= 8 hPa). In Table 1, subglottal system = OFF specifies this omission of the subglottal system. The frequency range of our simulations comprises up to 24 kHz. This unusually wide frequency range permits the visualization of the higher resonances of the fricative models and thus helps us to correctly interpret the resonance modes. However, in high frequencies, say 8 kHz and above, the simulated transfer functions will not necessarily match the actual noise spectra. This is because cross modes or higher order modes (resonance modes that are not longitudinal with respect to the vocal tract) can also occur at such high frequencies; this is naturally not predicted by a simulation method that assumes a plane-wave propagation.

3.2. Natural resonance frequencies of fricative models

When the two-tube resonating system is excited by a source (monopole or dipole) located at the glottis (i.e., source location 1 in Figure 2), the transfer function (Figure 3, upper curve) is characterized only by poles that correspond to the natural frequencies of the model.

This all-pole function involves three series of resonances. The lowest resonance of this system is a Helmholz resonance, expected around 170 Hz. This resonance is not noticeable on the transfer function because of the significant damping of the low frequencies due to the glottal opening. The resonances appearing with the shortest interval are those of the 12.5 cm long back cavity. Since the both ends of the back cavity are almost closed, we expect n/2 wavelength resonances around 1.4, 2.8, 4.2 kHz and so on, which roughly fits the observed interval of 1.3 kHz in the simulated transfer function. The resonances of the front cavity of 2.5 cm long are superimposed on the back cavity resonances. Since both ends of this cavity are roughly open, we expect multiples of half wavelength resonances around 7, 14, 21 kHz, and so on. Actually, the first resonance is observed around 6.5 kHz in the simulation. This value is slightly lower than expected because of the effect of lip radiation. The second and third resonances overlap the back cavity resonances around 13 and 19.5 kHz, respectively, making their relative intensity greater.

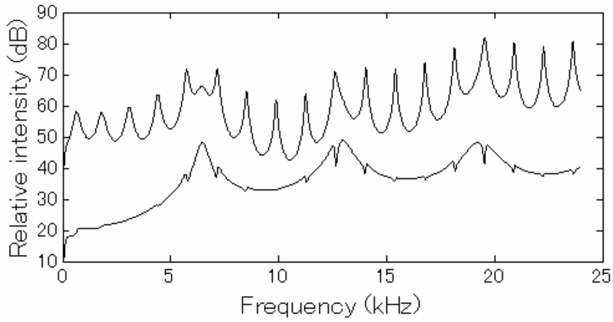


Figure 3. Transfer functions of a simple /s/ model. The upper curve (displayed with +30 dB offset for clarity) corresponds to the source location 1 (glottis), as shown in Figure 2. The lower curve corresponds to the source location 2 (at the junction of the back and front cavities).

3.3. Source located in front of a vocal tract cavity

When an excitation source is located downstream in a vocal tract cavity, this cavity produces pole-zero pairs in the transfer function. For example, when the noise source is located at the entrance of the front cavity as indicated by the triangle corresponding to the incisor position in Figure 2, the back cavity produces pole-zero pairs. The paired poles and zeros cancel each other in a theoretical condition where the interaction between the front and back cavities is minimal, that is, when the difference in area between these cavities is large. In real conditions, the vocal tract cavities are not completely independent, and the pole-zero pairs will not cancel each other completely. The narrower the constriction, the closer will be the distance between the paired pole and zero, so that their influence on the overall spectrum will tend to be negligible; this is seen in Figure 3, bottom curve. For straight tube models, an area ratio of 1 to 10 is usually considered to be sufficiently large that the interaction between two adjacent cavities can be ignored.

3.4. Source located midway within a cavity

When an excitation source is located midway within a cavity, this cavity produces free poles and free zeros (Figure 4). In the present example, this applies to the front cavity when the source is located at 8 mm from the constriction exit (source 3 in Figure 2).

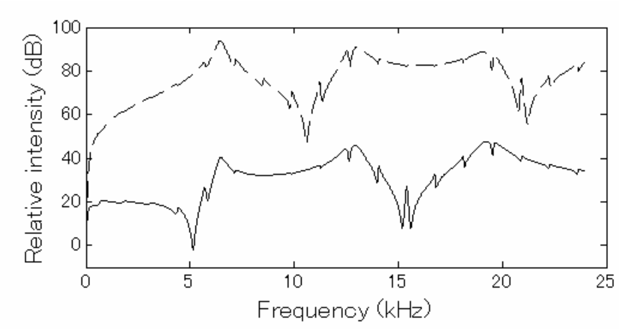


Figure 4. Transfer function of a simple /s/ model excited by a source located midway within the front cavity (8 mm behind the lip opening, numbered 3 in Figure 2). Top: Flow source equivalent to monopole. Bottom: Pressure source equivalent to dipole. The relative intensity is defined with respect to the first frequency sample (10 Hz) set to 0 dB (this applies to all of our transfer functions).

The zeros occur due to multiple reflections of sound waves in the cavity at the frequencies where the phase reversal takes place between forward and backward waves to cancel each other. Because of the inversion of phase in the forward/backward direction, dipole pressure sources (correctly oriented with respect to the resonance mode) do not cause the same response as monopole flow sources (Figure 4). Notice that the frequency of free poles remains the same independently of the source types, as well as the source locations (cf. Figure 3). Also, the frequency interval of the free zeros is not related to the frequency of poles, but to the position of the excitation source within the front cavity.

If the noise source were distributed along a cavity, as is the case for the long and narrow constriction characterizing [x], the frequency of free zeros would also be distributed along the frequency axis. If this were the case, the influence of free zeros on the overall spectrum would be insignificant in comparison with that of free poles, whose frequency remains constant regardless of the location of the sources. In [s] and [ʃ], since the source is likely to be localized around the incisors (Shadle 1985; 1991), it is more likely that the noise spectrum would be affected by sharp free zeros.

3.5. Vocal tract side branches

The location of the source is not the only factor that can introduce zeros in the spectrum of fricative-like configurations. In some articulatory settings, typically in English /ʃ/ and Polish /ʂ/, the front cavity is enlarged by a sublingual space created by a raised tongue tip combined with a retracted position of the tongue. In such a case, especially when the sublingual cavity is large, the lowest resonance arising from the front cavity will be best modeled by considering the sublingual space as a side branch. A side branch in an acoustic system creates the condition for a sound wave to bifurcate and reunify at the observation point with phase reversal, where zeros appear in the transfer function in the same manner as occurs with a source located midway within a vocal tract cavity.

The configuration schematized in Figure 5 mimics the case of a post-alveolar fricative, where the main tract comprises three tubes: the back cavity, the constriction and the front cavity. Its total length is now 17 cm, the constriction being 2.5 cm long (same as the previous model, Figure 2) and the front cavity measuring 2 cm. A side branch of 1.5 cm is connected at the posterior end of the front cavity. The source is located 1 cm behind the lip exit, as indicated by the triangle.

Theoretically, a side branch introduces additional poles and zeros in the transfer function (e.g. Jackson, Espy-Wilsn and Boyce 2001). However, since the area of the sublingual cavity is likely to be about the same as that of the front cavity at the branching point, a strong interaction is expected between the side branch and the front cavity. The sublingual cavity is therefore likely to modify the pole and zero frequencies of the front cavity, as seen in this simulation (Figure 5, bottom), where the lowest peak shifts from about 3 kHz to 2 kHz between the not-branching (broken line) and branching (plain line) configurations. It is worth noting that the effect of the sublingual cavity is very similar to that of lengthening the front cavity. The dotted line represents the transfer function of a straight model, without a side branch, where the front cavity measures 3.5 cm (the total length and constriction length are kept the same). The main difference is found in the frequency of the back cavity pole-zero pairs. Note, however, that when the source is located behind the front cavity, a straight model would produce front cavity poles only, whereas a branching model will produce free poles and zeros.

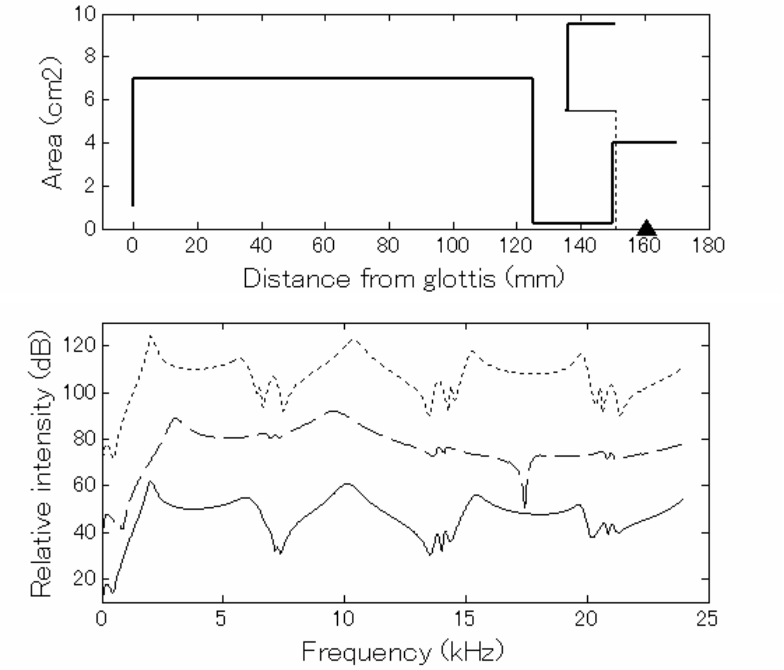


Figure 5. Top: area function of a branching post-alveolar fricative model. The side branch (15 mm long) is connected to the main tract at the posterior end of the front cavity. Bottom: transfer functions. The transfer function of the model without a side branch is given by the broken line. That of a straight (non-branching) model whose front cavity measures 35 mm is given by the dotted line. The solid line is for the branching model. All of the transfer functions are obtained with a pressure source located 10 mm behind the lip exit. The curves are shifted by 30 dB offsets for a better visibility.

Another kind of side branch consists of the interdental space. These lateral cavities are morphologically delimited by the tongue sides, upper and lower teeth, and cheeks. These side branches are connected to the main vocal tract at the lateral sides of the front cavity. Although their volume and thus the magnitude of their resonances are assumed to be rather small, they may have some impact on the spectrum during the vowel-fricative transitions, creating formant discontinuities due to the sudden coupling/decoupling of their poles and zeros.

3.6. Generalizations

To summarize, the transfer function of fricatives is likely to be made up of the following constituents:

a. Pole-zero pairs of independent vocal tract cavities (e.g. back cavity) excited by a source located downstream (within the front cavity), having a minor influence on the overall spectral shape;
b. An appreciable shift of front cavity poles and zeros due to the connection of a sublingual cavity side branch, in postalveolars;
c. Free zeros and free poles arising from cavities with a noise source located within it. In typical [s], [ʃ], or [ṣ] configurations, the spectra of sibilant fricatives are likely to be characterized by a sharp cut-off at the left side of the first fricative formant because of a free zero of this kind. Other sharp dips are also likely to occur in frequencies above the main spectral peak or peaks;
d. Free poles, not accompanied by zeros, arising from cavities excited by a source located upstream. In an [ʃ]/[ṣ]-like model, consisting of three tubes instead of two (wide back cavity tube, narrow constriction tube, and wide front cavity tube), sources located within the constriction tube will make the front cavity produce free poles. Some articulations of [s] can also be represented by the three-tube model, with different lengths. Here, the third tube, wider than the second constriction tube, represents the lip cavity. With a noise source located at the incisors (at the end of the constriction tube), the lip tube, despite its short length, will be responsible for free poles of high, but audible frequencies.

4. Vocal tract shapes of sibilant fricatives

Recently, acoustic studies on fricatives have benefited from the advancement of medical imaging techniques, especially magnetic resonance imaging (MRI). In earlier times, the vocal tract area function of fricatives was difficult to estimate from the sagittal profile derived from x-ray data, and needed to be completed with separate data such as x-ray tomography (e.g. Shadle 1991). Moreover, the derived area functions were optimized so that the calculated vocal-tract transfers better matched the measured spectra of the corresponding fricatives (Fant 1970). In the investigation of the acoustic system of fricatives, data of particularly high accuracy are needed because of the presence of constrictions with a small dimension. This is because small errors in the small structures can greatly affect the estimated formant frequencies. For example, in Narayanan, Alwan and Haker (1995), three-dimensional data on English fricatives were published for the first time, including, in particular, the shape as well as the volume of the sublingual cavity in /ʃ/. However, the accuracy of the extracted vocal tract contour in the front cavity region needed improvement, especially for the purpose of acoustic studies.

More recently, we acquired a large set of voiceless sibilant fricative MRI data (30 subjects, Toda 2009), from which we have extracted a few samples here. In this study, priority was given to the accuracy of the data in the front cavity region comprising the complex teeth contour as well as the narrow tongue constriction.

In order to correctly apprehend the vocal tract shape of sibilant fricatives, the teeth contour can by no means be ignored. It used to be, however, a challenge to combine the teeth contour with the MR images (usually, the teeth appears in the same brightness as the air, and thus their contour is not visible in MR images). Several approaches have been experimented with, including combining hardware dental casts to MRI data, applying a coating medium (Narayanan, Alwan and Haker 1995), or using post-hoc image processing (Takemoto et al. 2004). In the present study, a manual post-hoc image processing procedure was adopted. It consisted in extracting the negative image of the teeth from a set of MRI data where the tongue and the lips, tightly pressed on the teeth, served as contrasting medium. The ‘numerical tooth casts’ thus extracted were then inserted into the fricative MRI data. The main advantage of this method is the high precision with which the teeth contour can be combined with the articulatory MRI data.

The Polish sibilants will be analyzed in the present study. The Polish language possesses three voiceless sibilants with contrastive ‘places of articulation’, which constitute an interesting sample of sibilant sounds that the human vocal tract is able to produce in robust ways. In particular, Polish possesses two kinds of postalveolar sibilant fricatives, /ṣ/ and /ç/, radically different from one another in their tongue shape, /ç/ being strongly palatalized, as reported in Halle and Stevens (1997) based on Wierzchowska’s (1965) data. From the acoustic point of view, the centers of gravity of the frication noise of the two sounds overlap with each other (Jassem 1995: 3 male subjects; Nowak 2006: 1 female subject; 1 male subject in Zygis and Hamann 2003), or they are lower for /ṣ/ (1 female subject in Zygis and Hamann 2003); for both fricatives, the frication noise seems not far different from that of English /ʃ/. Nevertheless, the frication noise appears to be perceptually distinctive when these fricatives are presented in isolation (Lisker 2001, Nowak 2006). The F2 transitions, as shown in Figure 6, are very different between /ṣ/ and /ç/, in that the latter has a higher locus. The transitional information, especially of the following vowel, has a strong effect on native listeners’ judgement about the identity of the sibilant (Nowak 2006).

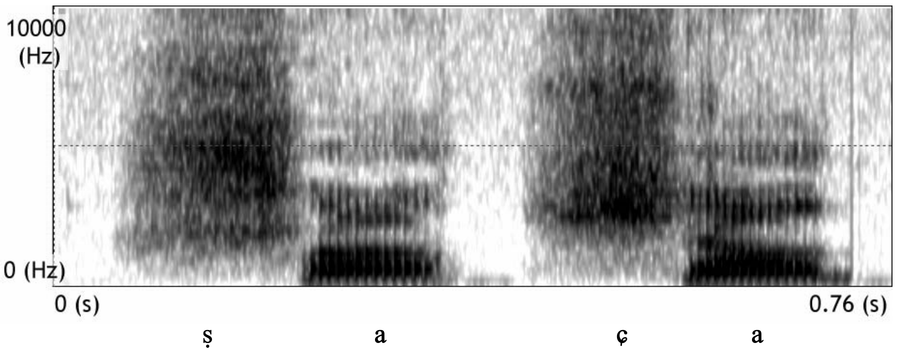


Figure 6. Spectrogram of the initial syllables /ṣa/ and /ça/ extracted, respectively, from the words **sza**tan (‘devil’) and **sia**dam (‘I hit), uttered by subject P2.

According to the UPSID database (Maddieson and Precoda 1989), tongue shape contrast (involving palatalization, dental/alveolar or palatal/retroflex/ palato-alveolar contrasts) is one of the most common ways of contrasting sibilant fricatives in the world’s languages, following voiced/voiceless and anterior/non-anterior contrasts. For instance, in the languages having an inventory of two or more sibilant fricatives (304 languages out of 451), 191 make use of a contrast concerning ‘place’ (in the general sense covering tongue position as well as tongue shape), whereas, by comparison, 147 make use of a voiced/voiceless contrast. Among the 191 languages having a place-related feature, 180 languages possess an anterior/non-anterior contrast, and 39 a tongue shape contrast. In languages having more than 3 sibilants, anterior/non-anterior and tongue shape contrasts can be combined.

Therefore, it is of interest to consider whether the variation in tongue shape results in specific characteristics of the frication noise spectrum in addition of that of formant transitions. In particular, Halle and Stevens (1997) suggest that the presence of a long and thin palatal channel should reduce the acoustic coupling of the back cavity in /ç/, in opposition to /ṣ/. Also, according to the same authors, the palatal channel should provide its own formants and antiformants, which also contribute to the spectral difference between /ç/ and /ṣ/.

4.1. Method

Two native Polish male subjects participated in the experiment. The data acquisition required two sessions: (1) acoustic recordings and (2) MRI data collection. The acoustic recordings are described in the following section (5.1.1.).

4.1.1. MRI acquisition

The MRI data were collected at the Brain Activity Imaging Center, ATR (Kyoto, Japan), with a Shimadzu-Marconi Eclipse 1.5T scanner. The MRI parameters of the acquisition are given in Table 2. The subjects lay down in supine position in the scanner. In this study, in order to obtain the most *natural* data possible, special efforts were made to keep the scan duration as brief as possible (23 seconds). A high sensitivity custom-made coil (Takano et al. 2003) was used to obtain high resolution data for the orofacial region with a good signal-to-noise ratio in spite of the short acquisition time. The subjects were presented words or word sequences containing the target fricatives in various phonetic contexts (2 or 3 vowel contexts – open, closed and rounded – for each fricative), and were told to produce a sustained fricative *as if* it was realized in those contexts (therefore we call them *virtual* contexts). They were instructed to keep producing the fricative, during the whole duration of the scan, and maintain the same posture when they could not sustain it to the end. These instructions are to prevent the blurring of the MR images due to movement. The subjects were already accustomed to the corpus, and trained for the sustained phonation during the acoustic recording session. They rehearsed again immediately before the MRI session took place. The corpus was printed in large characters and glued inside the scanner so that the subjects could refer to it during the whole MRI session.

Table 2. MRI acquisition parameters

scan direction	*field of view*	*resolution (1 pixel =)*	*slice thickness and spacing*
sagittal	128 mm × 128 mm × 45 mm	0.25 mm × 0.25 mm	1.5 mm
TE	*TR*	*scan duration*	*sequence name*
3.3 ms	10 ms	23 s	Fast 3D

The consistency of the articulation within the sibilant phonemes was verified through the comparison of the raw MRI data corresponding to the various phonetic contexts. The sustained fricatives /s/, /ç/ and /ṣ/ presented in the words *sadza* [sɑdzɑ] (‘lampblack’), *siadam* [ɕɑdɑm] (‘I hit’) and *szatan* [ṣɑtɑn] (‘devil’) will be analyzed in this study.

In addition, special *tooth scans* were acquired in order to extract the subjects’ numerical dental casts. For that scan, the subjects were instructed to put their tongue and lips tightly close to their teeth. The same acquisition parameters as in the fricative scans were used in order to insure the uniformity of the data and facilitate the post-processing.

4.1.2. Tooth cast extraction and insertion

The numerical tooth casts were extracted from the MRI tooth scans through brightness inversion and by manually selecting the regions of interest, including bones (mandible and hard palate).

The upper and lower tooth casts were then merged manually into the fricative scans by means of medical imaging software (Intage-rv), through rotations and translations. The details of the bones were of great help in the insertion of these dental casts into each of the sibilant dataset with good accuracy. The precision of this procedure is estimated to be around 0.5 mm and 1 degree.

4.1.3. Measurement of vocal tract area functions

The vocal tract area function of sibilants was measured. Considering the vocal tract to be roughly horizontal in the front oral cavity region, the area of the airway was measured on coronal images (perpendicular to the original sagittal images) at an interval of 1 mm. This method lacks accuracy in two measures: (a) the length of the lip tube, where the side boundaries of the airway at its extremity are not well captured because of its curvature (For the subject P1, no lip cavity could be measured at all); and (b) the length of the front cavity comprising the sublingual portion and the length of the tongue constriction, because of their angle with respect to the coronal slices. Both of these measurement errors would lead to shorter lengths than the actual vocal tract, and thus a shift of the resonances (in the longitudinal modes) towards higher frequencies in the estimated transfer functions. However, we assume that the overall resonance pattern will not be significantly affected by these inaccuracies, since the structural relationships among the vocal tract cavities are preserved.

4.2. Results

The mid-sagittal images as well as coronal images showing the constriction and the tongue dorsum are given in Figures 7-8. From these figures, it appears that subjects P1 and P2 produced the three Polish sibilants with certain common characteristics.

/s/ has a very front constriction, where the tongue is in contact with the internal surface of the lower incisors. The tongue is the most retracted in /ʂ/, with the narrowest pharyngeal cavity, especially for P2. A large sublingual cavity is noticeable for /ʂ/ on both the sagittal as well as coronal images. Although P1 exhibits a sublingual cavity also for /ɕ/, its volume is smaller than in /ʂ/. It can also be remarked that in /ʂ/, the depth of the sublingual cavity greater than the distance between the narrowest point of the tongue constriction to the teeth. The lowest front cavity resonance (other than Helmholtz resonance) should therefore involve the vertical dimension of the front cavity, including the sublingual portion. P2 realizes a gradual lip protrusion in the order /s/ < /ç/ < /ṣ/, whereas no particular lip gesture is noticeable for P1. The increasing front cavity size /s/ < /ç/ < /ṣ/ can also be seen on the area functions given for P2 (Figure 9).

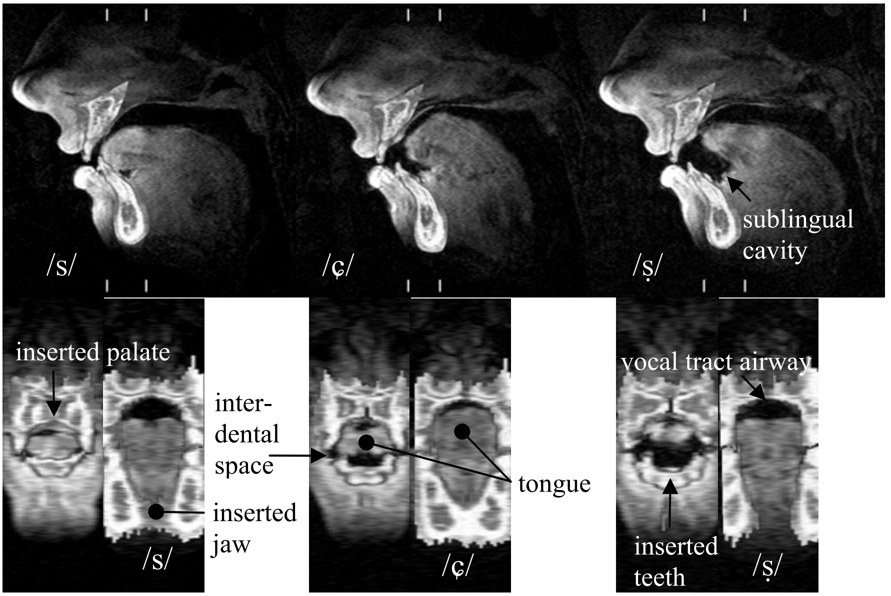


Figure 7. Mid-sagittal (top) and coronal (bottom) views of the Polish sibilants /s/ (left), /ɕ/ (center) and /ṣ/ (right), uttered by P1. Sustained fricatives produced in a virtual [ɑ] context. The position of coronal slices are indicated by tick marks in the corresponding sagittal images.

The palatalized sibilant /ç/ is characterized by a strong tongue dorsum doming in both subjects. This sagittal doming is accompanied by a doming in the coronal plane also, opposed to a grooved (P2 /s/) or flat (P1 /s/ and /ṣ/) tongue shape. As a consequence, a narrow *palatal channel* is created along the tongue constriction, with the narrowest point of constriction at its front end. The back cavity begins as far back as the post-palatal (P1) or velar (P2) region. This palatal channel sets /ç/ apart from the two other sibilants, as can be seen in the area functions.

The cross-sectional shape of the tongue constriction is elliptic rather than circular in all of the fricatives, as seen in the coronal slices (Figures 7 and 8). Its cross-sectional area varies according to the speaker and the sibilant type (ranging from about 5 to 25 mm^2), and no systematic relationship could be established with the other articulatory and morphological characteristics.

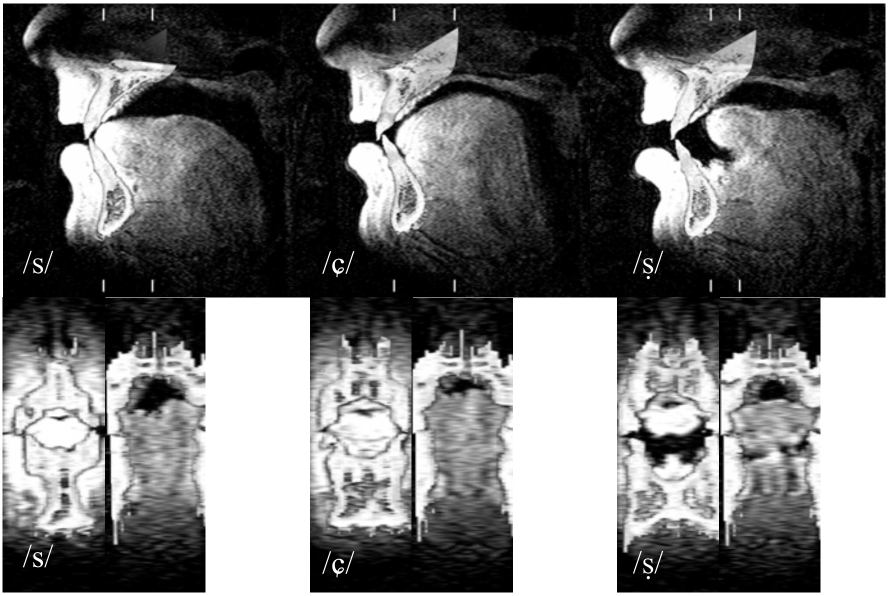


Figure 8. Mid-sagittal (top) and coronal (bottom) views of the Polish sibilants /s/ (left), /ç/ (center) and /ṣ/ (right), uttered by P2. Sustained fricatives produced in an imaginary [ɑ] context. The position of coronal slices are indicated by tick marks in the corresponding sagittal images.

In all the data, the jaw position is high, so that the vocal tract contains a second constriction involving the incisors. The area of this second constriction is comparable to that of the tongue constriction, as seen on the area function (Figure 9). Because of this tooth constriction, the vocal tract portion that was previously designated as the *front cavity* appears as two distinct parts, the oral part (hereafter ‘front *oral* cavity’) comprising the sublingual portion (if any) as well as the cavity between the tongue constriction and the tooth constriction, and the lip cavity.

Although P1’s occlusion type has not been formally examined, he presents a very advanced position of the upper incisors with respect to the lower incisors (i.e. a large overjet). As a consequence, a small cavity is observed between the constriction formed by the upper incisors and the lower lip and a second constriction formed by the tongue and the alveolar ridge for /s/.

These teeth-inserted high resolution data show that the horizontal alignment of the upper and lower incisors varies according to the sibilant for both subjects, with a greater overjet (antero-posterior distance between the upper and lower incisors) for /ç/ than for /s/ and /ʂ/. This might be interpreted as the consequence of the articulatory requirements for tongue positioning in /ç/ with respect to the palate. We cannot, however, exclude the possibility that it is also related to noise source generation, where the lower teeth, as an obstacle, should be positioned differently with respect to the airstream that originates from constrictions of various shapes and distances depending on the fricative.

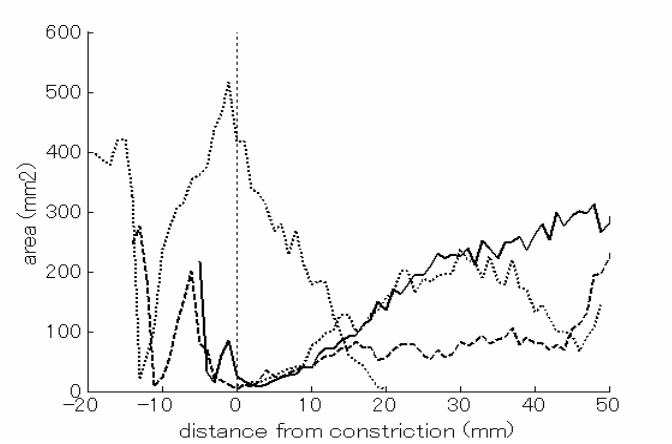


Figure 9. Area functions for P2; /s/ (plain line), /ç/ (broken line) and /ʂ/ (dotted line). The lip end is located at the left of the figure. The area functions are aligned at the tongue constriction (thin dotted line). For this speaker, only /ʂ/ involved a sublingual cavity. This cavity is represented in the continuity of the front cavity, behind the constriction, in parallel with the back cavity. A second constriction, formed by the teeth, is observed in front of the tongue constriction at -3 mm (/s/), -11 mm (/ç/) and -13 mm (/ʂ/), respectively.

5. Formant-cavity affiliation in realistic models

A close observation of the vocal tract for Polish sibilant fricatives raises a question as to how the respective vocal tract cavities are reflected in peak positions of the frication noise. Also, if the formant-cavity affiliation were established, then it could enable us to account for the acoustic goal of the articulatory targets.

Although the acoustic structure of the variants of [s] merits detailed examination, this paper will focus on the non-anterior sibilants /ç/ and /ṣ/, on account of space limitations. This contrast has been discussed in Halle and Stevens (1997), and attracted our attention because the spectral difference between the two is not clear cut, even though their articulatory targets are distinctive, as described in the previous section. Moreover, if native as well as non-native listeners are able to categorize the fricative from isolated frication noise (Lisker 2001, Nowak 2006), there should be consistent cues that distinguish /ç/ from /ṣ/.

5.1. Method

5.1.1. Acoustic recordings

The acoustic signal was recorded in a soundproof room. The subjects P1 and P2 produced the isolated words *siadam* /çɑdɑm/ and *szatan* /ṣɑtɑn/, followed by the sustained fricatives /ç/ and /ṣ/ respectively, as a part of a larger corpus not analyzed here. This part of the corpus was produced in supine position in order to simulate the MRI condition.

The time-averaged spectra of the sustained fricatives are calculated by averaging the power spectra obtained with as many 5 ms hamming windows (with 4.85 ms overlap) as necessary in order to cover the central 80 % portion of the fricatives. The averaging over many windows is based, within the source-filter theory framework, on the assumption that the noise source is random, thus potentially different in every window; however, the formant structure is assumed to be constant in a sustained utterance. So, the larger the number of samples, the less the effect of accidental source fluctuation on the averaged spectrum. The 5 ms time window, which is shorter than usual, is however considered to be long enough for our purpose with regard to its frequency definition (200 Hz).

5.1.2. Acoustic modeling

For the acoustic modeling we used the VTF_fric program (see section 3).

5.1.2.1. Original models with artificially extended back cavity

The vocal tract area functions were measured from the lips to about the velar region, as shown above in Figure 9, but the pharyngeal portion was discarded because of poor contrast in the MRI data. Therefore, the original vocal-tract area function covers the front cavity and the constriction (including the palatal channel, if any), but only a short portion of the back cavity.

At this stage, the vocal tract area functions cannot be used in simulation with either a closed or open glottis, because of the unrealistically short back cavity that results in erroneously higher resonance frequencies.

One possible solution to this problem is to eliminate the effect of sound reflections at the up-stream end of the short back cavity. An alternative approach consists in filtering out the back cavity resonances. This latter approach was chosen.

To do so, the rearmost section of the vocal tract was lengthened to about 1 m so that the resonances arising from this part of the model would be densely spaced. Specifically, the spacing ranged from 161 to 182 Hz depending on the sibilant model. The densely distributed resonances of the back cavity can therefore be easily separated via cepstral smoothing from the sparsely spaced resonances arising from the other cavities.

Various source locations were used with the original models. Since pole frequencies do not change with respect to the source position or source type, and since the source position as well as its precise composition (monopole, dipole, etc.) are not known, we calculated all-pole functions by positioning the source at the glottis when the models did not have a side branch.

Exciting the models at their rearmost point, which is not very realistic because the noise source is instead expected in the front, has the effect of emphasizing the relative amplitude of the resonances that arise from cavities located near the source. Therefore, the relative intensity of front cavity and palatal channel resonances might be inverted, while their respective frequencies are assumed to be correct.

Two other source locations, at the teeth and at the constriction, were used in the various models, including those in which the sublingual cavity was treated as a side branch. Our aim was not to examine the effect of all the exhaustive source locations, but to be able to observe the *frequency* of vocal tract resonances (invariant with respect to source location), so that they could be compared between original and truncated models (see below), and to interpret their cavity affiliation. By alternating source locations, we made certain not to miss a resonance, even if an antiresonance accidentally masked it in one of the source settings.

5.1.2.2. Truncated models

In fricatives, the tongue constriction is very narrow (in order to meet the aeroacoustic requirements for the generation of a turbulent source), which minimizes the acoustic coupling between the anterior and posterior parts of the vocal tract separated by the constriction. Therefore, the acoustic properties of the anterior and posterior vocal tract portions should not differ much when estimated separately.

In order to examine the regions of the vocal tract that affect the spectral peaks, we ran simulations with truncated vocal tract models, as in e.g. Fant (1970: 182–184). *Anterior* and *posterior* vocal tract models were constructed by cutting the original area function at the narrowest point of tongue constriction. The anterior models contain the front cavity (the truncated end was treated as closed), and the posterior models contain the palatal channel, if any. In posterior models, the rear portion was lengthened in the same manner as in the original models. In branching configurations, the front cavity models have the sublingual cavity as well as the front cavity connected to one another, and their all-pole function was calculated by locating a source at the bottom of the sublingual cavity (the rearmost section of the model).

In addition, the anterior models were further truncated. In the previous section, where the vocal tract shape was described in detail, it was observed that the upper and lower incisors are put close together so that the lip cavity is separated from the front oral cavity by a tooth constriction. Lip models were therefore created by truncating the vocal tract at the narrowest point of tooth constriction.

5.2. Results

5.2.1. /ç/

The actual noise spectra of the sustained utterances, as well as the cepstral-smoothed transfer functions of the various /ç/ models, are shown in Figure 10. Note that the natural noise spectrum is not equivalent to the calculated transfer functions. The former results from the combination of the source and transfer characteristics. As we mentioned in section 2, the source intensity is likely to be falling towards high frequencies, and this explains a difference of slope between the natural spectra and the simulation results. In addition, our original models do not feature a realistic back cavity. However, in the natural spectrum, back cavity resonances are expected to appear as regularly spaced pole-zero bumps of small magnitude superimposed on the other resonances. This should partly explain the higher complexity of the natural spectra when compared to simulation results.

As shown at the left in Figure 10 for P1, the spectral prominence of the natural sound, which ranges from 2.5 to 7.5 kHz (with two characteristic peaks), coincides with the prominence with two peaks in the original model's transfer function. It is likely that these peaks come from the anterior and posterior vocal tracts, respectively, as the anterior (broken line) and posterior (dotted line) models' functions exhibit peaks near those frequencies. It can be assumed therefore that the natural peak originates in the clustering of the front cavity and palatal channel resonances.

Likewise, the spectral prominence (2.5-7 kHz) of P2's actual sounds with two major peaks corresponds well to the prominence made of three peaks in the original model's all-pole function. Contrary to P1, the second peak, near 5 kHz, involves the anterior vocal tract, while the first and third peaks (around 2.5 and 7 kHz) involve the posterior vocal tract. The reversed order of the front cavity and palatal channel resonances between the two speakers can be explained by a longer front cavity, which involves a sublingual cavity, and a shorter palatal channel in P1 with respect to P2, as can be seen in the MR images (Figures 7 and 8).

To summarize, the spectral prominence of /ç/ is likely to involve the combination of front cavity and palatal channel resonances (although the palatal channel resonance might be the lower or the higher of the two, depending on the subject). Since the dimensions of the front oral cavity as well as those of the palatal channel can be adjusted independently, it can be suggested that the subjects deliberately cluster those resonances in order to enhance the spectral prominence (which exhibits as much as 30 dB gain with respect to the low frequency dip) to achieve the spectral saliency.

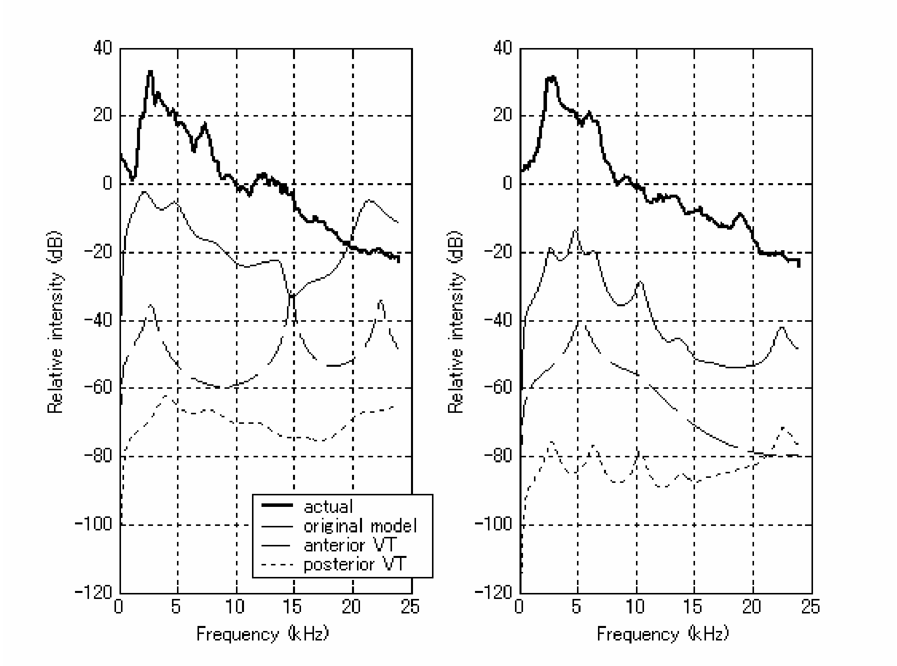


Figure 10. The actual averaged spectra (thick lines), and the simulated transfer functions (thin lines) for subject P1 (left) and P2 (right) /ɕ/. The original models (plain lines) as well as posterior vocal tract models (dashed lines) shown here were excited with a source located at the glottis, whereas the anterior vocal tract models (broken lines) were excited with a source located at their backmost section.

5.2.2. /ṣ/

The results for /ṣ/ are shown in Figure 11. The natural spectra of /ṣ/ also exhibit a spectral prominence in which several peaks can be identified. The lowest peak (around 2 kHz) is somewhat lower in frequency than that of /ç/, while the second peak is the most prominent for P2. (In word initial position, contrary to the sustained utterances, the first peak was the most prominent also for P2.)

When the original model's transfer functions are compared to the natural spectra, the frequency of the spectral prominences roughly coincide with each other; however, there is a *missing* peak in the simulation results with respect to the natural spectrum, i.e., the peak around 5 kHz that is the most prominent in P2's natural spectrum. Otherwise, the first and second peaks of the original model's function can be put in correspondence with the first and third peaks of the natural spectrum, observed around 2 and 7.5 kHz, respectively.

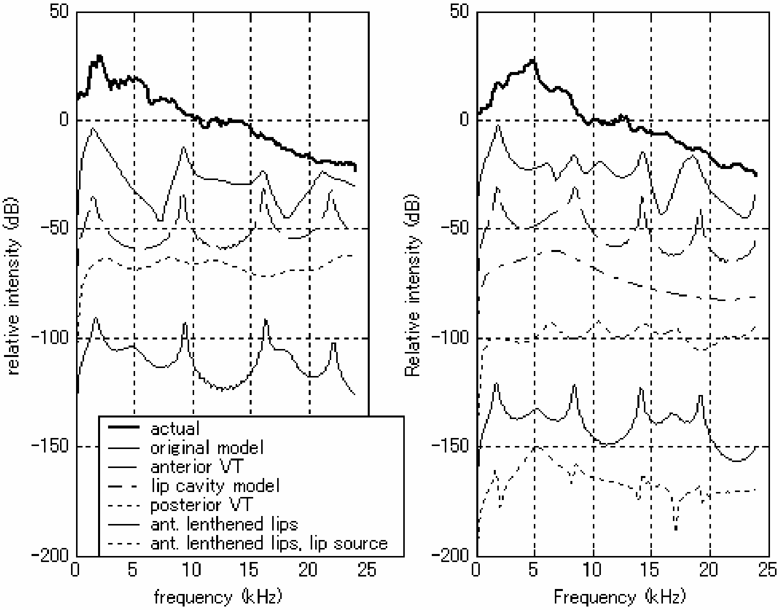


Figure 11. The actual averaged spectra of /ʂ/ (thick lines), and the simulated transfer functions (thin lines) for subject P1 (left panel) and P2 (right panel). This figure shows the simulation results for the original models (upper plain lines) excited by a glottal source (P1) and a constriction source (P2), respectively. The results for anterior vocal tract models (broken lines) and lengthened lips anterior vocal tract models (bottom plain lines), as well as P2's lip cavity model (dotted-dashed line, right panel) are obtained with a source located at the backmost section. The simulation results for P2's lengthened anterior model excited with a source located within the lips are also shown (lower dotted line). The posterior vocal tract models (upper dotted lines) were excited at the glottis.

In P1, the original model's peaks coincide best with the anterior vocal tract model's peaks. In P2, the posterior vocal tract contributes to some extent to the spectral structure of the original function. In general, in comparison to

/ɕ/, the involvement of the posterior vocal tract seems to be rather small in /ʂ/.

Looking more closely at P2’s anterior vocal tract function, a broad resonance can be noticed, which forms a shoulder (around 7 kHz) at the left of the second peak. Because of its wide bandwidth, a lip resonance is suspected (a radiation load is applied at the lip end). In order to verify this hypothesis, the acoustic properties of the lip model were examined. The dash-dotted line corresponds to the lip model’s result. A single peak is observed near the frequency of the broad banded resonance observed in the anterior vocal tract transfer function, thus corroborating our hypothesis. Therefore, we in fact have the right *number* of resonances corresponding to the three formants (around 2, 5 and 7 kHz, respectively) that constitute the natural utterance’s spectral prominence.

We consider that the lip issue needs to be examined in greater detail, as a strong lip protrusion characterizes P2’s MRI data for /ʂ/, and we would like to be able to interpret its acoustic function. In fact, the lip gesture accompanies English and French /ʃ/, which are not far different from Polish /ʂ/. It has been proposed (e.g. Stevens et al. 2004) that the acoustic function of lip protrusion, at least in English /ʃ/, is to lower the front cavity resonance, thereby increasing the acoustic contrast with /s/. However, in our articulatory data, the lip cavity appears well separated structurally from the front oral cavity. Furthermore, our simulation results indicate that the lip cavity produces a formant of its own, distinct from those arising from the front oral cavity. These observations lead us to propose another interpretation. The function of lip protrusion would be, instead, to strengthen the spectral prominence by clustering the lowest lip resonance with a front oral cavity resonance, in order to enhance the spectral energy in this frequency region.

Is this plausible? We have mentioned that the area of some lengths of the lips could not be measured because of their lateral curvature. In P2’s area function, the measured lip cavity is 6 mm long. On his mid-sagittal profile, however, it clearly exceeds 1 cm. The frequency of the lip cavity resonances are therefore largely overestimated in our simulation. If the anterior vocal tract model is altered, with the lip cavity lengthened by 5 mm, the lip resonances move down, as shown in the anterior vocal tract all-pole function (lower plain line). When the source is put within the lip cavity, the relative intensity of the first lip resonance becomes much closer to the actual spectrum, as shown in the lowest dotted curve. Note that it is not unrealistic to assume a lip noise source, created by the air jet exiting from the tooth constriction and directed against the lower lip.

A similar correction can be applied to P1's /ʂ/. For this speaker, no lip cavity could be measured, and thus the models contain no lip cavity. This is clearly far from the reality, even though no lip protrusion is observed in /ʂ/ for this speaker. If an artificial lip cavity of 1 cm of length and 4 cm^2 of area is appended to P1's front cavity model, an additional peak appears around 5 kHz, which nicely matches the second formant of the actual spectrum, without significantly affecting the other peaks.

To summarize, /ʂ/'s spectral prominence involves mainly the front cavity resonances (corresponding to the actual F1 and F3). In addition, it is very likely that the lip cavity is responsible for the second fricative formant of /ʂ/.

5.3. Discussion

From these results, the following observations can be made:

(1) The first fricative formant for /ɕ/ tends to be higher in frequency than that of /ʂ/. The size of the front oral cavity (larger for the latter) is largely responsible for this difference.

(2) In both /ɕ/ and /ʂ/, the spectral prominence is made of a set of resonances arising from distinct cavities. In our results, the first front oral cavity resonance is clustered with the first palatal channel resonance in /ɕ/, whereas the lip resonance is clustered with the front oral cavity resonances in /ʂ/.

If speakers aim to enhance the spectral saliency of these sounds, they will tend to bring the resonances arising from different cavities close to one another, by adjusting the lengths of these cavities. If the clustering patterns described in (2) could be generalized, then we would predict that:

(3) In /ɕ/, the front oral cavity resonance can be brought very close to the palatal channel resonance (this is still low enough so that this sound is distinct from /s/). Therefore, /ɕ/'s prominence will tend to be compact.

(4) In /ʂ/, if the goal is to make a very low fricative formant, the front oral cavity is the only vocal tract region that leads to this goal. However, due to the physical limitations of lip protrusion, the lip resonance cannot be grouped with the first front cavity resonance as close as the palatal channel resonance does in /ɕ/. Moreover, the lip

resonance possesses a wider bandwidth than the palatal channel resonance. Therefore, /ʂ/'s spectral prominence will tend to be more diffuse.

These predictions seem to be verified by our acoustic data. It is however difficult to make a generalization until the acoustic properties of frication noise are quantitatively examined by means of ensemble-averaged or long-term spectra from a larger set of Polish speakers.

Finally, we would like to add a few words about the coupling of the back cavity, which was assumed to be partly responsible for the spectral difference between the Polish sibilants /ɕ/ and /ʂ/ in Halle and Stevens (1997:190). Halle and Stevens (1997: 189) argued that the long and thin palatal channel prevents the back cavity resonances from coupling effectively in /ɕ/, in contrast to /ʂ/. A back cavity resonance would have been responsible for /ʂ/'s characteristic peak in the vowel's F2 region. However, it cannot easily be supposed that a pole-zero pair of the back cavity would be able to significantly shape the noise spectrum of /ʂ/, given that the constriction is very narrow, measuring about 0.1 to 0.15 cm^2 in our MRI data (see Badin 1989: 52-53). In these conditions, not only are closely bound back cavity poles and zeros expected to considerably attenuate each other, but in addition, the back cavity would be very weakly excited by the noise source. Indeed, the long-term spectral structures of sustained utterances of /ʂ/ do not exhibit such kind of regular formant-antiformant pairing big enough to shapen the overall spectral shape.

Instead, in the light of our results, it can be suggested that the larger number of peaks observed in the lower frequencies for /ʂ/ are mainly due to the longer front oral cavity (including the sublingual cavity), which leads to lower resonances, combined with the lip resonances.

6. Conclusion

This study aimed at interpreting analytically the spectral structure of sibilant fricatives. First, the acoustic properties of simple sibilant-like configurations were examined through acoustic simulations, by varying the straight/branching configuration and the source location. The spectral events specific to sibilant fricatives were then summarized.

Second, the articulation of Polish sibilants /s/, /ç/ and /ʂ/ was examined in detail using high-resolution MRI data from two subjects. The data showed the existence of a second narrow constriction located at the incisors, in addition to the constriction involving the tongue. The effect of this second constriction on a possible lip noise source remains to be investigated.

Finally, the formant-cavity affiliation for these subjects' non-anterior sibilants /ɕ/ and /ṣ/ was investigated, by simulating the acoustic properties of original and truncated models constructed from the MRI data. The main outcome of this experiment was to show that in both /ɕ/ and /ṣ/, the spectral prominence is likely to be made up of a bunch of resonances arising from different vocal tract cavities: front oral cavity and palatal channel for /ɕ/; and front oral cavity and lip cavity for /ṣ/. The spectral consequence of the absence/presence of a palatalized articulation is a lower (/ṣ/) or higher (/ɕ/) frequency of the spectral prominence, as well as its relative diffuseness (/ṣ/) or compactness (/ɕ/). Contrary to Halle and Stevens's (1997) proposal, our articulatory data as well as simulation results do not support a significant coupling of back cavity resonances in /ṣ/. Instead, the spectral difference between /ɕ/ and /ṣ/ is better explained by the difference in size of the front oral cavity, combined with other resonances (palatal channel and lip cavity resonances, respectively).

Acknowledgement

The authors acknowledge the financial support of the Telecommunications Advancement Organization of Japan, as well as the Future and Emerging Technologies (FET) programme within the Sixth Framework Programme for Research of the European Commission, under FET-Open contract no. 021324. We are grateful to the Polish subjects, who provided sustained efforts in order to permit the acquisition of high quality data, and also to Yasuhiro Shimada, Ichiro Fujimoto, and Yuko Shakudo at the ATR Brain Activity Imaging Center for their technical expertise.

Notes

1. Shadle's (1991) work is based on the articulatory data of a Russian subject (Fant 1960), but aims to be valid for [s] and [ʃ]'s in general, in other languages as well, e.g. English. Therefore, we use brackets '[]', thus referring to a phonetic entity typical of IPA categories, independently to its status within the phonological system of individual languages. We use slashes '//' to refer to a phoneme in a particular language.

2. We follow the notation of Halle and Stevens (1997) to refer to the Polish fricative traditionnaly called ‘palatoalveolar’, and which is also commonly transcribed by /ʂ/. This /ṣ/ notation, referring to an apical post-alveolar fricative (see Ladefoged and Maddieson 1996: 150–164), marks its difference from sub-apical or sub-laminal retroflex fricatives like, e.g., in Dravidian languages.

References

Badin, Pierre
1989 Acoustics of voiceless fricatives: production theory and data. *STL-QPSR* 3/1989: 33–55.

Curle, N.
1955 The influence of solid boundary upon aerodynamic sound. *Proc. Roy. Soc. Lond.* A 231: 505–514.

Fant, Gunnar
1970 Reprint. *Acoustic Theory of Speech Production.* Second printing. The Hague and Paris: Mouton. Original edition: The Hague: Mouton, 1960.

Flanagan, James L.
1972 Second edition. *Speech Analysis, Synthesis and Perception*. Berlin, Heidelberg and New York: Springer-Verlag. First edition: Berlin and Heidelberg, Springer-Verlag, 1965.

Halle, Morris and Kenneth N. Stevens
1997 Reprint. The postalveolar fricatives of Polish. In S. Kiritani, H. Hirose and H. Fujisaki (eds.) *Speech Production and Language: In Honor of Osamu Fujimura*. Berlin: Mouton de Gruyter: 177–193. Originally in *Speech Communication Group Working Papers 7*, Research Laboratory of Electronics, MIT, 1991: 77–94.

Harris, Katherine S.
1958 Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1: 1–17.

Jackson, Michel T.-T., Carol Espy-Wilson and Susanne E. Boyce
2001 Verifying a vocal tract model with a closed side-branch. *Journal of the Acoustical Society of America* 109 (6): 2983–2987.

Jassem, Wiktor
1995 The acoustic parameters of Polish voiceless fricatives: an analysis of variance. *Phonetica* 52: 251–258.

Lee, Sook-Hyang, Mary E. Beckman and Michel Jackson
1994 Jaw target for strident fricatives. *Proceedings of the ICSLP*: 37–40. Yokohama, Japan.

Lisker, Leigh
2001 Hearing the Polish sibilants [s ś š]: phonetic and auditory judgements. *Travaux du cercle de linguistique de Copenhague* vol. XXXI: 226–238.

Maddieson, Ian and Kristin Precoda
1989 Updating UPSID. *UCLA Working Papers in Phonetics* 74: 104–111.

Maeda, Shinji
1982 A digital simulation method of the vocal-tract system. *Speech Commuication* 1: 199–229.

Mooshammer, Christine, Philip Hoole and Anja Geumann
2007 Jaw and order. *Language and Speech* 50(2): 145–176.

Narayanan, Shrikanth and Abeer Alwan
2000 Noise source models for fricative consonants. *IEEE Transactions on Speech and Audio Processing* 8(2): 328–344.

Narayanan, Shrikanth, Abeer Alwan, and Katherine Haker
1995 An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America* 98 (3): 1325–1347.

Nowak, Paveł M.
2006 The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics* 34: 139–152.

Pastel, Leah M. P.
1987 Turbulent noise sources in vocal tract models. M.S. dissertation. Massachusetts Institute of Technology.

Shadle, Christine H.
1985 The acoustics of fricative consonants. PhD Dissertation, MIT.

Shadle, Christine H.
1991 The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics* 19: 409–424.

Stevens, Kenneth N.
1998 *Acoustic Phonetics.* Cambridge, MA and London: MIT press.

Stevens, Kenneth N., Zhiquiang Li, Chao-Yang Lee and Samuel Keyser
2004 A note on Mandarin fricatives and enhancement. In G. Fant, H. Fujisaki, J. Cao and Y. Xu (eds.), *From Traditional Phonology to Modern Speech Processing*. Beijing: Foreign Language Teaching and Research Press, 393–403.

Takano, Sayoko, Kiyoshi Honda, Shinobu Masaki, Yasuhiro Shimada and Ichiro Fujimoto
2003 Translation and rotation of the cricothyroid joint revealed by phonation-synchronized high-resolution MRI. *Proceedings of Eurospeech* 2003: 2397–2400, Geneva.

Takemoto, Hironori, Tatsuya Kitamura, Hironori Nishimoto and Kiyoshi Honda
2004 A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions. *Acoustical Science and Technology* 25 (6): 468–474.

Toda, Martine
2009 Etude articulatoire et acoustique des fricatives sibilantes. Doctoral dissertation, Université Paris III, France.

Wierzchowska, B.
1965 *Wymowa Polska*. [Polish pronunciation] Warszawa: Panstwowe Zaklady Wydawnictwo Szkolnych.

Żygis, Marzena and Silke Hamann
2003 Perceptual and acoustic cues of Polish coronal fricatives. *Proceedings of the 15th International Congress on Phonetic Sciences*: 395–398, Barcelona.

Subject index