ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT•

Journal of Phonetics 33 (2005) 363–381

Journal of
Phonetics

www.elsevier.com/locate/phonetics

Patterns of tongue movement

Khalil Iskarous*

Haskins Laboratories, 300 George Street, New Haven, CT 06511, USA

Received 18 March 2002; received in revised form 27 August 2004; accepted 21 September 2004

Abstract

This study investigates tongue dynamics during speech production. The focus is on the deformation of the whole midsagittal edge of the tongue in transitions between lingual segments, rather than the movement of individual points of the tongue, which is the traditional focus of tongue dynamics research. Based on the analysis of 600 lingual transitions from an X-ray database of speech, it is shown that there are only two basic patterns of tongue movement, the pivot and the arch, which are independent of the starting and ending segments of a transition. It is then argued that the acoustic effect of these patterns of tongue deformation is to make the acoustic signal as articulatorily-transparent as possible.

---

1. Introduction

The tongue is a highly flexible organ that is able to deform in a complex fashion, allowing it to form constrictions that vary widely in location and degree. This versatility allows the tongue to accomplish the vocal tract configurations required for most of the phonetic contrasts used by the world's languages—all vowels and most consonants have a lingual specification. Also, different parts of the tongue are able to move in parallel, allowing for simultaneous execution of several segments (Öhman, 1966; Perkell, 1969; Fowler, 1981). The tongue's capacity for parallel transmission of several segments underlies the high speed of speech. Since the tongue is so crucial in the achievement of linguistic contrasts, it is important to find basic principles of organization that govern its movement during speech production. Moreover, principles of tongue kinematics

---

*Tel.: +1 203 865 6163/234; fax: +1 203 865 8963.
E-mail address: iskarous@haskins.yale.edu (K. Iskarous).

0095-4470/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2004.09.001

during speech could provide insight into the mechanics of lingual coarticulation and phonological phenomena, such as assimilation, harmony and coalescence. Principles of movement would be evident in how the tongue moves in a large variety of different segmental transitions, e.g. [ke], [da], [su], [ai], and [tg]. If the same kinematic pattern occurs in all of these transitions and others, it would indicate a generalization about tongue movement in the speech task independent of the initial and final shapes of the tongue. We therefore look for recurring movement patterns since they are potential indicators of principles of tongue kinematics.

There has been a long tradition of study of tongue dynamics with the aim of uncovering the basic principles of its movement. Some imaging modalities such as X-ray cineflurography, MRI, and ultrasound produce images of complete sections of the tongue, allowing researchers to view the complex shapes, but these images are difficult to quantify and compare, because the shapes are not classifiable on a single quantitative scale. Progress has been made by tracking the movement of a few points of the tongue, and using the quantitative scale of point displacement to describe the movement of the tongue (Öhman, 1966; Houde, 1968; Perkell, 1969; Kent & Moll, 1972; Munhall, Ostry, & Flanagan, 1991; Mooshammer, Hoole, & Kühnert, 1995; Munhall & Jones, 1995; Gracco & Löfqvist, 1999). In one of the earliest studies, Houde (1968) found that the paths of individual fleshpoints are curved. This was surprising, because if the path is influenced only by the initial and final targets, one would expect a straight line path. Loops of tongue points have been reported to occur in a large variety of transitions and their presence seems to be independent of the specific segments. Much of the literature on tongue dynamics since then has attempted to provide an explanation for this deviation of the paths from straight segments, and the search for principles of tongue dynamics has become entangled with the search for an explanation for the curved paths.

Houde's (1968) data were VCV sequences, so his explanation for the curvature of the paths centered around the reaction of fleshpoints of the tongue to pressure forces behind the closure. This type of explanation was also suggested by Kent and Moll (1972). Another aerodynamic explanation was provided by Coker (1976), who suggested that the curvature is due to active expansion of the supraglottal cavity to prolong voicing during the intervocalic consonant. However, based on an investigation of a more varied set of intervocalic consonants, Mooshammer et al. (1995) showed that the velar nasal [ŋ] also shows curved paths, and therefore pressure behind the closure cannot be the cause of path curvature. Also curved paths occur when the intervocalic consonant is voiceless, so an explanation based on maintenance of voicing is untenable (Mooshammer et al., 1995; Munhall et al., 1991). Another way of approaching the tongue loops problem is by attempting to find a coordinate system in which the paths assume a simpler shape (Munhall et al., 1991; Munhall & Jones, 1995). If such a coordinate system can be found, it may be the one in which speech is planned, and finding such a system would be a major step in understanding tongue dynamics. Munhall et al. (1991) placed the tongue in a coordinate system provided by the jaw, and mathematically removed the influence of the jaw from the paths of tongue fleshpoints. The result, however, was that the paths were still curved. Therefore the curvature is due to tongue motion itself. More recently, Löfqvist and Gracco (2002) have suggested that the curvature of fleshpoint trajectories is due to an optimization constraint that requires movements to be smooth. Inspection of the data in Löfqvist and Gracco (2002), however, shows that there is a great deal of variation between subjects in the shape of the loops, and that some loops are indeed nonsmooth. Moreover, the loop shapes of the tracked tongue points are different. So even if a smoothness constraint explains why there is trajectory curvature, there needs to be an explanation of why the trajectories of points of the tongue assume different shapes.

From a motor control perspective, it seems strange that the controlled variables in tongue movement would be the displacement of the individual points of the tongue, since there are simply too many points to control. That is, there are too many degrees of freedom in the description of the tongue when the control variable is the position of individual points on the tongue surface. The linguistic task of the tongue would seem to require that the trajectories of the points of the tongue would be *coordinated* to achieve the simultaneous and time-varying constrictions of speech production—a level of analysis requiring fewer degrees of freedom. To find an explanation for trajectory curvature and the principles of tongue movement organization, we may therefore have to consider a higher level of tongue movement analysis. The theories of Articulatory Phonology and Task Dynamics (Browman & Goldstein, 1990; Saltzman & Munhall, 1989) distinguish between two levels of speech production dynamics, and therefore offer an appropriate framework for investigating different levels of analysis for tongue movement. In these theories, several individual articulators (articulator level) pursue their individual trajectories to achieve a single constriction requirement enforced at a higher level (gestural level). Covariation of the articulators at the articulator level occurs to preserve an invariance at the gestural level. Several experiments have shown that when the positioning of individual articulators is perturbed, other articulators alter their positions so that at the gestural level, the location and size of a constriction would remain invariant despite the perturbation (Kelso, Saltzman, & Tuller, 1986). Loops provide a high-dimensional description of tongue motion, and can therefore be identified with the articulator level of tongue kinematics. To understand their shapes we need to find the appropriate gestural level, which would provide more coarse-grained control variables.

Factor-analytic studies of tongue motion (Harshman, Ladefoged, & Goldstein, 1977; Maeda, 1990) analyze the correlation in movement between different points of the tongue along the vocal tract, and therefore provide a good basis for modeling the gestural level of tongue movement, since they achieve a reduction in dimensionality of description. Data for the analysis is usually provided by a set of tongue shapes from different segments. Each shape is divided into a number of segments and each segment acts as a variable for statistical analysis. Each tongue shape is seen to score highly on some variables (where the constriction lies), and low on others. And the variation in the data can be characterized by a covariance matrix, each entry of which indicates the covariation of these two segments. The idea in factor analysis is to find the variables that account for the greatest amount of covariation. This is done by extracting the eigenvectors (principal components) and eigenvalues of the covariance matrix. Each principal component indicates a particular linear combination of the variables which accounts for an amount of the overall variation proportional to the square of the eigenvalue, and each eigenvector accounts for covariation amongst the variables that the previous eigenvectors have not accounted for, therefore the eigenvectors are orthogonal or uncorrelated. If the lowest few eigenvectors account for most of the covariance in the data, that means that there are a few sections of the vocal tact which are the most responsible for contrasts between different tongue shapes. That is why Harshman et al. (1977) looked at factor analysis as a source of features emergent from speech production data. Any particular vocal tract shape can then be synthesized by weighing the different factors and summing the contributions. Even though the factors are usually devised from static data of individual vowels and consonants, the shape of the tongue during the transition between any two segments can be synthesized by interpolating between the weights for the segmental shapes in the factor-weight space. Different interpolation rules would lead to different patterns of tongue movement. That is, there is no reason internal to the factor-analytic approach to prefer one sort of transition to another. The factor-analytic approach identifies a higher level of analysis, but is compatible with different predictions about the *kinematics* at that level. Loop analysis picks a low-level description of tongue motion but pays a great deal of attention to the kinematics at that level. Factor analysis, on the other hand, picks a higher level of analysis, but predicts little about its kinematics.

In this paper, an empirical approach to tongue kinematics is taken at a level of analysis analogous to the factor level. This process has been initiated in Maureen Stone's work (Stone, 1991, 1992; Unser & Stone, 1992), where she investigates three-dimensional tongue dynamics using ultrasound imaging. Unser and Stone (1992) show several examples of tongue motion during transitions between lingual segments. One transition is between [ʃ] and [a], and the authors note that the tongue seems to rotate around a pivot point through the transition. Such a rotation would involve the collective action of all points of the tongue to accomplish a very simple action and may therefore indicate an important place to look for general principles of tongue motion. But it is unclear how widely these patterns are used, because they are mentioned only with reference to a few lingual transitions. It is also not clear in what frame of reference the rotations occur. The goal of the current work is to look for generalizations about how the tongue deforms in a wide variety of transitions and to investigate the frame of reference of rotation, as well as the significance of the rotational pattern.

2. Methods

This work is based on the analysis of 600 articulatory transitions between speech segments with lingual targets. These transitions were extracted from a database of Lateral X-ray films of speech production. The films were made in 1974 at the Radiology Department in L'Hôtel-Dieu Hospital in Quebec City under the direction of Dr. Claude Rochette of the University of Laval. A previous set of films that was made using the same procedures was analyzed by Rochette (1973). The films, which run a total length of about 55 min (50 frame/s), have been placed on a videodisc by researchers from the ATR Human Information Processing Laboratories (Kyoto, Japan) and Queens University (Ontario, Canada) (Munhall, Vatikiotis-Bateson, & Tohkura, 1994a, b; Tiede & Vatikiotis-Bateson, 1994). For the purpose of this work, one hundred and fifty transitions were analyzed from each of four subjects, for a total of 600 transitions. The first two subjects are native speakers of Canadian French, and both are male. The other two subjects are native speakers of Canadian English, the first of whom is female, and the second male. The major factor in choosing the subjects was the visibility of the tongue and the ease of distinguishing midsagittal from lateral edges. A total of 3695 frames was analyzed.

An X-Windows program XRS (X-ray speech) was developed by the author for the interactive extraction of the edge of the tongue edge from X-ray images. An example frame is given in Fig. 1, where an X-ray image is seen together with a curve fitting the tongue. To extract the edge of the tongue, palate and posterior pharyngeal wall, the user manipulates control points that influence the shape of a curve until it fits the edge of the structure. The control points can be seen in the

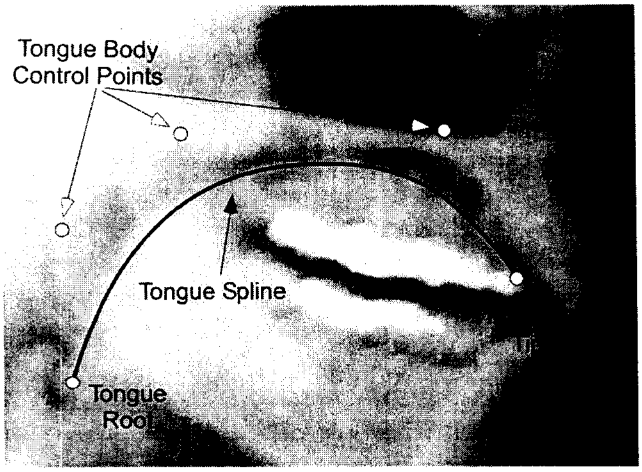

Fig. 1. X-ray frame with tongue spline superimposed. Control points for the spline are indicated as circles. Two control points control the positions of the endpoints and three control points determine the shape of the spline.

Table 1
Number of transitions in each category

	Alveolar	Palatal	Velar	Uvular	Pharyngeal
Alveolar		38	24	35	44
Palatal	37	6	13	38	30
Velar	34	33	14	36	35
Uvular	41	29	10	3	13
Pharyngeal	28	31	21	8	

figure. The curve used is a cubic B-spline curve (Farin, 1997), which means that the x and y coordinates of the curve are a piecewise-cubic function of a spatial parameter t that varies from 0 to 1 as the curve proceeds from beginning to end. The coefficients of the function are the control point locations, so as the user interactively manipulates the positions of the control points, the x and y functions change and the curve shape changes. Cubic functions were chosen since lower degree curves are not flexible enough to fit areas of high curvature in the tongue, and those of higher degree are too peaked for the contour of the tongue and other structures. Further details can be found in Iskarous (2001). The midsagittal distance function, the distance from tongue to palate and posterior pharyngeal wall at each section of the vocal tract, was then estimated based on a polar-rectangular grid fitted to each subject.

The database of 600 transitions analyzed include transitions of the following form: CV, VC, VV, and CC. *V* varied over the inventory of Canadian English and French vowels, and *C* varied over all consonants in the two languages that have a lingual target, including alveolar, palatal, velar, and uvular. Table 1 presents the number of each type of transition analyzed. The classification is in terms of Wood's (1979) segment classification system (Wood, 1979, 1997) in which both vowels and consonants are classified in terms of the place of constriction. This system was chosen, because it is a much more accurate description of tongue shapes for vowels than the traditional method of classifying vowels from the highest point of the tongue in a Height × Backness space. The vowels are classified into palatal [i,ɪ,e,ɛ], velar [u,ʊ], uvular [o, ɔ], and pharyngeal [a,æ,ɑ], and consonantal classification is the same as the IPA classification. The only exception found in the present work is the vowel [u], which is sometimes more uvular than velar. Each vowel was labeled according to the actual configuration of the tongue. Also, the transitions were collected randomly from within and across words, as well as from different prosodic contexts. This is an artifact of using an already existing database that was collected for other purposes, rather than balanced experimental data collected to test a specific hypothesis. There is no claim however that the patterns to be discussed later are insensitive to prosodic factors.

The parsing of a sentence into a set of transitions was done with the aid of the AVID program, a Macintosh program for film analysis, and XRS developed by the author. The beginning and end of a transition were synchronized with sign changes in velocity or acceleration of the spline edge, e.g., a change from upward to downward motion or changes from frontward to backward motion. So a transition was determined as the period between configurations of maximal constriction of two consecutive segments. Acceleration sign change was used when the tongue moved in one direction, but slowed down then speed up again.

3. Results

One method of studying tongue dynamics is to investigate the sequence of configurations of the midsagittal edge of the tongue from the initial to the final configuration of a transition. Fig. 2 shows such a sequence of B-spline approximations of the tongue edge for the sequence [ai] produced by Subject 1. Some qualitative observations can be gleaned from the sequence of tongue configurations in Fig. 2. The pharyngeal portion of the tongue moves gradually from a very retracted position during the [a] to an advanced position during [i]. At the same time, the most curved portion of the tongue during the [a], which is at the uvular region in Frame 1, becomes somewhat flatter at [i], as the curvature has become concentrated at the palatal region in Frame 10, through the forward and upward movement of the blade and body of the tongue. That is, the

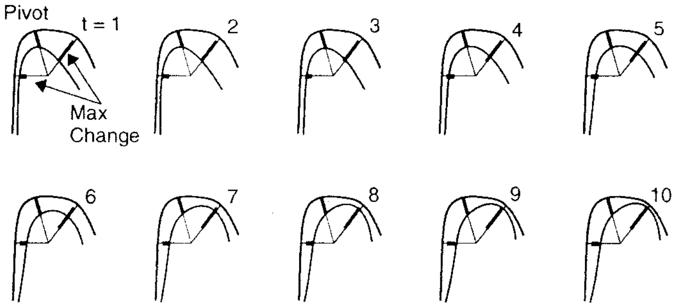

Fig. 2. Tongue splines for 10 frames of the transition [ai] for Subject 1. Three locations of the vocal tract are highlighted: palatal, uvular, and pharyngeal.

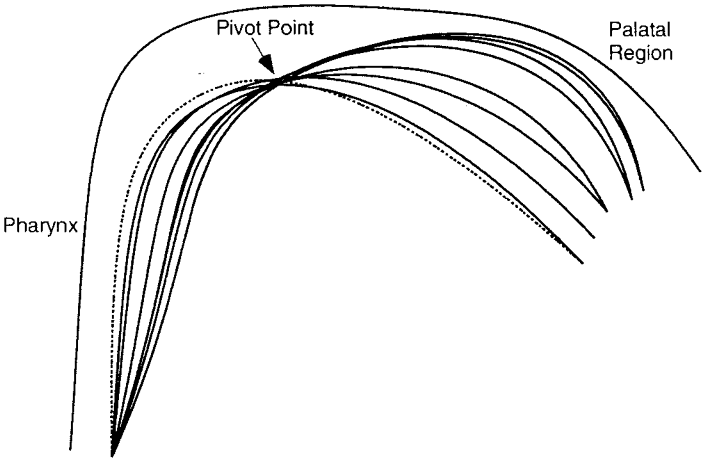

Fig. 3. Superimposed splines for [ai]. First frame is dotted.

shape of the midsagittal edge has drastically changed during the transition, through the orchestrated movement of all points of the tongue (except at the periphery).

Instead of looking at individual points of the tongue, and asking how they move, we need to focus on individual sections of the vocal tract, and ask how the midsagittal distance function changes at these sections.[1] Such an investigation can be carried out by superimposing the tongue edge approximation from all the frames in one transition. This allows us to see the evolution of the tongue during an entire transition in a single static figure. Fig. 3 shows the superimposed B-splines from the frames for the same [ai] shown separately in Fig. 2.

As can be seen in Fig. 3, in the palatal and pharyngeal regions, a great deal of midsagittal distance function change occurs. At sections intermediate between pharyngeal and palatal, there is less change. Indeed, for sections between the palatal and pharyngeal regions, less and less change occurs, the farther that section is from the palatal and pharyngeal regions. A point is then reached where there is virtually no change in midsagittal distance function. This point will be called the pivot point, and the general pattern of change seen in the figure will be called the pivot pattern. More examples will be given later in this section. The section of the tongue where there is little to no change is the middle gridline highlighted in Fig. 2, and the two sections where there is maximal change in midsagittal distance function are highlighted by the outside gridlines in that figure.

At first glance, it may seem that the pivot point coincides with a point of the tongue that does not change its position throughout the transition. But the movement of the tongue seen in Fig. 2 does not support this idea. The location at which the pivot point occurs (for this particular transition) is in the uvular region, where movement can be seen to occur. Indeed, the whole back of the tongue deforms from an overall back position to an overall forward tongue position. This cannot happen if some point of the tongue is stationary in the uvular region of the vocal tract

---

[1] Since the midsagittal distance function is a major component in determining the area function (Heinz & Stevens, 1964; Yehia, 1997), we expect generalizations about the effect of tongue movements on midsagittal distance function to have implications for area function change as well.

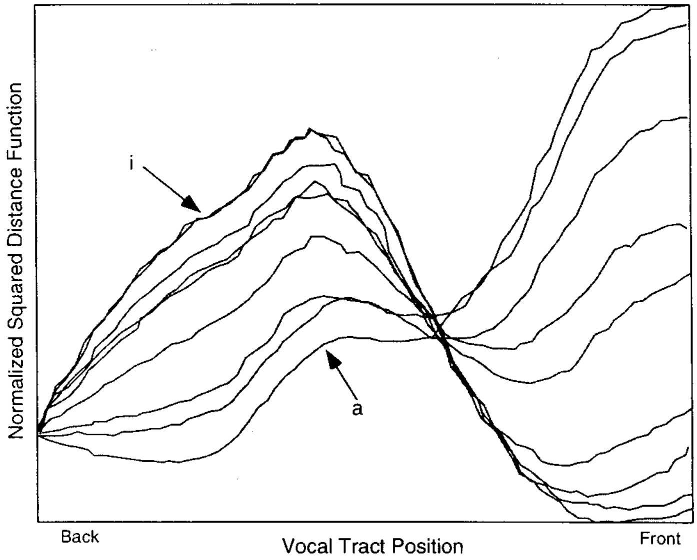

Fig. 4. Square of the midsagittal distance function for the [ai] frames.

from the beginning to the end of the transition. Indeed the back of the tongue can be seen to pass right through the uvular region. The pivot point is a stationary point in the midsagittal distance function, not a stationary point of the tongue. Fig. 4 shows the square of the midsagittal distance function (which is proportional to the area function)[2] for each frame of the same token discussed above. As can be seen, there is a point in the vocal tract, indicated by a vertical line, where the midsagittal distance function hardly changes. Fig. 5 shows the maximal change in midsagittal distance function at each gridline. This is calculated by measuring the distance between every two splines at each gridline, and obtaining the maximum. The vertical line indicates the pivot point, where almost no change has occurred.

How can the midsagittal distance function be stationary at a particular point, when the tongue itself is not stationary? Points of the tongue seem to move nearly orthogonally to the superior and posterior structures (palate and pharyngeal wall) at two locations in the vocal tract (palatal and pharyngeal in this case), but they seem to move nearly parallel to the superior and posterior structures at another location (uvular in this case). The result is that the midsagittal distance function changes maximally at certain locations, and hardly changes at all at others. One can visually appreciate the pattern by looking at how different points of the tongue pass through the uvular region in Fig. 2, while maintaining a fixed distance to the superior and posterior structures. This is in stark contrast to other sections of the vocal tract where the distance between the tongue and superior/posterior structures monotonically increases or decreases.

---

[2] It is known that the area function is a complicated function of the square of the midsagittal distance function (Heinz & Stevens, 1964). However, the alpha and beta parameters needed to calculate the area function from the midsagittal distance function are subject dependent and are unknown for the current subjects. The square of the midsagittal distance function is therefore used only as a rough estimate for the area function.

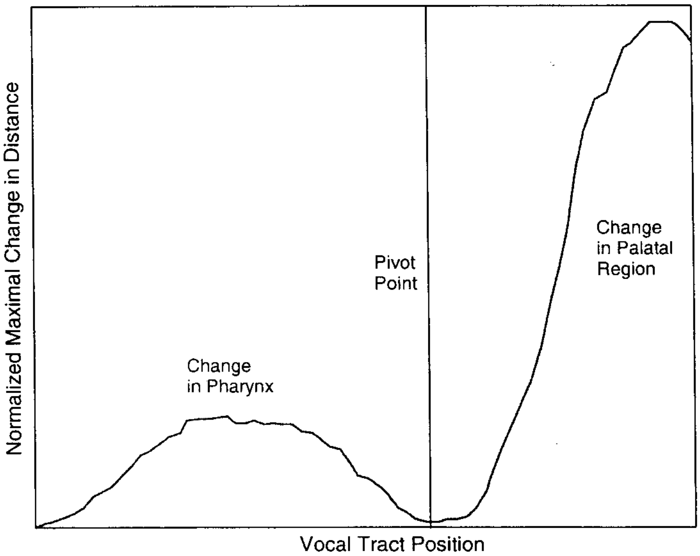

Fig. 5. Maximal change in midsagittal distance function. Two areas of maximal change surround an area with almost no change, marked with a vertical line.

A pivot transition can be characterized by three vocal tract sections, the pivot point where the midsagittal distance function changes minimally and two other sections, where the midsagittal functions changes maximally. For the token already discussed, the three sections can be seen in Fig. 2. The movement of each point of the tongue may be curved in some complex way, but the overall effect of the pivot pattern of tongue movement is to couple the many degrees of freedom of the points of the tongue so that two sections of the vocal tract see maximal change and one section sees minimal to no change. From the perspective of points of the tongue, the transition dynamic is a very complex process, but from the viewpoint of the midsagittal distance function and area function—the task space of speech production (Saltzman & Munhall, 1989)—the dynamic is simple. The pivot pattern is not used only in this one transition, it is used in the transitions between a wide variety of segments. Examples are given in Fig. 6. The subject number is given after the transition label to the right of each example.

It is clear from Fig. 6 that the pivot point does not always occur in the same place along the vocal tract. It shifts around depending on the places of maximal constriction in the transition. This also supports the claim that the pivot point is not a mechanical pivot, i.e., a point of the tongue that does not move. No theory of tongue muscle control would claim that in each transition, the muscles could be controlled in such a way that one point of the tongue (different for each transition) remains stationary while those on each side of it would move.

On many occasions, a transition involves only a localized change in mid-sagittal distance function. That is, the midsagittal distance function changes at one region, but the magnitude of change becomes less and less outside the region, until there are portions in the vocal tract where no change in mid-sagittal distance function occur. This dynamic pattern will be referred to as the arch pattern, because the tongue seems to simply arch or advance itself in a single region. In some

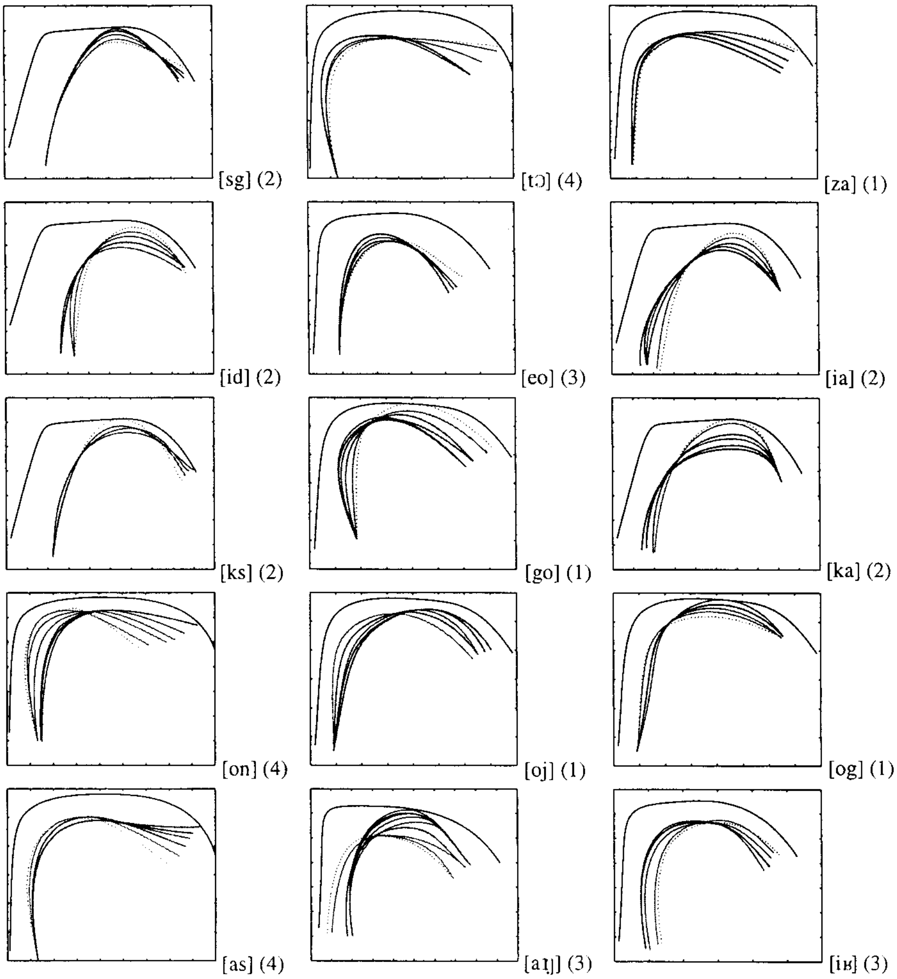

Fig. 6. Examples of the pivot pattern from a variety of transitions. The first frame in each transition is dotted. Each transition is marked by the subject number (in brackets).

transitions, the tongue de-arches or withdraws in a single region. Figs. 7 and 8 show cases of arching, and Figs. 9 and 10 give two examples of de-arching. The change in Fig. 7 is localized to the uvular and upper pharyngeal areas, and fades away from there. In Fig. 8, the tongue tip advances to the alveolar ridge, causing a change only in the alveolar and palatal areas. The de-archings in Figs. 9 and 10 are in the velar and palatal areas.

Of the 600 transitions analyzed, 86 did not fit either of the dynamic patterns. That is approximately 14%. All the exceptions were re-analyzed to make certain that the errors were not in the tracing procedure, but almost all remained as exceptions. This, however, does not eliminate the tracing procedure as the cause of the exceptions, since many of them were indeed difficult to analyze, and a second try at the tracing did not improve the analysis. Four examples are given in Fig. 11. In all of them, there seems to be an underlying pivot pattern, but the change at the pivot

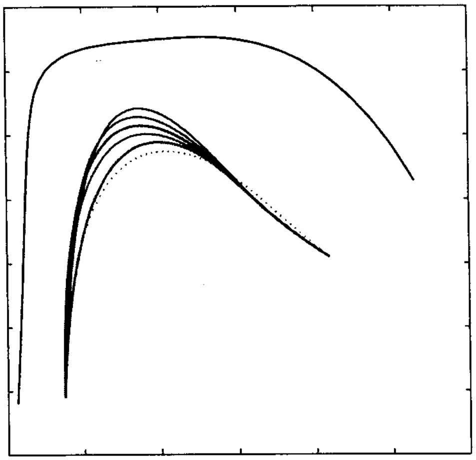

Fig. 7. Transition [ao] by Subject 3. Outline of vocal tract is in black. Vocal tract facing right.

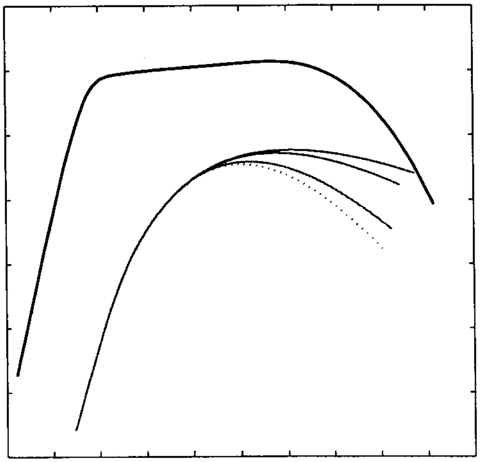

Fig. 8. Transition [æd] by Subject 2.

point is of the same order as the change at least one of the places of articulation. For instance, in the first example, [ag], the change in midsagittal distance function at the "pivot point" in the uvular region is of the same order as the change in the pharyngeal region. None of the obvious nonlinguistic factors like language, speaker, or gender seemed to be the cause for exceptions. The exceptions were also randomly distributed with respect to the initial and final places in the transitions, so that it does not seem to be the cause. Further empirical research is needed to establish whether the exceptions are due to general biological variation or some specific factor not examined here.

Due to expected biological variation, small change in midsagittal distance function does occur at locations in the vocal tract where no linguistic task is specified, e.g., at the pivot in pivot cases.

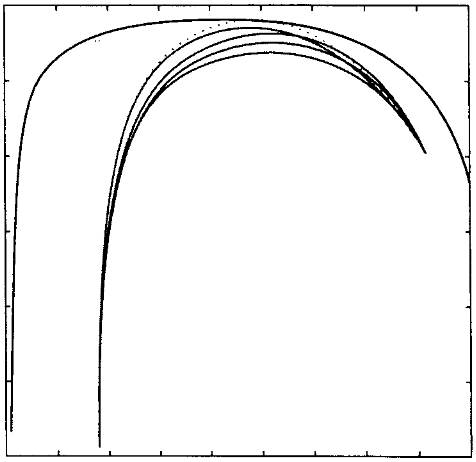

Fig. 9. Transition [gy] by Subject 4.

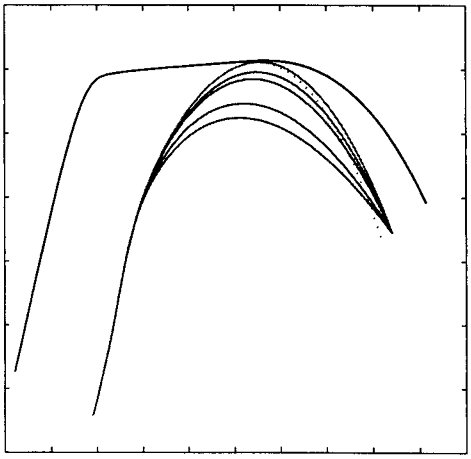

Fig. 10. Transition [kw] by Subject 2.

To quantitatively distinguish between the patterns and exceptions, I computed the ratio of change in the midsagittal distance function at the points of maximal change to change at the point/region of minimal change for the pivots, arches, and exceptions. For pivots, the maximal change was calculated as an average of the changes at the two maxima and the minimum was calculated as the change at the pivot. For the arch and the exception the ratio calculated was of the maximal and minimal points of change in the transition. The ranges of these ratios for the patterns and the exceptions are shown in Fig. 12. As can be seen, even when change does occur at the place of minimal change it is of a much smaller order than at the linguistically targeted areas.

Fig. 13 gives the breakdown of the data into pivots, arches, and exceptions. The classification is by the segment types at the beginning and end of a transition. Pivots are more likely to occur

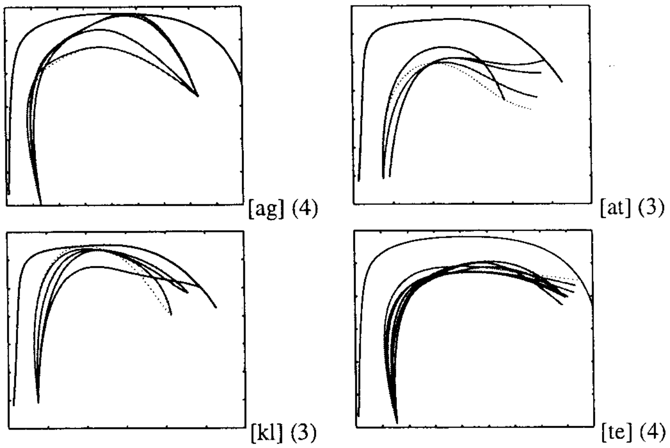

Fig. 11. Exceptions to the pivot and arch patterns. Each transition is marked by the subject number (in brackets).

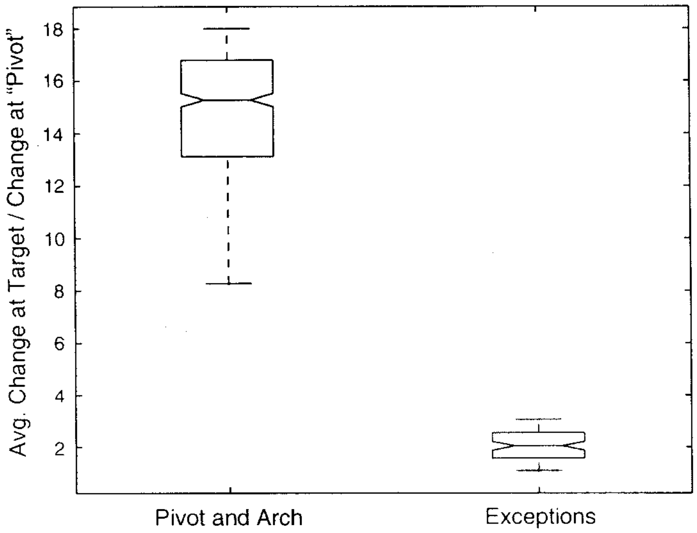

Fig. 12. Range of ratios of maximal change to minimal change. The boxes contain the interquartile range of the data (quartiles 1–3).

when the transition is between segments achieved by portions of the tongue that are not very tightly coupled (e.g., alveolar followed by pharyngeal or palatal followed by uvular), whereas the arch is used more often where the transition is accomplished by the same portion of the tongue or by two portions that are tightly coupled (e.g., velar followed by palatal or uvular followed by uvular). As mentioned earlier, the transitions were collected from a variety of prosodic contexts. The details of the patterns that may be affected by prosodic patterns have still not been investigated however.

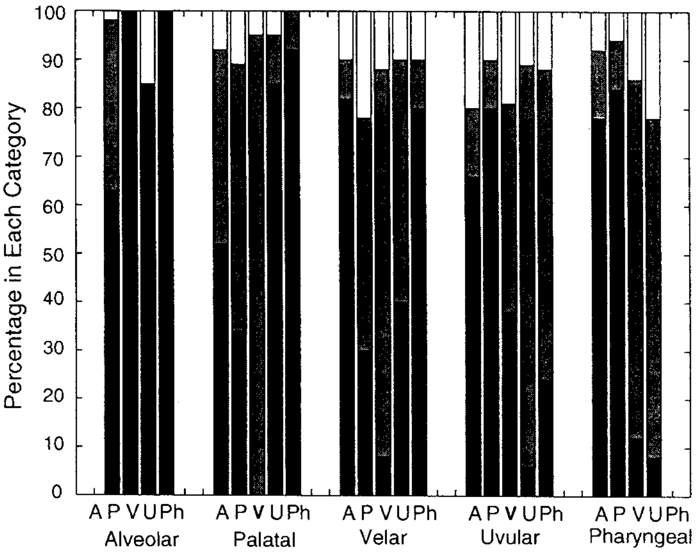

Fig. 13. Breakdown of the data by transition type into pivots (black), arches (gray), and exceptions (white). Each grouping of the bars shows the transition from one place of articulation to all the others. *Abbreviations:* A = Alveolar, P = Palatal, V = Velar, U = Uvular, Ph = Pharyngeal.

4. Discussion

The two dynamic patterns are evidence for a physical constraint on tongue movement in speech production. Change in the midsagittal distance function is limited to the locations of the vocal tract where there are explicit speech tasks. For pivoted cases, the tasks are at two regions in the vocal tract, whereas for the arch cases, only one region is involved. This seems like a fairly trivial generalization, until one considers the fact that points of the entire tongue are moving throughout the transition, and could in principle affect the midsagittal distance function throughout the vocal tract. But that does not occur. The tongue is an elastic continuum with a very large number of degrees of freedom along its midsagittal edge. Once the constraints of muscle geometry have been accounted for by reducing the number of degrees of freedom of the tongue, we are still left with a continuous high degree of freedom actuator. But, for the task of speech production, these degrees of freedom seem to be reduced enormously in that an entire transition can be described with very few parameters: locations and degrees of constrictions in start and end configurations. The reduction in degrees of freedom is possible only if the midsagittal distance function is constrained to vary at very few locations, and is achieved by clamping the function at some other point of the vocal tract (a single section for the pivot pattern, but a whole region in the vocal tract for the arch pattern).

Feature systems for describing the vowels and consonants of the world's languages use static descriptions for the segments. One class of featural systems specify height/backness for vowels and place/degree for consonants (Ladefoged & Maddieson, 1996) and another specifies all segments in terms of place and degree of constriction (Wood, 1979), but they agree on static specifications. This can be taken to suggest that the linguistic control of the vocal tract during speech production is concerned only with the positional targets that the articulators reach. The dynamic portion of speech production, which incidentally lasts far longer than the fleeting moments of zero velocity and direction changes, is then relegated to the fields of biology and physics. The data presented here suggests, in comparison, that the linguistic task governs tongue motion throughout its duration—the dynamic is governed as much as the static target. In other words, the view that language stops at the phonological level, and biology takes over from there seems untenable. Language persists all the way down to the details of the articulatory process, providing a linguistic status to the biological action of the articulators. This is consistent with much recent work in the Laboratory Phonology framework, where the distinction between phonetics and phonology is seriously questioned (Pierrehumbert, Beckman, & Ladd, 2001).

The fleshpoints of the tongue move in complicated paths, but the abstract task-based description is simple: change in midsagittal distance function occurs only where there is an explicit speech task. The theories of Articulatory Phonology and Task Dynamics provide an important framework in which to view the results of this study. Task Dynamics postulates two levels of representation in motor control, as mentioned earlier. At the articulator level, several actuators conspire through the covariation of their motion to achieve invariant tasks at the higher gestural level (Saltzman & Munhall, 1989). Applied to the tongue, its fleshpoints are the low level actuators at the articulator level whose motion acts to achieve the pivot and dynamic invariants by coordinating their motions in such a way as to disrupt the area function in certain locations, but not at others. The theory of Articulatory Phonology, a theory of the gestural level in speech, postulates that the units of speech production, as well as phonological contrast, are discrete yet dynamic gestures that overlap during speech production (Browman & Goldstein, 1990). Pivoting can naturally be seen as the overlap of two dynamic gestures from different phonetic segments. Discreteness is evident in the division of the vocal tract, at the gestural level, into two regions divided by a pivot point. At the articulatory level this discreteness is not present, since points over the whole tongue are moving. The theory of gestures and the relation between gestures and articulator motions are therefore able to account for the empirical results reported here.

The distinction between the arch and the pivot can be also be profitably viewed from the perspective of Articulatory Phonology: the pivot pattern is a case in which the release of the first constriction temporally overlaps the formation of the second constriction, but the release and formation of the constrictions are occurring at separate spatial locations. In the case of arching, there is both temporal and spatial overlap. The release of the first constriction temporally overlaps the formation of the second constriction, but release and formation of constrictions are occurring at one region, the extreme overlap leads to one blended motion.

Other theories that are able to account for the results are factor-analytic models and the discrete regions and modes theory (DRM). As discussed earlier, in factor-analytic models tongue shapes are represented as linear combinations of shape-factors that are extracted as the principal components of a covariance matrix computed over empirical tongue data (Harshman et al., 1977; Maeda, 1990). Tongue movement can be represented then as a trajectory in weight space. If the trajectory is linear, the transition will be pivoted, since the pivot point in a shape factor is represented by a point where the factor has a zero. And a linear operator acting on a zero will preserve that zero. To provide an example of linear vs. nonlinear weight transitions, I used three simulated factors that are sinusoidal in shape, and generated linear and nonlinear transitions from these factors. The right panels of Fig. 14 show the weight transitions, and the left panels present

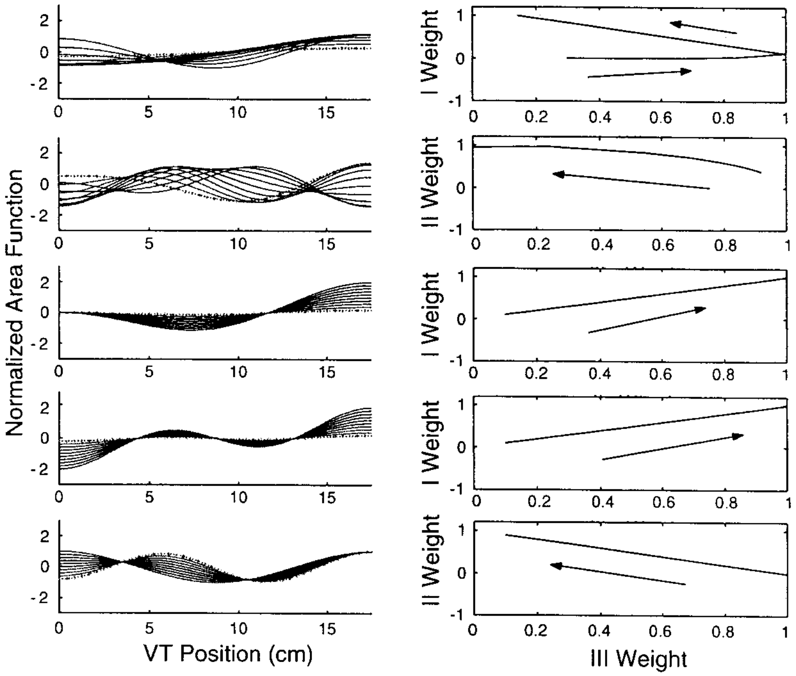

Fig. 14. Simulation of tongue motion as an interpolation of factor weights. Left panels show changes in normalized area function resulting from trajectories in factor space shown in the right panels.

the simulated changes in tongue shape. The top two panels show nonlinear transitions, which are clearly non-pivoted, whereas the bottom three show linear transitions leading to pivoted transitions. DRM researchers have hypothesized that the change in area function during speech production takes place by either the transversal or longitudinal constriction shifts (Carré & Mrayati, 1995). In the transversal strategy, an inter-segmental transition is formed by reducing the area of one region while expanding another. In the longitudinal constriction shift, a reduction in area is simply moved between contiguous regions to the left or right. Pivoting is evidence for the transverse strategy, however the methods used in this work cannot confirm whether pivot points co-occur with points in the vocal tract where formant sensitivity functions vanish, as would be predicted in DRM. No evidence for the longitudinal strategy was found. The data presented here therefore confirm predictions of a wide variety of theories from Articulatory Phonology and Task Dynamics to Factor Analysis and DRM. It may seem surprising that the same set of data can provide evidence for a variety of theories, but there is actually no competition between these theories. Each of them focuses on a different aspect of production.

What I have described here is the empirical support for two patterns of tongue movement. The approach has been descriptive, since the first step in this research is simply to establish that there is an empirical generalization. Future simulation studies are necessary to establish the cause of these patterns, but I outline here where the explanations for the patterns may lie. There are several biomechanical factors that could lead to the patterns: jaw motion, hydrostatic structure of the tongue, and the geometry of tongue muscle force generation. The jaw is a rigid body whose major contribution to speech is in the form of rotation, a kinematic pattern involving pivots, so jaw motion is a possible contributor to the pattern. However, jaw movement by itself would not be able to generate pivots in areas other than the uvular region, where the jaw's rotational pivot occurs. I have provided data for pivots in many other areas of the vocal tract, therefore other causes must also be involved. Due to hydrostatic volume conservation in the tongue, a contraction in one direction is accompanied by expansions in others, exactly as we see in pivoting. Many instances of pivoted transitions are therefore due to this cause, but all the data for this work come from midsagittal data, and the tongue is volume-preserving, not area-preserving, so it is impossible to confirm this conjecture based on this data. Three-dimensional modeling and empirical studies are necessary to demonstrate the role of volume-conservation in pivoting. Another possible biomechanical explanation is the muscular geometry of the tongue. The directions in which forces are generated due to muscle contraction play a major role in deciding which points on the surface of the tongue are going to move orthogonally to the superior and posterior structures, and which points are going to move parallel to them. Finite element models of the tongue that simulate the geometry of muscle forces seem to achieve pivoting in their simulations (Perrier, Loevenbruck, & Payan, 1996). But it is not clear whether this is indeed due to the muscular geometry or the specific implementation of the equilibrium point hypothesis used in that work, where the equilibrium positions of the tongue for different segments are represented as a weighted combination of shape factors, and the dynamics is simulated by a linear interpolation in weight space, which has previously been argued to lead to pivots. Simulation studies are crucial for establishing the involvement of all the biomechanical causes discussed, and for deciding whether there is a need to resort to new control strategies to achieve the required tongue movements.

The discussion of biomechanical factors is not, however, inconsistent with the earlier discussion of the linguistic control of tongue movement. It only seems inconsistent in a dualist framework, where biomechanical and control explanations compete with each other. There is evidence however that, at least in the case of human motor control, the controller knows a great deal about the plant it is controlling and that the controller fully uses the plant's biomechanical capabilities to achieve the desired kinematics (Kugler & Turvey, 1987). The usefulness of the usual division between controller and plant have even been questioned for engineering control systems (Willems, 1997). It is fully possible therefore that the pivot and arch are enforced by linguistic control, but that they would be realized through the use of the biomechanical capabilities of the tongue discussed.

As to the effect of the patterns, it seems that they do influence the dynamic acoustic signal generated during a transition (Iskarous, 2001). If the area function were allowed to be affected throughout the vocal tract during a transition, then the resulting acoustics would contain information in it about portions of the vocal tract that have nothing to do with those that are being signaled linguistically. Pivoting and arching allow the acoustics to be as much as possible about the regions of the vocal tract that are linguistically targeted, since these are the only regions that see a change in their midsagittal distance functions. The result is dynamic formant patterns that are maximally indicative of the regions of the vocal tract they signal. In other words, dynamic acoustic signals generated through pivot and arch transitions are as articulatorily-transparent as possible, since no acoustic contamination is introduced by area function changes in nontarget areas of the vocal tract. Pivot and arch transitions pack maximum information in the acoustics about gestural targets throughout the transitions, since the gestures are overlapped across time.

Indeed, evidence from the last 25 years of speech perception research (Strange, 1989) suggests that perceivers are especially sensitive to the dynamic portions of the acoustic signal, where greatest overlap occurs, and that they are indeed able to isolate different gestural sources of information in the signal (Mann & Repp, 1981; Fowler & Brown, 2000). Pivoting and arching therefore indicate a principle of tongue movement that structures the acoustic signal in such a way that the articulatory information is as available as possible to the perception system.

Acknowledgements

This work was supported by NIH-NIDCD grant DC-02717. I would like to thank Jennifer Cole, Doug Whalen, Mark Hasegawa-Johnson, Carol Fowler, Molly Homer, Jonathan Harrington, Phil Hoole, and two anonymous reviewers for thoughtful comments. This paper is based on the author's 2001 University of Illinois at Urbana-Champaign dissertation, funded by The Army Research Laboratory grant DAAL01-96-2-0003.

References

Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston, & M. Beckman (Eds.), *Papers in laboratory phonology I: between the grammar and physics of speech* (pp. 341–376). Cambridge: Cambridge University Press.

Carré, R., & Mrayati, M. (1995). Vowel transitions, vowel systems, and the distinctive region model. In C. Sorin, J. Mariani, & H. M. J. Schoentgen (Eds.), *Levels in speech communication: relations and interactions* (pp. 73–89). Amsterdam: Elsevier Science.

Coker, C. (1976). A model of articulatory dynamics and control. *Proceedings of the IEEE, 64*, 452–460.

Farin, G. (1997). *Curves and surfaces for computer-aided geometric design: a practical guide* (4th ed.). San Diego: Academic Press.

Fowler, C. (1981). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*, 113–133.

Fowler, C., & Brown, J. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception and Psychophysics, 62*, 21–32.

Gracco, V., & Löfqvist, A. (1999). Coarticulation of tongue and jaw movements in speech. *The Journal of the Acoustical Society of America, 105*, 1354–1360.

Harshman, R., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *The Journal of the Acoustical Society of America, 62*, 693–707.

Heinz, J., & Stevens, K. (1964). On the derivation of area functions and acoustic spectra from cineradiographic films of speech. *The Journal of the Acoustical Society of America, 36*, 1037.

Houde, R. (1968). *A study of tongue body motion during selected speech sounds*. Technical report, Speech Communications Research Laboratories, Inc., Santa Barbara.

Iskarous, K. (2001). *Dynamic acoustic-articulatory relations*. Ph.D. thesis, University of Illinois at Urbana-Champaign.

Kelso, J., Saltzman, E., & Tuller, B. (1986). The dynamical perspective on speech production: data and theory. *Journal of Phonetics, 14*, 29–60.

Kent, R., & Moll, K. (1972). Cinefluorographic analyses of selected lingual consonants. *Journal of Speech and Hearing Research, 15*, 453–473.

Kugler, P., & Turvey, M. (1987). *The information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Lawrence Erlbaum.

Ladefoged, P., & Maddieson, I. (1996). *Sounds of the WorldŠs languages*. Oxford: Blackwell.

Löfqvist, A., & Gracco, V. (2002). Control of oral closure in lingual stop consonant production. *The Journal of the Acoustical Society of America*, *111*, 2811–2827.

Maeda, S. (1990). Compensatory articulation during speech, evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In Hardcastle, W., & Marchal, A. (Eds.), *Speech production and speech modeling*, NATO ASI Series (pp. 131–149). Dordrecht: Kluwer Academic Publisher.

Mann, V., & Repp, B. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America*, *69*, 548–558.

Mooshammer, C., Hoole, P., & Kühnert, B. (1995). On loops. *Journal of Phonetics*, *23*, 3–21.

Munhall, K., & Jones, J. (1995). The spatial control of speech movement. In F. Bell-Berti, & L. Raphael (Eds.), *Producing speech: contemporary issues. For Kathrine Safford Harris.* New York: AIP Press.

Munhall, K., Ostry, D., & Flanagan, J. (1991). Coordinate spaces in speech planning. *Journal of Phonetics*, *19*, 293–307.

Munhall, K., Vatikiotis-Bateson, E., & Tohkura, Y. (1994a). X-ray film database for speech research. *The Journal of the Acoustical Society of America*, *98*, 1222–1224.

Munhall, K., Vatikiotis-Bateson, E., & Tohkura, Y. (1994b). *X-ray film database for speech research.* Technical report TR-H-116, ATR Human Information Processing Laboratories, Kyoto.

Öhman, S. (1966). Coarticulation in vcv sequences: spectrographic measurements. *The Journal of The Acoustical Society of America*, *39*, 151–168.

Perkell, J. (1969). *Physiology of speech production.* Cambridge: MIT Press.

Perrier, P., Loevenbruck, H., & Payan, Y. (1996). Control of tongue movements in speech: the equilibrium point hypothesis perspective. *Journal of Phonetics*, *24*, 53–75.

Pierrehumbert, J., Beckman, M., & Ladd, D. (2001). Conceptual foundations of phonology as a laboratory science. In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological knowledge.* Oxford: Oxford University Press.

Rochette, C. (1973). *Les Groupes de Consonnes en Francais: etude de l'Enchainement Articulatoire a l'aide de la Radiocinematographie et de l'Oscillographie*, Vol. I. Paris: C. Klincksieck.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*, 333–382.

Stone, M. (1991). Toward a model of three-dimensional tongue movement. *Journal of Phonetics*, *19*, 309–320.

Stone, M. (1992). Cross-sectional tongue shape and linguopalatal contact patterns. *Journal of Phonetics*, *20*, 253–257.

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *The Journal of the Acoustical Society of America*, *85*, 2135–2153.

Tiede, M., & Vatikiotis-Bateson, E. (1994). Extracting articulator movement parameters from a videodisc-based cineradiographic database. In *ICSLP 94 proceedings*, pp. 45–48.

Unser, M., & Stone, M. (1992). Automated detection of the tongue surface in sequences of ultrasound images. *The Journal of the Acoustical Society of America*, *91*, 3001–3007.

Willems, J. (1997). On interconnections, control, and feedback. *IEEE Transactions on Automatic Control*, *42*, 326–339.

Wood, S. (1979). A radiographic analysis of constriction locations for vowels. *Journal of Phonetics*, *7*, 25–43.

Wood, S. (1997). A cinefluorographic study of the temporal organization of articulator gestures: examples from greenlandic. *Speech Communication*, *22*, 207–226.

Yehia, H. (1997). *A study on the speech acoustic-to-articulatory mapping using morphological constraints.* Ph.D. thesis, Nagoya University.