Journal of Phonetics (2000) **28,** 233–272
doi:10.1006/jpho.2000.0116
Available online at http://www.idealibrary.com on IDEAL®

A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss

Joseph S. Perkell

Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511 50 Vassar St., Cambridge, MA 02139-4307, U.S.A.; Department of Brain and Cognitive Sciences, MIT, MA, U.S.A. and Department of Cognitive and Neural Systems, Boston University, Boston, MA, U.S.A.

Frank H. Guenther

Department of Cognitive and Neural Systems, Boston University, Boston, MA, U.S.A. and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511 50 Vassar St., Cambridge, MA 02139-4307, U.S.A.

Harlan Lane

Psychology Department, Northeastern University, Boston, MA, U.S.A. and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511 50 Vassar St., Cambridge, MA 02139-4307, U.S.A.

Melanie L. Matthies

Department of Communication Disorders, Boston University, Boston, MA, U.S.A. and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511 50 Vassar St., Cambridge, MA 02139-4307, U.S.A.

Pascal Perrier

Institut de la Communication Parleé, Institut National Polytechnique de Grenoble, Grenoble, France

Jennell Vick, Reiner Wilhelms-Tricarico and Majid Zandipour

Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511 50 Vassar St., Cambridge, MA 02139-4307, U.S.A.

Accepted 3rd August 2000

A theory of the segmental component of speech motor control is presented, followed by supporting data. According to the theory, speech movements are programmed to achieve auditory/acoustic goals. The goals are determined partly by "saturation effects", which are basic characteristics of speakers' production systems that make it possible to produce a sound output that has some relatively stable acoustic properties despite a somewhat variable motor input. The programming of articulatory movements to achieve auditory goals utilizes an internal model (or "mapping") of relations between articulatory configurations

0095-4470/00/030233 + 40 $35.00/0

and their acoustic consequences. The internal model is acquired and maintained with the use of auditory feedback. The supporting data for this theory come from experiments on speakers with normal hearing, cochlear implant users and a patient with neurofibromatosis-2.

1. Introduction

In this paper, we describe a set of hypotheses that are derived from a theory of the segmental component of speech motor control, and we present supporting data from speakers with normal hearing as well as speakers who have experienced a major change in hearing status. The latter are postlingually deafened speakers who received cochlear implants (CIs) and a hearing speaker who suffered hearing loss due to bilateral vestibular schwannomas (also called bilateral acoustic neuromas, neurofibromatosis type-2, or NF-2). We also outline a framework for computer modeling of speech motor control. The framework embodies the set of hypotheses in a format allowing the generation of specific predictions that can be used to guide further investigations of speech motor control in both normal-hearing and clinical populations.

One intention of presenting this material in the current context is to show how it supports the main premise of this issue of the journal: data from clinical populations can provide an invaluable contribution to the development and testing of a scientific theory—enhancing its explanatory power and its potential for eventual clinical application. Another purpose is to suggest that there can be a two-way interaction between research and clinical interests. Researchers and clinicians alike must deal with the role of inter-speaker variability in communication behavior and must make decisions about whether a meaningful change, due to experimental or therapeutic intervention, has occurred. Whether one is pursuing theory-based or therapeutic objectives, it is essential to understand individual variability in respiratory, phonatory, and articulatory behavior—and the contributions of underlying anatomical, physiological and perceptual factors to this variability.

We have found individual differences that can pose challenges to interpretation, but they can also help to illuminate theory and may give insights into the diagnosis and treatment of communication disorders. For example, we have shown that postlingually deafened adults differed from one another in their speech breathing while deaf and differed again in the respiratory changes they made when they received a cochlear implant. An analysis of those differences revealed an underlying regularity: speakers with breathy voice and speakers with pressed voice converged on normative values from opposite directions once they had auditory feedback (Lane, Perkell, Wozniak, Manzella, Guiod, Matthies, MacCollin & Vick, 1998).

In the following sections, we first present the set of hypotheses that guides our experimental work. We then describe a computer-modeling framework that embodies these hypotheses. Next, examples of supporting data from our own research on speakers with normal hearing and with hearing loss are provided. After that, we examine our hypotheses in the context of the work of others, and we conclude by summarizing our main points with emphasis on the crucial role of clinical data in the development and testing of our hypotheses. Portions of the research described herein have been presented in other publications, including Perkell, Matthies, Lane, Guenther, Wilhelms-Tricarico, Wozniak & Guiod (1997) and the publications cited below.

2. A theory of the segmental component of speech motor control

2.1. *Theoretical framework*

2.1.1. *Overview*

Our theoretical framework can be characterized as a collection of inter-related hypotheses, some of which have also been suggested by other researchers (as documented in the following subsections). First, we hypothesize that words in the lexicon are represented as sequences of segments and the segments are represented in the central nervous system as spatio-temporal auditory goal regions. Motor control mechanisms for the production of speech sound sequences are based on the auditory goals, which, for the current purposes, can be equated to *acoustic goals*. Second, we hypothesize that the speech motor control system utilizes an *internal model* of the relationship between vocal-tract shapes and their acoustic consequences. This internal model circumvents the need for continuous direct auditory feedback during mature speech production. Third, we hypothesize that the acoustic goals are determined partly by *saturation effects*, which are nonlinear quantal relations between muscle activity levels and the vocal-tract area function or between the area function and the sound output. These saturation effects reduce the need to keep the internal model precisely calibrated with the help of auditory feedback. Fourth, we hypothesize that the relationship between time-varying motor commands and the kinematic characteristics of articulatory movements is influenced by *biomechanical constraints*. These biomechanical constraints include articulatory collisions (one kind of saturation effect), characteristics of individual speakers' anatomy, and more general dynamical properties of the production mechanism. Finally, we hypothesize that biomechanical constraints are taken into account when planning movements for the production of an intelligible sound sequence. This process affords an *economy of effort*, in which sufficient perceptual contrast is achieved with minimal effort.

Our theoretical framework is expanded and explained in the following subsections with reference to Fig. 1, which is a schematic representation of hypothesized acoustic and articulatory patterns as a function of time for two separate repetitions of the VCV utterance /ugi/. The figure shows the interplay of (1) a motor-equivalent trading relationship of two articulatory movements in achieving an acoustic goal; (2) economy of effort; and (3) saturation effects as they influence articulatory trajectories. The top panel shows, for one acoustic dimension, the planned acoustic trajectory (dashed line) and the actual acoustic trajectories for the two repetitions (solid line and dotted line) as they pass through acoustic-temporal goal regions (shaded shapes—explained below) for each of the three sounds. (The acoustic trajectory is actually multi-dimensional; the single dimension shown here represents a schematic view of the first formant frequency in order to highlight our main points.) The bottom part of the figure shows the corresponding time histories of tongue-blade height, tongue-body height, and lip protrusion.

In order to achieve economy of effort, the acoustic trajectory is planned so it passes through the parts of the acoustic goal regions for /u/, /g/ and /i/ that are closest to one another. The /u/ is produced with a combination of tongue-body raising and lip protrusion, each of which act to lower F_2. (F_1 is shown in the top panel of the figure for

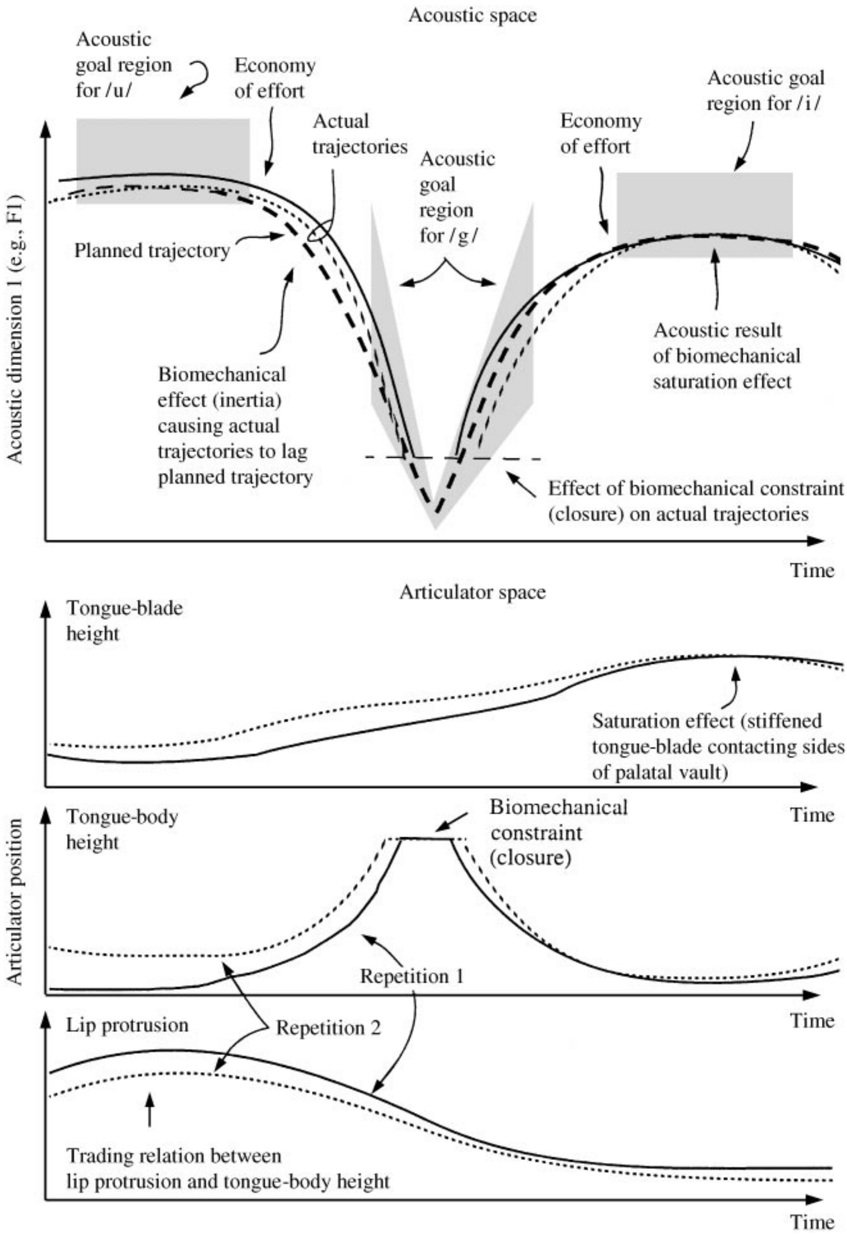


Figure 1. A schematic illustration of the interplay of motor equivalence in achieving acoustic goals, economy of effort and saturation effects as they influence articulatory trajectories. The top panel shows the planned acoustic trajectory along one dimension (dashed line) and the actual acoustic trajectories for the two repetitions of the utterance /ugi/ (solid line and dotted line). The bottom part shows the corresponding time histories of tongue-blade height, tongue-body height, and lip protrusion. See text for details.

clarity of illustration; however, F_2 lowering is a more salient acoustic correlate for /u/ and is the parameter for which we have found motor equivalence for /u/.) In repetition 1, the tongue body is relatively low, so an increased amount of lip protrusion is planned to achieve the desired value of F_2. For economy of effort at the articulatory level, the lip protrusion is only increased enough to place the acoustic trajectory just within the /u/ region, near its edge. In repetition 2, the same principle operates, but lip protrusion is less and tongue-body height is greater. Thus, in the two repetitions, there is a motor-equivalent trading relation between the two articulations, which operates in concert with economy of effort to make the planned and actual acoustic trajectories pass through the edge of the /u/ region.

As the acoustic trajectory approaches the /g/ closure, one kind of biomechanical constraint—the dynamical properties of the articulators—makes the actual acoustic trajectory lag the planned acoustic trajectory. Then, a different kind of biomechanical constraint—a saturation effect due to tissue contact—prevents the actual acoustic trajectory from following the planned trajectory during /g/ closure. A contact saturation effect also facilitates the production of /i/: if the tongue blade is stiff when it is pressed up against the sides of the hard palate to produce the constriction for the vowel, the constriction area and resulting acoustics do not vary much, in spite of variation in the pressing force. These ideas are developed further below.

2.1.2. *The goals of speech movements are regions in acoustic space*

We hypothesize that segmental speech movements are programmed to achieve sequences of phoneme-related goals that consist of regions in a multi-dimensional auditory perceptual space (cf. Jakobson, Halle & Fant, 1969; Guenther, Hampson & Johnson, 1998; Savariaux, Perrier, Orliaguet & Schwartz, 1999) — the "task space" in motor control terminology. Inclusion of the temporal dimension (cf. Fowler, 1980; Nearey & Assmann, 1986; Nearey, 1995) accounts for some aspects of segmental timing and allows for the incorporation of time-varying properties of auditory goals, as proposed by some investigators (cf. Guenther, 1995*a*; Strange & Bohn, 1998; Jenkins, Strange & Trent, 1999) and illustrated for /g/ in Fig. 1. To a first approximation, the nontemporal dimensions of the goal regions can be described by auditorily based transformations of acoustic parameters such as spectral peaks, sound amplitude and fundamental frequency (cf. Zwicker & t'Hart, 1980; Miller, 1989). In the following we make the simplifying assumption that the hypothesized auditory goals of speech movements are approximated by regions in a multi-dimensional acoustic-temporal space (e.g., the shaded regions in Fig. 1), with parameters that can be computed from measurable acoustic characteristics such as sound level, formant frequencies and f_0.[1]

In general, there are several acoustic goals for each speech sound. For example, the main goals for each vowel are a voiced sound source and a particular region in formant space. For many consonants, one goal is abrupt acoustic parameter transitions associated with a diminution in sound level relative to surrounding vowels (e.g., Halle & Stevens, 1990); another (in the case of obstruents) is a noise source with certain spectral characteristics. As illustrated for /g/ in Fig. 1, we speculate that another is a particular

[1]The relation between acoustics and auditory perception may involve decision processes as well as transformations (cf. Kingston, Macmillan, Dickey, Thorburn & Bartels, 1997; Macmillan, Kingston, Thorburn, Dickey & Bartels, 1999).

region in formant space toward which formant trajectories point at the times of vocal-tract closure and release (cf. Delattre, Liberman & Cooper, 1955).[2]

Different styles of speaking could be generated with selective manipulations of the target regions. Fast speech could be generated by selectively compressing the time scale of target regions, and casual speech could be generated by selectively expanding some acoustic dimensions, such as formant ranges.

The claim that all segmental goals are acoustic (auditory) is a strong one. We hypothesized previously that the goals are combinations of articulatory and acoustic parameters (Perkell, Matthies, Svirsky & Jordan, 1995), and in other frameworks primarily articulatory goals have been proposed (cf. Browman & Goldstein, 1989; Saltzman & Munhall, 1989). The elaboration of acoustic goals as a major feature of the current theory is supported by accumulating results of (1) motor equivalence and (2) speech sensorimotor adaptation experiments on normal-hearing speakers (see Section 4), (3) control modeling, and (4) findings on the speech of hearing-disordered speakers—CI users and NF-2 patients.

2.1.3. *An "internal model" is used in the control of speech movements*

If the task space for speech motor control is acoustic, how does the motor control system program articulatory movements so that they follow a planned acoustic trajectory, and how can the speaker be sure that the trajectory is being followed? It is unlikely that auditory feedback is used for closed-loop error correction in the intra-segmental control of individual articulatory movements, because the feedback delay is too large. This delay would include the time needed for perceptual processing of the sound's acoustic–phonetic properties and for making corrections to motor commands, as well as the delay between muscle activation and the generation of muscle tension in producing corrective movements. Such combined durations would probably be longer than some brief speech movements (cf. Kawahara & Williams, 1996; Perrier, Ostry & Laboissière, 1996), and many movements for vowels begin during relatively quiet consonant intervals, when little is available in the way of audible cues about movement trajectories. Moreover, many deafened people continue speaking intelligibly for decades following total hearing loss (cf. Cowie & Douglas-Cowie, 1983; Lane & Webster, 1991; Section 3, below).

We hypothesize that, as with other types of skilled movement, speech production avoids such problems of feedback control with the use of an "internal model". The term "internal model" has been used in a variety of ways (cf. Jordan, 1990, 1996), and it is likely that the brain uses more than one kind of internal model when planning speech movements.[3] The internal model that we will concentrate on in this article (see Fig. 3) is a learned model of how acoustic signals correlate with vocal tract configurations. Specifically, it is defined as a mapping of information about vocal-tract shape, from orosensory feedback and an *efference copy* of the outflow command, into the expected acoustic consequences (Guenther, 1992; Guenther *et al.*, 1998; Houde & Jordan, 1998). In

[2]The consonant /g/ is used in this schematic example for ease of illustration; however, it is well known that the formant loci or /g/ describe two regions in formant space, depending on the vowel context (front or back). This issue presents a dilemma for the current theory which will need to be resolved.

[3]For example, the brain could be utilizing internal models of the forward dynamics, inverse dynamics, and inverse kinematics of the vocal tract in addition to the forward kinematic internal model currently used in our framework.

our framework, the internal model is learned during speech acquisition with the help of auditory and orosensory feedback. With maturation, the model becomes increasingly accurate (cf. Kent, 1976); as it does so, auditory feedback is used mostly for maintaining the model's parameter settings.

A full treatment of the internal model, including its computational mechanisms (cf. Kawato, Furukawa & Suzuki, 1987; Guenther, 1992; Jordan & Rumelhart, 1992; Jordan, 1996) and hypotheses about the responsible neural machinery are beyond the scope of this discussion. For the current purpose, it suffices to say that with an internal model there is no need for continuous direct auditory feedback in controlling speech movements. As shown below in a number of examples, this idea has important implications for the interpretation of data from CI users and NF-2 patients.

2.1.4. *Auditory feedback is used to train and maintain the internal model and to provide information for regulating suprasegmental aspects of speech*

An internal model relating somatosensory information such as muscle lengths to their acoustic consequences must be learned by a developing infant. This is because the parameters that define the internal model differ from one individual to the next due to differences in characteristics such as the sizes and shapes of the speech articulators and the surrounding vocal-tract structures. Furthermore, these parameters must change with time as the speech articulators grow and muscle strengths change. In our framework, the internal model is learned during the process of speech acquisition with the help of auditory, somatosensory and perhaps visual feedback. With maturation, the model becomes robust in the sense that peripheral feedback is used more intermittently—mostly for maintaining the model's parameter settings. The robust nature of the model is evidenced by the good intelligibility of the speech of people with postlingually-acquired profound deafness, even many years after the onset of profound deafness (Cowie & Douglas-Cowie, 1983; Lane & Webster, 1991).

The speech movement control system must combine the influences of robust segmental mechanisms that do not rely on closed-loop auditory feedback with more labile suprasegmental mechanisms that may use relatively simple auditory information closed-loop. That is, the planned acoustic trajectory is influenced by adjustments in suprasegmental parameters that affect intelligibility, such as the average sound level, speaking rate, the degree of prosodically based f_0 and SPL inflection (Perkell, Lane, Svirsky & Webster, 1992; Lane, Wozniak, Matthies, Svirsky, Perkell, O'Connell & Manzella, 1997) and clarity of individual sounds (Picheny, Durlach & Braida, 1986). To guide such adjustments, the speaker uses auditory information to assess factors such as the ambient noise level or quality of a phone line (along with other information such as the listener's view of the speakers face and the listener's knowledge of the language and subject matter). On the basis of this information, the speaker makes relatively rapid adjustments in the average "postural" (baseline) settings of the underlying respiratory, laryngeal and supraglottal mechanisms. Many of these same mechanisms are also under segmental control. For example, the postural settings underlying speech SPL include average levels of subglottal air pressure and degree of vocal-fold adduction, which interact with the state of the vocal folds as they determine the voiced/voiceless distinction. Therefore, the control system must include a means of accounting for the interactions of segmental mechanisms and suprasegmental (postural) mechanisms.

2.1.5. *"Saturation effects" help to define some acoustic goals and simplify motor programming*

To some extent, speech sounds with more stable acoustic goals, such as the point vowels (Stevens, 1972, 1989), are favored in the world's inventory of speech sounds (Ladefoged & Maddieson, 1996). It has been shown that a number of acoustically stable goals can be produced with some variability in the underlying articulation because of "quantal" effects (Stevens, 1972, 1989). A quantal effect is defined as a nonlinear relation between input and output parameters in different measurement domains, in which a continuous change in the input parameter corresponds to a region of rapid change and then a leveling off, or "saturation", of the output parameter. (We call this a "saturation effect" since the term is more widely applicable—see Fujimura & Kakita, 1979.) Thus, there is a region in which the output parameter is stable even with some inaccuracy of the input parameter. Saturation effects are often found in cases where one articulator comes into contact with another. In these cases, which we refer to as biomechanical saturation effects, continuous changes in the motor commands to the muscles (i.e., muscle activation levels) lead to a saturation (stabilization) in the position(s) of the articulator(s). Typically, this also results in a saturation of some acoustic parameter(s). For example, as schematized in the left panel of Fig. 2, with continuous contraction of the muscles that produce labial or lingual closure, the constriction area decreases until it saturates at zero when the closure is complete. A dotted ellipse in the figure encloses a region in which closure is complete over a range of motor command values. This effect is enhanced by a nonlinear relation between sound level and constriction area (middle panel). As a result, the level of the acoustic energy radiating from the vocal tract (through the vocal-tract walls when there is closure) remains constant over a range of closure-producing motor command levels (right panel—dotted ellipse). This kind of saturation effect can help to achieve the stable acoustic goal of a low sound level at closure for obstruent consonant sounds (as with /g/ in Fig. 1) despite somewhat imprecise amounts of muscle activity. Biomechanical saturation effects may also help to determine the acoustic goals for the vowels /i/ and /ɑ/ (Perkell, 1996). For example, as mentioned above, pressing the sides of a stiffened tongue

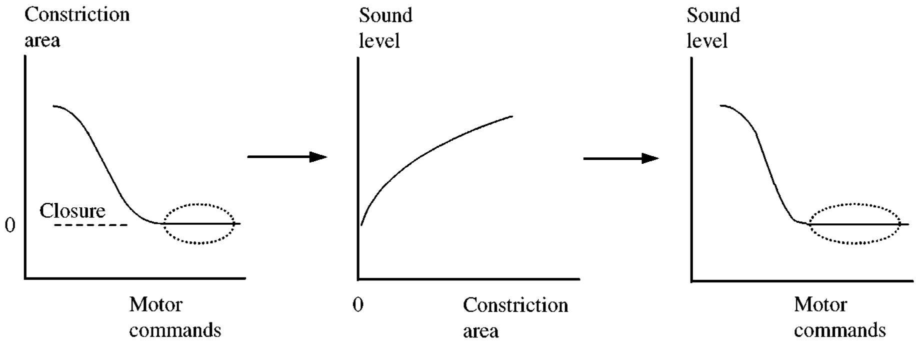


Figure 2. The left panel shows a schematic nonlinear, quantal relation in which continuous variation of the motor command input for a vocal-tract closing movement, on the horizontal axis, results in a region of stability or "saturation" (enclosed in an ellipse) of the constriction area at closure. The middle and right panels show how this effect propagates through to yield a stable region of low sound level output for stop consonants over some range of variation of the motor input.

blade against the hard palate may result in a stable vocal tract area function in the palatal region (and corresponding formant pattern) for the vowel /i/, even with a variable amount of contraction of the posterior genioglossus muscles that are responsible for the pressing action (Fig. 1; Fujimura & Kakita, 1979; also see Badin, Perrier, Boë & Abry, 1990).

In addition to biomechanical saturation effects, there are saturation effects that occur in relations between the vocal tract area function and the acoustic output, which are essentially acoustic in nature. As noted above, these are called quantal effects by Stevens (1972, 1989). For example, because of acoustic cavity interactions, formants of the vowels /i/, /ɑ/ and /u/ are relatively insensitive to some variation in constriction location (cf. Stevens, 1972, 1989; Perkell & Nelson, 1985; Beckman, Jung, Lee, de Jong, Krishnamurthy, Ahalt, Cohen & Collins, 1995). Stevens presents a number of additional examples. Some sounds, such as the more centralized vowels /ɛ/ and /æ/, may not be characterized by either acoustic or biomechanical saturation effects.

The net result of the two kinds of saturation effect is stability in aspects of the acoustic output for certain speech sounds (e.g., stop consonants and the vowels /i/ and /ɑ/), in spite of some variability in their motor input. It should be possible to produce those sounds adequately even with some imprecision or drift in the parameters of the internal model, as might occur with prolonged profound deafness. Thus, the existence of saturation effects may mitigate some consequences of degradation of the internal model. On the other hand, speech sounds that are not characterized by strong saturation effects, such as the centralized vowels, should be more susceptible to distortion with some degradation of the internal model after prolonged deafness. These ideas are illustrated in Sections 3 and 4 by observations of the speech of CI users.

2.1.6. *Biomechanical constraints affect the kinematics of articulatory movements*

As schematized in Fig. 1, we hypothesize that the speech motor control mechanism plans an acoustic trajectory that passes through the target regions for each phoneme in the string of words that makes up the planned utterance. However, this planned trajectory is virtual—that is, the actual acoustic trajectory may or may not fully realize the planned trajectory. We have hypothesized that one of the acoustic goals for each stop consonant is a region in formant space toward which formant trajectories point at times of closure and release. Therefore, the actual formant trajectories for consonants cannot follow the virtual trajectory completely into the target region because the articulators come into contact with one another or the vocal tract wall—shown in Fig. 1 for /g/. As described above, these articulatory "collisions" exemplify the biomechanical saturation effect that helps to define another of the acoustic goals for stop consonants—a relatively low sound level.

Apart from saturation effects, other kinds of biomechanical constraints can also affect movement trajectories. Recent modeling and experimental work has illustrated how the masses and viscoelastic properties of the articulators could partly account for differences in observed mandibular movements that otherwise might have been attributed to central commands (cf. Laboissière, Ostry & Feldman, 1996; Perrier *et al.*, 1996; Ostry, Gribble & Gracco, 1996). Such a dynamical filtering action would account for the lag between the planned and actual trajectories in the approach to the /g/ closure in Fig. 1. In a demonstration of a possible anatomical influence, modeling work by Payan & Perrier (1997) has shown how an observed bi-phasic (double-peaked) velocity profile of a speaker's tongue movement could be due to the complex arrangement of the tongue musculature, rather than a bi-phasic control signal. It is not clear how much of observed kinematic behavior is due to central planning, which takes into account perceptual requirements and biomechanical constraints, and how much is due directly to dynamical and anatomical properties of the vocal tract (Perkell, 1997). To understand the relative contributions of central planning and biomechanical constraints, it is important to quantify those constraints; we are addressing this issue with biomechanical vocal tract models in our modeling framework described below.

2.1.7. *Movements are planned to achieve an "economy of effort"*

In addition to saturation effects, segmental goals, at least for vowels, are hypothesized to be influenced by a compromise between economy of articulatory effort and sufficient perceptual contrast (cf. Lindblom & Engstrand, 1989; Schwartz, Boë, Vallée & Abry, 1997*a*, *b*). By "economy of effort", we simply mean that the motor control system attempts to minimize the physical cost (which could be the amount of energy, and/or the time, or some other quantity—Nelson, 1983) of achieving acoustic goals. For example, the amount of energy required to produce the utterance in Fig. 1 is minimized by moving the articulators just enough to reach the edge of each acoustic target region, rather than expending the extra energy that would be required to reach the center of each acoustic target region (also see Keating, 1988).

To some extent, the relation between economy of effort and perceptual contrast can predict how vowel sounds will be distributed in formant space, depending on the number of sounds in the particular language (Lindblom & Engstrand, 1989). The goal for each vowel will occupy a region (as opposed to a point) in acoustic space (Mermelstein, 1973), and the amount of coarticulation that is observed will depend on the sizes of the target regions (Manuel, 1990; Guenther, 1995*a*). Findings that exemplars of a sound near the center of its region are perceived as being more "prototypical" than those near the edge (Fairbanks & Grubb, 1961; Volaitis & Miller, 1992; Hodgson & Miller, 1996) indicate that the regions have internal structure.

2.2. *A modeling framework that embodies the theory*

We have incorporated much of the above theory into a computer-implemented modeling framework. It is increasingly recognized that in order to reach a thorough understanding of speech motor control strategies, it is essential to quantify hypothetically important constraints such as biomechanical saturation effects, the influences of vocal tract anatomy, and the filtering effects of the dynamical properties of the articulators. If, as we have reasoned, such constraints influence articulatory movements, then motor planning and control strategies must take them into account. Since current techniques cannot measure these constraints directly, the most effective available approach is to implement and test biomechanical—physiological models of the vocal tract and drive them with control models. Implementing our hypotheses in a quantitative model allows us to test our hypotheses in a rigorous fashion.

Fig. 3 shows an overview of our modeling framework, which covers the speech production processes involved in transforming a phoneme string into movements of the speech articulators and a sound output. The framework is comprised of three interrelated models (indicated by the three large shaded boxes in the figure). In the model description

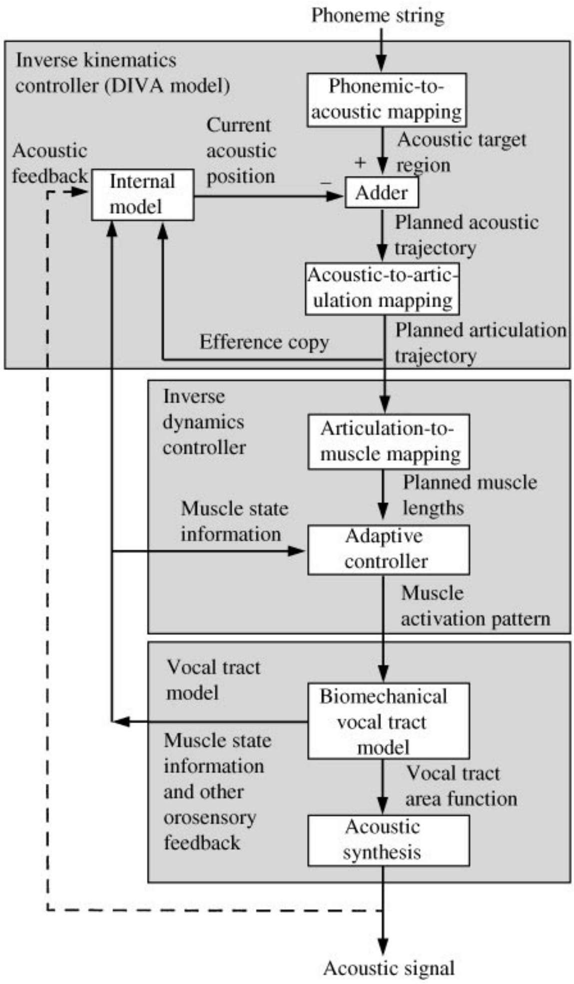


Figure 3. Overview of the proposed modeling framework, which consists of three interrelated models. The inverse kinematics controller takes as input a phoneme string and uses an acoustic movement-planning scheme to produce a planned articulation trajectory that will implement the desired phoneme string. The inverse dynamics controller takes as input the planned articulation trajectory and outputs a muscle activation pattern that will produce the desired articulator movements. The vocal tract model transforms the muscle activation pattern into movements of muscles and tissue in a simulated vocal tract. The resulting vocal-tract area function is used to synthesize an acoustic output. See text for details.

that follows, terms in boldface refer to specific processing blocks and those in italics refer to signal vectors in the figure.

The **Inverse Kinematics Controller**[4] (Fig. 3, top level) takes as input a phoneme string and outputs a planned temporal sequence of articulator positions (*planned articulation trajectory*) to implement the desired phoneme string. The inverse kinematics controller in

[4]The problem of specifying a kinematic trajectory for a set of articulators to carry out a task-space goal (here, an acoustic-space goal) is commonly referred to as the inverse kinematics problem in robotics and motor control.

our modeling framework is an implementation of the DIVA model of speech production, which is an adaptive neural network model that provides an integrated account of a large number of speech production phenomena in normal speakers (Guenther, 1995*a*, *b*; Guenther *et al.*, 1998). When a phoneme string is specified to the model for production, a learned **Phonemic-to-Acoustic Mapping** transforms the phoneme string into a sequence of *acoustic target regions*. The current position of the vocal tract in acoustic space is determined using an **Internal Model** that maps *orosensory information* from the vocal tract and an *efference copy* of the outflow signal into an estimate of the acoustic signal that is expected to arise from the current vocal tract shape. Acoustic feedback is needed to tune the internal model during babbling and to keep it updated later, but (as indicated by the dashed line) direct acoustic feedback is replaced by the internal model's estimate of the acoustic signal during normal speech to avoid long feedback delays. The difference between the target region's position in acoustic space and the current position in acoustic space (calculated by the **Adder** in the figure) defines the desired movement direction in acoustic space. This *planned acoustic trajectory* is then transformed through the **Acoustic-to-Articulation Mapping** into a *planned articulation trajectory* that is used to achieve the desired movement in acoustic space. The use of a mapping from desired acoustic space directions into corresponding articulator directions, rather than a mapping from desired acoustic positions into corresponding articulator positions, allows the controller to account for the high degree of motor equivalent behavior seen in normal speech, such as the ability to compensate for bite blocks or downward perturbations of the lip or jaw (Guenther, 1994, 1995*a*; Guenther *et al.*, 1998).

The **Inverse Dynamics Controller**[5] (Fig. 3, middle level) takes as input the *planned articulation trajectory* and outputs the time-varying *muscle activation pattern* (such as that measured by EMG) that is needed to achieve the desired articulator movements. The inverse dynamics controller first translates the desired articulation trajectory into a higher-dimensional *muscle length trajectory* (i.e., translates each articulation into target lengths of the several muscles that carry out the articulation), indicated by the **Articulation-to-Muscle Mapping** in Fig. 3. This planned *muscle length trajectory* is then transformed by an **Adaptive Controller** into a *muscle activation pattern* that drives muscle contractions in the vocal tract model.

The **Vocal Tract Model** (Fig. 3, bottom level) consists of a **Biomechanical Vocal Tract Model** that transforms the *muscle activation pattern* into movements of muscles and articulators in a computer-simulated vocal tract. The resulting *vocal tract area functions* are used to synthesize an acoustic signal at the **Acoustic Synthesis** stage in the figure. The vocal-tract model currently implemented in our framework is a 2-D model that incorporates anatomical, physiological, and biomechanical properties of the vocal tract (Payan & Perrier, 1997). This 2-D model will eventually be replaced by a more realistic 3-D model of the vocal tract (Wilhelms-Tricarico, 1995, 1996).

3. Supporting data

We developed our theoretical framework in order to integrate observations made from experiments on speech production. In this section, we present a number of such

[5]The problem of specifying a set of muscle forces or actuator torques to carry out a planned kinematic trajectory is commonly referred to as the inverse dynamics problem in the robotics and motor control literatures.

observations from our own research. Some of these examples come from speakers with normal hearing, and many of them come from cochlear implant users and an NF-2 patient. Further examples and methodological details can be found in the cited references.

3.1. *Acoustic goals*: *saturation effects and motor equivalence*

3.1.1. *Some acoustic goals are defined by saturation effects*

Fig. 4 shows the locations of three X-ray microbeam pellets on the tongue (blade, mid and rear) for two speakers (a) and (b) in multiple (*n*) repetitions of the vowel /i/ in a variety of phonetic environments (Perkell & Nelson, 1985). The outline of the subject's hard palate is shown above each data set; anterior is to the right. An ellipse is drawn around each group of points and is centered at its mean location. The inclination of the major axis of the ellipse corresponds to the angle of the principal component of variance for the group, and the length of each axis is equal to two standard deviations. The long axis of the distributions of the blade and mid points is parallel to the outline of the hard palate. We hypothesize that the length of this distribution reflects a quantal effect for /i/ that tends to produce relatively stable formant values over a range of constriction

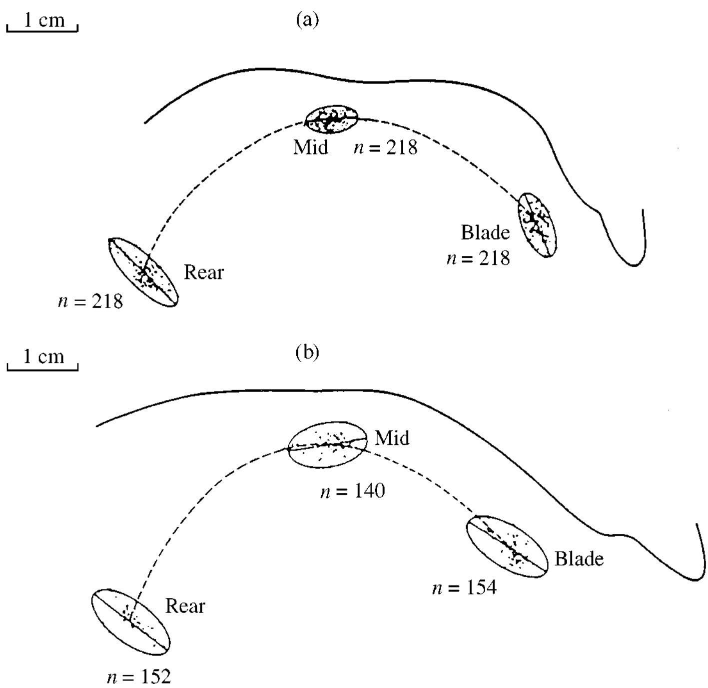


Figure 4. The locations of three X-ray microbeam pellets on the tongue (blade, mid and rear) for two speakers (a) and (b) in multiple (*n*) repetitions of the vowel /i/ in a variety of phonetic environments. The outline of the subject's hard palate is shown above each data set; anterior is to the right. An ellipse is drawn around each group of points and is centered at its mean location. The inclination of the major axis of the ellipse corresponds to the angle of the principal component of variance for the group, and the length of each axis is equal to two standard deviations along the axis.

locations (Stevens, 1972, 1989). The more limited length of the minor axis is presumably due to the above-mentioned biomechanical saturation effect, in which movement of the tongue dorsum perpendicular to the hard palate is restricted by a stiffened tongue blade that is braced against the lateral walls of the palate. Recall that another part of the hypothesized mechanism for production of an /i/ is a variable amount of contraction of the posterior genioglossus muscle as it pushes the tongue upward and forward. This hypothesis is supported by the orientation of the long axis of the rear pellet distributions, which are approximately parallel to the genioglossus muscle fibers in that region of the tongue. As mentioned above, there are other saturation effects, biomechanical and acoustic, that help to define some phonemic acoustic goals. We speculate that the vowels /i/ and /ɑ/ are particularly stable, partly because of combinations of acoustic saturation effects for place of constriction (quantal acoustic effects—Stevens, 1972, 1989) with biomechanical saturation effects for degree of constriction (Perkell, 1996). The vowel /u/ is characterized by an acoustic saturation effect for place of constriction (Stevens, 1972, 1989), but there is no obvious saturation effect, either acoustic or biomechanical, which could compensate for the sensitivity of the vowel's acoustics to changes in constriction degree. As described below, a motor equivalence strategy may provide such compensation.

3.1.2. *There is motor equivalence of tongue and lip articulations in production of the vowel /u/*

In speech production, it is possible to produce approximately the same acoustic transfer function with somewhat different area functions (Atal, Chang, Mathews & Tukey, 1978); therefore, two independently controllable vocal-tract constrictions may make "motor-equivalent" contributions to achieving the same acoustic goal. The term "motor equivalence" refers to the observation that, in multiple tries, the same goal is reached in more than one way. If two independent actions contribute to achieving a single goal, then in different repetitions, one action could contribute more and the other less, and vice versa. The existence of such a motor-equivalent trading relation would be evidenced by covarying contributions of the two actions in multiple attempts to reach the goal, and it would support the idea that the goal is the controlled variable—as opposed to parameter values of the contributing actions. We hypothesize that motor equivalence in the contributions of two vocal-tract constrictions to the same acoustic transfer function reflects a motor control strategy that helps keep acoustic variation within perceptually acceptable regions in acoustic space, such as those schematized in Fig. 1. Thus, findings of this kind of motor equivalence would support the hypothesis that acoustic goals are controlled variables in speech production.

We have obtained preliminary evidence of such articulatory-to-acoustic motor equivalence, in production of the vowel /u/ (Perkell, Matthies, Svirsky & Jordan, 1993; Perkell *et al.*, 1995), as schematized in Fig. 1. The plots in Fig. 5 present examples of data from these studies. They show locations of movement transducer coils on the tongue body (TB), upper lips (UL), lower lips (LL) and lower incisors (LI) for multiple repetitions of /u/ produced in a carrier phrase by two subjects (2 and 6). In the vowel /u/, tongue-body raising and lip rounding both contribute to decreasing F_2, an important acoustic correlate of the vowel. We found significant negative correlations, indicating trading relations between values of tongue-body height and upper-lip protrusion, in such data from four of six male speakers of American English, including Subject 2. Data from one

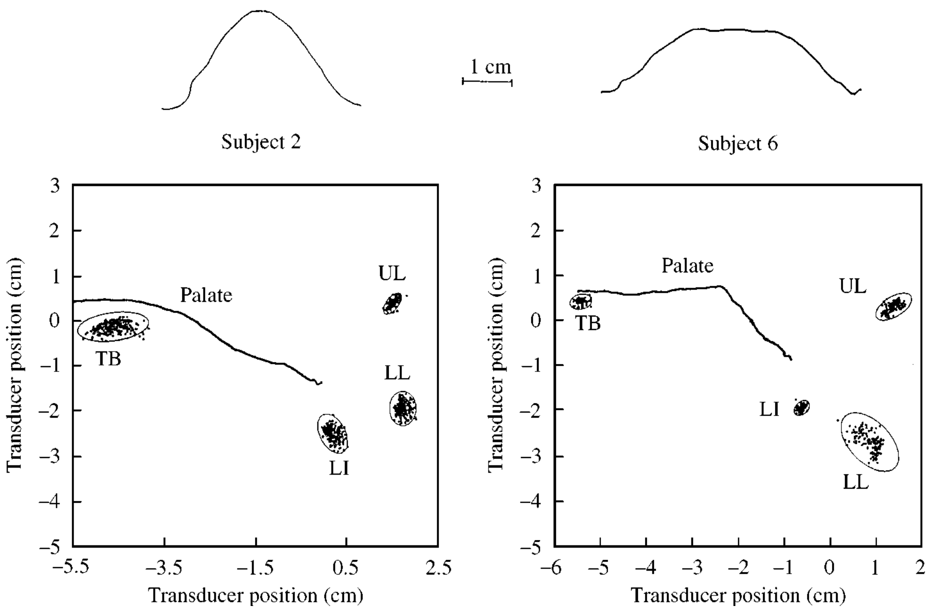


Figure 5. The lower part of each half shows midsagittal locations (cm) of movement transducer coils on the tongue body (TB), upper lips (UL), lower lips (LL) and lower incisors (LI) for multiple repetitions of /u/ produced in a carrier phrase by two male speakers of American English. The upper part of each half shows the coronal contour of the speaker's hard palate, close to the place of maximum constriction for the vowel /u/.

of the two subjects who did not evidence a trading relation (Subject 6) are shown in the right half of the figure. Above each data plot is a coronal outline of the subject's hard palate in the molar region, made from a cut dental cast. The outlines show that the palate of Subject 6 is unusually broad and flat in the region of the /u/ constiction. With such morphology, even modest vertical tongue displacements would create large percentage changes in the area function at the constriction, which would be expected to result in unacceptably large formant differences. Consistent with this idea, the distribution of tongue body points for /u/ of Subject 6 (Fig. 5) was the smallest among the subjects. While Subject 6 could also have been using motor equivalence, his small TB distribution may not have contained enough variation for a significant correlation. Thus, individuals may differ in their expression of motor equivalence, depending on the shapes of their vocal tracts.

3.1.3. *There are motor equivalent trading relations in production of the two types of /r/*

We have performed another type of motor equivalence experiment for the American English phoneme /r/ (Guenther, Espy-Wilson, Boyce, Matthies, Zandipour & Perkell, 1999). For this experiment seven speakers pronounced five repetitions of /r/ in different phonetic contexts. The contexts were designed to elicit productions of /r/ between two different tongue configuration extremes: "bunched" as in the utterance /wagrav/, and "retroflexed" as in the utterances /warav/ or /wabrav/. The positions of three EMMA transducers on the tongue were recorded (Perkell, Cohen, Svirsky, Matthies, Garabieta

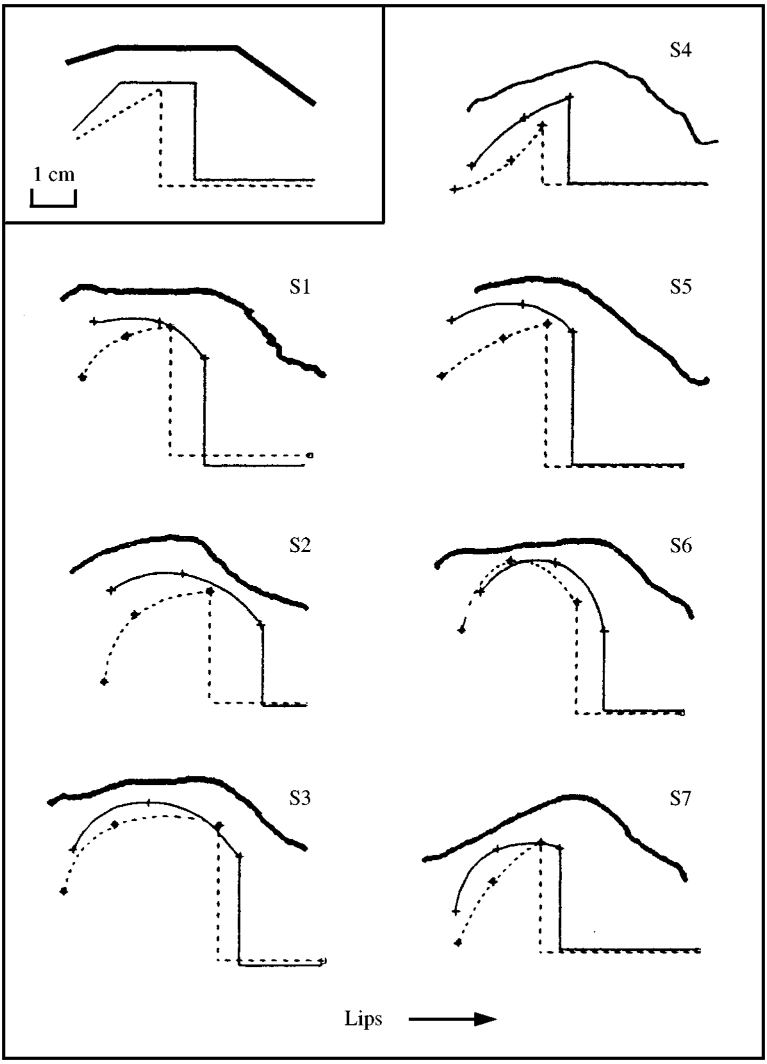


Figure 6. The average positions of three EMMA transducers on the tongue for /r/ in two contexts, as produced by seven subjects (S1–S7). The positions for /r/ following a velar consonant are connected by solid curves, and those for the /r/ following a dento-alveolar consonant or the vowel /ɑ/ are connected by dashed curves. The curves are extended as straight-line segments downward and forward from anterior transducer location. Tradeoffs between front cavity length and constriction length and/or area are evident in all seven subjects.

& Jackson, 1992), along with the acoustic signal. Fig. 6 shows the resulting articulatory data for the seven subjects (S1–S7). For each subject, the average transducer positions for the five repetitions of /r/ in each of the two contexts are indicated by small symbols. The transducer positions for the /r/ following a velar consonant are connected by solid curves, and those for /r/ in a second context (either /wɑrɑv/ or /wɑbrɑv/, depending on the subject) are connected by dashed curves. The curves are extended as straight-line segments downward and forward from anterior transducer location, for use in approximating the size of the front cavity.

All seven subjects showed tradeoffs between the front cavity length and the constriction length and/or area, which helps to maintain a low F_3 for /r/ produced with different tongue configurations in the two different contexts. Furthermore, statistical prediction of the time-varying F_3 trajectories from the tongue point trajectories showed that inclusion of covariances of the transducer positions significantly reduced the F_3 variability. These results indicate that the acoustic goal of a low F_3 for /r/ is achieved with different tongue configurations in different contexts, implying that the goal for /r/ is acoustic rather than articulatory.

3.1.4. *Motor equivalence helps sharpen acoustic phonetic boundaries*

We have run additional motor equivalence studies with the sounds /u/, /r/ and /ʃ/, in which we manipulated phonetic context and speaking rate to create differences in how "canonical" the target sounds are. These results show that when motor equivalence is found, it is among less canonical sets of tokens, i.e., tokens that are near boundaries shared by other sounds in acoustic space (Perkell, Matthies & Svirsky, 1994; also see Fig. 1, Section 3.3.3). Such results may indicate that the motor equivalence mechanism has the effect of "sharpening" acoustic–phonetic boundaries by coming into play only as necessary. For example, if some source of variation such as the preceding or following consonant were to cause the tongue body to be low enough to endanger the phonetic identity of an intended production of the vowel /u/, an acceptable /u/ could be produced with increased lip rounding. According to the principle of economy of effort, the increased lip rounding would be just enough to place the token just within the /u/ acoustic target region, near the edge (see Fig. 1). Without the motor-equivalent compensation, the token would have fallen outside of the /u/ region, contributing to a less distinct boundary between /u/ and the adjacent speech sound. A control mechanism that utilizes motor-equivalent trading relations would thus increase acoustic stability, particularly in situations where such stability does not result from saturation effects. Where biomechanical or acoustic saturation effects do exist, there is a reduced need for motor equivalence in achieving acoustic stability. This point is developed further in a discussion of data on the /s/-/ʃ/ contrast in Section 3.3.4.

Motor-equivalent behavior might be the speech production counterpart of discrimination enhancement or "perceptual sharpening" at phoneme boundaries in studies of categorical perception (cf. Liberman, Harris, Hoffman & Griffith, 1957; Liberman, Harris, Kinney & Lane, 1961; Pisoni, 1973) and increased discrimination abilities away from perceptual "prototypes" in studies of the perceptual magnet effect (cf. Kuhl, 1991; Iverson & Kuhl, 1995, 1996).

3.2. *The role of auditory feedback in adult speech production*

To acquire normal speech, children need to hear. On the other hand, people with normal hearing who have acquired speech and then become profoundly deaf in adulthood continue to speak intelligibly, albeit with some degree of abnormality variable across speakers, for decades following the onset of profound deafness (Cowie & Douglas-Cowie, 1983; Lane & Webster, 1991). In terms of our theoretical framework, these observations indicate that postlingually deafened speakers are using a robust internal model of the production process that they acquired when they could hear. The observations also indicate that insight into the role of auditory feedback in speech production could be gained by studying speakers who experience a change in hearing status. There is indeed a large literature on the effects of short-term changes in hearing status, particularly on suprasegmental parameters (reviewed in Lane & Tranel, 1971); for the most part, however, this literature does not address the nature of the internal model for segmental control and its interaction with suprasegmental control.

We have studied the role of auditory feedback in two groups of people who experienced a change in hearing after having acquired normal speech: (1) cochlear implant (CI) users who, having acquired speech with at least adequate hearing, became profoundly deaf as adults, and then after some time gained a form of prosthetic hearing, and (2) patients with bilateral vestibular schwannoma (NF-2—for neurofibromatosis type 2; see Parry, Eldridge, Kaiser-Kupfer, Bouzas, Pikus & Patronas, 1994), who lost their hearing due to tumors growing on the auditory nerves. As is suggested in the theoretical overview and indicated by the following data from these studies, auditory feedback has two roles in mature speech production: (1) to maintain parameters, which we call "phonemic settings," of the segmental mechanisms that produce phonemic distinctions, and (2) to assure adequate intelligibility by monitoring the acoustic environment and helping to guide any needed changes in "postural settings" that underlie average values of SPL, f_0, speaking rate and degree of prosodically related inflection of SPL and f_0.

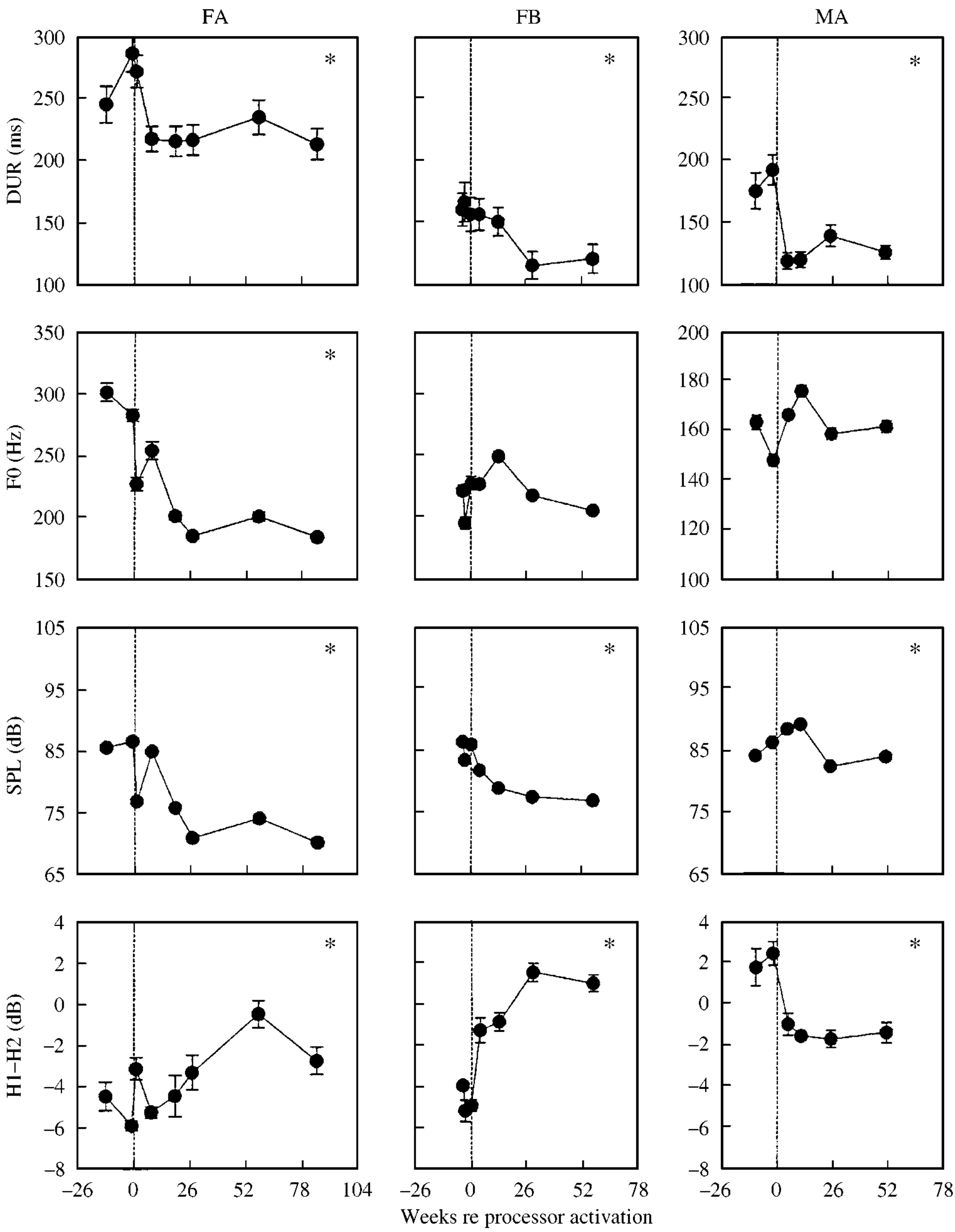
3.2.1. *Postural changes are made rapidly in response to changing acoustic transmission conditions*

The following examples illustrate changes in suprasegmental parameters such as average speaking rate, sound level and degree of prosodically related inflection of f_0 and SPL that occur with a change in hearing status. We hypothesize that speakers modify the underlying physiological "postures" in response to perceived changes in acoustic transmission conditions.

Fig. 7 shows longitudinal data averaged across nine nondiphthongized vowels of American English as spoken in carrier phrases by two female (FA and FB) and one male (MA) cochlear implant user (Perkell *et al.*, 1992). (Data from these and additional subjects are shown in several of the following figures, using the same identification method throughout.) The three columns contain the data from the three subjects. The rows (from top to bottom) show vowel duration (DUR), f_0, SPL, and $H_1 - H_2$ (the

Figure 7. (Opposite.) Longitudinal data averaged across vowels of American English as spoken in carrier phrases by two female (FA and FB) and one male (MA) cochlear implant users. The rows (from top to bottom) show vowel duration (DUR—ms), f_0 (Hz), SPL (dB), and $H_1 - H_2$ (dB—the difference in amplitude between the first two spectral harmonies for selected vowels). In each panel: each data point represents the mean parameter value across three repetitions of the set of vowels recorded in one session; the error bars indicate one standard error about the mean; and the vertical line indicates the time that the subject first received prosthetic hearing from activation of the speech processor of the implant. The two data points to the left of the line are from preactivation (no hearing) sessions and the points to the right of the line are from sessions with prosthetic hearing. The horizontal axis shows time in weeks from the time of processor activation. The asterisks indicate that there was a statistically significant difference ($p < 0.05$) between the mean of data from the two preactivation sessions and the last two postactivation sessions.

difference in amplitude between the first two spectral harmonics for selected vowels—an indirect relative measure of glottal abduction and a correlate of breathiness). In each panel: each data point represents the mean parameter value across three repetitions of the set of nine vowels recorded in one session; the error bars indicate one standard error about the mean; and the vertical line indicates the time that the subject first received prosthetic hearing from activation of the speech processor of the cochlear implant. Thus, the two data points to the left of the line are from preactivation (no hearing) sessions and



the points to the right of the line are from sessions with prosthetic hearing, with the first postactivation session occurring within a few days of processor activation. The horizontal axis shows time in weeks from the time of processor activation. The asterisks indicate that there was a significant difference between the mean of data from the two preactivation sessions and the last two postactivation sessions (t-test for matched pairs; $p < 0.05$). Following the onset of prosthetic hearing, vowel duration decreased (i.e., rate increased) in all three subjects; f_0 dropped for FA (in whom it had been abnormally high); SPL dropped in all three subjects; and $H_1 - H_2$ (i.e., breathiness) increased for two subjects and decreased for one. Most of these changes occurred within the first 26 weeks and many occurred within the first few weeks.

Fig. 8 shows average values (in percentage of maximum) of "mean successive differences" (MSD) of f_0 from syllables in phrases of the Rainbow Passage (Fairbanks, 1960). To obtain MSD, each syllable f_0 in a phrase is expressed as a percent of f_0 maximum in the phrase (Lane *et al.*, 1997) and the absolute differences among successive syllables are computed and averaged. Each pair of data points connected by a line shows averaged data from a different speaker. The points connected by solid lines and labeled MD, FA, FB and FC come from cochlear implant users who changed from a no-hearing (Pre) status to a prosthetic hearing status (Post). The points connected by a dashed line and

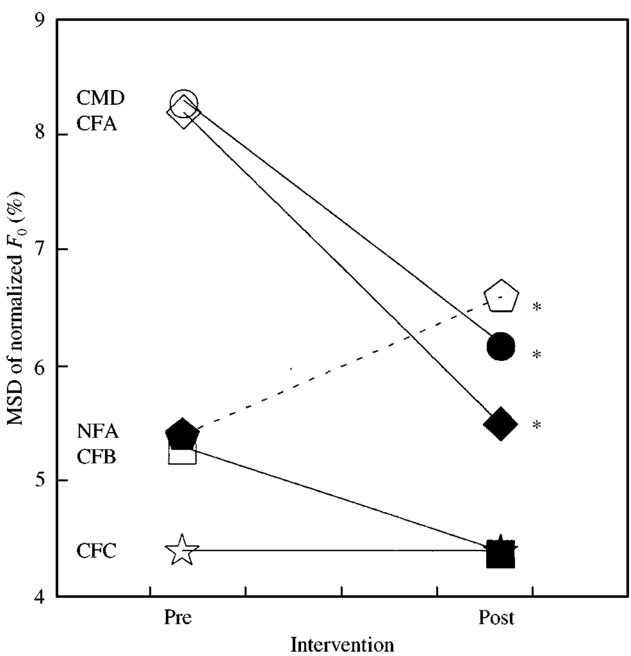


Figure 8. Average values (in percentage of maximum) of "mean successive differences" (MSD) of f_0 from syllables in phrases of the Rainbow Passage (see text for details). MSD is an average of the absolute values of syllable-to-syllable inflection of f_0, normalized by the maximum f_0 in the phrase. Points connected by solid lines and labeled CMD, CFA, NFA and CFB come from cochlear implant users who changed from a no-hearing (Pre) status to a prosthetic hearing status (Post). Points connected by a dashed line and labeled CFC come from an NF-2 patient who changed from a hearing (Pre) status to a severe but not complete hearing loss (Post) status. The asterisks indicate significant pre–post differences.

labeled FD come from an NF-2 patient who changed from a hearing (Pre) status to a severe but not complete hearing loss (Post) status. (More details on the NF-2 patient are given below.) The asterisks indicate significant pre–post differences (*t*-test for matched pairs; $p < 0.05$). The figure shows that syllable-to-syllable inflection of f_0 is increased in the absence or loss of hearing (open symbols). Note that the two deafened adults who did not reduce f_0 inflection with prosthetic hearing (FB, FC), had preimplant MSD values similar to those obtained from the NF-2 subject (FD) before her hearing loss. Generally, MSD of SPL values behaved in the same way as f_0. We infer that the inter-syllable inflection of f_0 and SPL—presumably related to conveying prosodic information—belong to the category of postural parameters that are adjusted to assure intelligibility. The amount of this inflection is increased in the presence of hearing loss or with degraded acoustic transmission conditions (Lane *et al*., 1997).

Fig. 9 shows values of average vowel SPL (top two panels) and $H_1 - H_2$ (bottom two panels) from two cochlear implant users, a male (MA—on the left) and a female (FB—on the right) (Svirsky, Lane, Perkell & Webster, 1992). These data were collected in single "on–off" recordings made a number of months after the subjects had been using their implants, when the parameters had more or less stabilized. The subjects' speech processors had been left off for 24 h prior to the recording session. During the session, the speech material was first recorded with the processor off (unfilled circles); then twice with the processor on (filled circles) and twice more with the processor off (unfilled circles). The error bars show one standard error about the mean. The horizontal axis shows time in minutes. The unfilled triangles at the left of each panel show average preimplant parameter values measured before the subjects had received any prosthetic hearing, and the filled ones show average values from the most recent longitudinal recording made with prosthetic hearing. In general, the values in the off condition are closer to pre-hearing values and the values in the on condition are closer to the most recent hearing values. The changes at the processor off-to-on transition are more pronounced than at the following on-to-off transition in three of the four cases. Individual token data were variable within conditions, but there were a number of obvious rapid changes (within a few tokens) between conditions, especially from off to on (Svirsky *et al*., 1992). For FB, the $H_1 - H_2$ changes probably reflect changes in average glottal aperture that she made to help implement her SPL changes. The picture is less clear in MA's case, perhaps because he was making greater use of subglottal pressure in regulating SPL.

Considered together, such data indicate that these postural parameters (average duration, f_0, SPL, $H_1 - H_2$, inflection of f_0 and SPL) change relatively rapidly with a change in hearing status (Perkell *et al*., 1992; Svirsky *et al*., 1992; Lane *et al*., 1997), in a way that reflects what speakers with normal hearing presumably do all the time to assure intelligibility: when transmission conditions become more uncertain or demanding, people speak more slowly, with increased sound level and exaggerated inflection of f_0 and SPL contours.

3.2.2. *Phonemic settings are relatively stable*

The following example presents data about phonemic settings for vowels. Phonemic settings are hypothesized to be parameters of the internal model that govern the production of phonemic contrasts. Once learned, phonemic settings generally are resistant to change, especially in adulthood.

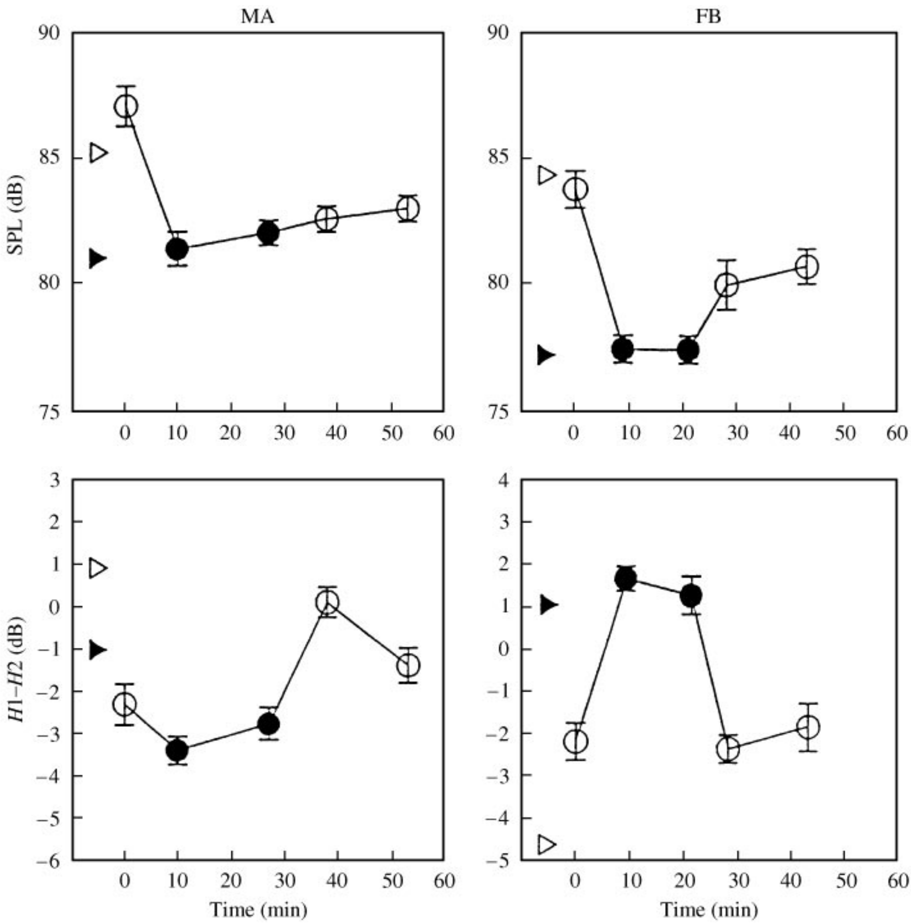


Figure 9. Values from an "on–off" experiment of average vowel SPL (dB—top two panels) and $H_1 - H_2$ (dB—bottom two panels) from two cochlear implant users, a male (on the left) and a female (on the right). The data were collected in single recordings made a number of months after the subjects had been using their implants. The subjects' speech processors had been left off for 24 h prior to the recording session. The speech material was first recorded with the processor off (○○○); then twice with the processor on (●●●) and again twice with the processor off (○○○○). The error bars show one standard error about the mean. The horizontal axis shows time in minutes. The unfilled triangles at the left show average preimplant parameter values (without hearing); the filled triangles show average values from the most recent longitudinal recording made with prosthetic hearing: ○, SP off; ●, SP on; ▷, pre-activation; ▶, post-activation.

The stability of the phonemic settings for vowels is evidenced by the predominantly normal vowel formant patterns seen in Fig. 10. The figure shows sets of average F_1 and F_2 values arranged by vowel for cochlear implant users FA, FB and MA. The small squares connected by dotted lines show normative values from Peterson & Barney (1952). The values indicated by unfilled circles are preimplant, and those indicated by filled circles are postimplant (averaged from the last two recordings for each subject). The error bars indicate one standard error about the mean. For the most part, overall vowel formant patterns (relations of formant values to one another among the vowels) appear

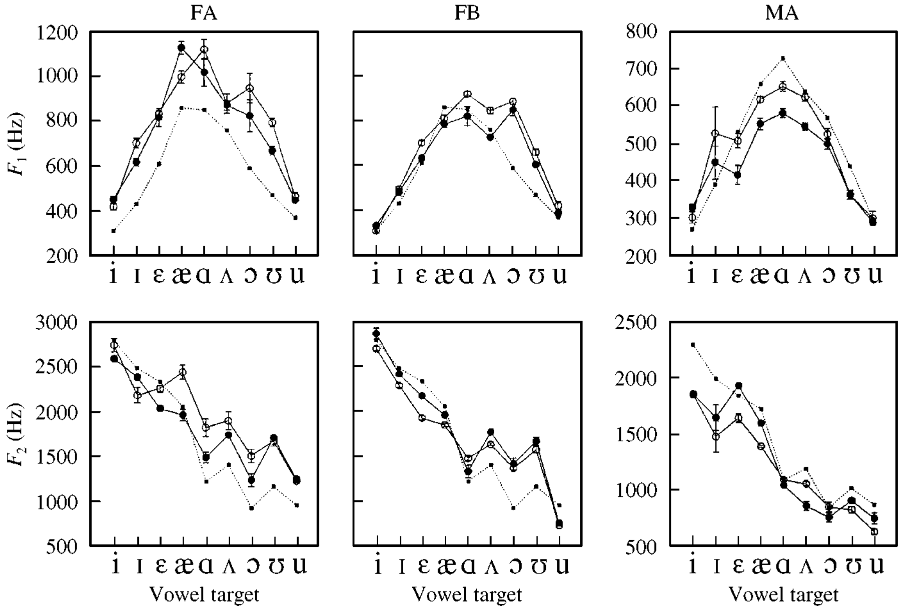


Figure 10. Average F_1 and F_2 values (Hz), arranged by vowel, for CI users, FA, FB and MA. The set of values indicated by small squares and connected by dotted lines are normative values from Peterson & Barney (1952). The values indicated by unfilled circles are preimplant, and those indicated by filled circles are postimplant (averaged from the last two recordings for each subject).

to be relatively congruent with the normative patterns, even years after the onset of profound hearing loss (Perkell *et al.*, 1992). However, there are several exceptions to this observation. Three of these are associated with large amounts of variability, as indicated by the error bars: FA's F_1 for /ɔ/ and MA's F_1 and F_2 for /ɪ/. In addition, FB's F_1 for /ɔ/ is relatively high and MA's F_2 for /i/ is low, pre- and postimplant. (The F_1 results are discussed further below.) For FA, 18 years after the onset of her profound hearing loss, preimplant F_2 values among the front vowels /i/, /ɪ/, /ɛ/ and /æ/ were somewhat disordered with respect to the Peterson and Barney data, primarily due to relatively high values for /ɛ/ and especially /æ/. As indicated by the filled circles, after about a year with prosthetic hearing, these F_2 values are more in line with the Peterson and Barney pattern. Thus, FA's abnormal preimplant F_2 pattern was "corrected" toward the normative pattern after some months of implant use, but FB's value of F_1 for /ɔ/ and MA's F_2 for /i/ were not. Although there are relatively higher F_2 values for most of FA and FB's back vowels, which indicates some overall tongue fronting (both pre- and postimplant), the shapes of the patterns are congruent with the Peterson and Barney data.

FA's vowel identification scores improved postimplant (Perkell *et al.*, 1992), with approximately the same time course as the changes toward a normative pattern in values of F_2 for /ɛ/ and /æ/ (over an interval of roughly 26 weeks postimplant). Thus, perceptual learning and motor learning seem to have proceeded together.

In sum, based on the data available so far (Perkell *et al.*, 1992; Cowie & Douglas-Cowie, 1992), the relative positioning of vowels within the vowel space seems largely unaffected by hearing loss, although we did observe some small shifts. We believe that these data reflect the relative stability of speakers' internal models once they have matured, prior to the onset of hearing loss.

3.2.3. *There are interactions between postural changes and phonemic settings*

Since postural and phonemic settings can influence the same articulatory structures, there can be interactions between them. Such interactions, or the lack thereof, can reveal further details about the control mechanism. We consider two cases of interactions and one case of independence: interactions between speaking rate and vowel F_1, and an interaction between average glottal abduction and vowel-specific differences in SPL; and independence of several postural parameters from VOT.

As discussed above, the top half of Fig. 10 shows F_1 values by vowel for three implant users, FA, FB and MA pre- and postimplant, in comparison with data from Peterson & Barney (1952). A comparison between the unfilled (pre) and filled (post) circles indicates some compression along the F_1 dimension postimplant, primarily due to decreased F_1 values for the low vowels. This change in F_1 pattern with prosthetic hearing might seem counterproductive, since it could lead to some diminution of phonemic contrasts. However, correlations between F_1 and vowel duration supported the explanation that the F_1 compression was a by-product of increased speaking rate postimplant, which would allow for less time for tongue lowering for the low vowels (Perkell *et al.*, 1992).

The left half of Fig. 11 shows values of SPL by vowel for cochlear-implant user FB. Normative data from Lehiste & Peterson (1959) are shown as small squares connected by dotted lines; preimplant values are shown with open circles and postimplant values, with filled circles. Error bars indicate one standard error about the mean. Preimplant, the relation among vowels is somewhat similar to the normative data, with the highest, most closed vowels /i/ and /u/ having the lowest values, about 3 dB below maximum.

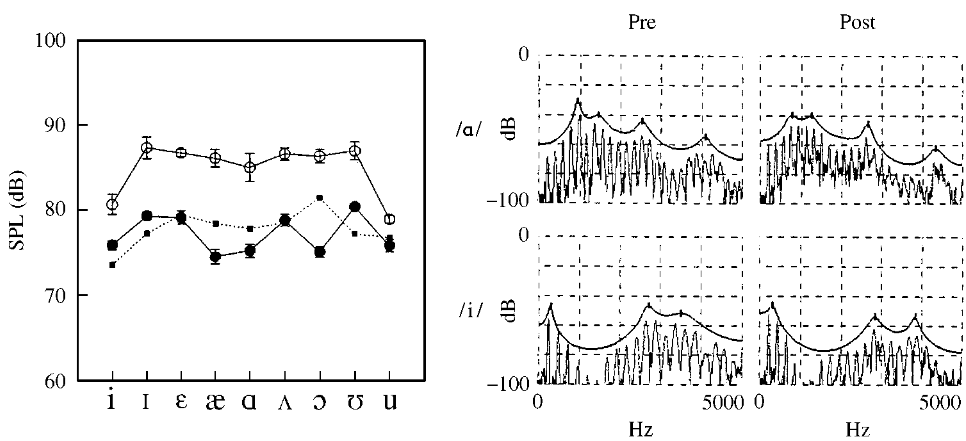


Figure 11. Left half: Values of SPL (dB) by vowel for cochlear-implant user FB. Normative data from Lehiste & Peterson (1959) are shown in small squares connected by dotted lines; preimplant values are shown with open circles and postimplant values, with filled circles. Error bars indicate one standard error about the mean. Right half: Acoustic spectra for the vowels /ɑ/ and /i/ produced by FB pre- and postimplant.

However, in the postimplant data, the open vowels /æ/, /ɑ/ and /ɔ/ have lower SPL values than those for /i/ and /u/. The explanation for this rather unexpected finding is illustrated by the acoustic spectra for the vowels /ɑ/ and /i/, shown in the right half of Fig. 11 and FB's data on $H_1 - H_2$ and F_1. As seen in the right half of the figure, the relative amplitude of F_1 for /ɑ/ is about 10 dB lower postimplant than preimplant; this is not the case for /i/. In Fig. 10, upper half, middle panel, it can be seen that FB's postimplant F_1 frequencies for /æ/, /ɑ/ and /ɔ/ are within about 75 Hz of one another, a little above 800 Hz. According to Klatt & Klatt (1990), the median first subglottal resonance in females has a value of about 750 Hz. A 6-dB postimplant increase of FB's values of $H_1 - H_2$ (lower right panel, Fig. 9, left-hand symbols) implies that her average glottal aperture (a postural parameter) increased postimplant. Thus, the unexpected postimplant pattern of vowel sound levels is probably due to increased subglottal coupling, with a resulting spectral attenuation in the region of 800 Hz from the first subglottal resonance that acts as a supra-glottal anti-resonance. (Presumably, spectra for /æ/ and /ɔ/ would show the same effect.) As did some other implant users, FA apparently increased her average glottal aperture following implantation to help lower her sound level (Lane *et al*., 1998).

The previous two examples (rate-related compression along the vowel F_1 dimension and anomalous amplitude patterning of low vowels) illustrate interactions between postural and phonemic (or vowel-specific) parameters with potentially adverse consequences for the control and realization of segmental contrasts; however, the effects we have seen so far are small and appear to have had minimal impact on the contrasts.

There can be additional cases in which speech parameters might be expected to covary because of mechanical and/or aerodynamic interdependence. If they do not covary, it suggests that the segmental control mechanism is actively overriding the interdependence. For example, the NF-2 patient's VOT (adjusted for speaking rate—see Lane, Wozniak, Matthies, Svirsky & Perkell, 1995) for voiceless consonants—a segmental mechanism—did not change longitudinally, in spite of changes in the aerodynamically related postural parameters of SPL, f_0 and $H_1 - H_2$ measured in following vowels. Unless VOT is maintained with active control, it could increase as a passive effect of the decrease in subglottal pressure and/or the increase in glottal aperture that a speaker would use to lower sound level when gaining prosthetic hearing (cf. Stevens, 1998, pp. 80–82).

3.2.4. *The robustness of the internal model may depend on how mature it is prior to hearing loss*

Another aspect of the robustness of the internal model is addressed by our observations of two cochlear implant users who had lost hearing in childhood (data from one of the subjects, MB, appeared in Perkell *et al*., 1992). Informal observation indicated that their speech as adults was much less intelligible than those who became deaf as adults. This is compatible with evidence that long-term intelligibility after the onset of profound hearing loss is related to the age at which the loss occurs (Cowie & Douglas-Cowie, 1983). While the speech of the two early-deafened subjects changed with the acquisition of prosthetic hearing, many of those changes were not in the direction of normal (also see Plant & Oster, 1986). This sample of two cochlear implant users is much too small for firm conclusions, but the results indicate tentatively that if speech acquisition with hearing does not progress enough before the onset of profound hearing loss, the resulting internal model is not mature enough to be recalibrated readily with the acquisition of some prosthetic hearing in adulthood.

3.3. *Interrelations among the hypotheses: observations of the /s/-/ʃ/ contrast*

In the following subsections, we present examples of data and observations of the /s/-/ʃ/ contrast from subjects with normal hearing, CI users and the NF-2 patient. Taken together, these examples illustrate possible interrelations among phonemic acoustic goals, stability of the internal model, motor equivalence and saturation effects.

Fig. 12 is a schematic illustration of midsagittal sections through the anterior part of the oral cavity and the lips as they may be configured to produce the consonants /ʃ/ (on the left) and /s/ (on the right). An /s/ is produced by creating a short midsagittal groove in anterior tongue blade and positioning the blade against the dento-alveolar ridge to produce a narrow constriction. Turbulence noise is generated by airflow through the constriction; the noise is enhanced as the air stream strikes the lower incisors (Shadle, 1985). The noise excites the small resonant cavity anterior to the constriction, resulting in a spectrum that has energy concentrated in a high-frequency range, 3–6 kHz for males and 4.5–7 kHz for females. It has been observed that this articulation may include contact between the under side of the tongue blade and the lingual aspect of the mandibular alveolar ridge and incisors (Perkell, Boyce & Stevens, 1979). To produce a /ʃ/, the tongue blade is shaped to form a longer constriction, which is positioned somewhat further back along the anterior hard palate, so that a sublingual cavity may be formed. The result, in contrast to /s/, is a larger cavity anterior to the constriction, which contributes to the production of a lower spectral center of gravity for /ʃ/. The effect may be enhanced by some lip rounding, which further enlarges the anterior resonant cavity.

3.3.1. *The sibilant contrast is generally stable, even with prolonged hearing loss*

Matthies, Svirsky, Lane & Perkell (1994) measured spectral median and symmetry for /s/ and /ʃ/ produced in carrier phrases by five cochlear implant users. The measurements, which reflect acoustically and perceptually important differences between /s/ and /ʃ/ (also see Forrest, Weismer, Milenkovic & Dougall, 1988), were made preimplant, within

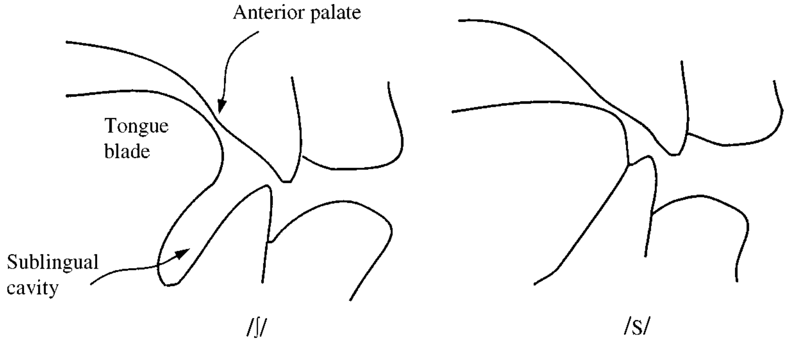


Figure 12. A schematic illustration of midsagittal sections through the anterior part of the oral cavity and the lips as they may be configured to produce the consonants /ʃ/ (on the left) and /s/ (on the right).

a few months after implant and 6 months postimplant. Preimplant, four of the five subjects had higher values of spectral median for /s/ than for /ʃ/, higher values of symmetry for /ʃ/ than /s/ and clear separation between the /s/ and /ʃ/ values. These results indicate a good distinction between the two consonants preimplant—even decades following the onset of profound deafness. On the other hand, the fifth subject had reversed values of the two measures preimplant, consistent with a perceptual impression that her sibilants were quite distorted. After months of implant use with auditory feedback, that subject's spectral median and symmetry values for the two sibilants were no longer reversed. The good preimplant distinctions between the sibilants in the other four subjects indicate that their segmental control mechanisms for the production of /s/ and /ʃ/ are generally quite resistant to change, even in the prolonged absence of auditory feedback.

3.3.2. *The contrast can collapse if the internal model becomes invalid*

The nature of the control of sibilant production is further illustrated by data from the NF-2 patient, FD (see above), who had tumor-removal surgery that severed her remaining auditory nerve (Perkell, Manzella, Wozniak, Matthies, Lane, Svirsky, Guiod, Delhorne, Short, MacCollin & Mitchell, 1995). At the time of surgery, she received an auditory brainstem implant;[6] although the implant was multi-channel, it effectively provided only auditory envelope and not spectral cues (Perkell *et al.*, 1995). Fig. 13 plots spectral median *vs.* week from the onset of hearing loss (OHL) for /s/ and /ʃ/. During the period before OHL and continuing for over 70 weeks post OHL, FD maintained a good contrast between the two sounds. At week 72, she had surgery to anastomose her left hypoglossal nerve to the facial nerve, in an attempt to restore some facial function that had also been lost at the time of tumor removal surgery. The anastomosis surgery denervated some tongue muscles on the left side, producing a slight tongue weakness that effectively altered a functional property of the vocal tract. Without auditory feedback about the sibilant contrast to help the control mechanism develop a compensatory adaptation to the tongue weakness, the contrast gradually collapsed. In terms of our theoretical framework, the anastomosis surgery invalidated her internal model by changing a characteristic of the low-level control of the "biomechanical plant" (the vocal tract in the block diagram in Fig. 1). Due to her deafness, it was then impossible to update the internal model by making auditorily based corrections to its parameter settings. Since people with normal hearing are capable of compensating for significant changes in vocal-tract morphology, we assume that if FD had adequate hearing, she would have been able to compensate for the surgery, even though it resulted in some slight tongue weakness.

3.3.3. *Some speakers demonstrate motor equivalence for /ʃ/*

Fig. 14 is a schematic illustration of a possible motor equivalent trading relation that can be used in the production of /ʃ/. The double-headed arrows indicate that if the tongue blade is positioned in a slightly anterior location, lip rounding may be increased and vice

[6]The auditory brainstem implant has been developed (and FD was implanted) at the House Ear Institute, Los Angeles, CA. It consists of an electrode array placed on the cochlear nucleus, trans-cutaneous electromagnetic signal transmission and an external microphone and signal processor. The electrode array has seven active electrodes and one reference electrode, forming seven channels that are stimulated with an F0 F1 F2 F5 strategy, intended to provide spectral, amplitude and temporal information, including voicing.

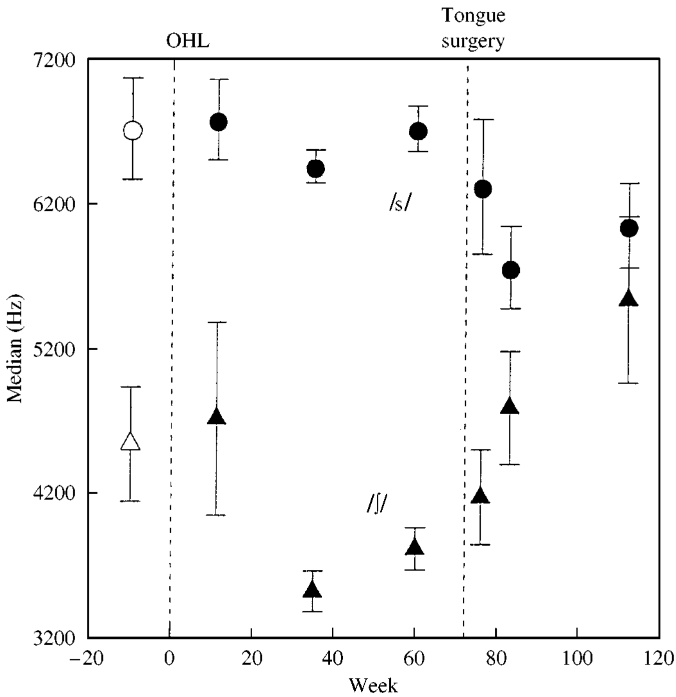


Figure 13. Spectral median (in Hz) for /s/ (circles) and /ʃ/ (triangles) *vs.* weeks after the onset of hearing loss (OHL) for the NF-2 subject, FD. The right-hand dotted line indicates the time that surgery was performed to anastomose the left hypoglossal nerve to the facial nerve.

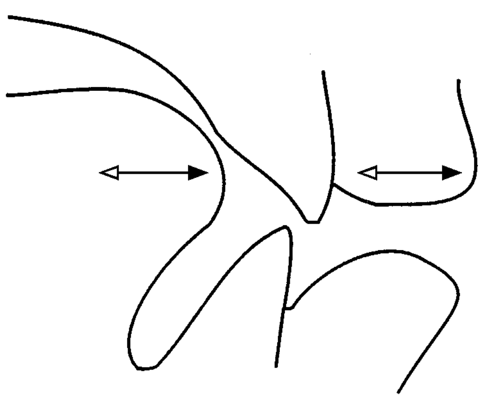


Figure 14. A schematic illustration of a possible motor equivalent trading relation that could be used in the production of /ʃ/. The double-headed arrows indicate that if the tongue blade is positioned in a slightly anterior location, lip rounding may be increased and vice versa.

versa. This behavior would serve to maintain the volume of the cavity anterior to the constriction and assure a low spectral center of gravity. Therefore, motor equivalence for /ʃ/ would be evidenced by positive correlations of tongue-blade fronting and upper-lip protrusion in multiple repetitions. We conducted an experiment with eight subjects in which we looked for such correlations. The results showed that when there were significant correlations, they were only positive, consistent with motor equivalence. However, there were individual differences, with many possible cases not showing motor equivalence (no significant correlation). Such individual differences may be related to whether or not speakers use a saturation effect with the lingual articulation to help maintain the /s/-/ʃ/ contrast.

Speakers who do not show motor equivalence of /ʃ/ to enhance its contrast with /s/ may use a saturation effect instead. Fig. 15 is a schematic illustration of how a saturation effect is hypothesized to enhance the contrast. The upper part of the figure shows a schematic relation between the horizontal tongue-blade location and the sublingual

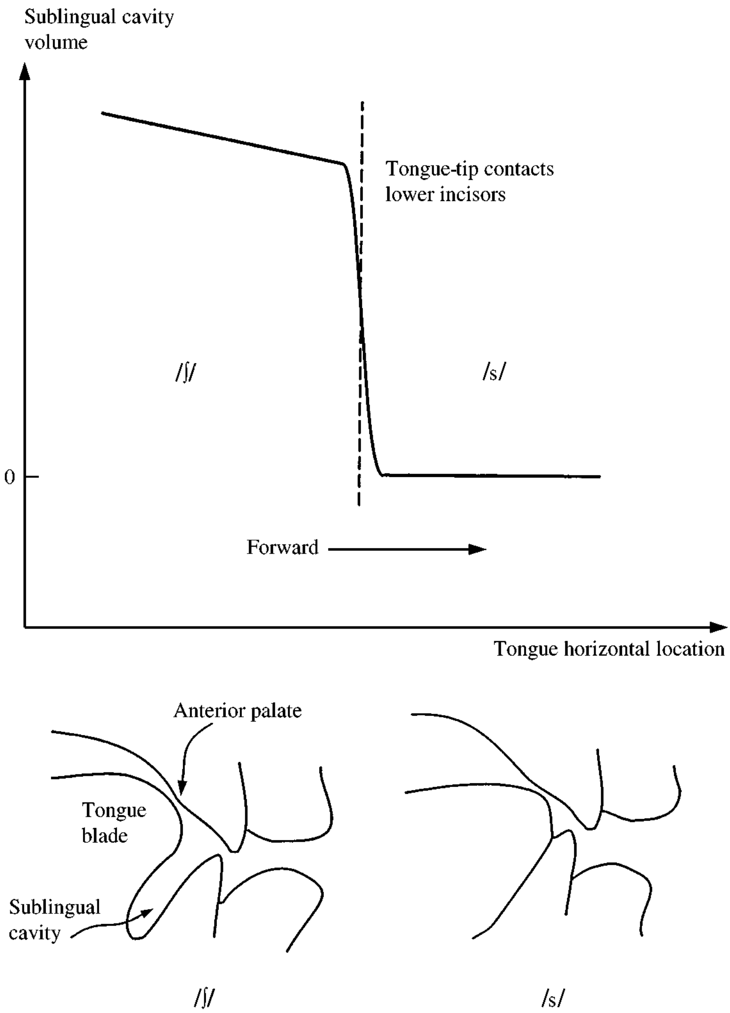


Figure 15. A schematic illustration of how a saturation effect is hypothesized to enhance the /s/-/ʃ/ contrast. The upper part of the figure shows a schematic relation between the horizontal tongue-blade location and the sublingual cavity volume. The lower part shows diagrams of articulations of the two sounds, as in Fig. 12.

cavity volume. Starting from a /ʃ/ articulation, as the tongue blade is moved continuously forward, the volume of the sublingual cavity decreases gradually and then suddenly saturates at zero when the tongue-tip contacts the lower incisors and /s/ is produced. Speakers who use such a mechanism by employing contact between the tongue blade and lower incisors for /s/ would not need to use motor equivalence for /ʃ/ to maintain the sibilant contrast since their contrast would be adequate without motor equivalence.

3.3.4. *Short-term drift in /ʃ/ production without hearing may be due to a speaker's vocal-tract morphology and his inability to use a saturation effect for /s/*

Fig. 16 shows data that illustrate unexpected short-term drift of the phonemic mechanism in a cochlear implant user. The left half of the figure plots spectral median for /ʃ/ (a phonemic parameter) and SPL for a following vowel (a postural parameter) versus time during an experiment with a male CI user, MC, in which his speech processor was on at the beginning and end of the experiment (500–1000 and 2000–2500 s) and off for two 500 ms segments in the middle of the experiment (1000–2000 s). The solid and dashed lines represent the same data with different amounts of smoothing, to show the rapidly varying fluctuations and more slowly varying trends (Matthies, Svirsky, Perkell & Lane, 1996). When MC's speech processor was turned off, vowel SPL increased rapidly; it remained high until the final "on" condition, when it dropped rapidly (consistent with findings presented for FB in Fig. 5 and Svirsky *et al.*, 1992). On the other hand, the sibilant spectral median did not change rapidly at the on-to-off transition. During the off condition, the /ʃ/ spectral median gradually drifted upward (toward /s/ values). It decreased abruptly at the off-to-on transition and then fluctuated somewhat. Further

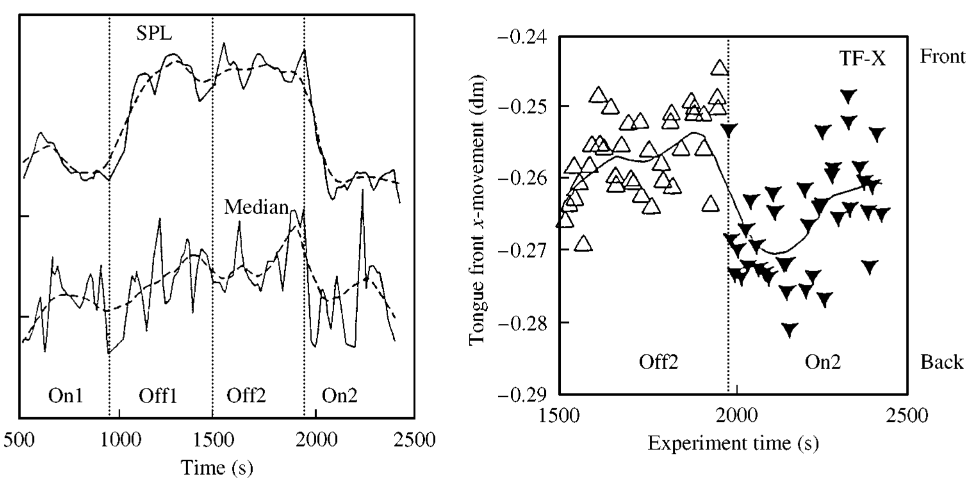


Figure 16. Left half: Values of spectral median (Hz) for /ʃ/ and SPL (dB) for a following vowel *vs.* time (s) during an experiment with a male cochlear implant user, MC, in which his speech processor was on at the beginning and end of the experiment (500–100 and 2000–2500 s) and off in the middle of the experiment (1000–2000 s). The solid and dashed lines show different degrees of smoothing of the same data. Right half: the horizontal position (in decimeters) of a movement transducer coil on MC's tongue-blade *vs.* time (in s) for repetitions of /ʃ/ in the second half of the experiment (1500–2500 s), spanning the off-to-on transition.

insight about the upward drift and abrupt decrease in the spectral median is gained from examining the underlying articulation.

The right half of Fig. 16 shows the horizontal position of a movement transducer coil on the speaker's tongue blade *vs.* time for repetitions of /ʃ/ during the second half of the experiment, spanning the off-to-on transition. The vertical axis shows forward tongue-blade displacement in decimeters, and the horizontal axis shows time in seconds. The solid line through the data represents about the same amount of smoothing as the dashed lines in the left half of the figure; it shows the more slowly varying trend. The gradual forward movement during the off condition presumably is responsible for the upward drift of the spectral median. The first /ʃ/ produced after the speech processor is turned on follows the trend from the preceding off condition (filled symbol on dotted line); however, by the next repetition, the subject seems to have detected the acoustic distortion and has pulled his tongue back by 2 mm, in an apparent overcompensation for the detected error. Following this overcompensation, there is again some forward movement in what might be an attempt to re-calibrate the articulation.

What could account for this drift of a presumably stable phonemic parameter? Observation of this speaker's dental cast revealed that his dental arches were the narrowest of the approximately 100 we have examined. His extremely narrow dental arches could have made the consonant acoustics especially sensitive to small articulatory changes. It may be difficult for him to produce an /s/ with contact between the tongue blade and lower incisors and to produce the /ʃ/ articulation (as described above) without fairly close auditory monitoring. Without a reliable saturation effect and without the help of auditory feedback for maintenance purposes when his speech processor was off (possibly in combination with the perturbing presence of the movement transducer coil on his tongue blade), the segmental control parameters degraded slightly and caused the unexpectedly rapid articulatory and acoustic drift of a segmental parameter that generally is resistant to change (Matthies *et al.*, 1996).

3.3.5. *Saturation effects may mask changes to the internal model for segmental control*

In the theoretical overview, we suggested that the existence of saturation effects may mask some degradation of the internal model for segmental control. In other words, saturation effects may be partly responsible for the stable nature of some phonemic mechanisms. As suggested above, the long-term stability seen for the sibilant contrast in most of the cochlear implant subjects may be aided by the use of a saturation effect. As another (hypothetical) example, it should be easy to maintain the production of a bilabial stop even with some degradation of the internal model from prolonged deafness, if the controller's main goals for this consonant are a relatively low sound level (achieved simply with lip closure) in combination with a region of formant values that can only be targeted with lip closing movements. Information about how such clearly quantal goals are produced is represented in the internal model; this information should be less susceptible to degradation than knowledge about articulatory-to-acoustic relations that involve more graded contrasts. When saturation effects are weak or lacking, there is more of a tendency for segmental changes to occur in the absence of auditory feedback, especially with perturbing influences to articulation as suggested above for MC's production of /ʃ/.

Along analogous lines, we observed earlier that FB's vowel F_2 values (a phonemic parameter), especially for /ɛ/, were abnormal preimplant, and they improved gradually postimplant—in parallel with improvement in her vowel recognition scores. These observations are consistent with the idea that the production of /ɛ/ involves a graded contrast that is not constrained by obvious acoustic or articulatory saturation effects (see also Svirsky & Tobey, 1991). FB's F_2 values for /ɛ/ also changed rapidly in an on–off experiment (Svirsky *et al*., 1992). Along the lines suggested in the preceding paragraph, we speculate that when we observe rapid changes of segmental parameters in on–off experiments, we are more likely find them for graded contrasts that are not strongly constrained by saturation effects. Further, as suggested in the following example, such rapid segmental changes may be the result of the subject's alternating or drifting between two sets of parameter values of the internal model.

3.3.6. *There can be more than one set of parameter settings of the internal model*

In a refinement of the "on-off" paradigm described above in Section 3.2.1, we investigated the speed of changes in speech production parameters by seven cochlear implant users, in response to switching the speech processors of their implants on and off a number of times in a single experimental session (Perkell, Zandipour, Vick, Matthies, Lane & Gould, submitted). The subjects repeated short utterances ("a said", "a shed", "a sad", "a shad") many times in semi-random order. The transitions between hearing (on) and nonhearing (off) states were introduced in the silent intervals between utterances; the number of utterances between transitions was varied to minimize subject anticipation of the transition. Using the times of on–off or off–on transitions as line-up points, values of median and symmetry of sibilant spectra and vowel F_1, F_2, duration and SPL were each averaged over repetitions of each utterance and compared across the transitions, separately for the on-to-off and off-to-on transitions. The speakers' vowel SPL and duration had changed reliably by the first sound in the utterance immediately following the transition, indicating that they were using subtle aspects of the processor output to detect its state even before they spoke.

Reliable changes in vowel formant values and sibilant spectra were less prevalent, but just as immediate. Without hearing, SPL was greater and durations longer but phonetic contrasts (/s/-/ʃ/, /ɛ/-/æ/) were reduced. Thus, the changes in the postural variables, of a kind that would enhance intelligibility, may have offset the presumed loss of intelligibility due to reduced segmental contrasts. This trading relation and the speed of the changes in both kinds of variables suggests that they are part of the same rapidly-acting mechanism, one that attempts to maintain intelligibility in the face of changing acoustic transmission conditions. We speculate that the speakers were switching between two sets of parameter settings of the internal model. One set, associated with no auditory input, had resulted from gradual deterioration of the original, hearing parameter settings after years of deafness. The other set, associated with auditory input, had resulted from the use of prosthetic hearing (which is very different from normal hearing) to recalibrate the original settings that were established during normal speech and language acquisition (also see Houde & Jordan, 1998).

3.4. *Summary*

In this section, examples of motor equivalence for the production of /u/ and /r/ were used to support the idea that acoustic goals are the basic programming units in speech motor control. Articulatory data for production of the vowel /i/ were used to illustrate the idea that some acoustic goals are defined by saturation effects. Since temporal factors preclude the use of auditory feedback closed loop for achieving acoustic goals, it was hypothesized that the control scheme also includes the use of a robust internal model of the relation between articulatory states and the resulting acoustics. The apparent robustness of the internal model for speech sounds that are defined by saturation effects may be due partly to the fact that the saturation effects allow for imprecision or drift in motor commands. Data from cochlear implant users and an NF-2 patient who experienced changes in hearing status were interpreted as evidence for two roles of auditory feedback: (1) to maintain phonemic settings of the internal model, and (2) to monitor acoustic transmission conditions and guide changes in postural parameters that are needed to assure intelligibility. It was suggested that speech sounds such as the vowels /ɛ/ and /æ/ that are not characterized by saturation effects may be more labile with respect to hearing status, but, in general, stability characterizes most sounds.

Several observations from the /s/-/ʃ/ contrast were used to illustrate inter-relations among these ideas. The contrast was shown to be generally stable, even in people who had experienced years of profound hearing loss. It was shown that the contrast can collapse if the internal model becomes invalid (e.g., due to surgery) and hearing is not available to maintain it. Inter-subject differences were observed in motor equivalence findings for /ʃ/; these may be due to whether or not speakers use a hypothesized saturation effect in producing /s/. Rapid drift was shown in measures of /ʃ/ production by a CI user when his implant was turned off. This finding may be explained by the speaker's very narrow dental arches, which could make it difficult for him to maintain the contrast without close auditory monitoring. Finally, several CI users exhibited very rapid changes in spectral median and symmetry for /ʃ/ with short-term changes in hearing status in an on–off experiment. The changes indicated more canonical productions of /ʃ/ and therefore greater contrast with /s/ with hearing than without. The rapidity of the changes indicate that for some speakers, there may be two sets of parameter settings of the internal model, a hearing set, associated with greater phonemic contrast, and a nonhearing set, associated with diminished contrast. In combination, these results lead to the speculation that the stability of the /s/-/ʃ/ contrast may vary among speakers for a variety of reasons, including differences in anatomy and the usefulness of amount of experience with prosthetic hearing. In some speakers, such as those who use a saturation effect for producing /s/, the contrast may be stable for years without the need for auditory monitoring; other speakers, such as people with atypical oral morphology, may use auditory feedback to closely monitor and maintain the contrast.

4. Discussion: results from other studies

The results from additional studies provide further support for the views described above. These studies include motor equivalence experiments, the use of steady-state perturbations such as bite blocks and artificial palates, a recent set of experiments on "sensorimotor adaptation" in speech production, and some additional examples from cochlear implant users.

Findings of some other motor-equivalence studies are also consistent with the idea that acoustic goals are used in speech motor programming. Based on cineradiographic data of the vowels /i/ and /a/ as pronounced by two speakers of French, Maeda (1991) found complementary contributions (i.e., trading relations) of relatively independent (horizontal) tongue and (vertical) jaw displacements to each of the vowel productions. Savariaux, Perrier & Orliaguet (1995) studied production of the vowel /u/ using X-ray (articulatory) and acoustic measurements, and a lip tube to impose an abnormally large labial cross-sectional area in 11 speakers of French. Seven of the subjects produced rearward tongue movements that partially or completely compensated for the perturbation; most of the compensation was evident in the subjects' first attempts. These results were interpreted as evidence that the speakers' behavior was "... directed toward an enhancement of the acoustic end product" (Savariaux *et al.*, 1995, p. 2440). De Jong (1997) found some trading relations between tongue and lip movements in the production of the rounded back vowels /ʊ/ and /o/ by three speakers of American English; however, the results were complicated and differed among the subjects. Thus, while there are several findings consistent with the use of acoustic goals, inter-subject differences should be explored further.

A number of experiments have been performed in which bite blocks or artificial palates have been used to induce "steady-state" compensatory articulations. When considered collectively, the results of these experiments indicate that speakers are aiming to achieve acoustic goals and that achieving those goals may be aided by the use of saturation effects. These studies have examined compensatory adjustments in the productions of vowels (most often the point vowels /i/, /ɑ/ and /u/) and consonants (most often /t/ and /s/) after immobilization of the mandible at unusual degrees of opening by insertion of a bite block (cf. Lindblom, Lubker & Gay, 1979; Fowler & Turvey, 1980; McFarland & Baum, 1995) or after insertion of a thick artificial palate (Hamlet & Stone, 1976; Flege, Fletcher & Homiedan, 1988; McFarland, Baum & Chabot, 1996). For the most part, rather good compensations have been observed, indicating that speakers readily use unaccustomed patterns of muscle contraction to achieve phonemically appropriate sounds. In the earliest studies it was thought that compensations (of vowels to a bite block) were achieved immediately, and therefore without the benefit of auditory feedback (cf. Lindblom *et al.*, 1979; Fowler & Turvey, 1980). However, other studies have revealed that such compensations are not complete, and that they improve with practice (Hamlet & Stone, 1976; Flege *et al.*, 1988; McFarland & Baum, 1995; McFarland *et al.*, 1996). It has also been found that practice may help with compensation for kinematic perturbations (Munhall, Löfqvist & Kelso, 1994).

In terms of our theoretical framework, the reasonably good (but imperfect) initial compensations would be accomplished with the help of saturation effects, along with an internal model that is robust but does not cover the perturbing conditions of the experiment. The improvement of the compensation with practice would be due to the use of auditory feedback to refine the internal model's parameters under the perturbing conditions.

It follows from the above reasoning (and the examples of labile /ʃ/ and /ɛ/ production in the absence of auditory feedback) that differences in the strength of saturation effects may make some sounds more vulnerable to changes in auditory feedback than others. Svirsky & Tobey (1991—referred to earlier) performed an "on–off" study of vowel production with two cochlear implant users (similar to the one we conducted described in Section 3.2.1). Our overview generally agrees with Svirsky and Tobey's interpretation of their results: the central vowels, which are not characterized by strong saturation effects, may be more vulnerable to deprivation of auditory spectral information.

For the most part, however, phonemic instability is the exception. There were relatively few abnormalities of phonemic parameters in the four postlingually deafened speakers examined by Perkell *et al.* (1992) and the larger study by Cowie & Douglas-Cowie (1992). And, as observed above, most postlingually deafened speakers' sibilant contrasts were good, even before they received cochlear implants. Among our subjects, F_2 abnormalities were rare and the one corrected with implant use involved FB's more central vowels, /ɛ/ and /æ/, which hypothetically lack strong saturation effects.

Properties of the internal model have also been revealed in a set of "sensorimotor adaptation" experiments, in which Houde (1997) and Houde & Jordan (1998) observed changes in the productions of vowels whispered by subjects whose auditory feedback had been altered. The feedback alteration shifted the first two formants along the /i, ɪ, ɛ, æ, ɑ/ axis and had the effect of changing vowel identity. The subjects were unaware of the shifts, which were introduced gradually over a number of trials. Subjects compensated for the formant shift in their auditory feedback by producing a shift in the opposite direction. The amount of compensation was, on average, about half the amount of the feedback shift. For example, when subjects were initially pronouncing "bed" and then gradually were fed back an increasingly "bead"-like acoustic result, they ended up pronouncing "bad". It was found that the effect generalized to different consonant environments and to different vowels. For example, training with "get" yielded compensations in test words "geck", "guess", "debt", and "pet" and also in test words "geet", "git", "gat" and "got". These results are consistent with the notion that speakers are using phonemic acoustic goals in combination with an internal model. Within such a framework, the results show that parameters of the internal model can be changed by modifying auditory feedback.

The issue of adaptation was addressed in a different way by the Hamlet and Stone (1976) study, in which it was found that adaptation of vowel productions continued for 2 weeks in subjects with a thick artificial palate in place. Re-adaptation was examined in one subject, who compensated much more rapidly to re-insertion of the palate a month later than during the initial 2-week adaptation period (Hamlet *et al*., 1976). Houde (1997) re-tested his sensorimotor adaptation subjects 1 month after the initial adaptation session. When simply hearing the feedback signal without any formant alterations, the subjects immediately produced small vowel modifications in the same direction as in the initial session a month previously. Both of these findings are compatible with the idea that an alternative set of values of articulatory-to-acoustic mappings was acquired (as part of the internal model) during the adaptation period and was retained for later recall. As suggested above, when cochlear implant users make rapid modifications of phonemic parameters in on–off experiments (as with the above examples of /ɛ/), they may be switching back and forth between two sets of parameter values of the internal model, one set from when they lacked hearing, and a later set, learned with the help of auditory input from the implant.

5. Summary and conclusions

The data cited above, along with observations by other researchers, are consistent with the following hypotheses. Motor programming for the production of speech segments is based on acoustic goals. An internal model of the relation between vocal-tract shape and the sound output is acquired and then maintained with the help of auditory feedback. Auditory feedback is also used to monitor acoustic transmission conditions to help make situation-dependent, rapid adjustments in postural parameters underlying average sound level, rate, f_0 and prosodic inflection of f_0 and SPL—to assure adequate intelligibility. Segmental mechanisms are usually stable in the absence of auditory feedback, but this may depend partly on the availability of saturation effects that help to determine phonemic categories. Such stability may also depend on individual factors, such as how well established a recalibrated internal model has become from experience with a cochlear implant (which can be influenced by the quality of the prosthetic hearing) or anatomical determinants of saturation effects—which influence the difficulty of parameter maintenance under adverse conditions, such as perturbations to articulation.

We caution that the chosen examples of data from CI users and NF-2 patients are ones that most clearly illustrate our points. Generally, such data are very variable, within and across subjects. Studies of speakers with normal hearing almost always show inter-speaker differences as well. However, we have not found any examples of data that are clearly inconsistent with the theoretical overview presented here. Results from additional subjects, along with more deeply probing paradigms that further address inter-subject differences, should help to refine the theory and make future conclusions less speculative.

As pointed out in the introduction, data from speakers with hearing disorders and from speakers with normal hearing have proved crucial to the development of our theoretical framework. For refinement of this theory, a number of issues need to be explored further. For example, we would like to understand how individual differences in vocal-tract anatomy contribute to the role of saturation effects in articulatory-to-acoustic relationships. We are also working to further the understanding of the complex interactions among the perceptual, respiratory, phonatory, and articulatory subsystems that are reflected in the data from cochlear implant users. These areas of research have potential clinical as well as theoretical application. For example, individuals with hearing loss and individuals who have multiple production deficits represent challenging clinical cases, in which the interaction among speech production subsystems—including individual manifestations of saturation effects—could affect therapy designs and outcomes. Perhaps most importantly, it can be argued that the most effective clinical advances will be based on a sound theoretical account of normal speech production and its disorders.

In summary, we believe that the examples in this paper represent only a few of the many possible intersections of speech science and communication disorders. The need to explain and inter-relate our sometimes puzzling results from subjects with hearing disorders and from those with normal hearing, as well as the need to give an account of inter-subject differences in both populations, have moved our thinking into new areas. This process has proved to be invaluable in helping the theory to evolve and it should have potential for new clinical approaches.

This work was supported by grants DC01925, DC01291, DC02525, DC02852 and DC00361 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We thank Mario Svirsky (currently at the Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN) and Jane Wozniak for their many valuable contributions to this work, John Gould for his help with the manuscript, and Jay Moody and Sridhar Kalluri for their helpful comments. We are especially grateful to the cochlear implant users and NF-2 patients who have served as our subjects.

References

Atal, B. S., Chang, J. J., Mathews, M. V. & Tukey, J. W. (1978) Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique, *Journal of the Acoustical Society of America*, **63,** 1535–1555.

Badin, P., Perrier, P., Boë, L. J. & Abry, C. (1990) Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences, *Journal of the Acoustical Society of America*, **87,** 1290–1300.
Beckman, M. E., Jung, T., Lee, S., de Jong, K. J., Krishnamurthy, A. K., Ahalt, S. C., Cohen, K. B. & Collins, M. J. (1995) Variability in the production of quantal vowels revisited, *Journal of the Acoustical Society of America*, **97,** 471–490.
Browman, C. & Goldstein, L. (1989) Articulatory gestures as phonological units, *Phonology*, **6,** 201–251.
Cowie, R. I. & Douglas-Cowie, E. (1983) Speech production in profound post-lingual deafness. In *Hearing science and hearing disorders* (M. E. Lutman & M. P. Haggard, editors), pp. 183–231. New York: Academic Press.
Cowie, R. I. & Douglas-Cowie, E. (1992) *Postlingually acquired deafness: speech deterioration and the wider consequences*. New York: Mouton de Gruyter.
de Jong, K. J. (1997) Labiovelar compensation in back vowels, *Journal of the Acoustical Society of America*, **101,** 2221–2233.
Delattre, P. C., Liberman, A. M. & Cooper, F. S. (1955) Acoustic loci and transitional cues for consonants, *Journal of the Acoustical Society of America*, **27,** 769–773.
Fairbanks, G. (1960) *Voice and articulation drill book*. New York: Holt, Rinehart & Winston.
Fairbanks, G. & Grubb, P. (1961) A psychophysical investigation of vowel formants, *Journal of Speech and Hearing Research*, **4,** 203–219.
Flege, J. E., Fletcher, S. G. & Homiedan, A. (1988) Compensating for a bite block in /s/ and /t/ production: palatographic, acoustic, and perceptual data, *Journal of the Acoustical Society of America*, **83,** 212–228.
Forrest, K., Weismer, G., Milenkovic, P. & Dougall, R. N. (1988) Statistical analysis of word-initial voiceless obstruents: preliminary data, *Journal of the Acoustical Society of America*, **84,** 115–123.
Fowler, C. A. (1980) Coarticulation and theories of extrinsic timing, *Journal of Phonetics*, **8,** 113–133.
Fowler, C. A. & Turvey, M. T. (1980) Immediate compensation in bite-block speech, *Phonetica*, **37,** 306–326.
Fujimura, O. & Kakita, Y. (1979) Remarks on quantitative description of the lingual articulation. In *Frontiers of speech communication research* (B. Lindblom & S. Öhman, editors), pp. 17–24. New York: Academic Press.
Guenther, F. H. (1992) *Neural models of adaptive sensory-motor control for flexible reaching and speaking*. Doctoral dissertation, Boston University, Boston.
Guenther, F. H. (1994) A neural network model of speech acquisition and motor equivalent speech production, *Biological Cybernetics*, **72,** 43–53.
Guenther, F. H. (1995*a*) Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production, *Psychological Review*, **102,** 594–621.
Guenther, F. H. (1995*b*) A modeling framework for speech motor development and kinematic articulator control, *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 2, 92–99.
Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M. & Perkell, J. S. (1999) Articulatory tradeoffs reduce acoustic variability during American English /r/ production, *Journal of the Acoustical Society of America*, **105,** 2854–2865.
Guenther, F. H., Hampson, M. & Johnson, D. (1998) A theoretical investigation of reference frames for the planning of speech movements, *Psychological Review*, **105,** 611–633.
Halle, M. & Stevens, K. N. (1990) Knowledge of language and the sounds of speech. In *Music, language, speech and brain* (J. Sundberg, L. Nord & R. Carlson, editors), pp. 1–19. Basingstoke, Hampshire: MacMillan.
Hamlet, S. L. & Stone, M. L. (1976) Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses, *Journal of Phonetics*, **4,** 199–218.
Hamlet, S. L., Stone, M. L. & McCarthy, T. (1976) Persistence of learned motor patterns in speech, *Journal of the Acoustical Society of America*, **60,** S66 (A).
Hodgson, P. & Miller, J. L. (1996) Internal structure of phonetic categories: evidence for within-category trading relations, *Journal of the Acoustical Society of America*, **100,** 565–576.
Houde, J. F. (1997) *Sensorimotor adaptation in speech production*. Doctoral dissertation, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA.
Houde, J. F. & Jordan, M. I. (1998) Sensorimotor adaptation in speech production, *Science*, **279,** 1213–1216.
Iverson, P. & Kuhl, P. K. (1995) Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling, *Journal of the Acoustical Society of America*, **97,** 553–562.
Iverson, P. & Kuhl, P. K. (1996) Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/, *Journal of the Acoustical Society of America*, **99,** 1130–1140.
Jakobson, R., Halle, M. & Fant, G. (1969) *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.
Jenkins, J. J., Strange, W. & Trent, S. A. (1999) Context-independent dynamic information for the perception of coarticulated vowels, *Journal of the Acoustical Society of America*, **106,** 438–448.
Jordan, M. I. (1990) Motor learning and the degrees of freedom problem. In *Attention and performance XIII* (M. Jeannerod, editor), pp. 796–836. Hillsdale, NJ: Lawrence Erlbaum.
Jordan, M. I. (1996) Computational aspects of motor control and motor learning. In *Handbook of perception and action* (H. Heuer & S. Keele, editors), Vol. 2, pp. 71–120. New York: Academic Press.
Jordan, M. I. & Rumelhart, D. E. (1992) Forward models: supervised learning with a distal teacher, *Cognitive Science*, **16,** 307–354.

Kawahara, H. & Williams, J. C. (1996) Effects of auditory feedback on voice pitch trajectories: characteristic responses to pitch perturbations. In *Vocal fold physiology* (P. J. Davis & N. H. Fletcher, editors), pp. 263–278. San Diego: Singular.

Kawato, M., Furukawa, K. & Suzuki, R. (1987) A hierarchical neural-network model for control and learning of voluntary movement, *Biological Cybernetics*, **57,** 169–185.

Keating, P. A. (1988) The window model of coarticulation: articulatory evidence, *UCLA Working Papers in Phonetics*, **69,** 3–29.

Kent, R. D. (1976) Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies, *Journal of Speech and Hearing Research*, **19,** 421–447.

Kingston, J., Macmillan, N. A., Dickey, L. W., Thorburn, R. & Bartels, C. (1997) Integrality in the perception of tongue root position and voice quality in vowels, *Journal of the Acoustical Society of America*, **101,** 1696–1709.

Klatt, D. H. & Klatt, L. C. (1990) Analysis, synthesis and perception of voice quality variations among male and female talkers, *Journal of the Acoustical Society of America*, **87,** 820–857.

Kuhl, P. K. (1991) Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not, *Perception and Psychophysics*, **50,** 93–107.

Laboissière, R., Ostry, D. J. & Feldman, A. G. (1996) The control of multi-muscle systems: human jaw and hyoid movements, *Biological Cybernetics*, **74,** 373–384.

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

Lane, H., Perkell, J., Wozniak, J., Manzella, J., Guiod, P., Matthies, M., MacCollin, M. & Vick, J. (1998) The effect of changes in hearing status on speech sound level and speech breathing: a study conducted with cochlear implant users and NF-2 patients, *Journal of the Acoustical Society of America*, **104,** 3059–3069.

Lane, H. & Tranel, B. (1971) The Lombard sign and the role of hearing in speech, *Journal of Speech and Hearing Research*, **14,** 677–709.

Lane, H. & Webster, J. (1991) Speech deterioration in postlingually deafened adults, *Journal of the Acoustical Society of America*, **89,** 859–866.

Lane, H., Wozniak, J., Matthies, M. L., Svirsky, M. A. & Perkell, J. S. (1995) Phonemic resetting vs. postural adjustments in the speech of cochlear implant users: an exploration of voice onset time, *Journal of the Acoustical Society of America*, **98,** 3096–3106.

Lane, H., Wozniak, J., Matthies, M. L., Svirsky, M. A., Perkell, J. S., O'Connell, M. & Manzella, J. (1997) Changes in sound pressure and fundamental frequency contours following changes in hearing status, *Journal of the Acoustical Society of America*, **101,** 2244–2252.

Lehiste, I. & Peterson, G. E. (1959) Vowel amplitude and phonemic stress in American English, *Journal of the Acoustical Society of America*, **31,** 428–435.

Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. (1957) The discrimination of speech sounds within and across phoneme boundaries, *Journal of Experimental Psychology*, **54,** 358–368.

Liberman, A. M., Harris, K. S., Kinney, J. & Lane, H. (1961) The discrimination of relative onset-time of the components of certain speech and nonspeech patterns, *Journal of Experimental Psychology*, **61,** 379–388.

Lindblom, B. & Engstrand, O. (1989) In what sense is speech quantal?, *Journal of Phonetics*, **17,** 107–121.

Lindblom, B., Lubker, J. & Gay, T. (1979) Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation, *Journal of Phonetics*, **7,** 147–161.

Macmillan, N. A., Kingston, J., Thorburn, R., Dickey, L. W. & Bartels, C. (1999) Integrality of nasalization and F1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision-rule interactions, *Journal of the Acoustical Society of America*, **106,** 2913–2932.

Maeda, S. (1991) On articulatory and acoustic variabilities, *Journal of Phonetics*, **19,** 321–331.

Manuel, S. (1990) The role of contrast in limiting vowel-to-vowel coarticulation in different languages, *Journal of the Acoustical Society of America*, **88,** 1286–1298.

Matthies, M. L., Svirsky, M. A., Lane, H. & Perkell, J. S. (1994) A preliminary study of the effects of cochlear implants on the production of sibilants, *Journal of the Acoustical Society of America*, **96,** 1367–1373.

Matthies, M. L., Svirsky, M. A., Perkell, J. S. & Lane, H. (1996) Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant, *Journal of Speech and Hearing Research*, **39,** 936–946.

McFarland, D. H. & Baum, S. R. (1995) Incomplete compensation to articulatory perturbation, *Journal of the Acoustical Society of America*, **97,** 1865–1873.

McFarland, D. H., Baum, S. R. & Chabot, C. (1996) Speech compensation to structural modifications of the oral cavity, *Journal of the Acoustical Society of America*, **100,** 1093–1104.

Mermelstein, P. (1973) Articulatory model for the study of speech production, *Journal of the Acoustical Society of America*, **53,** 1070–1082.

Miller, J. D. (1989) Auditory-perceptual interpretation of the vowel, *Journal of the Acoustical Society of America*, **85,** 2114–2134.

Munhall, K. G., Löfqvist, A. & Kelso, J. A. S. (1994) Lip–larynx coordination in speech: effects of mechanical perturbations to the lower lip, *Journal of the Acoustical Society of America*, **95,** 3605–3616.

Nearey T. M. (1995) Evidence for the perceptual relevance of vowel-inherent spectral change for front vowels in Canadian English, *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 2, pp. 678–681.
Nearey, T. M. & Assmann, P. F. (1986) Modeling the role of inherent spectral change in vowel identification, *Journal of the Acoustical Society of America*, **80,** 1297–1308.
Nelson, W. L. (1983) Physical principles for economies of skilled movements, *Biological Cybernetics*, **46,** 135–147.
Ostry, D. J., Gribble, P. L. & Gracco, V. L. (1996) Coarticulation of jaw movements in speech production: is context sensitivity in speech kinematics centrally planned?, *Journal of Neuroscience*, **16,** 1570–1579.
Parry, D. M., Eldridge, R., Kaiser-Kupfer, M. I., Bouzas, E. A., Pikus, A. & Patronas, N. (1994) Neurofibromatosis 2 (NF-2): clinical characteristics of 63 affected individuals and clinical evidence of heterogeneity, *American Journal of Medical Genetics*, **52,** 450–461.
Payan, Y. & Perrier, P. (1997) Synthesis of V–V sequences with a 2D biomechanical model of the tongue controlled by Equilibrium Point Hypothesis, *Speech Communication*, **22,** 185–205.
Perkell, J. S. (1996) Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modeling, *Journal of Phonetics*, **24,** 3–22.
Perkell, J. S. (1997) Articulatory processes. In *The handbook of phonetic sciences* (W. Hardcastle & J. Laver, editors), pp. 333–370. Cambridge, MA: Blackwell.
Perkell, J. S., Boyce, S. E. & Stevens, K. N. (1979) Articulatory and acoustic correlates of the [s-ʃ] distinction. In *Speech communication papers presented at the 97th meeting of the acoustical society of America* (J. J. Wolf & D. H. Klatt, editors), pp. 109–113. New York: Acoustical Society of America.
Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I. & Jackson, M. (1992) Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements, *Journal of the Acoustical Society of America*, **92,** 3078–3096
Perkell, J. S., Lane, H., Svirsky, M. A. & Webster, J. (1992) Speech of cochlear implant patients: a longitudinal study of vowel production, *Journal of the Acoustical Society of America*, **91,** 2961–2979.
Perkell, J. S., Manzella, J., Wozniak, J., Matthies, M. L., Lane, H., Svirsky, M. A., Guiod, P., Delhorne, L., Short, P., MacCollin, M. & Mitchell, C. (1995) Changes in speech production following hearing loss due to bilateral acoustic neuromas, *Proceedings of the XIIIth International Congress of Phonetic Sciences*, **3,** 194–197.
Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F. H., Wilhelms-Tricarico, R., Wozniak, J. & Guiod, P. (1997) Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models, *Speech Communication*, **22,** 227–250.
Perkell, J. S., Matthies, M. L. & Svirsky, M. A. (1994) Articulatory evidence for acoustic goals for consonants, *Journal of the Acoustical Society of America*, **96,** 3326 (A).
Perkell, J. S., Matthies, M. L., Svirsky, M. A. & Jordan, M. I. (1993) Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: a pilot motor equivalence study, *Journal of the Acoustical Society of America*, **93,** 2948–2961.
Perkell, J. S., Matthies, M. L., Svirsky, M. A. & Jordan, M. I. (1995) Goal-based speech motor control: a theoretical framework and some preliminary data, *Journal of Phonetics*, **23,** 23–35.
Perkell, J. S. & Nelson, W. L. (1985) Variability in production of the vowels /i/ and /a/, *Journal of the Acoustical Society of America*, **77,** 1889–1895.
Perkell, J., Zandipour, M., Vick, J., Matthies, M., Lane, H. & Gould, J. (submitted) Rapid changes in speech production parameters in response to a change in hearing, *Journal of Phonetics*.
Perrier, P., Ostry, D. J. & Laboissière, R. (1996) The equilibrium point hypothesis and its application to speech motor control, *Journal of Speech and Hearing Research*, **39,** 365–378.
Peterson, G. E. & Barney H. L. (1952) Control methods used in a study of the vowels, *Journal of the Acoustical Society of America*, **24,** 175–184.
Picheny, M. A., Durlach, N. I. & Braida, L. D. (1986) Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech, *Journal of Speech and Hearing Research*, **29,** 434–446.
Pisoni, D. (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels, *Perception and Psychophysics*, **13,** 253–260.
Plant, G. & Oster, A. (1986) The effects of cochlear implantation on speech production. *Quarterly Progress and Status Report*, Speech Transmission Lab., Stockholm, **1/1986,** 65–84.
Saltzman, E. L. & Munhall, K. G. (1989) A dynamical approach to gestural patterning in speech production, *Ecological Psychology*, **1,** 333–382.
Savariaux, C., Perrier, P. & Orliaguet, J. P. (1995) Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: a study of the control space in speech production, *Journal of the Acoustical Society of America*, **98,** 2428–2842.
Savariaux, C., Perrier, P., Orliaguet, J. P. & Schwartz, J. L. (1999) Compensation strategies for the perturbation of French [u] using a lip tube: II. Perceptual analysis, *Journal of the Acoustical Society of America*, **106,** 381–393.
Schwartz, J. L., Boë, J. L., Vallée, N. & Abry, C. (1997*a*) The dispersion–focalization theory of vowel systems, *Journal of Phonetics*, **25,** 255–286.

Schwartz, J. L., Boë, J. L., Vallée, N. & Abry, C. (1997*b*) Major trends in vowel system inventories, *Journal of Phonetics*, **25,** 233–254.
Shadle, C. (1985) *The acoustics of fricative consonants*. Research Laboratory of Electronics, Technical Report 506, Massachusetts Institute of Technology, Cambridge, MA.
Stevens, K. N. (1972) The quantal nature of speech: evidence from articulatory-acoustic data. In *Human communication—a unified view* (P. B. Denes & E. E. David, editors), pp. 51–56. New York: McGraw-Hill.
Stevens, K. N. (1989) On the quantal nature of speech, *Journal of Phonetics*, **17,** 3–45.
Stevens, K. N. (1998) *Acoustics phonetics*. Cambridge, MA: MIT Press.
Strange, W. & Bohn, O.-S. (1998) Dynamic specification of coarticulated German vowels: perceptual and acoustical studies, *Journal of the Acoustical Society of America*, **104,** 488–504.
Svirsky, M. A., Lane, H., Perkell, J. S. & Webster, J. (1992) Speech of cochlear implant patients: results of a short-term auditory deprivation study, *Journal of the Acoustical Society of America*, **92,** 1284–1300.
Svirsky, M. A. & Tobey, E. A. (1991) Effect of different types of auditory stimulation on formant frequencies in multichannel cochlear implant users, *Journal of the Acoustical Society of America*, **89,** 2895–2904.
Volaitis, L. E. & Miller, J. L. (1992) Phonetic prototypes: influence of place of articulation and speaking rate on the internal structure of voicing categories, *Journal of the Acoustical Society of America*, **92,** 723–735.
Wilhelms-Tricarico, R. (1995) Physiological modeling of speech production: methods for modeling soft-tissue articulators, *Journal of the Acoustical Society of America*, **97,** 3085–3098.
Wilhelms-Tricarico, R. (1996) A biomechanical and physiologically-based vocal tract model and its control, *Journal of Phonetics*, **24,** 23–38.
Zwicker, E. & t'Hart, E. (1980) Analytical expressions for critical-band rate and critical band width as a function of frequency, *Journal of the Acoustical Society of America*, **68,** 1523–1525.