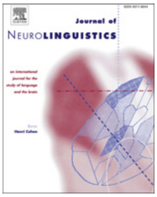
Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling



Movement goals and feedback and feedforward control mechanisms in speech production

Joseph S. Perkell

Speech Communication Group, Massachusetts Institute of Technology, Research Laboratory of Electronics, Room 36–591, 50 Vassar St., Cambridge, MA 02139-4307, United States

ARTICLE INFO

Article history:
Received 18 June 2009
Received in revised form 12 February 2010
Accepted 22 February 2010

Keywords:
Speech production
Speech perception
Speech motor control
Sensory goals
Feedback control
Feedforward control
Phonemic goals

ABSTRACT

Studies of speech motor control are described that support a theoretical framework in which fundamental control variables for phonemic movements are multi-dimensional regions in auditory and somatosensory spaces. Auditory feedback is used to acquire and maintain auditory goals and in the development and function of feedback and feedforward control mechanisms. Several lines of evidence support the idea that speakers with more acute sensory discrimination acquire more distinct goal regions and therefore produce speech sounds with greater contrast. Feedback modification findings indicate that fluently produced sound sequences are encoded as feedforward commands, and feedback control serves to correct mismatches between expected and produced sensory consequences.

1. Introduction

Speech production is the primary mode of human communication. It is the means by which discretely specified linguistic messages are converted to an acoustic signal that can be perceived and understood by a listener. Speech is produced by the combined actions of the respiratory system, the larynx and the supra-laryngeal vocal tract. Each of these systems is very complicated and has a unique set of biomechanical and physiological properties. For example, the vocal tract includes the mandible, tongue, pharynx, velum, nasal cavity, oral cavity, dentition and the lips, which are all markedly different from one another in terms of their structure and many aspects of their function. The actions of

E-mail address: perkell@mit.edu
URL: http://www.rle.mit.edu/perkell

0911-6044/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jneuroling.2010.02.011

these systems are controlled and coordinated to produce utterances comprised of sequences of words, which consist of strings of speech sounds that speakers can pronounce at rates of up to about 15 per second in fast speech. The movements of the vocal-tract articulators are produced by contractions of over 50 paired muscles, most of which are arranged symmetrically on either side of the midsagittal plane.

The articulatory movements for speech sounds are overlaid on the production of more slowly varying prosodic variables that convey important linguistic information beyond word meaning (cf., Fry, 1955; Lehiste, 1970). Prosodic structure is signaled by short-term variations in sound and syllable durations, in fundamental frequency of the vibration of the vocal folds in the larynx during voiced sounds, and in sound level. (Sound level is controlled mainly by the sub-glottal pressure generated by expiratory muscles and by elastic forces in the lungs.) These prosodically induced suprasegmental variations are expressed by actions of some of the same muscles that produce articulatory movements for sequences of individual speech sounds.[1]

The research in our laboratory has focused mainly on one part of the speech communication process: the control of movements of the supra-laryngeal articulators in the production of individual speech sounds and sequences of speech sounds. We have pursued questions common to the study of the control of many types of movement: what are the movement goals or targets, and what are the roles of feedback and feedforward mechanisms in controlling those goal-directed movements? Speech movements produce a time-varying acoustic signal with properties that are determined by variables that can be specified in several hierarchical domains, including: amounts of muscle tension; changes in muscle lengths and gestures (movements) of vocal-tract structures; the overall vocal-tract shape or cross-sectional area as a function of vocal-tract length (which determines the resonant properties of the vocal tract, viz. its acoustic transfer function); aerodynamic events and aeromechanical interactions that are produced by the passage of air through vocal-tract constrictions; and acoustic properties of the radiated sound. Movement goals, sometimes referred to as "controlled variables," presumably can be specified in terms of parameters in any of these domains.

Currently, there are two prevalent theories about the domain of the primary controlled variables for speech sounds. One theory posits that the primary controlled variables are articulatory gestures (e.g. Browman & Goldstein, 1989), which result in vocal-tract shapes that are perceived directly by the listener (Fowler, 1986). We have pursued an alternative theory motivated by the assumption that the speaker's objective is to produce sound sequences with acoustic patterns that are intelligible to the listener. This view leads us to hypothesize that the controlled variables for speech movements are time-varying patterns of auditory and somatosensory sensation. Thus in the first theory, the goals are specified primarily in terms of articulatory parameters (gestures, vocal-tract shapes) and in the second, they are specified primarily in terms of sensory parameters (e.g., auditory trajectories).

Only the sensory-based theory has been implemented in the form of a quantitative model that is grounded in neurophysiology in the sense of relating model components directly to brain regions and neural function. The model, called DIVA,[2] is a neurocomputational model of relations among sensory goals, brain activity, the speech motor output and the resulting auditory and somatosensory sensations. (c.f., Guenther, 1994, 1995; Guenther, Ghosh, & Tourville, 2006; Guenther, Hampson, & Johnson, 1998). Thereby, DIVA provides a unique basis for making quantitative tests of hypotheses about the neural mechanisms of speech motor control and their behavioral (kinematic and acoustic) consequences. Most of our more recent studies have been inspired directly by DIVA while some earlier ones were not. Regardless of this distinction, DIVA comprises an extremely useful framework for an organized account of our research. Thus, it is helpful to begin with a brief description of the model.

A simplified schematic diagram of the DIVA model is shown in Fig. 1. Each box in the diagram corresponds to a set of neurons. Neurons in the model are meant to correspond roughly to small

[1] In this paper, the word. "phoneme" is used to refer to a "speech sound" and the two terms are used interchangeably. Languages utilize limited numbers of phonemes as the basic building blocks of semantically meaningful units such as words. In the current context, the concept of a phoneme as a speech sound has a more physical connotation than it usually does in linguistic descriptions.

[2] "DIVA" stands for a basic characteristic of the model's functionality, "Directions Into Velocities of Articulators".

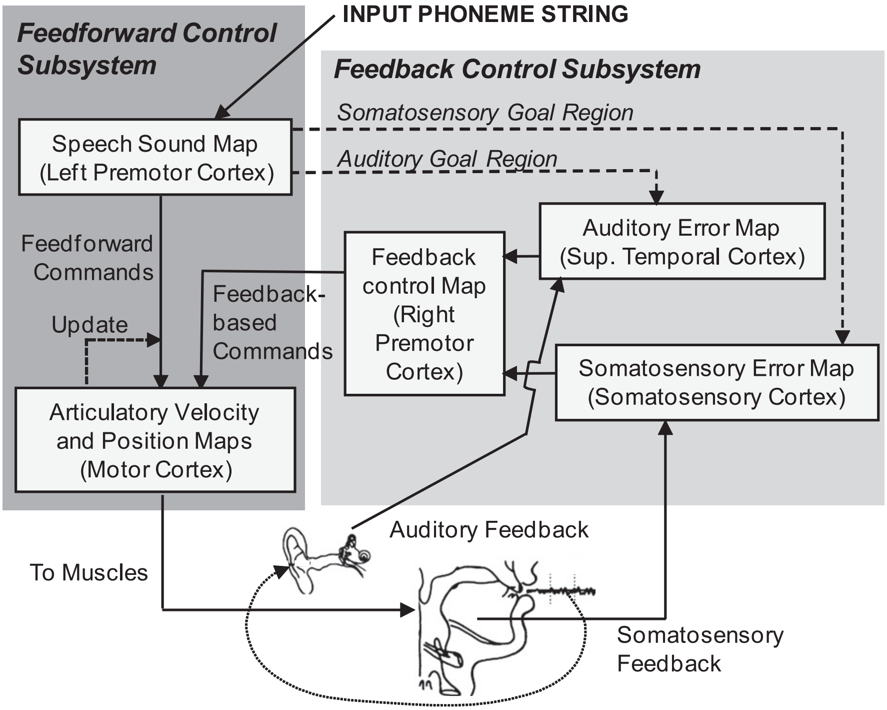

INPUT PHONEME STRING
Feedforward Control Subsystem
Feedback Control Subsystem
Somatosensory Goal Region
Auditory Goal Region
Speech Sound Map (Left Premotor Cortex)
Auditory Error Map (Sup. Temporal Cortex)
Feedback control Map (Right Premotor Cortex)
Somatosensory Error Map (Somatosensory Cortex)
Feedforward Commands
Update
Feedback-based Commands
Articulatory Velocity and Position Maps (Motor Cortex)
Auditory Feedback
To Muscles
Somatosensory Feedback

Fig. 1. A schematic block diagram of the DIVA model of speech motor planning and its interactions with an articulatory synthesizer.

populations of neurons in the brain that act as processing units. Thus, a "set" or "map" of neurons in the model is a set of neurons that represent a particular type of information in a particular brain region. Arrows in the diagram correspond to synaptic projections that form mappings from one type of neural representation to another. Several mappings in the network are tuned during a babbling phase in which semi-random articulator movements lead to auditory and somatosensory feedback. The model's synaptic projections are adjusted to encode sensory-motor relationships based on this combination of articulatory, auditory, and somatosensory information. The model posits additional forms of learning wherein (i) auditory targets for speech sounds are learned through exposure to the native language (Auditory Goal Region in Fig. 1), and (ii) feedforward commands between premotor and motor cortical areas are learned during "practice" in which the model attempts to produce a learned sound (Feedforward Commands in Fig. 1).

In the model, production of a phoneme or syllable starts with activation of a *Speech Sound Map* cell corresponding to the sound to be produced. These cells are hypothesized to reside in the left inferior frontal cortex (specifically posterior Broca's area - left Brodmann's Area 44 - and adjoining left ventral premotor cortex - BA 6). Activating a cell sends signals through tuned synapses, which encode sensory expectations for the sound, to bilateral auditory cortical areas where they are compared to incoming sensory information. Any discrepancy between expected and actual sensory information constitutes a production error, which is registered in the Auditory Error Map and/or Somatosensory Error Map. Projections from these error maps bilaterally send the error information to a Feedback Control Map that is hypothesized to lie in the right hemisphere lateral premotor cortex and posterior inferior frontal gyrus. This map generates corrective movements via bilateral projections to motor cortex.

Additional projections from speech sound map cells to the motor cortex form a feedforward motor command. This command constitutes the "motor program" for the current syllable and is read out in a feedforward fashion, i.e. without waiting to hear or feel the sensory consequences of the ongoing movement.

Dimensions in the auditory domain consist of parameters such as auditory-based transforms of formant frequencies[3] and amplitudes, fundamental frequency of vocal-fold vibrations, durations of sounds or parts of sounds, spectral characteristics of noise sounds (such as the frequency of the spectral mean[4] of the noise), sound level, etc. Dimensions in the somatosensory domain include spatial patterns of feedback of articulator contact, sensations reflecting levels of air pressure and air flow, muscle lengths and changes in muscle lengths, muscle tensions, angle of the temporomandibular joint, etc.

When DIVA is trained to control an articulatory speech synthesizer and is then given a sequence of phonemic goals to produce, it generates patterns of movements and a natural-sounding speech output with widely-observed properties such as: economy of effort, coarticulation,[5] motor equivalence, aspects of speech acquisition and responses to perturbations (cf. Guenther et al., 2006).

Since the functions of the model's components correspond to specific patterns of brain activation and functional connectivity, simulations run with the model can be compared to the results of brain-imaging and behavioral experiments with speakers, thereby providing quantitative tests of hypotheses about neurologically mediated relations between phonemic representations of speech sound sequences and the resulting sound output. Thus the DIVA model affords us an important means of dealing with the complexity of speech motor control, by providing a basis for a coherent and focused set of hypotheses that can be tested with a systematic program of experimentation. The remainder of this paper gives a number of examples of behavioral studies (many from our laboratory) with results that support DIVA-based hypotheses. When experimental results have not supported such hypotheses, they have led to modifications of the model that have been or will be tested in subsequent experiments.

2. Phonemic goals in speech production

2.1. Influences on phonemic goals

It is widely held that general properties of speakers' production and perception mechanisms help to define phonemic goal regions for language systems. Some of these properties are characterized by quantal acoustic effects (Stevens, 1989). A quantal acoustic effect consists of a relationship in which a continuous change in some articulatory parameter produces an acoustic parameter that is characterized by stable regions, separated from one another by regions of rapid change. Examples of such nonlinear quantal effects are found for the relation between formant frequencies and the constriction location for the point vowels /i/, /a/ and /u/ in which one or more formant frequencies is relatively stable over a range of constriction locations for each of the vowels (Stevens, 1989). Additional examples are found in the relation between the amount of separation of the vocal folds in the larynx when air is flowing outward between them and the type of sound that is generated: beginning with a widely spread configuration, as the folds are approximated there is first silence, then turbulence noise, then voicing and finally silence again when the folds are in tight closure. Such quantal relations are hypothesized to be bases of a number of distinctive features (c.f., Stevens, 1989, 1998). In some phonological theories, unique combinations of distinctive features serve to define individual phonemes and characterize linguistically meaningful phonemic contrasts (c.f., Chomsky & Halle, 1968; Jakobson, Fant, & Halle, 1951).

Based on work with a three-dimensional computer simulation of a biomechanical model of the tongue, Fujimura and Kakita (1979) hypothesized an additional source of nonlinear, quantal relations between articulation and acoustics in the form of a biomechanical "saturation effect." This modeling work indicated that if the tongue blade is stiffened and the tongue is moved upward and forward in production of the

[3] Formant frequencies are the frequencies of acoustic resonances of the vocal tract.

[4] The spectral mean is sometimes referred to as the "spectral center of gravity" (cf. Matthies et al., 1996).

[5] Because auditory targets in DIVA are regions (as opposed to points) that consist of trajectories in multi-dimensional auditory-temporal and somatosensory-temporal spaces, the model can and does move toward the closest point of the next target, which naturally results in coarticulation and the expression of economy of effort (see Perkell et al., 2000, pp. 235–237).

vowel /i/, when the stiffened blade is pressed against the lateral walls of the hard palate, the acoustically critical cross-sectional area of the resulting constriction remains relatively stable regardless of continuously increasing contraction of the muscles that raise the tongue.[6] Such saturation effects occur whenever two articulators are pressed against one another to produce a closure or narrow constriction in the vocal tract, as in virtually all consonants. For example, when the lips are approximated in the production of a voiceless stop consonant, there is a sudden transition to silence, which continues into the stop interval while the lip-closing movements may continue briefly and cause some compression of the lips (cf., Löfqvist & Gracco, 1997). Thus, like quantal acoustic effects, biomechanical saturation effects also are associated with nonlinear relations between articulation and acoustic cues.[7] Saturation effects are expressed in nonlinear relations between articulation and the vocal-tract geometry, which in turn influences the acoustics, whereas quantal effects are expressed more directly in nonlinear relations between the vocal-tract geometry and the acoustics (Perkell et al., 2000; Perkell et al., 1997).

By making it possible to produce relatively stable acoustic cues with some variation in underlying motor commands, quantal and saturation effects are hypothesized to contribute to determining the inventories of sounds for languages of the world (Stevens, 1989, 1998). Presumably as languages evolve, there would be a tendency to select and retain sounds that contrast with one another by virtue of containing distinct acoustic cues that can be produced relatively reliably with somewhat imprecise motor commands. However, not all perceptually relevant acoustic cues are associated with clear nonlinear relations between articulation and acoustics and there are other likely influences on sound inventories. Based on work with an articulatory-to-acoustic model of the vocal tract (Lindblom & Sundberg, 1971), it has been hypothesized that the configurations of vowel inventories of languages containing up to about seven vowels are determined approximately by a compromise between the amount of acoustic contrast among the vowel sounds and a measure of articulatory effort, i.e. a principle of “clarity” vs. “economy of effort” (Lindblom, 1990; Lindblom & Engstrand, 1989).[8] There are also social influences on the formation of local dialects that groups of speakers may use to distinguish themselves from one another (e.g., Labov, 1966). Neither the principle of clarity vs. economy of effort nor social influences on sound systems are quantal in nature, so if they operate in combination with nonlinear effects as languages evolve, the actual manifestation of any particular mechanism in natural speech may be difficult to identify instrumentally. (Also see Stevens & Keyser, 1989, 2010).

Mechanisms underlying quantal and saturation effects are not only hypothesized to play a role in determining sound systems of languages; they may also be employed by individual speakers to facilitate the production of stable acoustic cues with some imprecision in motor commands. Considering the variety of overlapping influences on sound systems and the fact that the production mechanisms of individual speakers differ from one another (e.g. in relative lengths of the pharyngeal and oral cavities), it is not surprising that it can be difficult to find experimental evidence for quantal mechanisms in the natural speech of individual talkers. Nevertheless, careful research by Stevens and colleagues has uncovered a number of subtle but convincing examples of quantal acoustic effects (Stevens, 1998). A specific example of individual speakers' use of a biomechanical saturation effect is discussed below in Sect. 2.2.1.

As described previously, movement goals in the DIVA model are specified in both auditory and somatosensory domains. During early learning, auditory goals dominate and act to shape feedforward commands, but after a fully learned set of feedforward motor commands is available and a somatosensory target is learned, then the somatosensory target is also part of the control process. Thus auditory targets are used initially by individuals to learn feedforward commands. While biomechanical saturation effects are hypothesized to have provided bases for some phonemic goals *as sound systems of*

[6] As discussed further in Sect. 2.2.1 (below), recent work by Buchaillard, Perrier P, and Payan (2009) indicates that controlling the articulation of /i/ is more complicated than proposed initially by Fujimura and Kakita (1979).

[7] Strictly speaking, “saturation effect” and “quantal effect” both refer to the same kind of nonlinear input-output relationship between parameters in two different domains. The two different terms are used in the current context to help distinguish between the mechanisms by which the effects are expressed (Perkell et al., 1997).

[8] In this case, articulatory effort was approximated by the amount of displacement of the articulators from a neutral position; other approximating metrics can include peak speed of articulator movement (cf., Nelson, 1983; Perkell, Zandipour, Matthies, & Lane, 2002).

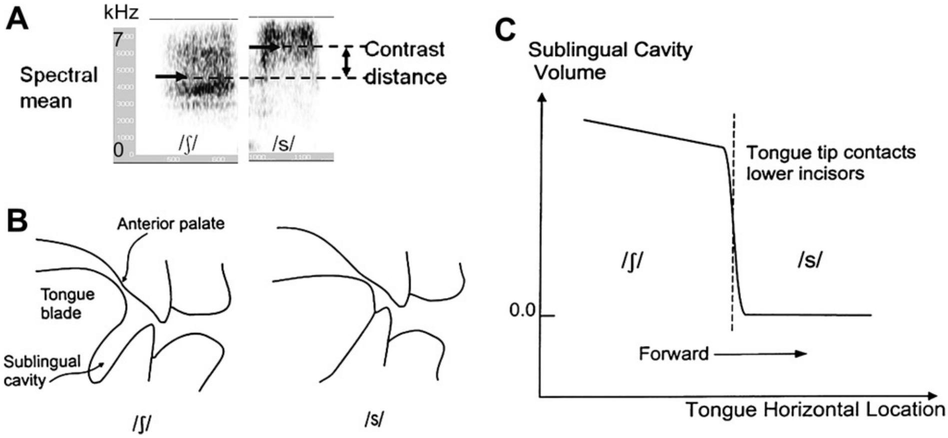


Fig. 2. A: Spectrograms of productions of the sounds /ʃ/ and /s/, illustrating the approximate frequencies of the spectral mean and the use of the spectral mean as a measure of the contrast distance between the two sounds. B: Schematic midsagittal drawings of articulations of /ʃ/ and /s/. C: Schematic illustration of a biomechanical saturation effect.

languages have evolved, their corresponding patterns of somatosensory feedback are acquired as somatosensory movement goals by individual language learners as a consequence of their first having learned auditory goals in attempts to mimic speech sounds of people in their environments.

2.2. *Mechanisms for producing relatively stable acoustic outputs*

2.2.1. *The use of saturation effects in producing the sibilant /s/ and the vowel /i/*

We have suggested that the sibilants have prominent sensory goals in the auditory/acoustic[9] domain and in the somatosensory domain (Perkell, Matthies, et al., 2004). Fig. 2A shows spectrograms of the sounds /ʃ/ (as in the word, "shed") on the left and /s/ (as in "said") on the right. A perceptually important difference between the two sounds is illustrated by the arrows that show the approximate locations of the spectral mean of the noise produced during each sound. The spectral mean is higher for /s/ than for /ʃ/; the amount of contrast or "contrast distance" between the two sounds can be quantified as the difference between their spectral means. We thus hypothesize that the auditory goals for the sibilants /s/ and /ʃ/ are noise sounds with spectral differences that can be characterized by contrastive, well-separated spectral means.

The noise source for both sounds is generated by turbulent airflow through a narrow constriction between the grooved tongue blade or tip and the hard palate; this source excites resonances (formants) of the small cavity anterior to the constriction. The articulatory configurations for the two sounds are illustrated by schematic midsagittal drawings in Fig. 2B. An /ʃ/ (on the left) is produced by creating the groove in the region of the tongue blade midline and pressing the grooved blade against the upper alveolar ridge. The positioning of the tongue blade leaves a sublingual space between the ventral side of the blade and the lower alveolar incisors. In contrast, an /s/ is produced with a shorter groove and contact between the tongue tip and the upper dento-alveolar ridge. The more anterior positioning of the tongue blade for /s/ is hypothesized to eliminate the sublingual cavity, thereby creating a substantially smaller resonant cavity anterior to the constriction for /s/ than for /ʃ/. This change contributes to the higher frequency spectral mean for /s/, presumably by causing a shift from excitation of F3 for /ʃ/ to F4 or F5 for /s/ (Perkell, Boyce, & Stevens, 1979). Fig. 2C is a schematic illustrating a hypothesized saturation effect and somatosensory goal for /s/. Beginning at the configuration for /ʃ/, as the constriction position is changed by gradually moving the tongue blade forward, the volume of

[9] It is assumed that auditory parameters consist of perceptually based transformations of acoustic parameters (e.g., Miller, 1989).

the sublingual cavity decreases gradually. When contact occurs between the tongue tip and lower incisors, the volume of the sublingual cavity drops abruptly to zero, and there is a corresponding rapid rise in the spectral mean of the sibilant noise. After contact occurs, the muscle activity moving the tongue blade forward can continue to increase, but of course, the volume of the sublingual cavity cannot decrease further.

We partially tested the hypothesis illustrated in Fig. 2C in 19 speakers of American English with an experiment that measured the acoustic signal and the presence or absence of contact between the tongue tip and lower alveolar ridge. For this purpose we glued a thin, custom-made contact sensor to the lower alveolar ridge (which sensed contact, but not pressure – see Perkell, Matthies, et al., 2004). The speakers pronounced multiple repetitions of the words *said, shed, sod* and *shod*, each embedded in the same carrier phrase. Across speakers, those who consistently showed contact for /s/ but not /ʃ/ tended to produce higher levels of contrast between the sibilants than those speakers who did not use the contact difference as consistently. We interpreted this result as supporting the contact hypothesis and indicating that contact between the tongue tip and lower alveolar ridge could be a somatosensory goal for /s/ that is associated with a saturation effect. In other words, the use of contact could be a mechanism for producing a relatively stable acoustic cue despite some variation in a part of the motor input.

Evidence consistent with the influence of both kinds of nonlinear relations (acoustic and biomechanical) has been found in measures of tongue articulation of the vowel /i/. Perkell and Nelson (1985) showed data from x-ray microbeam recordings of the positions of points on the tongue dorsum during two speakers' multiple repetitions of the vowel /i/ in a variety of contexts. In the palatal regions of both subjects there was greater variation of the point locations in a direction parallel to the vocal-tract midline than normal to the midline (also see Beckman et al., 1995). These results are compatible with use of the two mechanisms for stabilizing the acoustic output: the quantal acoustic effect for constriction location (Stevens, 1989, 1998) and the biomechanical saturation effect for constriction degree (Fujimura & Kakita, 1979). The coincidence of the two effects for this particular vowel may help account for its prevalence among languages of the world (cf., Ladefoged & Maddieson, 1996).

A recent modeling study by Buchaillard et al. (2009) has shed additional light on the hypothesized use of a saturation effect for producing the vowel /i/. The authors used a sophisticated, three-dimensional finite-element model of the tongue and the surrounding vocal-tract walls to examine the effects of contractions of individual muscles and combinations of muscles in the production of /i/ (among other articulations). They demonstrated the hypothesized global stabilization of the tongue body due to palatal contact. In addition, they showed that the acoustically critical cross-sectional area of the lingual groove was quite sensitive to variation in contraction of the anterior genioglossus. This modeling result highlights an important additional property of saturation effects: while they may allow for variation in aspects of movements that are constrained by articulatory contact (such as contraction of the posterior genioglossus to raise the tongue for /i/), other aspects of the same articulations (such as formation of the lingual groove in the palatal region) may have to be controlled with a fair amount of precision. The same principle should apply to the production the sibilants: the lingual groove for those sounds presumably must be formed with a certain degree of precision. (See Sect. 4.2.2 for further discussion of this point.).

According to the functionality of DIVA, all sounds have both somatosensory and acoustic/auditory goals. The goals for most vowels tend to be significantly more prominent in the auditory domain than in the somatosensory domain, with the notable exception of high front vowels and especially /i/, because of contact between the tongue blade and hard palate. Due to inter-articulator contact, the goals for most consonants tend to be more prominent in the somatosensory domain, especially goals for stop consonants. The sibilants are somewhat unusual in having prominent goals in both domains.[10]

[10] Thus, the relative prominence of auditory and somatosensory goals can vary among vowels (with high vowels having prominent patterns of tongue-palate contact as well as obvious auditory goals) and among consonants (e.g., stops that seem to be differentiated most obviously from one another by their patterns of articulatory contact, versus the sibilants with prominent goals in both domains).

2.2.2. The use of motor equivalence to stabilize the acoustic output for the vowel /u/ and the semivowel /r/

The vowel /u/ in American English is produced by forming two narrow constrictions in the vocal tract, one by rounding the lips and one by raising the tongue to a high position in the back of the oral cavity. In general, there is a many-to-one relation between articulation and acoustics: approximately the same sound can be produced with more than one configuration of the articulators (Atal, Chang, Mathews, & Tukey, 1978). In the case of /u/, the same sound can be produced with more tongue raising and less lip rounding, and vice versa. If the most prominent goal for vowels (particularly those that aren't produced with significant tongue-to-palate contact) is in the auditory domain, then in multiple repetitions of the vowel there could be a motor-equivalent trading relation between tongue-body raising and lip rounding, which can be controlled independently of one another. This hypothesis was supported by a study of several speakers, in which measurements were made of the positions of points on the articulators using an ElectroMagnetic Midsagittal Articulometer (EMMA) (Perkell et al., 1992). This finding of an articulatory trading relation for /u/ is consistent with the idea that the goal for the production of /u/ is in an acoustic/auditory frame of reference, as opposed to a spatial or motor frame (Perkell, Matthies, Svirsky, & Jordan, 1993).

A study of motor equivalence in the production of American-English /r/ by Guenther et al. (1999) found convergent results. American-English /r/ is characterized by a wide range of articulatory configurations across speakers; for the most part, these configurations fall in two categories, retroflexed, in which the tongue tip is raised upward toward the middle or posterior part of the hard palate, and bunched, in which the tongue tip is drawn down and backward while the middle part of the tongue blade is raised toward the palate (Delattre & Freeman, 1968). In general a speaker would tend to produce a retroflexed /r/ following /b/ (as in "bread") or following a low vowel ("Tara") and a bunched /r/ following /g/ (as in "grab"). These two variants of /r/are not contrastive, and the primary acoustic characteristic for both configurations is the same low value of the third formant frequency, F3. In order to look for articulatory trading relations between tongue configurations for the two sounds, Guenther et al. (1999) recorded the positions of three points on the tongue and the acoustic signal from seven speakers as they pronounced nonsense words such as /wagrav/ (bunched), /wabrav/ and /warav/ (retroflexed). The positions of the tongue points were used to infer the approximate size of the constrictions for /r/ and the shapes of vocal-tract cavity anterior to the constriction. Although these parameters varied across speakers, tongue configurations in each speaker were consistent with the two types of /r/. On producing the acoustic cue for /r/ (low F3) each speaker yielded evidence of a trading relation: a longer front cavity for the more retroflexed configuration was compensated for by the longer and narrower constriction of the more bunched configuration. When the articulatory positions were used to predict the value of F3, the variability of F3 was greatly reduced by including the covariances among the articulatory measures in the prediction, indicating that the speakers used articulatory trading relations to reduce acoustic variability. These results reinforce the idea that the goal for /r/ is in the auditory domain.

2.3. The role of auditory feedback in learning and maintaining phonemic goals

On the one hand, it is well known that hearing is crucial for normal speech acquisition; if a child is born without useful hearing, he or she will have a very difficult time learning how to speak fluently. On the other hand, if an individual is born with relatively normal hearing, learns how to speak and then becomes deaf postlingually, that person is usually able to speak intelligibly for decades without having any useful hearing (Cowie & Douglas-Cowie, 1983). In terms of the function of the DIVA model, this ability to speak in the absence of hearing is largely due to the speaker's having previously acquired a robust system of feedforward commands for producing speech (as with any highly skilled form of movement). Although such individuals' speech remains intelligible, it can gradually develop anomalies and some decrements in contrasts among speech sounds (Cowie & Douglas-Cowie, 1992; Lane & Webster, 1991; Plant, 1984), which can be interpreted as the result of deterioration of feedforward commands. If, after prolonged hearing loss that results in reduced contrasts, a speaker regains some hearing with a cochlear implant, those reduced contrasts will be partially restored over time. However, when the implant is first turned on and the individual experiences an auditory sensation that is very different from anything he or she has heard previously, there can be a slight, further decrement in sound contrasts compared to pre-implant levels (Lane, Matthies, et al., 2007). This observation indicates that the speaker must first establish new mappings between the novel auditory sensations coming from the implant and previously established phonemic goals for speech sounds. Once this remapping has progressed sufficiently, the feedback system can be used for retuning feedforward commands, usually resulting in progressive, parallel improvements in speech perception, production and intelligibility of the produced speech (cf. Lane, Matthies, et al., 2007; Matthies, Svirsky, Perkell, & Lane, 1996; Vick et al., 2001).

An illustration of the role of auditory feedback from a cochlear implant in retuning feedforward commands is provided by a recent study of the production and perception of /r/ by Matthies et al. (2008). In terms of the DIVA model, the mapping between motor commands for the production of bunched and retroflexed allophones of /r/, on the one hand, and the single auditory target of a low F3 on the other, would deteriorate with prolonged hearing loss but would be re-tuned with auditory feedback from an implant. Thus, postlingually deafened speakers should show larger acoustic variation among /r/ allophones than hearing controls, and auditory feedback from a cochlear implant should reduce that variation. To test this hypothesis, measures were made of phoneme perception and of the production of tokens containing /r/, stop consonants, and /r/ + stop clusters in hearing controls and in eight postlingually deafened adults pre- and post-implant. The hypothesis was supported. Pre-implant, the postlingually deafened speakers had greater acoustic variability for /r/ than the hearing controls. Post-implant, the variability for /r/ of seven of the eight implant users did not differ from the mean values obtained from a normal-hearing control group.

3. On the nature of sensory goals in speech

A basic neural linkage between action and perception has been found in neurophysiological studies of “mirror neurons” in premotor cortex in monkeys. These neurons are active both when an animal produces a particular trained action and when the animal perceives the same action being produced by another individual (cf. Ferrari, Fogassi, Gallese, & Rizzolatti, 2003; Rizzolatti & Craighero, 2004). Furthermore, the same group of investigators has found indirect evidence of mirror neuron-like function related to speech in non-invasive experiments with humans (cf., Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Kohler et al., 2002).[11] With further regard to speech, we already have mentioned the importance of hearing for speech acquisition and noted the improvements in speech that occur when people with profound hearing loss begin to use cochlear implants. Convergent findings also come from some studies of second-language learning. For example, Bradlow, Kahane-Yamada, Pisoni, and Tohkura (1999) trained Japanese speakers learning American English to improve their perception of the difference between /l/ and /r/ and found that productions of the two sounds improved, even without speech training (also see Jamieson & Rvachew, 1992; Rvachew, 1994).

In view of the links between speech perception and production discussed above, and the well-known variability among speakers in their production measures, the question arises whether differences among speakers’ productions may be attributable in part to differences among them in perceptual capacities. To address this question, we have conducted four studies in which we tested the following hypothesis about auditory goals: speakers with greater auditory acuity, i.e., who discriminate well between speech sounds with subtle acoustic differences, will produce those sounds with greater contrast than speakers who discriminate the same sounds less well.

[11] The speech sound map cells in DIVA (see Sect. 1) effectively constitute “mirror neurons.” In this sense, their function in DIVA has to do with learning in the motor system; i.e. these cells are activated during perception of a new sound in the learning of that sound’s auditory target; the same sound-specific cell is activated later in production of the sound. The DIVA model includes no explicit prediction regarding the function of speech sound map cells in perception. Findings of mirror neurons have been cited as providing support for the motor theory of speech perception (Liberman & Mattingly, 1985; Rizzolatti & Arbib, 1998; above-cited references from the Rizzolatti lab); Lotto, Hickok, and Holt (2009) have provided a critique of this view.

3.1. Cross-speaker relations between vowel production and perception

Two of the studies under this heading focused on the production and perception of vowels. In the first of these, the acoustic signal and movements of a point on the tongue dorsum were recorded while each of 19 speakers of American-English produced multiple repetitions of individual words comprising two vowel contrasts, *cod* – *cud* and *who'd* – *hood*, with each word embedded in a carrier phrase. The "articulatory contrast distance" for each subject was calculated as the average across the two word pairs of the mean Euclidean distances in the midsagittal plane between the mid-vowel tongue point locations, i.e., for /a/ vs. /ʌ/ and for /u/ vs. /ʊ/. Average "acoustic contrast distance" for each subject was calculated in the same way in the F1 × F2 plane. The same subjects participated in a perceptual experiment in which they were asked to discriminate between pairs of synthetic stimuli along two 7-step vowel continua, *cod-cud* and *who'd-hood*. Based on a median split of their discrimination scores, subjects were divided into two groups, high-discrimination (higher acuity) and low-discrimination (lower acuity) for each contrast. The high-acuity group produced significantly greater contrasts in both articulatory and acoustic spaces than did the low-acuity group, leading to the tentative conclusion that the more accurately a speaker discriminated a vowel contrast, the more distinctly the speaker produced the contrast (Perkell, Guenther, et al. (2004); also see Groenen, Maassen, Crul, & Thoonen, 1996; Newman, 2003).

Partly because of the previous study's limited number of vowels and relatively coarse-grained acuity measures, we have conducted a similar, more elaborate study on a new group of speakers (18 subjects). For the production measures, we used a greater variety of test words containing six different vowels in /pVp/, /tVt/ and /kVk/ contexts. Those words were pronounced in three different speaking conditions (*normal*, *clear* and *fast*), with and without emphatic stress. We recorded and analyzed the acoustic signal and, for each subject, we calculated average vowel spacing (AVS – a measure of overall vowel contrast – cf. Lane, Denny, et al., 2007) and average dispersion for each vowel around its centroid in F1, F2 space (both measured in mels). For the acuity experiment, we used more finely-graded stimulus sets (1001 steps in each of two continua), and a more sensitive approach to measuring discrimination (in terms of just noticeable differences, JNDs) than in the first study. There were significant cross-subject positive correlations between acuity (1/JND) and AVS in all three speaking conditions, reinforcing the findings of the earlier study. There were also negative correlations between acuity and dispersion values in all three conditions. Thus, speakers with greater auditory acuity for vowels, as defined above, not only produced greater vowel contrasts, but also had smaller vowel goal regions than speakers with lower acuity (Perkell, Lane, et al., 2008).

These findings can be interpreted with reference to the schematic diagram in Fig. 3. The figure shows hypothetical goal regions for the vowels /I/ and /ɛ/ in F2 × F1 space for a high-acuity speaker (gray, solid lines) and a low-acuity speaker (black, dashed lines). In accord with the results, the high-acuity speaker (relative to the low) has goal region centers further apart (greater contrast) and the regions are smaller (smaller dispersions). To motivate this view, we assume that as a child is acquiring speech, he or she comes to understand (subconsciously) that it is advantageous to be intelligible as possible. Thus, the child would try to produce speech sounds with the greatest possible contrast. In the case of vowels, this would result from learning acoustic goal regions that have the greatest possible inter-vowel distances and the smallest possible dispersions, as schematized in the figure for the high-acuity speaker. According to the way DIVA functions, a high-acuity speaker, when learning goal regions, would reject outlying exemplars (indicated by the X) as produced badly, whereas such sounds would be acceptable for a low-acuity speaker.

3.2. Cross-speaker relations between production and perception of sibilants

In the study of sibilant production discussed above in Sect. 2.1.1, we found that speakers who more consistently used contact between the tongue tip and lower alveolar ridge for /s/ but not /ʃ/ tended to produce the two sounds with more acoustic contrast than speakers who did not consistently show such a contact difference (Perkell, Matthies, et al., 2004). This finding led to the hypothesis that contact between the tongue tip and lower alveolar ridge was a somatosensory goal for /s/. In the same study,

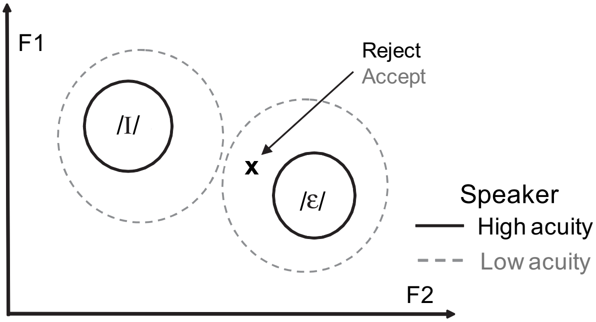


Fig. 3. Hypothetical goal regions for the vowels /I/ and /ε/ in F2 × F1 space for a high-acuity speaker (gray, solid lines) and a low-acuity speaker (black, dashed lines).

we also hypothesized that the sibilants have auditory goals, which consist of distinctive spectra for the two different types of noise sounds, with /s/ having a higher spectral mean than /ʃ/. Along the same lines as the vowel studies described above, we also measured the subjects' auditory acuity by using discrimination tests with seven-step synthetic stimuli along continua from *said* to *shed* and *sod* to *shod*. As with vowels, we found that speakers with higher acuity for the sibilant contrast tended to produce the sounds with greater contrast (Perkell, Matthies, et al., 2004). However, our theoretical framework left an open question: would speakers with greater somatosensory acuity produce the sibilants with greater contrast than speakers with lesser somatosensory acuity?

To answer this question, we conducted a study of somatosensory and auditory acuity in relation to produced contrast with the same group of subjects who participated in the second vowel study just described. The sibilant study consisted of three experiments. In the first, the somatosensory acuity of each subject's tongue tip was measured using JVP Domes™ (Van Boven & Johnson, 1994; also see Wohlert, 1996). These are a set of eight plastic domes, each 19 mm in diameter, containing grids consisting of equidistant bars and grooves with different widths: 0.35, 0.5, 0.75, 1.00, 1.25, 2.00 and 3.00 mm. Fig. 4 shows a drawing of a cross-section of the dome with 2.00 mm bar and groove widths. To test for somatosensory acuity, each dome is pressed against a subject's skin or mucosa with one of four distinct grid orientations and the subject was asked to identify the orientation. We built a custom holder with a handle and a cylindrical receptacle into which one of the domes could be inserted with the grid set at one of the four orientations. To allow for control over the duration and magnitude of applied pressure, the receptacle was mounted on the end of a strain-gauge cantilever beam contained inside the handle. The strain gauges were attached to a bridge amplifier, connected to an A/D channel of the computer that ran the experiment and recorded data. An algorithm was developed to control the experimental protocol, measure the applied force, and signal the experimenter when an appropriate range of application force had been reached and when the force application should end. The

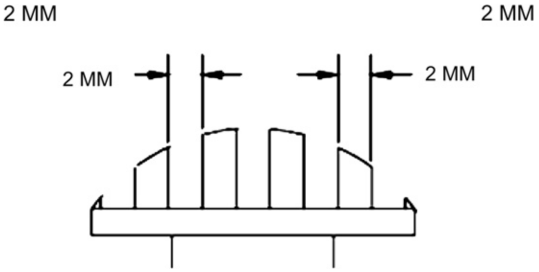


Fig. 4. A cross-sectional diagram of a JVP dome with 2.00 mm bar and groove widths.

experimenter pressed the dome against the subject's protruded tongue tip, and the subject identified the orientation of the grid. Each orientation was presented 10 times for each grid, with random orientation order for grid. The subject was unable to see the grid, but was given feedback of the result (correct/incorrect) after each trial. A score was calculated as the percent correct for each dome. The measure of acuity used for each subject was the subject's proportion correct for the dome on which he or she achieved the best score (which had either the 2 or 3 mm grid spacing, depending on the subject). These results varied considerably across subjects.

The second and third sibilant experiments measured the subjects' auditory acuity and sibilant contrast distance as described above. The cross-subject correlation between somatosensory proportion correct and contrast distance was significant ($r = 0.44$, $p < 0.04$; one-tailed), as was the correlation between auditory acuity (1/JND) and contrast distance ($r = 0.41$, $p < 0.05$; one-tailed). A multiple linear regression showed that a combination of somatosensory and auditory acuity measures also predicted contrast distance ($R = 0.42$, $p < 0.01$). These results support the hypothesis that speakers have somatosensory and auditory goals for the sibilants, and that speakers with higher acuity of both sensory modalities tend to produce greater sibilant contrasts (Ghosh et al., submitted for publication; Perkell, Matthies, et al., 2008).

As noted earlier in Sect. 2.1, when sensory goals are being acquired by the DIVA model, auditory goals must be learned first, since the model is initially trying to match the speech sounds it produces to externally specified acoustic goal regions (in the way children try to imitate auditory targets from speakers in the environment). Somatosensory goals are acquired as the somatosensory patterns that accompany successful productions of auditory targets. From this perspective, the intelligibility of postlingually deafened speakers can be attributed to at least two factors: a robust set of feedforward commands for speech sounds, syllables and words, and an intact set of somatosensory goals. Once somatosensory goals are established, they are not affected directly by hearing loss, so they can presumably be used by feedback control mechanisms to help correct some of the errors that may occur as feedforward commands begin to deteriorate with acquired deafness.

4. Feedback and feedforward mechanisms in speech motor control

The preceding sections have included general descriptions of the roles of auditory and somatosensory feedback in speech motor control. Experiments are described below that were designed to explore the ways in these feedback mechanisms operate and how they interact with feedforward commands.[12]

4.1. Feedback control: closed-loop error correction

For the purposes of this discussion, feedback-based motor control is considered to consist of four successive stages: (1) detection of a movement error by comparing feedback of peripheral sensory information about movement progress with the specification (sensory goal) for the intended trajectory or target, (2) computation of the necessary corrective command, (3) transmission of the corrective command to the muscles and (4) contraction of the muscles to correct the movement trajectory. Such feedback-based control is called "closed-loop" if all four of these stages are achieved during a single ongoing movement (of a single articulator or a synergism of multiple articulators). If closed-loop control is to be effective, the error correction has to be sufficiently complete before the end of the movement. Thus the accumulated delay due to transduction and neural transmission of sensory information to the central nervous system (CNS), computation of corrective motor commands, neural

[12] To study feedback in speech production, a great deal of research has been conducted in which the shape of the vocal tract was modified by inserting a bite block between the teeth (cf. Flege, Fletcher, & Homeidan, 1988; Fowler & Turvey, 1980; Lane et al., 2005; Lindblom, Lubker, & Gay, 1979), a tube between the lips (Savariaux, Perrier, & Orliaguet, 1995) or a prosthesis that altered the shape of the oral cavity (cf. Baum & McFarland, 1997; Hamlet & Stone, 1976; Hamlet, Stone, & McCarthy, 1976; Jones & Munhall, 2003). In such studies, the perturbations were left in place long enough for the subjects to consciously attempt to make compensatory adjustments. Although these studies have revealed a great deal about factors involved in speech motor control and learning, the focus here is on results from the use of more transient perturbations.

transmission of the motor commands to the muscles and conversion of the resulting muscle action potentials to muscle contraction and movement has to be shorter than the movement duration. It is generally believed that the accumulated delays make it impossible to use closed-loop feedback control of many types of movements, especially rapidly sequenced skilled movements of the extremities such as in typewriting or piano playing (with neural computation and transmission times that presumably are long in relation to intervals between key presses).

4.1.1. Somatosensory closed-loop feedback control

A number of studies have used transient perturbations to explore the use of somatosensory closed-loop feedback mechanisms in speech motor control. For example, Abbs and Gracco (1984) used a servo mechanism to mechanically follow lower-lip movements during lip closure for bilabial stops and unexpectedly arrest the lower-lip closing movement in randomly-selected trials. They observed increased displacement and velocity of the upper lip to complete the closure, accompanied by compensatory EMG signals beginning about 22–75 ms following the lip perturbation. With a 30–75 ms delay between the onset of EMG activation and the resulting movement, this result leads to an estimate of corrective movements within 50–150 ms of the perturbation. Shaiman (1989) used a similar apparatus to interrupt upward jaw movements during closure for the /d/ in /ædæ/ and the /b/ in /æbæ/ as produced by six speakers. Compensatory kinematic adjustments were found in lip and/or jaw movements during the bilabial closure, but not in labial movements during the dento-alveolar closure (also see Munhall, Löfqvist, & Kelso, 1994, on laryngeal responses to labial perturbations). Such results indicated that “the components of the motor system are flexibly assembled, based on the requirements of the specific task. That is, compensatory responses … occur only when such responses are functionally necessary.” (Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984, also Shaiman, 1989, p. 78).

Gomi and Honda (2002) investigated this issue further by perturbing lower-lip closing movements in two experiments with Japanese speakers. In the first experiment, the subjects pronounced sustained versions of the bilabial fricative /Φ/, which requires constriction, and /a/, for which the lips are relatively relaxed. In the second, a single subject pronounced the utterance “kono /aΦaΦa/ mitai,” and a downward perturbation load was applied to the lower lip at various onset times. Partial or complete downward compensation by the upper lip was found in both experiments. The compensatory movement was larger for /Φ/ than /a/, reinforcing earlier findings of sound-specific, temporarily assembled neuromuscular synergies. However, the initial component of the downward displacement of the upper lip frequently preceded the onset of upper-lip EMG. A muscle-linkage model was used to show that the initial part of the compensatory movement was likely due to a preplanned stiffness increase for the bilabial phoneme, “in order to robustly configure a labial constriction.” (p. 261).

While experiments like these show that closed-loop somatosensory control may function in speech, they do not directly address the question of whether it can act rapidly enough to correct errors in the speaker’s acoustic output. Tiede, Ito, and Ostry (2006) investigated this issue by perturbing jaw movements of three speakers either upward or downward on one of five randomly-selected repetitions of utterances such as “see red” while recording jaw movement and the acoustic signal. In the perturbed utterances, formant frequencies showed initial deviation from control trajectories and then compensation that began about 75 ms post-perturbation and was often nearly complete 120–140 ms after perturbation onset. The acoustic recovery occurred in spite of the fact that the jaw movement showed no compensatory response to the perturbation, leading to the inference that the acoustic compensation was due to adjustments of tongue movements. In this case, acoustic recovery was possible because the vowels lasted longer than the time needed for complete compensation.

Somatosensory feedback about speech movements can consist of several types of information conveyed by afferent nerve fibers from peripheral receptors to the CNS, including: touch, or tissue contact (from pressure sensors in the skin and vocal-tract mucosa), muscle length and rates of change of muscle length (from muscle spindles), muscle tension (from tendon organs) and joint angle (from receptors in the capsule of the temporo-mandibular joint).

4.1.2. Auditory closed-loop feedback control

The development of sufficiently fast digital signal-processing algorithms has made it possible to introduce controlled modifications of vowel formants in the auditory feedback of speakers (delivered over headphones) in nearly real-time (with virtually imperceptible delays of around 20 ms or less – cf. Houde & Jordan, 1998, 2002; Purcell & Munhall, 2006a; Villacorta, Perkell, & Guenther, 2007). Using such a computer algorithm, Tourville et al. (2005) shifted F1 either up or down every four trials in the auditory feedback of 11 speakers, while they slowly pronounced words containing the vowel /ε/ (e.g., *beck, bet, debt*). The subjects responded by pronouncing the perturbed words with F1 shifted in the opposite direction, with a variable latency that averaged approximately 130 ms. The subjects generally were unaware of the formant shift and unaware that they were compensating for it. Similar compensation delays have also been found in several experiments in which voicing fundamental frequency has been shifted (cf., Burnett & Larson, 2002; Larson, Burnett, Bauer, Kiran, & Hain, 2001). Such findings, which are compatible with closed-loop auditory feedback control as modeled in DIVA, can be observed in speakers if the sound lasts long enough.

4.2. Feedforward control

If speech is being produced at an average rate of approximately 10 sounds per second, as is the case in normal conversational interchanges, corrective movements that take about 130 ms are likely to be ineffective for the ongoing control of such fluent speech-sound sequences. Thus, it is widely believed that once speech is acquired and has matured, it operates almost entirely under feedforward control. Although feedback control is presumed to always be available to correct and detect errors (Guenther et al., 2006), circumstances that might shed light on this function in mature fluent speech occur very rarely outside the laboratory.[13]

As described earlier, feedback control operates in the DIVA model by detecting errors between expected and actual consequences of articulatory movements and generating corrective motor commands. If errors persist over a number of repetitions, the corrective commands become incorporated into future feedforward commands, which is the mechanism by which feedforward commands are learned and refined. The following paragraphs describe several experiments that illustrate aspects of this process.

4.2.1. Relations between feedback and feedforward control

A study reported by Honda, Fujino, and Kaburagi (2002) and Honda and Murano (2003) is particularly supportive of the interplay of feedback and feedforward control as hypothesized above. In this experiment, a mechanical perturbation was introduced unexpectedly, and the results indicated that error corrections based on auditory and somatosensory feedback were incorporated into subsequent feedforward commands. The mechanical perturbation consisted of unexpectedly increasing the thickness of the dento-alveolar portion of speakers' palates by rapidly inflating a small stiff balloon incorporated into a thin acrylic palatal prosthesis that was custom-made for each subject. Articulatory movements, tongue-muscle EMG and the acoustic signal were recorded while Japanese-speaking subjects pronounced multiple repetitions of an utterance consisting of a string of eight /ʃa/ syllables preceded by the syllable /iya/. The balloon was inflated on selected repetitions during the downward tongue movement for the /a/ in the initial /iya/. This was done under four different feedback conditions: (1) normal, (2) hearing blocked with masking noise, (3) tongue tactile sensation masked with topical anesthesia and (4) both hearing and tactile sensation masked. In the trials with palatal perturbation, the tongue showed active compensatory downward movements beginning 100 ms following initial tongue tip contact with the thickened palate under normal feedback (1). The initial responses were: almost the same under auditory masking (2), diminished and slightly delayed under tactile masking (3) and most markedly affected when both modalities were masked (4). A group of listeners rated whether each of the fricative sounds was produced correctly; the error scores were worst for the first syllable and generally improved for successive syllables in the string. The rate of improvement was highest

[13] This observation does not hold for some speech disorders, most notably stuttering, a condition that affects speech fluency. The development of other types of highly skilled, rapidly sequenced movements such as playing most kinds of musical instruments, handwriting or "touch typing" may have been facilitated by the fact that they also do not generally encounter unexpected perturbations.

with normal feedback (1) and successively lower with tactile masking (3), auditory masking (2) and combined masking (4). The similar (approximately 100 ms) latencies for the initial compensatory response under normal feedback and with only auditory masking indicated that somatosensory feedback control was used in generating the earliest compensatory response; the physiological measurements and listener judgments indicate that both feedback modalities were used to help make continuing refinements of feedforward commands for successive syllable repetitions.

It is suggested above that mature fluent production of sound sequences must be mainly under feedforward control because the movements for individual sounds generally don't last long enough to be guided on-line by feedback-based error correction. However, prosodic (suprasegmental) parameters such as F0, sound level and speaking rate vary over longer time scales, from one syllable to the next, or even longer. A difference in the control of segmental and prosodic parameters has been revealed in a study by Perkell, Lane, et al. (2007). In this study the timing of changes in parameters of speech production was investigated in six cochlear implant users when their implant microphones were switched off (hearing blocked) and on (hearing restored) a number of times in a single experimental session. The subjects repeated four short, two-word utterances, *dun said, dun shed, don sad* and *don shad*, one at a time in quasi-random order, with a variable number of tokens between switches to prevent anticipation of the switch. Ramp-shaped changes between hearing and non-hearing states were introduced by a voice-activated switch at the onset of the first vowel (V1, in *dun* or *don*) and were complete within 20 ms of vowel onset. Measures were made of the suprasegmental parameters of vowel SPL, duration and *F*0; measures of phonemic contrast were made of vowel contrast distance between pair members in the formant plane, /ʌ/ vs. /a/ (V1) and /æ/ vs. /ɛ/ (V2) and sibilant contrast difference in spectral means. Changes in parameter values were averaged over multiple utterances, lined up with respect to the switch.

Regardless of whether prosthetic hearing was blocked or restored, contrast measures for vowels and sibilants did not change systematically, in spite of the fact that the vowels were generally longer than 150 ms. However, some systematic changes in duration, SPL and F0 were observed during the vowel within which hearing state was changed (/ʌ/, /a/), or during the second vowel following the change (/æ/, /ɛ/); these changed values persisted through subsequent utterance repetitions until the next switch. The lack of consistent contrast changes is compatible with the operation of feedback control of phonemic movements in the DIVA model, which responds in order to correct perceived *errors* that are produced under feedforward control, but not to the *presence* vs. *absence* of feedback. The observed changes in suprasegmental parameters are consistent with other findings of short-latency changes in duration and sound level when hearing is blocked (e.g., by the occurrence of loud environmental noise) and restored: without sufficient hearing, people speak louder and more slowly than they do with hearing, presumably to try to maintain the intelligibility of their speech when they can no longer discern their own produced sound contrasts (Perkell, Denny, et al., 2007). Thus, suprasegmental parameters, which tend to vary over syllabic and longer durations, could be regulated with the participation of feedback control. Currently, DIVA includes no mechanism for control of the more slowly varying suprasegmental parameters.

4.2.2. Sensorimotor adaptation studies of feedforward control in achieving auditory goals

A series of "sensorimotor adaptation" (SA) studies have been conducted to further investigate the mechanisms by which feedback-based error corrections are incorporated into feedforward commands. These studies also employ feedback perturbations; however, in the SA paradigm, a perturbation of a vowel formant or formants is introduced in a speaker's auditory feedback on every trial during repeated productions of stimulus words. The initial perturbation is small and is increased gradually from one token to the next over a large number of repetitions while the subject's responses are recorded (Houde & Jordan, 1998, 2002; Max, Wallace, & Vincent, 2003; Purcell & Munhall, 2006b). The formant shifts are made in nearly real-time, usually using the same kind of apparatus as described above in Section 4.1.2 on auditory feedback control. In this paradigm the experiment is divided into four phases, each containing a number of repetitions: (1) baseline, with no-shift, (2) ramp, in which the magnitude of the shift is increased gradually from one trial to the next, (3) full-shift, in which the shift is maintained at the same maximum level for a number of trials, and (4) post-shift, in which the shift is removed. We have conducted two such studies in our laboratory.

Villacorta et al. (2007) report results of a sensorimotor adaptation experiment with 20 normal-hearing speakers. The subjects pronounced /CɛC/ words (e.g., *bet, get, peck*) while the first formant frequency (F1) of the vowel in their auditory feedback was shifted in nearly real time (18 ms delay), without their being aware of the shift. Ten of the subjects received upward shifts and the other 10, downward shifts. The subjects partially compensated for the shifts over many trials by modifying their productions so that F1 moved in the direction opposite to the shift. This compensation persisted for some minutes after the shift had been removed, thus providing evidence of "adaptation," viz., a temporary modification of feedforward commands. The result is compatible with the function of the DIVA model described above, in which auditory feedback is used to generate corrective commands that are then incorporated into modifications of feedforward commands for subsequent movements. The fact that compensation was not complete may be due at least in part to the influence of the vowel's somatosensory goal.[14] Compensation for auditory errors will introduce somatosensory errors; therefore, feedback-based compensation for the somatosensory errors will counteract and prevent complete auditory compensation from taking place.

The amount of compensation in this experiment varied among the subjects. Based on our earlier results indicating that speakers with greater auditory acuity have smaller auditory goal regions spaced further apart, we hypothesized that there was a relation between speaker acuity and the amount of compensation. This hypothesis is illustrated in Fig. 5, which schematizes how two speakers, differing in acuity, and therefore in the sizes of their goal regions for the vowel /ɛ/, might respond to a perturbation of F1. The high-acuity speaker (HI) has a smaller goal region. The perturbation of F1 is indicated by a dotted arrow pointing to the right, and the shifted value of F1, by a vertical broken line. The distance between the shifted value of F1 (vertical line) and the edge of the goal region is greater for the high-acuity speaker. Each speaker continues to compensate until the F1 of his or her auditory feedback (which includes the shift) moves into the goal region; therefore, in response to the shift in F1, the high-acuity speaker will produce a greater compensatory response (middle arrow) than the one with lesser acuity. To test this hypothesis, we measured the subjects' auditory acuity and compared the resulting values with the amount of compensation. There was a significant, positive cross-speaker correlation between acuity and amount of compensation; the correlation strength increased when the magnitudes of individual speaker's produced vowel contrasts were factored out. This combination of sensorimotor adaptation and acuity studies was simulated using subject-specific versions of the DIVA model, in which goal-region size differed according to measured subject acuity. In further support of our findings and inferences about goal-region size, the model versions with smaller goal regions adapted more than those with larger regions. The overall results lend validity to the modeling approach and reinforce other results relating speaker acuity to goal-region size and contrast distance (Perkell, Guenther, et al. 2004; Perkell, Matthies, et al., 2004; Villacorta et al., 2007).

Until very recently, almost all experimentation on the nature of phonemic goals and the roles of feedback and feedforward control in achieving them has focused on the production of steady state sounds – including the research cited above. However, many perceptually important aspects of the spectral structure of natural speech are time varying (for vowels, see Jenkins, Strange, & Trent, 1999; Nearey & Assmann, 1986). The time-varying nature of speech sounds is accounted for in the DIVA model by its implementation of auditory goals as regions in multi-dimensional auditory-*temporal* and somatosensory-*temporal* spaces. Thus the movement goals are effectively modeled as time-varying regions in these multi-dimensional spaces.

Cai, Ghosh, Guenther, and Perkell (submitted for publication) conducted a sensorimotor adaptation study that investigated the responses of speakers of Mandarin Chinese to time-varying perturbations of acoustic trajectories of Mandarin Chinese triphthongs (like diphthongs, but containing three vowel sounds instead of two), embedded in a carrier phrase. The triphthong /iau/ was chosen partly because of the large span of its trajectory in F1 × F2 space and its long duration, during which perturbations would presumably be more perceptually salient. Fig. 6A shows a spectrogram of a token of /iau/ with

[14] As noted earlier, all sounds are hypothesized to have both auditory and somatosensory goals. Thus while the goals for most vowels are considerably more prominent in the auditory than the somatosensory domain, somatosensory errors could develop and lead to compensatory feedback commands when speakers compensate for auditory perturbations.

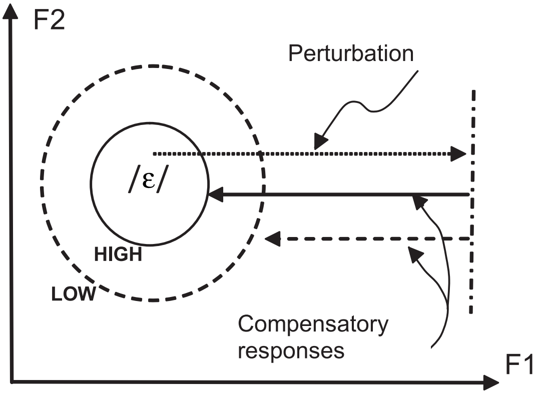


Fig. 5. Schematic diagram of goal regions and compensatory responses for /ɛ/ for a high-acuity speaker (solid circle) and a low-acuity speaker (dashed circle). F1 perturbation is indicated by the dotted arrow, and compensatory responses, by the solid and dashed arrows.

white tracks of F1 and F2 vs. time. Fig. 6B shows a schematic diagram of such a trajectory (Average trajectory) in the F1x F2 plane. The sets of arrows illustrate “inflate” and “deflate” perturbation vector fields, which had zero magnitudes at the beginning and end of the triphthong and maximum magnitudes at the point of greatest F1 excursion. The general format of the experiment was the same as in Villacorta et al. (2007). Twelve subjects experienced inflate perturbations; eleven subjects, deflate perturbations. The experiment was divided into baseline (called *start*), ramp, full-perturbation (*stay*) and no-perturbation (*end*) phases. Resulting trajectories, averaged across subjects, are shown for responses to the inflate perturbation in Fig. 6C and to the deflate perturbation in 6D. The average trajectories show compensation by both groups: decreased trajectory curvature in response to the inflate perturbation and increased curvature in response to the deflate perturbation. In both cases, the average *end* trajectories (no-perturbation) fall between the *start* and *stay* trajectories, indicating some temporary modification of feedforward commands as observed by Villacorta et al. (2007). Even though the differences between the averaged trajectories is relatively small, there were statistically significant effects of phase in both the *deflate* and *inflate* groups. These results are compatible with those of other sensorimotor adaptation studies and with DIVA’s specification of auditory goal regions for vowels in multi-dimensional auditory-*temporal* space.

Both Cai et al. (submitted for publication) and Villacorta et al. (2007) also included presentations of additional words which the subjects pronounced only in the presence of masking noise that effectively blocked their auditory feedback of vowel quality. These words contained vowels that were different from the ones to which the subjects adapted in response to formant shifts with auditory feedback. In the Villacorta et al. (2007) study, the other vowels were /i/, /æ/, /I/, /a/, /o/ and /^/. In Cai et al. (submitted for publication), the masked speech productions comprised Mandarin monophthongs and diphthongs that are sub-parts of the triphthong (/a/, /ia/, /au/); a triphthong that has a trajectory like /iau/ but in reverse order (/uai/); a triphthong with the same formant trajectory as /iau/ but a different tone (/iau51/); and a triphthong with a different via point in the formant plane (/iou/). These words were presented one at a time throughout the experiment (and pronounced under masking noise) in blocks that alternated with blocks of the words containing the vowels to which the subjects adapted (pronounced in response to altered feedback). In both studies, the adaptation learned under modified auditory feedback transferred to the different vowels that were pronounced only with hearing masked. The magnitude of the transfer generally decreased with increasing dissimilarity between the perturbed vowels and the unperturbed ones. (Also, see Houde & Jordan, 1998, 2002.) These results indicate a generalized effect of the formant shifts in auditory feedback that cannot be simulated with the current version of DIVA, which represents different sounds, syllables and words as independent entities. This effect will have to be taken into account in a future revision of the model.

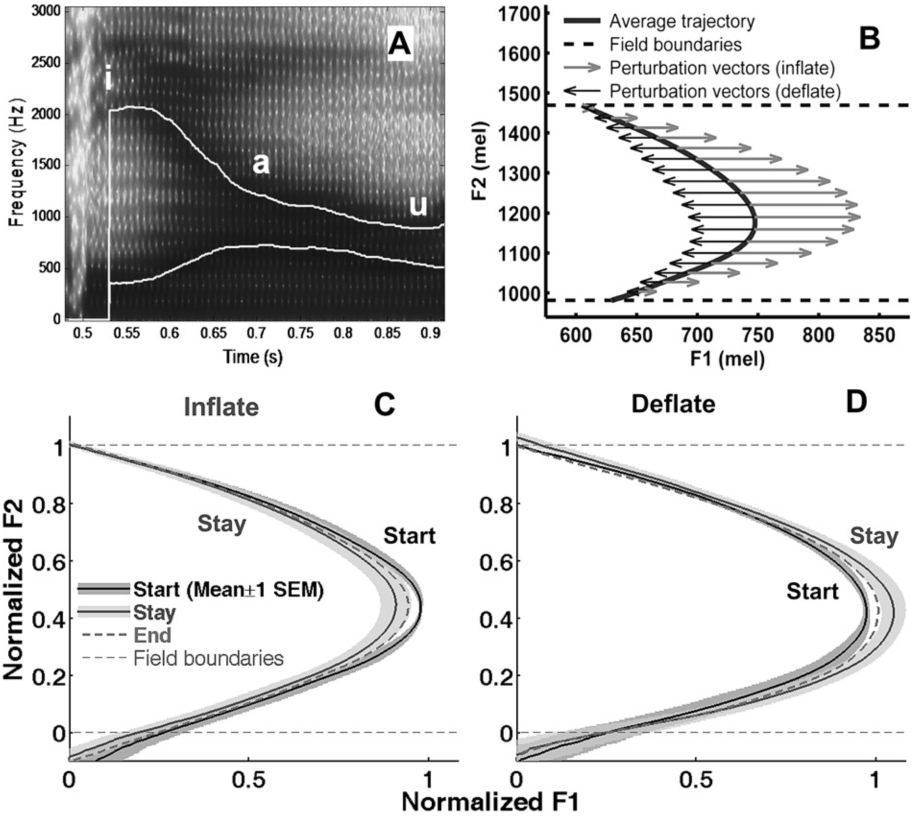


Fig. 6. A: A spectrogram of a token of /iau/ with white tracks of F1 and F2 vs. time. B: A schematic diagram of such a trajectory (Average trajectory) in the F1 × F2 plane. The sets of arrows illustrate “inflate” and “deflate” perturbation vector fields. C: Resulting trajectories, averaged across subjects, for responses to the inflate perturbation in panel C. D: Subject responses to the deflate perturbation.

A recent SA study by Shiller, Sato, Gracco, and Baum (2009) has not only shown compensatory changes in production of a speech sound, but also in its perception. This study was also unique in focusing on the production and perception of the consonant /s/. Shiller et al. (2009) used an apparatus that shifted the entire frequency spectrum downward during speakers’ repeated, prolonged productions of /sV/ utterances in a paradigm much like the above-described SA experiments for vowels. The effect of the perturbation was to decrease the spectral mean for /s/ toward that of /ʃ/. (The perturbation also lowered the fundamental frequency of the following vowel [15]). Shiller et al. determined their subjects’ perceptual /s/-/ʃ/ category boundaries with a labeling task (using synthetic stimuli varying in nine steps between /s/ and /ʃ/). The labeling task was performed before and after the SA paradigm, to look for possible category boundary shifts. There were three groups of 10 female subjects, all of whom performed the labeling task. Only two of the groups performed the production experiment. Altered Feedback (AF) subjects received perturbed auditory feedback. They compensated and adapted (showing carry-over when the perturbation was removed) by shifting the spectral mean of their /s/ productions upward away from /ʃ/ (opposite to the perturbation), presumably to maintain contrast distance between the two sounds. They also showed a category boundary shift toward /ʃ/ (in the same

[15] Shiller et al. (2009) do not discuss the possible effects of lowering F0 of the following vowel, nor the fact that the uttered sibilants were very long, about 600 ms.

direction as the SA perturbation) between the labeling test administered before the production experiment and the one following it. This shift effectively reduced the discrepancy between the perturbed acoustics of /s/ and its perceptual category representation by increasing the perceptual distance between the category boundary and the altered /s/ sounds heard by the subjects. Unaltered Feedback (UF) subjects received auditory feedback that was unperturbed. They showed no production effect, but exhibited a category boundary shift in the direction opposite to the AF subjects – toward /s/ – similar to a number of earlier studies that showed "selective adaptation" (cf. Eimas & Corbit, 1973). The third subject group, passive listeners, took the labeling test before and after hearing the utterances produced by a member of the AF group; they showed no categorical boundary shift.

These findings have important implications for the current theoretical framework. As Shiller et al. (2009) observe, their results appear to demonstrate plasticity within the auditory system in addition to adjustment of motor commands. Both kinds of plasticity, in perception and in production, may help reduce a speaker's perceived error between expected and produced auditory sensations. If this were the case, motor compensation would not have to be complete, as has been observed so far in all such SA experiments. As mentioned previously, incomplete motor compensation may also be explained partly by functionality inherent in the DIVA model: as motor commands and articulatory movements are modified to help correct auditory errors, discrepancies are introduced between expected and resulting somatosensory feedback. Compensating for these somatosensory errors would work in opposition to compensating for auditory errors. Only motor adaptation has been demonstrated thus far in simulations with the DIVA model; currently the model's auditory representation is non-adaptive (although see Guenther & Bohland, 2002 and Guenther & Gjaja, 1996 for neural models of plasticity within auditory cortical areas).

An additional observation with implications for our theory can be drawn from Shiller et al.'s (2009) results: speakers are able to make relatively small changes in the spectral mean of /s/. This observation may seem at odds with the hypothesized saturation effect for /s/ described above in Sect. 2.2.1. However, the use of a saturation effect for /s/ does not preclude variation in other aspects of the articulation, such as configuration of the groove in the tongue blade and the size and shape of the small resonant cavity formed by the lips; both kinds of articulatory variation could produce the observed within-category spectral modifications.

4.2.3. Compensatory adaptation of jaw-movement trajectories to persistent somatosensory perturbations

The experiments on feedback and feedforward control described thus far have focused on the roles of those mechanisms in achieving phonemic sensory goals for specific speech sounds. In a somewhat different vein, Ostry and colleagues have performed a series of studies of the role of somatosensory feedback in the control of opening and closing movements of the mandible (Nasir & Ostry, 2008; Tremblay, Houle, & Ostry, 2008; Tremblay, Shiller, & Ostry, 2003). Although the lower incisors can serve to enhance fricative noise when they are in the path of the airstream anterior to constrictions for the sibilants (Shadle, 1991), the mandible has little or no direct effect on the vocal-tract shape and resonant properties for other speech sounds. However, it does have an important indirect role in helping to position the tongue and lower lips (which are attached to the mandible) for many vowels and consonants. Mandible movements can thereby be influenced simultaneously by coarticulation of several sounds in a sequence, each of which mainly involves the actions of one or two primary articulators such as the lips, the tongue blade and the tongue body.

Ostry et al. used a computer-controlled robot to apply relatively small forward-directed perturbations to mandibular movements, perpendicular to the path of its opening and closing movements in the midsagittal plane. The magnitude of the perturbation depended on the velocity of the opening and closing movements, resulting in a perturbation field that looked similar to a less-curved version of the *inflate* acoustic perturbation used by Cai et al. (submitted for publication, Fig. 6B, above). The jaw-movement perturbation experiments were divided into phases much like acoustic sensorimotor adaptation experiments (approximately analogous to the above-described baseline, ramp, full perturbation and post-perturbation phases). Subjects in experiments by Tremblay et al. (2003) produced repetitions of the utterance /siæt/ with and without voicing, as well as non-speech opening and closing jaw movements. Compensation and adaptation of mandibular opening and closing trajectories (like flatter versions of the acoustic results of the auditory experiments of Cai et al., submitted for publication, shown in Fig. 6C) were found for both the vocalized and silent speech movements, but not for the non-speech movements. There were no measurable effects on the acoustics of the vocalized vowels, which is not surprising considering the presumably minor, indirect effect of small variations in the antero-posterior position of the jaw on the vocal-tract area function. In a later study, Tremblay et al. (2008) used three groups of subjects who pronounced different /sVs/ utterances in the same kind of paradigm, but also included testing for transfer of compensation and adaptation from a perturbed vowel to an unperturbed one. They found compensation and small amounts of adaptation for the perturbed vowels, but no transfer to unperturbed vowels. Nasir and Ostry (2008) ran the paradigm on groups of speakers with profound hearing loss and hearing controls and found compensation and adaptation in both groups.

Taken together, these findings provide evidence of somatosensory targets and feedforward control of jaw movement trajectories in the repeated productions of specific sound sequences, but not in the production of non-speech movements for which there are no phonetically meaningful targets (cf. Nelson, Perkell, & Westbury, 1984). However, unlike the auditory-based results cited earlier (Sect. 4.2.2), the effect on speech movements of the mandible did not transfer (under masking noise) from utterances with the jaw perturbed to different, unperturbed utterances. This lack of transfer may be related to the lack of meaningful acoustic-phonetic targets for the mandible, and its lower-level role in supporting sequences of overlapping movements of the lips, tongue blade and tongue body. In contrast, the auditory-based transfer of acoustic effects from perturbed to unperturbed speech sounds reflects the function of a higher-level feedforward control system for reaching linguistically relevant auditory goals (controlled variables) with movements of those primary articulators.

4.3. Feedforward control and sequencing of words

Another very important component of speech motor control that has not yet been explored in much depth is the set of mechanisms by which phonemic and larger units are sequenced and concatenated. Some insight into one aspect of this process was provided by a study of a single subject who pronounced word pairs such as *perfect memory* while his speech was being recorded and movements of points on his tongue were being captured by an x-ray microbeam (Fujimura, Kiritani, & Ishida, 1973). The word pairs were pronounced in two different ways, as a sequence of two separate words from a list, e.g., *perfect, memory* or as though they were part of a fluent continuous phrase, *perfect memory*. (1) In the list productions, (a) largely overlapping closing movements of the tongue body were evident for the /k/ and /t/ in *perfect* and a separate, later movement was evident for lower-lip closing for the /m/ in *memory* and (b) the corresponding acoustic signal contained a clear (audible) release burst for the /t/ at the end of *perfect*. (2) In the phrasal production, (a) there were still closing gestures for all three consonants, (b) the closing movements for the three consonants overlapped considerably, and (c) because of the overlap of the /m/ with the preceding /t/, there was no audible release burst for /t/. The finding of a /t/ gesture in the phrasal production that did not produce any audible acoustic cues has been interpreted as evidence that the basic underlying phonemic units are articulatory gestures (Browman & Goldstein, 1990).

We repeated this study by recording the articulatory movements and acoustic signals from 21 speakers as they pronounced the same kind of utterances materials (Tiede, Perkell, Zandipour, & Matthies, 2001; Tiede et al., 2007) and found that they all produced the same patterns that were observed by Fujimura et al. (1973), viz., inaudible tongue tip gestures for /t/ at the end of the first word in a smoothly concatenated pair. An alternative to the gestural explanation of these remarkably consistent results is possible under the theoretical framework modeled by DIVA. From this perspective, each frequently used word becomes encoded and stored as a cohesive, fluent sequence of feedforward commands; these "motor plans" are concatenated when words are produced in succession in running speech. Because of biomechanical limitations (particularly during fluent speech), the acoustic correlates of these motor plans might not be fully realized. Thus, it would not be surprising to find inaudible gestures at the boundaries of concatenated words if they overlap when being produced in sequences in natural speech, as found by Fujimura et al. (1973) and Tiede et al. (2001, 2007).

5. Summary

This paper is concerned with the nature of phonemic goals in speech production and the roles of auditory and somatosensory feedback and feedforward control in achieving those goals. Results of a number of studies are consistent with the following summary.

Phonemic goals for speech movements, the highest-level controlled variables for speech sounds, consist of regions that are defined in multi-dimensional auditory-temporal and somatosensory-temporal spaces. It is generally held that properties of speakers' production and perception mechanisms, especially ones that contribute to nonlinear relations between articulatory movements and the resulting acoustics, help to define sound systems for languages by allowing for the reliable production of relatively stable, perceptually contrastive acoustic cues despite some variation or imprecision in some underlying motor commands. Languages would tend to select and retain such speech sounds because of their acoustic stability; however, it is likely that there are additional influences on phonemic inventories that are not quantal in nature, such as a compromise between clarity and economy of effort and social factors. It is also likely that some components of articulatory commands have to be relatively precise.

Saturation and quantal effects can also be used by individual speakers to help simplify motor control for the production of intelligible speech. However, speakers differ in the shapes of their vocal tracts so they may differ as well in how they conform to the language-specific auditory goals for speech sounds have evolved under a number of influences, ranging from clarity and effort constraints to social factors.

Evidence for stabilization of the acoustic output with speakers' use of saturation and quantal effects has been found in experiments on the production of the sibilant /s/ and the vowel /i/. These results are thought to reflect somatosensory goals for the two sounds, consisting partly of patterns of contact of the tongue tip with the lower alveolar ridge for /s/ and contact of the stiffened tongue blade with the lateral aspects of the hard palate for /i/. Evidence of motor control mechanisms for stabilization of the acoustic output has been found in the form of motor-equivalent trading relations between tongue-body raising and lip rounding for the vowel /u/ and between bunched and retroflexed configurations of the tongue for the semivowel /r/. These findings are thought to reflect the use of auditory goals for those two sounds.

Auditory feedback is crucial for the normal acquisition of fluent speech. If an individual learns to speak with adequate hearing and becomes deaf as an adult, that person's speech usually remains intelligible for decades, reflecting the speaker's having a robust set of feedforward commands that was learned with the use of auditory feedback. Nevertheless, the speech of such individuals and presumably the underlying feedforward commands often degrade somewhat with time. If hearing is then restored with a cochlear implant, parallel ongoing improvements are observed in the speaker's accuracy of perception and clarity of produced speech. An example of this process was demonstrated in a study of allophonic variation in the production of bunched and retroflexed versions of /r/. Prior to receiving a cochlear implant, postlingually deafened speakers with profound hearing loss showed greater than normal variability of acoustic measures of their productions of the two allophones; after a year of implant use the variability diminished to within normal ranges.

Several of the cited studies investigated the nature of sensory goals by examining cross-subject relations between production and perception. Measures of speakers' produced vowel and sibilant contrasts were compared with measures of their auditory acuity (1/JND) for those contrasts. For both sound types, speakers with greater acuity tended to produce larger contrasts. These findings led to the inference that speakers with higher auditory acuity have auditory goal regions that are smaller and spaced further apart than speakers with lower acuity. In production of the sibilants, speakers who showed more consistent contact between the tongue tip and lower alveolar ridge for /s/ but not /ʃ/ had larger sibilant contrasts than speakers who did not consistently show this contact difference between the two sounds. This finding is compatible with the hypothesis that such contact for /s/ corresponds to a somatosensory goal that involves the use of a saturation effect. Measures of speakers' produced sibilant contrasts and the somatosensory acuity of their tongue tips also showed a cross-speaker correlation: those with greater acuity in the somatosensory domain produced greater contrasts.

It is hypothesized that all sounds have goals in both sensory domains. Most vowels and vowel-like sounds have more prominent auditory goals. Consonants, for which contact at their primary places of articulation is a defining property, have more prominent somatosensory goals. The sibilants have prominent goals in both domains: turbulence noise with distinct acoustic spectra and patterns of articulatory contact. During speech sound acquisition with auditory feedback, somatosensory goals are acquired as patterns of somatosensory sensation that accompany successful implementation of auditory goals. If profound hearing loss occurs after robust feedforward control of fluent speech has been acquired, the use of somatosensory goals, which are not affected directly by hearing loss, may provide an additional, important means of maintaining intelligibility.

Unexpected modifications of speakers' somatosensory feedback and of their auditory feedback have shown closed-loop error correction; however, the latencies of those corrections (120–150 ms) make it improbable that closed-loop feedback control is used in the production of mature fluent speech, which contains many sounds that are considerably shorter than such latencies. Therefore, auditory feedback must be used mainly for acquiring and maintaining the feedforward control of speech sounds, syllables and words. Under most circumstances, fluent speech production rarely contains errors (mismatches between expected and produced auditory or somatosensory sensations) large enough to require feedback correction. One real-life exception to this observation occurs when people first receive full dentures, in which case auditory feedback control (if available) can be used successfully to correct and update feedforward commands over a period of several days.

Further insight into interactions between feedback and feedforward control has been provided by a series of sensorimotor adaptation experiments, in which shifts of vowel formants or sibilant spectra were introduced gradually in speakers' auditory feedback as they produce many repetitions of textually prompted utterances. The speakers compensated by producing sounds with acoustic parameters shifted in the opposite direction without being aware of the perturbation or anything unusual in their responses. This has been shown for steady state American-English vowels, time-varying Mandarin-Chinese triphthongs and the Canadian-English sibilant /s/. In the case of the steady state vowels, the amount of compensation was positively correlated with speaker acuity, which adds support to the idea that more acute speakers have smaller goal regions for vowels that are spaced further apart. With all three sound types, speakers' compensation persisted for a while after the perturbation was removed, demonstrating adaptation – a temporary modification of feedforward commands. When vowels other than the perturbed ones were pronounced during those experiments in the presence of masking noise, they were modified by the speakers (although to a lesser extent) in the same way as the perturbed ones. Thus, there appeared to be a generalized adaptation of feedforward commands for vowels in response to feedback modification of only one vowel in the speaker's auditory space. Speakers' compensation and adaptation to acoustic modification of /s/ spectra was accompanied by shifts in their perceived category boundary between /s/ and/ʃ/; the sibilant production and perception effects both tended to help maintain the speakers' perceived relations between production and perception.

Compensation and adaptation has been demonstrated in opening and closing movements of the mandible in response to somatosensory perturbations: small anterior displacements of jaw position with a computer-controlled robot. However, the perturbations did not affect vowel acoustics, and the movement compensation and adaptation did not generalize to movements for utterances containing different vowels. These findings reflect the main role of the mandible in speech: supporting the phoneme-specific acoustically critical movements of the tongue body, tongue blade and lower lip, which may overlap (coarticulate) with one another in productions of sequences of linguistically meaningful sounds.

Finally, productions have been observed of phonemically significant word-final articulatory gestures (e.g., raising of the tongue tip for /t/) that did not generate an audible acoustic cue (a release burst), because the word-final gesture was overlapped by a word-initial gesture of a following word when the two words were concatenated in production. This observation is consistent with DIVA's storage and output of the motor representation of each of the words as a continuous stream of feedforward commands for movements that could overlap at word boundaries when the words are pronounced naturally as part of a fluent sequence.

By necessity, findings like those described in this paper are from relatively artificial laboratory situations and not from natural, running speech. Experiments on speech motor control often require multiple repetitions of a small corpus of utterances that are recalled or read by subjects from textual prompts and may include speaking more rapidly or slowly than normal. In addition, the instrumentation not only includes a microphone, but sometimes a set of headphones that may transmit modified auditory feedback or block it with masking noise, and often devices that are inserted in the mouth, or glued to the articulators, or placed so they purposefully displace articulators. Experimenters usually try to control for such disturbing influences as much as possible, but there is no way that the data collected under such circumstances can be considered to be from natural speech.

In spite of such shortcomings the findings reviewed here demonstrate a coherent set of mechanisms and principles that are likely to be operating in natural speech. The hypothesized mechanisms and principles are, furthermore, consistent with the function of a neurocomputational model of speech motor control which has been implemented in computer simulations, DIVA, that can convert phonemic inputs to fluent sequences of articulatory movements and sound. This model serves as a basis of hypotheses for future experiments, the results of which can either validate the model's function or provide information to be used in revising the model. Since the model's components and function have correlates in brain activity, it has also been tested with brain-imaging experiments, some of which have investigated the same phenomena that have been demonstrated by the behavioral findings described above. For example, Tourville, Reilly, and Guenther (2008) performed an auditory perturbation, functional magnetic resonance imaging (fMRI) experiment that corroborated psychophysical findings of compensation for the same kind of perturbation, while also identifying the brain regions involved in the compensation (i.e., the auditory feedback network). (Also see Guenther and Vladusich, this issue.) Thus the model makes it possible to investigate the neural mechanisms underlying quantitative observations of linguistically meaningful articulatory movements and sounds from individuals with normal and disordered speech production.

Acknowledgements

The research described in this paper that was performed in our laboratory has been supported mainly by two grants from the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health: R01-DC01925, "Constraints and Strategies in Speech Production" and R01-DC03007, "Effects of Hearing Status on Adult Speech Production." This work was performed with the indispensible collaboration of a number of current and former members and affiliates of the Speech Communication Group of the Research Laboratory of Electronics at M.I.T. Current collaborators on the research include Shanqing Cai, Satra Ghosh, Frank Guenther, Miwako Hisagi, Harlan Lane, Melanie Matthies, Lucie Ménard, Ken Stevens and Mark Tiede. Previous collaborators include Suzanne Boyce, Marc Boucek, Marc Cohen, Margaret Denney, Bob Hillman, Eva Holmberg, Kiyoshi Honda, Michael Jordan, Nicole Marrone, Emi Murano, Win Nelson, Pascal Perrier, Joyce Rosenthal (formerly Manzella), Mario Svirsky, Reiner Wilhelms-Tricarico, Jennell Vick, Virgilio Villacorta, John Westbury, Jane Wozniak (formerly Webster) and Majid Zandipour. We have also received invaluable advice from Ken Stevens on acoustic phonetics; help with data collection and analysis from Ellen Burton (formerly Stockmann), John Gould, Peter Guiod and Alexandra Hanson; computer systems and technical support from Seth Hall and administrative support from Arlene Wint. Cochlear implant users were referred by Don Eddington at the Massachusetts Eye and Ear Infirmary. This manuscript has benefitted from helpful comments and suggestions by an anonymous reviewer.

References

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures – orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology, 51*, 705–723.

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.

Baum, S. R., & McFarland, D. H. (1997). The development of speech adaptation to an artificial palate. *Journal of the Acoustical Society of America, 102*, 2353–2359.

Beckman, M. E., Jung, T., Lee, S., de Jong, K. J., Krishnamurthy, A. K., Ahalt, S. C., et al. (1995). Variability in the production of quantal vowels revisited. *Journal of the Acoustical Society of America, 97*, 471–490.

Bradlow, A. R., Kahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. *Perception and Psychophysics, 61*, 977–985.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251.

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

Buchaillard, S., Perrier, P., & Payan, Y. (2009). A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning. *Journal of the Acoustical Society of America, 126*, 2033–2051.

Burnett, T. A., & Larson, C. R. (2002). Early pitch-shift response is active in both steady & dynamic voice pitch control. *Journal of the Acoustical Society of America, 112*, 1058–1063.

Cai, S., Ghosh, S. S., Guenther, F. H., Perkell, J. S. Adaptive auditory feedback control of the production of the formant trajectories in the Mandarin triphthong /iau/ and its patterns of generalization, submitted for publication.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, MA.: MIT Press.

Cowie, R. I. D., & Douglas-Cowie, E. (1983). Speech production in profound postlingual deafness. In M. Lutman, & M. P. Haggard (Eds.), *Hearing science & hearing disorders* (pp. 183–230). London: Academic Press.

Cowie, R. I. D., & Douglas-Cowie, E. (1992). *Postlingually acquired deafness: Speech deterioration & the wider consequence*. New York: Hawthorne.

Delattre, P., & Freeman, D. C. (1968). A dialect study of American /r/ by x-ray motion picture. *Linguistics, 44*, 29–68.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*, 99–109.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience, 15*, 399–402.

Ferrari, P. F., Fogassi, L., Gallese, V., & Rizzolatti, G. (2003). Mirror neurons responding to the observation of ingestive & communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17*, 1703–1714.

Flege, J. E., Fletcher, S. G., & Homeidan, A. (1988). Compensating for a bite block in /s/ & /t/ production: palatographic, acoustic, perceptual data. *Journal of the Acoustical Society of America, 83*, 212–228.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3–28.

Fowler, C. A., & Turvey, M. T. (1980). Immediate compensation in bite-block speech. *Phonetica, 37*, 306–326.

Fry, D. B. (1955). Duration & intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27*, 765–768.

Fujimura, O., & Kakita, Y. (1979). Remarks on quantitative description of lingual articulation. In B. Lindblom, & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 17–24). San Diego: Academic Press.

Fujimura, O., Kiritani, S., & Ishida, H. (1973). Computer controlled radiography for observation of movements of articulatory & other human organs. *Computers in Biology & Medicine, 3*, 371–384.

Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A, Tiede, M., Ménard, L., Guenther, F. H., Lane, H., Perkell, J. S. An investigation of the relation between sibilant production & somatosensory & auditory acuity. *Journal of the Acoustical Society of America*, submitted for publication.

Gomi, H., & Honda, M. (2002). Compensatory articulation during bilabial fricative production by regulating muscle stiffness. *Journal of Phonetics, 30*, 261–279.

Groenen, P., Maassen, B., Crul, T., & Thoonen, G. (1996). The specific relation between perception & production errors for place of articulation in developmental apraxia of speech. *Journal of Speech & Hearing Research, 39*, 468–482.

Guenther, F. H. (1994). A neural network model of speech acquisition & motor equivalent speech production. *Biological Cybernetics, 72*, 43–53.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, rate effects in a neural network model of speech production. *Psychological Review, 102*, 594–621.

Guenther, F. H., & Bohland, J. W. (2002). Learning sound categories: a neural model and supporting experiments. *Acoustical Science and Technology, 23*, 213–220.

Guenther, F. H., Espy-Wilson, C., Boyce, S., 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., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling & imaging of the cortical interactions underlying syllable production. *Brain & Language, 96*, 280–301.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America, 100*, 1111–1121.

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.

Guenther, F.H., & Vladusich, T. A neural theory of speech acquisition and production, Journal of Neurolinguistics, this issue.

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).

Honda, M., Fujino, A., & Kaburagi, T. (2002). Compensatory responses of articulators to unexpected perturbation of the palate shape. *Journal of Phonetics, 30*, 281–302.

Honda, M., Murano, E. Z. (2003). Effects of tactile & auditory feedback on compensatory articulatory response to an unexpected palatal perturbation. In: *Proceedings of the 6th Speech Production Seminar, Sydney* (pp. 97–100).

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279*, 1213–1216.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: compensation & adaptation. *Journal of Speech, Language, Hearing Research, 45*, 295–310.

Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis*. Cambridge, MA.: MIT Press.

Jamieson, D. G., & Rvachew, S. (1992). Remediating speech production errors with sound identification training. *Journal of Speech Language Pathology & Audiology, 16*, 201–210.

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.

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: the role of auditory feedback. *Journal of the Acoustical Society of America, 113*, 532–543.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception & Performance, 10*, 812–832.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297*, 846–848.

Labov, W. (1966). *The social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

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

Lane, H., Denny, M., Guenther, F. H., Matthies, M. L., Ménard, L., Perkell, J. S., et al. (2005). Effects of bite blocks & hearing status on vowel production. *Journal of the Acoustical Society of America, 118*, 1636–1646.

Lane, H., Denny, M., Guenther, F. H., Matthies, M. L., Perkell, J. S., Stockmann, E., et al. (2007). On the structure of phoneme categories in listeners with cochlear implants. *Journal of Speech, Language & Hearing Research, 50*, 2–14.

Lane, H., Matthies, M., Denny, M., Guenther, F. H., Perkell, J. S., Stockmann, E., et al. (2007). Effects of short- & long-term changes in auditory feedback on vowel & sibilant contrasts. *Journal of Speech Language & Hearing Research, 50*, 913–927.

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

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F0 responses to pitch-shift onset & offset conditions. *Journal of the Acoustical Society of America, 110*, 2845–2848.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge: MIT Press.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech revised. *Cognition, 21*, 1–36.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production & speech modeling* (pp. 403–439). Netherlands: Kluwer Academic Publishers.

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 & a model of speech motor programming by predictive simulation. *Journal of Phonetics, 7*, 147–161.

Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, larynx movement. *Journal of the Acoustical Society of America, 50*, 1166–1179.

Löfqvist, A., & Gracco, V. L. (1997). Lip & jaw kinematics in bilabial stop consonant production. *Journal of Speech Language & Hearing Research, 40*, 877–893.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences, 13*, 110–114.

Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J. S., Burton, E., Vick, J., et al. (2008). Perception & production of /r/ allophones improve with hearing from a cochlear implant. *Journal of the Acoustical Society of America, 124*, 3191–3202.

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

Max, L., Wallace, M. E., Vincent, I. (2003). Sensorimotor adaptation to auditory perturbations during speech: acoustic & kinematic experiments. In: *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1053–1056).

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.

Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience, 10*, 1217–1222.

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.

Nelson, W. L., Perkell, J. S., & Westbury, J. R. (1984). Mandible movements during increasingly rapid articulations of single syllables: preliminary observations. *Journal of the Acoustical Society of America, 75*, 945–951.

Newman, R. S. (2003). Using links between speech perception & production to evaluate different acoustic metrics. *Journal of the Acoustical Society of America, 113*, 2850–2860.

Perkell, J. S., Boyce, S. E., & Stevens, K. N. (1979). Articulatory & acoustic correlates of the /s-sh/ distinction. Papers presented at the 97th meeting of the Acoustical Society of America. In J. J. Wolf, & D. H. Klatt (Eds.), *Speech communication* (pp. 109–113). New York: American Institute of Physics.

Perkell, J. S., Cohen, M., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. T. (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., Denny, M., Lane, H., Guenther, F. H., Matthies, M. L., Tiede, M., et al. (2007). Effects of masking on vowel & sibilant contrasts in normal-hearing speakers & postlingually deafened cochlear implant users. *Journal of the Acoustical Society of America, 121*, 505–514.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., et al. (2000). A theory of speech motor control & supporting data from speakers with normal hearing & with profound hearing loss. *Journal of Phonetics, 28*, 233–372.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., & Tiede, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America, 116*, 2338–2344.

Perkell, J. S., Lane, H., Denny, M., Matthies, M. L., Tiede, M., Zandipour, M., et al. (2007). Time course of speech changes in response to short-term changes in hearing state. *Journal of the Acoustical Society of America, 121*, 2296–2311.

Perkell, J. S., Lane, H., Ghosh, S. S., Matthies, M. L., Tiede, M., Guenther, F. H., et al. (2008). Mechanisms of vowel production: auditory goals & speaker acuity. In: *Proceedings of the Eighth International Seminar on speech production, Strasbourg, France* (pp. 29–32).

Perkell, J. S., Matthies, M. L., Ghosh, S. S., Maas, E., Hanson, A., Guenther, F. H., et al. (2008). Auditory & somatosensory goals for sibilants. *Journal of the Acoustical Society of America, 123*, 3459, (A).

Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F. H., Wilhelms-Tricarico, R., Wozniak, J., et al. (1997). Speech motor control: acoustic goals, saturation effects, auditory feedback & internal models. *Speech Communication, 22*, 227–250.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising & 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., Tiede, M., Lane, H., Zandipour, M., Marrone, N., et al. (2004). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination & use of an articulatory saturation effect. *Journal of Speech, Language & Hearing Research, 47*, 1259–1269.

Perkell, J. S., & Nelson, W. L. (1985). Variability in production of the vowels /i/ & /a/. *Journal of the Acoustical Society of America, 77*, 1889–1895.

Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H. (2002). Economy of effort in different speaking conditions I: a preliminary study of intersubject differences & modeling issues. *Journal of the Acoustical Society of America, 112*, 1627–1641.

Plant, G. (1984). The effects of an acquired profound hearing loss on speech production. *British Journal of Audiology, 18*, 39–48.

Purcell, D. W., & Munhall, K. G. (2006a). Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America, 119*, 2288–2297.

Purcell, D. W., & Munhall, K. G. (2006b). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *Journal of the Acoustical Society of America, 120*, 966–977.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trend in Neurosciences, 21*, 188–194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech & Hearing Research, 37*, 347–357.

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.

Shadle, C. H. (1991). The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics, 19*, 409–424.

Shaiman, S. (1989). Kinematic & electromyographic responses to perturbation of the jaw. *Journal of the Acoustical Society of America, 86*, 78–88.

Shiller, C. M., Sato, M., Gracco, V. L., & Baum, S. (2009). Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America, 125*, 1103–1113.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3–46.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language, 65*, 81–106.

Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics, 38*, 10–19.

Tiede, M., Perkell, J. S., Zandipour, M., & Matthies, M. L. (2001). Gestural timing effects in the 'perfect memory' sequence observed under three rates by electromagnetometry. *Journal of the Acoustical Society of America, 110*, 2657, (A).

Tiede, M., Shattuck-Hufnagel, S., Johnson, B., Ghosh, S. S., Matthies, M. L., Zandipour, M.& et al. (2007). Gestural phasing in /kt/ sequences contrasting within & cross word contexts. In: *Proceedings of the 16th International Conference of Phonetic Sciences (ICPhS07), Saarbrücken, Germany* (pp. 521–524).

Tiede, M. K., Ito, T., Ostry, D. J. (2006). Compensatory response to unexpected jaw perturbation triggered by formant transitions during speech. In: *Proceedings of the Seventh International Seminar on Speech Production, Ubatuba, Brazil* (pp. 217–224).

Tourville, J. A., Guenther, F. H., Ghosh, S. S., Reilly, K. J., Bohland, J. W., Nieto-Castanon, A. (2005). Effects of acoustic & articulatory perturbation on cortical activity during speech production. In: *11th Annual Meeting of the Organization for Human Brain Mapping* (p. S49).

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage, 39*, 1429–1443.

Tremblay, S., Houle, G., & Ostry, D. J. (2008). Specificity of speech motor learning. *Journal of Neuroscience, 5*, 2426–2434.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Letters to Nature, 423*, 866–869.

Van Boven, R. W., & Johnson, K. O. (1994). The limit of tactile spatial resolution in humans: grating orientation discrimination at the lip, tongue, finger. *Neurology, 44*, 2361–2366.

Vick, J., Lane, H., Perkell, J. S., Matthies, M. L., Gould, J., & Zandipour, M. (2001). Speech perception, production & intelligibility improvements in vowel-pair contrasts in adults who receive cochlear implants. *Journal of Speech, Language & Hearing Research, 44*, 1257–1268.

Villacorta, V., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics & its relation to perception. *Journal of the Acoustical Society of America, 122*, 2306–3219.

Wohlert, A. B. (1996). Tactile perception of spatial stimuli on the lip surface by young & older adults. *Journal of Speech & Hearing Research, 39*, 1191–1198.