

The Sensorimotor Control of Speech Production

Joseph Perkell^{1,2,3}, Frank Guenther^{3,1}, Harlan Lane^{4,1}, Melanie Matthies^{5,1}, Yohan Payan⁶
 Pascal Perrier⁷, Jennell Vick¹, Reiner Wilhelms-Tricarico¹, Majid Zandipour¹

1 – Speech Communication Group, Research Laboratory of Electronics, M.I.T., Rm. 36-591,
 50 Vassar St., Cambridge, MA, U.S.A. 1+617 253-3223 (tel.), 1+617 258-7864 (fax); perkell@speech.mit.edu;
 2 – Dept. of Brain and Cognitive Sciences, M.I.T.; 3 – Dept. of Cognitive and Neural Systems, Boston University;
 4 – Dept. of Psychology, Northeastern University; 5 – Dept. of Communication Disorders, Boston University;
 6 – TIMC/GMCAO, La Tronche, France; 7 – Institute de la Communication Parlee, INPG, Grenoble, France

ABSTRACT

A model of the sensorimotor control of speech production is presented. The model is being implemented as a set of computer simulations. It converts an input sequence of discrete phonemes into quasi-continuous motor commands and a sound output. A key feature of the model is that the goals for speech movements, at least for some kinds of sounds, are regions in auditory-temporal space. The model is designed to have properties that are as faithful as possible to data from speakers – including measures of brain function, speech motor control mechanisms, physiology, anatomy, biomechanics and acoustics. Examples of simulations and actual data from some of these domains are presented. The examples demonstrate properties of the model or they are consistent with hypotheses generated from it. Our long-range goal is to implement the model completely and test it exhaustively, in the belief that doing so will significantly advance our understanding of speech motor control.

1. INTRODUCTION

Communication via spoken language is a defining characteristic of human beings, and speech production is very likely the most complicated motor act performed by any species. Speech production entails the coordination of the many degrees of freedom of the respiratory, laryngeal and supraglottal articulatory systems to convert a discretely specified linguistic message to a quasi-continuous stream of sound that can be understood by a listener. In this process, the movements of a number of slowly moving structures are coarticulated to produce sequences of sounds that are transmitted to listeners at rates approaching 15 per second.

Our research is directed at modeling the brain activity and the motor, biomechanical and sensory processes involved in speech production. This multifaceted effort is collaborative in nature; it involves a number of investigators working on inter-related projects that are based at several institutions. Our approach is to use a combination of computational models that are supported and tested with brain imaging, psychophysical, physiological, anatomical and acoustic data. An overview of the current version of the overall model is schematized in Fig. 1. Our long-range goal is to completely implement this model in the form of computer simulations and to test and refine it by comparing simulation results with data from speakers.

2. A MODEL OF SPEECH MOTOR CONTROL

The model consists of a Speech Production System and a Speech Perception System. The model provides that the goals of at least some speech movements consist of regions in auditory-temporal space [1]. To a first approximation, we assume that the non-temporal dimensions of the goal regions can be described by auditorily based transformations of acoustic parameters such as spectral peaks, sound amplitude and fundamental frequency and that such transformations take place at various levels in the auditory-perceptual system. For the most part, we make the simplifying assumption that we can address experimental questions by using acoustic measures; in this paper, we use the terms “auditory” and “acoustic” interchangeably. The model’s major components are shown as

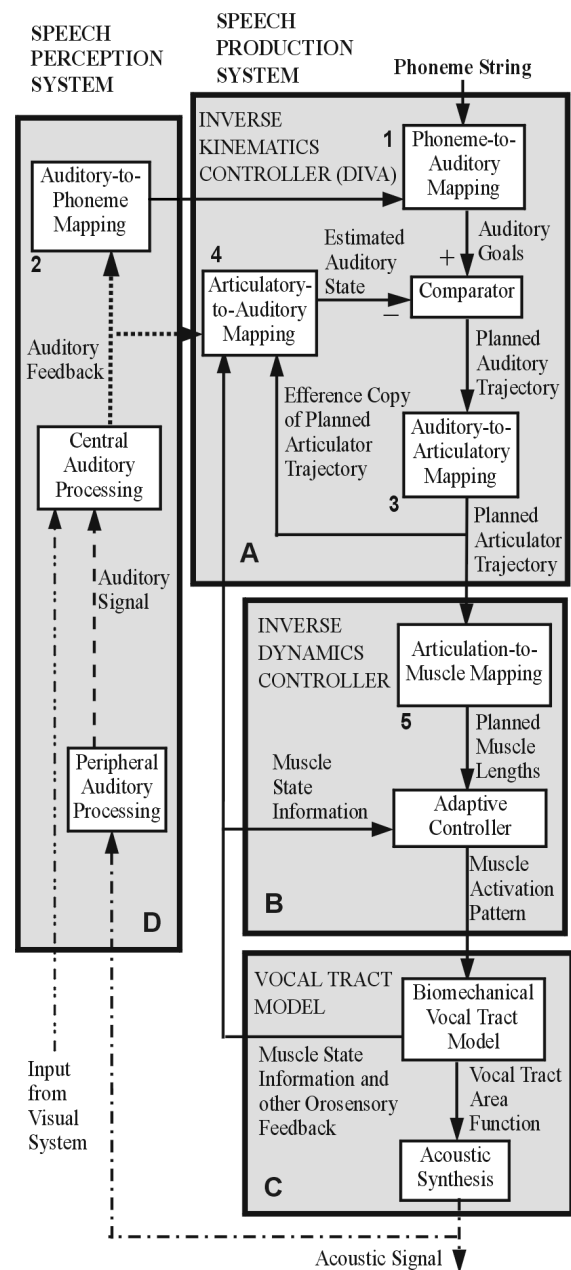


Figure 1: A block diagram of a model of the segmental component of speech motor control and the influence of feedback.

gray boxes in Fig. 1; the components contain several kinds of processing mechanisms (shown as white boxes), five of which are mappings, numbered 1-5.

The *Inverse Kinematics Controller* (component A) is the top-level component of the Speech Production System. It takes as input a phoneme string and outputs a planned temporal se

quence of articulator positions (Planned Articulation Trajectory) to implement the desired phoneme string. The problem of specifying such a kinematic trajectory for a set of articulators in the DIVA model of speech production which is an adaptive neural network model that provides an integrated account of many speech production phenomena, including movement variability due to context, motor equivalence, coarticulation, rate effects and aspects of speech sound acquisition [2, 1, 3].

When a phoneme string is specified to the system for production, a Phoneme-to-Auditory Mapping (block 1 in component A) transforms the phoneme string into a sequence of auditory goal regions. The current position of the vocal tract in auditory space is determined using an Articulatory-to-Auditory mapping (4) that maps orosensory information from the vocal tract and an efference copy of the outflow signal into the auditory signal that is expected to result from the current vocal tract shape. Auditory feedback is used to acquire and later update the Articulatory-to-Auditory mapping, but not to determine the current position of the vocal tract in auditory space – because of prohibitive feedback delays. The difference between the target position in auditory space and the current position in auditory space (calculated by the Comparator in component A) defines the desired movement direction in auditory space. This Planned Auditory Trajectory is then transformed through an Auditory-to-Articulation Mapping (3) into a Planned Articulation Trajectory that is used to achieve the desired auditory space movement.

Mapping 1 is a forward mapping and 2 is an inverse mapping of the same relationship – between phonemes and auditory goals; they are language- and *phoneme-specific*. For each abstract phoneme, there is a unique phonetic target region in auditory-temporal space. Thus, the phonemic auditory goals are independent of phonetic context.

Mappings 3 through 5 are phoneme-independent, or *systemic*; they depend on anatomical and physiological properties of the speaker’s production (and perception) systems, such as the size and shape of the speaker’s vocal tract. We assume that systemic mappings 3 and 5 are forward mappings and 4 is an inverse mapping of the same relationship – between parameters of a speaker’s articulations and their acoustic and auditory consequences. Thus, the phoneme-specific mappings specify *what* the auditory-phonemic goals are in the speaker’s language and the systemic mappings characterize *how* individual speakers use their production mechanisms (tongue raising and lowering, lip rounding, voicing, etc.) to achieve those auditory goals.

The *Inverse Dynamics Controller* (Fig. 1, component B) converts the planned articulation trajectory into a time-varying muscle activation pattern (such as that measured by EMG) that will achieve the desired articulator movements. The problem of specifying a set of muscle or actuator forces to carry out a planned kinematic trajectory is commonly referred to as the inverse dynamics problem in the robotics and motor control literatures. An Articulation-to-Muscle Mapping (5) in the inverse dynamics controller first translates the desired articulation trajectory into a higher-dimensional trajectory of planned muscle lengths (i.e., translates each articulation into target lengths of the several muscles that carry out the articulation). This planned muscle length trajectory is then transformed by an Adaptive Controller into a muscle activation pattern that will drive muscle contractions in the Vocal Tract Model.

The *Vocal Tract Model* (Fig. 1, component C) consists of a Biomechanical Vocal Tract Model that transforms the muscle activation pattern into movements of muscles and tissue in a

tors to carry out a task space goal (here, an auditory space goal) is referred to as the inverse kinematics problem in robotics and motor control. The inverse kinematics controller is a computer-simulated vocal tract. The resulting vocal tract area functions are used to synthesize an acoustic signal at the Acoustic Synthesis stage (in component C). We are currently working with a relatively simple 2-D model and a more complex 3-D model of the vocal tract. The vocal-tract models are designed to incorporate anatomical, physiological and biomechanical properties that are as realistic as possible, which will include adapting them to the anatomy (using MRI and dental casts) and detailed physiological and acoustic data from individual speakers. Those properties are reflected in the systemic mappings 3-5 in the model in Fig. 1.

The implementation of the model is far from complete, although numerous simulations of portions of the model have been carried out and reported elsewhere, e.g. [2, 1, 3]. We believe that this modeling approach has the potential to move us closer to understanding neural processes underlying relations between speech perception and production, and to a coherent account of speech processes in normal speakers and clinical populations. In the remainder of this paper, we present examples of how the model has been helpful in providing a systematic account of our experimental findings, and in formulating hypotheses about underlying mechanisms and predictions concerning outcomes of various interventions with several speaker groups.

3. EVIDENCE FOR TARGET REGIONS

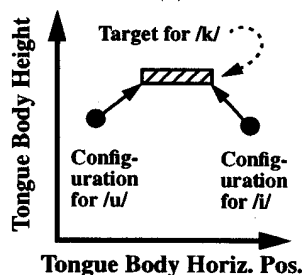


Figure 2: Schematic illustration of a convex-hull target region for /k/ in the DIVA model.

One key component of the DIVA model is the use of phoneme targets that are multidimensional regions, rather than points, in the planning reference frame. Guenther [2] showed how this *convex region theory* for the targets of speech production provides a unified explanation for a wide range of speech production phenomena, including findings on contextual variability, motor equivalence, anticipatory and carryover coarticulation, and speaking rate effects. For example, the model provides a simple and intuitive explanation for carryover coarticulation of tongue body position during production of the /k/ in “luke” and “leak”. It has been noted that English speakers utilize a more posterior tongue position in producing /k/ when it is preceded by a back vowel as compared to /k/ when preceded by a front vowel, e.g. [4]. The convex region theory explanation for this phenomenon is schematized in Fig. 2. When producing a phoneme, the DIVA model moves from the current configuration of the vocal tract to the closest point on the next phoneme’s target region. When the back vowel /u/ precedes /k/ as in “luke”, the tongue body is further back during /k/ than when the front vowel /i/ precedes /k/ as in “leak”.

4. ARTICULATORY EVIDENCE FOR AUDITORY/ACOUSTIC GOALS

We have conducted several experiments in which measurements of articulatory movements have provided evidence that the goals of speech movements are auditory in nature. In order to measure articulatory movements, we have developed the system shown in Fig. 3.

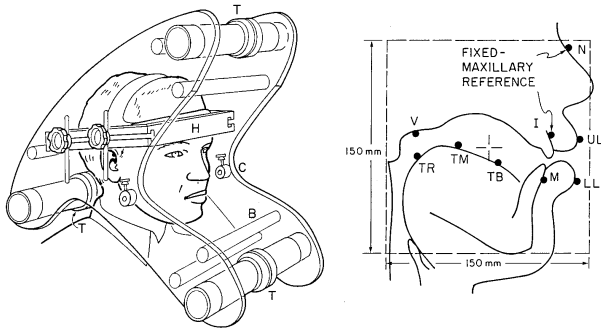


Figure 3: Left half – an Electromagnetic Midsagittal Articulometer movement transducer system. Right half – possible transducer locations.

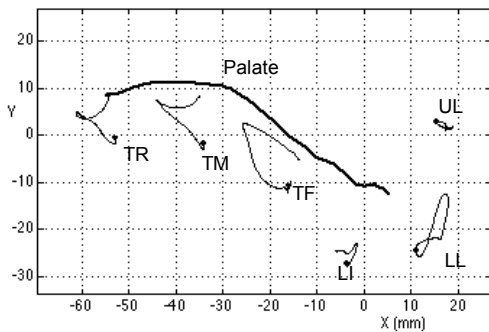


Figure 4: Trajectories for transducers on the tongue rear (TR), tongue mid (TM), tongue front (TF), lower incisor (LI), upper lip (UL) and lower lip (LL) for the sound sequence /ə#put#əg/.

The left half of Fig. 3 shows this ElectroMagnetic Midsagittal Articulometer (EMMA) system [5]. It consists of three transmitter coils, labeled T, mounted in a transmitter assembly that is placed over a subject's head. The right half of the figure shows possible locations of small (2 mm x 5 mm) transducer coils that can be glued to a speaker's articulators in the midline. Alternating magnetic fields generated by the transmitters induce voltages in the transducers; those voltages are converted to locations vs. time in the midsagittal plane. Figure 4 shows an example of transducer trajectories for the sound sequence /ə#put#əg/ (from "a pool again").

Motor equivalence for the vowel /u/ and the liquid /r/

We have obtained evidence for auditory/acoustic goals from speech motor equivalence experiments. The left half of Fig. 5 is a schematic illustration of motor equivalent behavior in production of the vowel /u/ in American English. An /u/ is produced by rounding the lips and raising the tongue body in the back part of the mouth. There is a many-to-one relation between articulations and the vowel acoustics, so, as illustrated by the double-headed arrows, it is possible to produce approximately the same vowel acoustics with more tongue raising and less lip rounding, or vice versa. Finding such a motor-equivalent trading relation between lip rounding and tongue raising in multiple repetitions of /u/ would support the hypothesis that the goal for /u/ is auditory/acoustic, rather than configurational.

The right half of Fig. 5 shows a plot of tongue vertical vs. lip horizontal locations for multiple repetitions of /u/ from a subject, demonstrating the schematized motor equivalent trading relation, which we interpret as support for an acoustic goal for the /u/. We have found such motor equivalence for /u/ in a number of speakers [6]

Figure 6 illustrates motor equivalence in the production of American English /r/ by seven speakers, who pronounced five repetitions of /r/ in different phonetic contexts [7]. The con-

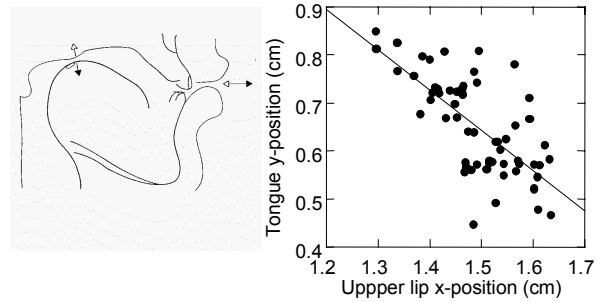


Figure 5: Left half – schematic illustration of the vocal-tract configuration for /u/. The double-headed arrows indicate a motor-equivalent trading relation between tongue raising and lip rounding. Right half – tongue vertical vs. lip horizontal locations for multiple repetitions of /u/, demonstrating the trading relation schematized in the left half.

texts were designed to elicit productions of /r/ between two different tongue configuration extremes: "bunched" as in the utterance /wagrav/, and "retroflexed" as in the utterances /warav/ or /wabrav/. The positions of three EMMA transducers on the tongue were recorded, along with the acoustic signal. Figure 6 shows the resulting articulatory data for the seven subjects (S1-S7). For each subject, the average transducer positions for the five repetitions of /r/ in each of the two contexts are indicated by small symbols. The transducer positions for the /r/ following a velar consonant are connected by solid curves, and those for /r/ in a second context (either /warav/ or /wabrav/, depending on the subject) are connected by dashed curves. The curves are extended as straight-line segments downward and forward from the anterior transducer location, for use in approximating the size of the front cavity.

All seven subjects showed trade-offs between the front cavity length and the constriction length and/or area, which helped to maintain a low F3 for /r/ (the primary acoustic cue) produced with different tongue configurations in the two different contexts. Furthermore, statistical prediction of the time-varying F3 trajectories from the tongue point trajectories showed that inclusion of covariances of the transducer positions significantly reduced the F3 variability. These results indicate that the acoustic goal of a low F3 for /r/ is achieved with different tongue configurations in different contexts, implying that the goal for /r/ is acoustic rather than articulatory.

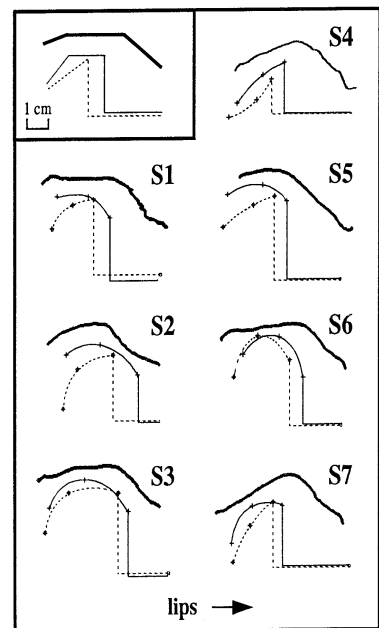


Figure 6: Articulations of /r/ in two different contexts (dotted vs. solid lines, below the palatal contour), produced by 7 subjects (S1-S7).

5. INVERSE DYNAMICS

Work is in progress on the inverse dynamics controller shown in Fig. 1 (component B). As mentioned above, the Inverse Dynamics Controller is responsible for transforming the output of the inverse kinematics controller (a trajectory of

planned articulator movements) into a muscle activation pattern that will drive movements of the vocal tract model to achieve the sequence of acoustic goals corresponding to the phoneme string being produced. The degrees of freedom, or dimensions, of the planned articulation trajectory correspond to quasi-independent articulations (e.g. movements of the tongue body, tongue blade, mandible, lips, and velum), which are implemented by sets of muscles, or “synergies”. The dimensions of this “articulation representation” are based on traditional phonetic descriptions: lip rounding/spreading, lip closing/opening, mandible opening/closing, tongue blade raising/lowering, pharyngeal narrowing/widening and velopharyngeal port opening/closing and two tongue-body articulations.

The articulation representation will have approximately 10 dimensions, compared to the 30+ dimensions of the muscle activation frame that forms the output of the inverse dynamics controller. The lower dimensionality of the articulation representation greatly reduces the complexity of the inverse kinematics control process: the inverse kinematics controller need not be concerned with individual muscle states since that knowledge is incorporated into the inverse dynamics control process. The first stage of the inverse dynamics controller, indicated by the block labeled *Articulation-to-Muscle Mapping* (5, component B, Fig. 1), consists of the decomposition of each articulation into corresponding lengths of the individual muscles that implement the articulation. In other words, this stage transforms the 10-dimensional articulation trajectory into a roughly 30-dimensional muscle length trajectory (the “planned muscle lengths” in Fig. 1). This process will incorporate knowledge of the musculature obtained from the anatomical and physiological literature and basic experiments with the model.

The second stage of the inverse dynamics control process is indicated by the *Adaptive Controller* block in Fig. 1, component B. This stage is responsible for transforming desired muscle lengths into the muscle activation patterns necessary to achieve these lengths in the muscles of the vocal tract model. The muscle activation pattern required to achieve a desired muscle length will depend on the current state of the muscle; this information is provided by the biomechanical vocal tract model and is shown as the *Muscle State Information* pathway in Fig. 1. The different articulations may impose different target lengths for the same muscle, since the articulations are only quasi-independent. We hypothesize that there is a sub-controller for each muscle that will assign a weight to the input from each articulation, to reconcile conflicting length targets.

6. BIOMECHANICAL VOCAL-TRACT MODELS

The inverse dynamics controller is being developed initially using a basic 2-D biomechanical vocal tract model. The performance of the inverse dynamics controller with the 2-D vocal tract model is being tested by providing it with simple inputs, specified in terms of desired articulations, and verifying that the appropriate movements are produced by the biomechanical vocal tract model.

Figure 7 is an illustration of the current version of the 2-D model, which is primarily a model of the tongue. The tongue model (an improved version of the model of Payan and Perrier, [8]) includes the main muscles responsible for shaping and moving the tongue in the midsagittal plane. Elastic properties of the tissues are accounted for by finite-element (FE) modeling of the tongue mesh in 2D defined by 221 nodes and 192 isoparametric elements. Muscles are modeled as force generators that (1) act on anatomically specified sets of nodes of the FE structure, and (2) modify the stiffness of specific elements of the model to account for muscle contractions within tongue tissues. Curves representing the contours of the lips, palate and pharynx in the midsagittal plane are added. The lower jaw and the hyoid bone are represented in this

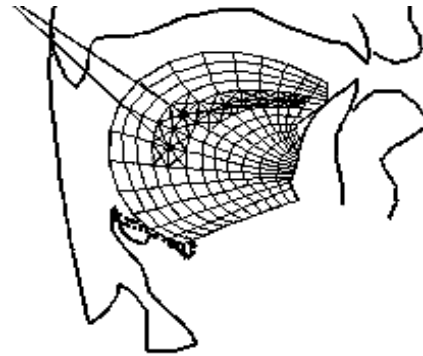


Figure 7: Illustration of a 2-D finite element (FE) model of the tongue, showing the FE mesh, and the styloglossus muscle (hatched elements and two linear force generators connected to the styloid process).

plane by rigid structures to which the tongue is attached. Changes in jaw opening (rotation and translation) and lip opening and protrusion are simulated using a second order system of differential equations.

Influences of vocal-tract anatomy and biomechanics on articulatory kinematics

In an experiment with the earlier version of this model, Payan and Perrier [8] simulated vowel-to-vowel articulatory movements of a speaker of French, using cineradiographic data of the speaker’s tongue movements and vocal-tract contours. Formant trajectories, calculated from simulated tongue shapes and the resultant area function, were compatible with the speaker’s vowel space. The simulation data were compared with the subject’s kinematic and acoustic data. In one example, a bi-phasic (double-peaked) velocity profile observed in movement data from the speaker could be simulated with a constant-rate shift (a simple ramp change) in the control parameter input to the model. It was shown that the bi-phasic nature of the velocity profile in the simulation was due to the complicated anatomical arrangement of the tongue musculature as represented in the model. This example helps to motivate one of the main objectives of our modeling approach: to use the model in delineating the separate influences on speech kinematics of vocal tract properties and control strategies.

A three-dimensional tongue model.



Figure 8: The lines show a block model of the tongue and floor of the mouth. Several muscles are represented by groups of blocks that are independent of the block model.

The vocal-tract anatomy of individual speakers influences how the speakers move their articulators to form different speech sounds. Therefore, to obtain maximally interpretable simulation results, we aim to model the vocal tract anatomy and articulatory behavior of individual speakers in three dimensions. Central to the development of a 3-D model is the generation of a finite element data structure that represents the

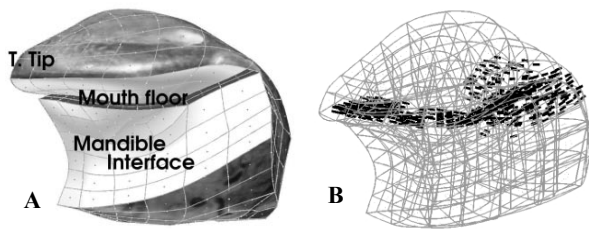


Figure 9: A – The finite element model of the tongue and floor of the mouth. B – The direction of fibers for the styloglossus muscle, indicated with short black lines. See text for details.

distribution and orientation of the muscular tissue of the oral articulators, including the tongue body, the floor of the mouth, and the pharyngeal walls. For the generation of a finite element model of the tongue and mouth floor, a geometrical block decomposition has been hand edited to fit data from the Visible Human Project; this block model is indicated by the lines in Fig. 8. Figure 8 also shows several muscles, each of which is represented by several blocks that are independent of the overall block model.

Figure 9A shows the finite element model that is obtained by subdividing the block representation shown in Fig. 8. (Further subdivision is possible if necessary.) The large light region shows the area of attachment to the mandible; all the nodes on this surface move with the rigid structure of the mandible. The directions of muscle fibers are specified throughout the finite element model at each point where stress fields are computed.

Figure 9B demonstrates the direction of muscle fibers for the styloglossus muscle. The direction field (shown by short black lines in the figure) is generated by calculating tangent vectors to the curvilinear coordinate lines of the muscle's block representation (shown in Fig. 8).

7. AUDITORY FEEDBACK IN ADULT SPEECH PRODUCTION

The model in Fig. 1 includes an auditory feedback pathway (left half of the figure) that is responsible for the learning and then the maintenance of the Articulatory-to-Auditory mapping (4). In order to characterize this mapping, we are conducting research on the role of auditory feedback in speech motor control. This work focuses on changes in speech that occur with a change in hearing.

It is well known that people born deaf usually have a very difficult time learning how to speak intelligibly. On the other hand, if someone is born with hearing, learns how to speak and then becomes deaf, that person is able to continue speaking intelligibly for decades without being able to hear. These basic observations support the idea that learning how to speak involves establishing mappings such as those shown in Fig. 1. We hypothesize that once the phoneme-specific mappings (1 and 2) are established, they are relatively resistant to change. On the other hand, the systemic mappings (3-5) must be able to change as the vocal-tract grows and is subject to other modifications, such as the wearing of dentures. While the speech production of postlingually deafened people is intelligible, it can sound abnormal to some extent, indicating that without the use of auditory feedback to maintain the systemic mappings, they can degrade somewhat.

We have been investigating the role of auditory feedback in adult speech production by observing changes in speech that occur in response to the loss of hearing due to disease or the acquisition of some hearing from a cochlear implant. These studies are directed mainly at understanding the function of mappings 3-5 in Fig. 1.

Stability of auditory goals for vowels

The stability of the phoneme-specific auditory goals for vowels is evidenced by the predominantly normal vowel for-

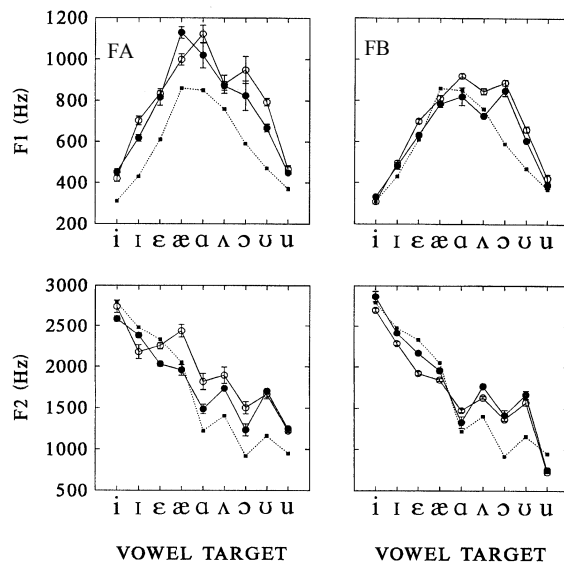


Figure 10: average F_1 and F_2 values (upper and lower panels) arranged by vowel for two female CI users (left and right panels).

mant patterns seen in for two female cochlear implant (CI) users in Fig. 9. The figure shows sets of average F_1 and F_2 values (upper and lower panels) arranged by vowel for the two speakers (left and right panels). The small squares connected by dotted lines show normative values from [9]. The values indicated by unfilled circles are pre-implant, and those indicated by filled circles are post-implant (1-2 years post-implant). The error bars indicate one standard error about the mean. For the most part, overall vowel formant patterns (relations of formant values to one another among the vowels) appear to be relatively congruent with the normative patterns, even years after the onset of profound hearing loss [10]. The most prominent exception to this observation is for FA. 18 years after the onset of her profound hearing loss, pre-implant F_2 values among her front vowels /i/, /I/, /ε/ and /æ/ were somewhat disordered with respect to the Peterson and Barney data, primarily due to relatively high values for /ε/ and especially /æ/. As indicated by the filled circles, after about a year with prosthetic hearing, these F_2 values are more in line with the Peterson and Barney pattern. Thus, FA's abnormal pre-implant F_2 pattern was "corrected" toward the normative pattern after some months of implant use.

Stability and change of auditory goals for the fricative consonants /s/ and /ʃ/

Figure 10 shows values of spectral median for /s/ (as in "said") and /ʃ/ (as in "shed") produced in carrier phrases by three of five cochlear implant users studied by Matthies et al. [11]. This measurement, which reflects acoustically and perceptually important differences between /s/ and /ʃ/, was made pre-implant, within a few months after implant and six months post-implant. Pre-implant, as exemplified by speakers 1 and 3, four of the five subjects had higher values of spectral median for /s/ than for /ʃ/ and clear separation between the /s/ and /ʃ/ values. These results indicate a good distinction between the two consonants pre-implant – even decades following the onset of profound deafness. The good pre-implant distinctions between the sibilants in four of the subjects indicate that their systemic and phoneme-specific mappings for the production of /s/ and /ʃ/ were generally quite stable, even in the prolonged absence of auditory feedback. On the other hand, the fifth subject (2) had reversed values of the two measures pre-implant, consistent with a perceptual impression that her sibilants were quite distorted and indicating an unusually extreme distortion of systemic mappings. After months of implant use with auditory feedback, Subject 2's spectral median and symmetry values were greatly improved.

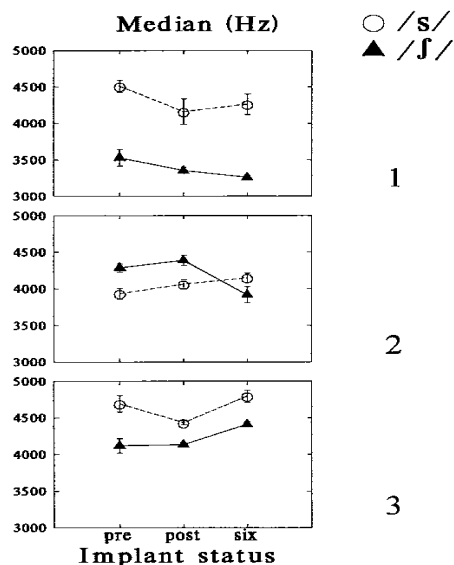


Figure 10: Sibilant spectral median for one male (3) and two female (1,2) CI users.

Invalidation of systemic mappings due to a change in the vocal tract

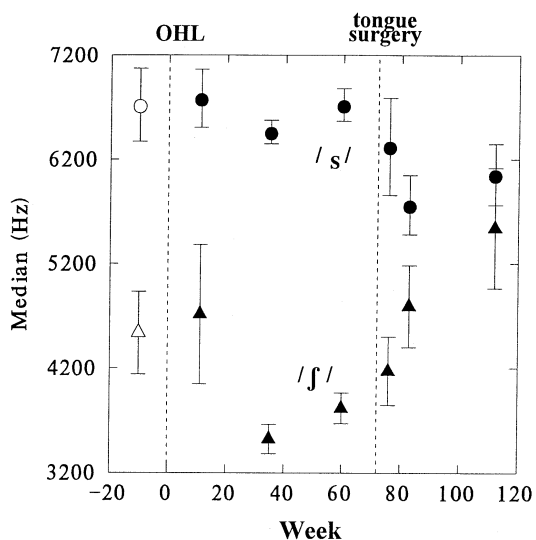


Figure 11: Spectral median for /s/ and /ʃ/ versus time in weeks from a subject who lost hearing (at time OHL) due to removal of an acoustic neuroma.

We have made another observation of the /s-/ʃ/ contrast, from a subject who lost hearing due to bilateral acoustic neuromas (NF-2). The subject had tumor-removal surgery that severed her remaining auditory nerve [12]. At the time of surgery, she received an auditory brainstem implant, which effectively provided her with auditory envelope but not spectral cues. Figure 11 shows spectral median versus week from the onset of hearing loss (OHL) for /s/ and /ʃ/. During the period before OHL and continuing for over 70 weeks post-OHL, FD maintained a good contrast between the two sounds. At week 72, she had surgery to anastomose her left hypoglossal nerve to the facial nerve, in an attempt to restore some facial function that had also been lost at the time of tumor removal surgery.

The anastomosis surgery denervated some tongue muscles on the left side, producing a slight tongue weakness that effectively altered a functional property of the vocal tract. Without auditory feedback about the sibilant contrast to help the control mechanism develop a compensatory adaptation to the

tongue weakness, the contrast gradually collapsed. In terms of the model in Fig. 1, the anastomosis surgery invalidated systemic mappings by changing a characteristic of the low-level control of the “biomechanical plant” (the vocal tract in Fig. 1). Due to the subject’s deafness, it was then impossible for her to update the mappings by making auditorily based adjustments. Since people with normal hearing are capable of compensating for significant changes in vocal-tract morphology (e.g. with the initial insertion of dentures), we assume that if FD had adequate hearing, she would have been able to compensate for the surgery, even though it resulted in some slight tongue weakness.

Relations between production and perception

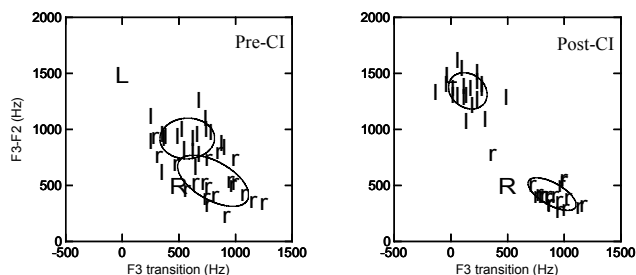


Figure 12: Separation between F3 and F2 at vowel beginning vs. extent of F3 transition from vowel beginning to midvowel for preceding /r/ and /l/ in a carrier phrase, spoken by a male CI user. (Capital letters indicate normative data.) Left half – pre-implant; right half – data from 6 and 12 months post implant.

As mentioned above, the model in Fig. 1 leads to the hypothesis that with prolonged profound hearing loss, the systemic mappings (boxes 3-5) can deteriorate somewhat. This deterioration can result in somewhat diminished phonemic contrasts. Conversely, the provision of some hearing with a CI may lead to enhancement of contrasts, which presumably are accompanied by increases in intelligibility.

To examine these relationships in detail, we gathered speech production, perception and intelligibility data for the liquids /r/ and /l/ spoken in carrier phrases by eight postlingually deaf adults, pre- and post cochlear implant (CI). Formant transition analysis for the CI speakers and two speakers with normal hearing indicated that /r/ and /l/ could be differentiated by the extent of the F3 transition and the distance in Hz between F2 and F3 at the C-V boundary. Speakers who had a limited contrast between /r/ and /l/ pre-cochlear implant and who showed improvement in their perception of these consonants with prosthetic hearing were found to demonstrate greatly improved production of /r/ and /l/ six months post-CI. The speech production changes noted in the acoustic analyses were corroborated by intelligibility improvements in the post-CI speech, as measured with a panel of normal-hearing listeners. Figure 12 shows an example of enhanced contrast between /r/ and /l/ for a male CI user, from pre- to post implant.

Rapid changes in phonemic settings

The longitudinal observations discussed above reveal that mappings can change slowly. However, we have also observed relatively rapid changes in the speech production of CI subjects in response to a change in hearing. The speed of such changes was investigated in seven cochlear implant users, in response to switching the speech processors of their implants on and off a number of times in a single experimental session. Using the times of on-off or off-on switches as line-up points for averaging, several parameters were compared across the switches. The speakers’ vowel SPL and duration had changed by the first utterance following the switch. There were equally rapid, significant changes in phonemic parameters: measures of vowel and sibilant contrasts. The magnitudes of

those changes were relatively small and they differed in size from one subject to the next. Most of these “subphonemic” changes were consistent with the prediction that contrasts are enhanced in the presence of hearing and diminished without hearing. The rapidity of such contrast changes is consistent with the function of a single control parameter in the model (not shown in Fig. 1) that is used to make rapid adjustments in speech clarity and rate [2].

8. CONCLUSIONS

Findings from modeling studies and studies of normal-hearing speakers and speakers with hearing loss are supportive of the model of sensorimotor control of speech production described in Section 2. Two important features of the model are 1) the goals of articulatory movements for some sounds are in the auditory-temporal domain and 2) the planning of speech movements depends on several mappings that are acquired and then maintained with the use of auditory feedback. Current research is aimed at more detailed quantitative modeling of the anatomy, biomechanics, motor control and brain function involved in speech production.

9. ACKNOWLEDGEMENTS

This work was supported by grants from the N.I.H., N.S.F. in the United States and the C.N.R.S. in France.

10. REFERENCES

- [1] F. H. Guenther, M., Hampson & D. Johnson: A theoretical investigation of reference frames for the planning of speech movements, *Psychological Review*, Vol. 105, 1998, pp. 611-633.
- [2] F. H. Guenther: Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, Vol. 102, 1995, pp.594-621.
- [3] D. Callan, R. Kent, F. H. Guenther, & H. K. Vorperian: An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *J. Speech, Language Hearing Res.*, Vol. 43, 2000, pp. 721-736.
- [4] R.D. Kent, & F.D. Minifie: Coarticulation in recent speech production models. *J. Phonetics*, Vol. 5, 1977, pp. 115-133.
- [5] J.S. Perkell, M.H. Cohen, M.A. Svirsky, M.L., Matthies, I. Garabietta, & M. Jackson: Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements, *J. Acoust. Soc. Am.* Vol. 92, 1992, pp. 3078-3096.
- [6] J.S. Perkell, M.L. Matthies, M.A. Svirsky & M.I. Jordan: Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study, *J. Acoust. Soc. Am.* Vol. 93, 1993, pp. 2948-2961.
- [7] F. H. Guenther, C.Y. Espy-Wilson, S.E. Boyce, M.L. Matthies, M. Zandipour, M., & J.S. Perkell: Articulatory tradeoffs reduce acoustic variability during American English /r/ production, *J. Acoust. Soc. Am.*, Vol. 105, 1999, pp. 2854-2865.
- [8] Y. Payan & P. Perrier: Synthesis of V-V sequences with a 2D biomechanical model of the tongue controlled by Equilibrium Point Hypothesis, *Speech Comm.*, Vol. 22, 1997, 185-205.
- [9] G.E. Peterson & H.L. Barney: Control methods used in a study of the vowels, *J. Acoust. Soc. Am.*, Vol. 24, 1952, pp. 175-184.
- [10] J.S. Perkell, H. Lane, M.A. Svirsky & J. Webster: Speech of cochlear implant patients: A longitudinal study of vowel production, *J. Acoust. Soc. Am.* Vol. 91, 1992, 2961-2979.
- [11] M.L. Matthies, M.A. Svirsky, J.S. Perkell, J.S. & H. Lane: Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant, *J. Speech and Hearing Res.* Vol. 39, 1996, pp. 936-946.
- [12] J.S. Perkell, J. Manzella, J. Wozniak, M.L. Matthies, H. Lane, M.A. Svirsky, P. Guiod, L. Delhorne, P. Short, M. MacCollin & C. Mitchell: Changes in speech production following hearing loss due to bilateral acoustic neuromas, Proceedings of the XIIIth International Congress of Phonetic Sciences, Vol. 3, Stockholm, 1995, pp. 194-197.