Part V Language, Speech and Hearing

Section 1   Speech Communication

Chapter 1 Speech Communication

Chapter 1. Speech Communication

Academic and Research Staff

Professor Kenneth N. Stevens, Professor Jonathan Allen, Professor Morris Halle, Professor Samuel J. Keyser, Dr. Melanie L. Matthies, Dr. Joseph S. Perkell, Dr. Stefanie Shattuck-Hufnagel, Dr. Mario A. Svirsky, Dr. Alice Turk, Dr. Yi Xu, Peter C. Guiod, Seth M. Hall, Joyce Manzella,

Visiting Scientists and Research Affiliates

Dr. Vladimir M. Barsukov, Dr. Corine A. Bickley, Dr. Suzanne E. Boyce,[1] Dr. Carol Espy-Wilson,[2] Dr. Richard S. Goldhor,[1] Dr. Robert E. Hillman,[3] Dr. Caroline Huang, Dr. Harlan Lane,[4] Dr. John Locke,[5] Dr. John I. Makhoul,[6] Dr. Sharon Manuel, Dr. Jorgen Pind,[7] Dr. Reiner Wilhelms-Tricarico, Dr. David R. Williams, Sergio Basalo,[8] Anna Esposito,[9] Helen Hanson,[10] Eva B. Holmberg,[11] Clay T. Mitchell, Jane W. Wozniak[3]

Graduate Students

Hwa-Ping Chang, Marilyn Y. Chen, Karen Chenausky, David M. Horowitz, Mark A. Johnson, Hong-Kwang J. Kuo, Sharlene A. Liu, Noel Massey, Katerina Nguyen, Kelly L. Poort, Lorin F. Wilde

Undergraduate Students

Howard Cheng, Laura C. Dilley, Helena Geng, Brett A. Geoffroy, John Hardy, Marnie L. Harker, Henry Lu, Rebecca Mallin, Rachel Molinaar, Laura M. Sever

Technical and Support Staff

D. Keith North, Arlene E. Wint

Sponsors

C.J. Lebel Fellowship
Dennis Klatt Memorial Fund
National Institutes of Health
    Grant R01-DC00075
    Grant R03-DC01721

National Institutes of Health
    Grant R01-DC01291
    Grant R01-DC00261[12]
    Grant P01-DC00361-06A1[12]
    Grant R01-DC00776[13]

---

[1] Audiofile, Inc., Lexington, Massachusetts.

[2] Boston University, Boston, Massachusetts.

[3] Massachusetts Eye and Ear Infirmary, Boston, Massachusetts.

[4] Department of Psychology, Northeastern University, Boston, Massachusetts.

[5] Massachusetts General Hospital, Boston, Massachusetts.

[6] Bolt, Beranek and Newman, Inc., Cambridge, Massachusetts.

[7] University of Iceland, Reykjavík, Iceland.

[8] Hospital de Clinicas, Universidad de la Republica, Montevideo, Uruguay.

[9] Deptartment of Physics, Salerno University, Italy.

[10] Harvard University, Cambridge, Massachusetts.

[11] MIT Staff Member and Massachusetts Eye and Ear Infirmary, Boston, Massachusetts.

[12] Under subcontract to Massachusetts Eye and Ear Infirmary.

[13] Under subcontract to Massachusetts General Hospital.

1.1 Studies of Normal Speech Production

1.1.1 Articulatory Evidence for Acoustic Goals for Consonants: a Pilot Study

In previous work, we found evidence for trading relations between tongue-body raising and upper lip protrusion (measured with an EMMA system) for the vowel /u/, reflecting a "motor equivalence" strategy that should help to constrain acoustic variation. Theoretically, analogous relations in the transformation between the area function and the acoustic transfer function are possible for the consonants /r/ and /ʃ/, which are also produced with two independently-controllable constrictions, formed by the tongue and by the lips. Such relations might occur more among tokens that are least prototypical, i.e., closest to perceptual boundaries. In a preliminary test of these hypotheses, a single speaker has produced the sounds /r/, /ʃ/ and /u/, embedded in contexts designed to induce differences in prototypicality. Motor equivalence was observed for the /u/ in /tut/ (least prototypical, with the highest $F2$ values) but not in /pub/ or /hud/. For /r/ and /ʃ/, anterior displacement of the tongue constriction was positively correlated with upper lip protrusion, providing initial support for the hypothesis that movement goals for the consonants also include acoustic components. These components are manifested in a tendency to maintain sufficient front-cavity length to help assure acoustic-phonetic distinctiveness. These findings were also confined to less prototypical tokens. A second subject has been recorded with the same paradigm.

1.1.2 Clarity versus Effort in Speech Production: Initial Results of a Pilot Study

The goal of this research is to test the hypothesis that there is a trading relation between clarity (measured acoustically and in intelligibility tests) and articulatory effort in speech production. We have recorded midsagittal-plane articulatory movements of points on the tongue body, tongue blade, lips and mandible (with an EMMA system), as well as the acoustic signal from a male talker in five speaking conditions: normal, clear, clear plus fast, fast, and slow. Different CVC words were embedded in two positions in a carrier phrase (called "one" and "two"). Kinematic analyses were performed as a function of speaking condition and position in the phrase. As an example result, CV movements of a point on the tongue body for the word "cab" have the following characteristics. Movement duration was shorter for position two; in both positions, it followed the order: slow > clear > normal > clear plus fast > fast. The range of distance traveled was greater for both slow and clear than for the other conditions. Peak velocity was higher for fast and clear plus fast than the other conditions. Midsagittal plane location at movement end was less variable for the tokens in position one (longer duration movements). A measure of effort was higher for clear, clear plus fast and fast, than for normal or slow, in both word positions. Acoustic analyses and the construction of intelligibility tests are in progress. Most recently, the same kind of recordings have been made for a second subject.

1.1.3 Articulatory Processes

A chapter called "Articulatory Processes" was prepared for a forthcoming book, *A Handbook of Phonetic Science*, edited by W.J. Hardcastle and J. Laver (Blackwell Publishers, Oxford, England). The chapter presents the following view of speech production.

Articulatory processes are actions of the respiratory, laryngeal, and vocal-tract systems that convert linguistic messages to an intelligible signal. The linguistic message is encoded in acoustic characteristics of individual sound segments and in superimposed prosodic patterns of sound level, fundamental frequency and duration. Important additional influences are listener needs, which are determined by factors such as the level of environmental noise and the listener's familiarity with the language, speaker, and subject matter. Kinematic adjustments are used to help convey some linguistically-contrastive information; however, the large amount of variability inherent in speech and the lack of truly comprehensive kinematic data currently make it difficult to provide a systematic account of articulatory kinematics. Some prelimi-

---

nary ideas about articulatory dynamics indicate that the biomechanical response properties of articulators and effort minimization strategies have roles in shaping the control of articulatory processes; on the other hand, it is very likely that perceptually-related factors are at least as important in this regard.

1.1.4 Modeling Vocal-fold Vibration

The goal of the modeling work is to use a relatively simple model of glottal function to infer approximate relative values of three important physiological parameters that cannot be measured directly: vocal fold stiffness, degree of adduction of the membranous folds, and size of an inter-arytenoid chink. It is hoped that the models will help us to understand more about the mechanisms that underlie our aerodynamic and acoustic measurements. We have extended our implementation of the two-mass model of Ishizaka and Flanagan (1972) to account for two frequently-made observations: (1) gradual closure that propagates along the folds in the antero-posterior direction and (2) the presence of a chink. We assume that stiffness parameters in the model are roughly analogous to vocal-fold stiffness, and rest positions of the masses are analogous to the degree of intra-membranous vocal-fold adduction. We intend to employ the model in an analysis-by-synthesis approach. Using inputs consisting of the measured value of subglottal air pressure and $F0$ and estimates of gender-specific values of parameters such as vocal-fold mass and length, values will be estimated for the model's stiffness parameters, rest positions of the masses, and chink size. The estimates will be adjusted so that the model's flow output matches the measured glottal airflow waveform.

The original model of Ishizaka and Flanagan contains a linear relation between the displacements of the two masses and the areas. For sufficiently large amplitudes of the oscillations of the masses, the minimum area between them is zero, and the closure becomes unrealistically abrupt. In our first modification, to simulate gradual closure, the positive ramp relation between mass displacement and glottal area was replaced by a function that maintains minimal opening areas greater than zero, and a smooth quadratic function which, for small openings, computes the areas from the oscillator amplitudes. Our second modification has been to introduce a chink which represents the opening between the arytenoid cartilages. In the equivalent circuit, the acoustic effects of this chink are represented by a parallel impedance running from below to above (across) the glottis. In preliminary simu-

lation experiments, waveforms were computed as the sum of air flow through the glottis and the chink. These waveforms were quantitatively within reasonable ranges, and they looked qualitatively very similar to observed waveforms.

1.1.5 Physiological Modeling of Speech Production

In the past year, Dr. Reiner Wilhelms-Tricarico joined the Speech Group to continue work on physiological modeling of speech production that he had begun at Ohio State University and continued at ATR Laboratories, Kyoto, Japan.

During this year, previously-developed finite element code for modeling the behavior of the tongue was installed on a DEC Alpha PC; an altered, more realistic muscle model was incorporated; and a number of computational improvements were implemented. Improvements were also made in programs for displaying the computational results. The possibility of building finite element code based on Matlab was evaluated and compared with the current C⁺⁺ programming approach. The result is that Matlab is just as good for this purpose and may be easier to maintain. A number of computational experiments were performed to clarify stability limitations of the current method. Theoretical studies were performed to determine what type of element needs to be used to avoid instabilities; results indicate the need to develop a mixed tri-quadratic type of element with four degrees of freedom of the pressure field. Implementation of this type of element in Matlab code has been initiated. The code makes use of vector and matrix commands wherever possible. The resulting code should be portable to any machine that runs Matlab, including the Cray supercomputer, which is the target platform once the described refinements are completed.

1.1.6 Stop Consonant Production: An Articulation and Acoustic Study

Articulation and acoustic data for stop consonant production have been examined with the aims of (1) describing the movements and coordination of the articulatory structures and (2) developing procedures for interpreting acoustic data in terms of articulatory movements. Movements in the midsagittal plane of points on the lower jaw, lips, tongue blade, and tongue body were measured using the EMMA system. Acoustic data were recorded simultaneously. Three normal hearing, normal speaking male subjects spoke single-syllable words /CVt/, composed of a voiceless stop

consonant C followed by the vowel /a/ or /i/ and the consonant /t/, placed in a carrier phrase. Half the tokens were preceded by the fricative consonant /s/. An estimate of the constriction cross-sectional area change with time following the release of each of the stops was calculated with the aid of the articulation data. For example, the cross-sectional area of the lips in the labial /p/ in **spot** increases by approximately 35 cm²/sec within the first 5-10 ms following the stop release. During the first few ms following the release, the constriction cross-sectional area change with time can be related to the corresponding acoustics in two important ways: (1) by modeling the vocal tract as a series of tubes of varying cross-sectional area (in the case of a labial stop, a Helmholtz resonator) and calculating the transition of the first formant frequency *F1*; and (2) by calculating the time course of the burst when the constriction cross-sectional area is used as a parameter in a circuit model of the vocal tract. The articulation data were also studied to determine the effects of phonetic context. Findings include a constraint on lower jaw position during /s/ production, maximum downward velocity for the lower jaw occurring approximately at the time of the consonant release, and a correlation between the distance articulators must travel and rates of movement. The inferences made from the detailed examination of the articulation and acoustic data will aid in developing a more comprehensive model of stop consonant production.

1.2 Speech Research Relating to Special Populations

1.2.1 Speech Production of Cochlear Implant Patients

This project is funded under a subcontract to the Massachusetts Eye and Ear Infirmary in collaboration with Drs. Harlan Lane, Donald K. Eddington, and Joseph Nadol.

This research aims to characterize the speech production of postlingually deafened adults before and after they receive cochlear implants, in order to contribute to the development of models of the role of hearing in speech. The findings can also contribute to evaluating and improving prostheses and to focusing speech therapy for implant users. Toward these goals, we have been measuring: vowel acoustics (vowel formants, *F0*, SPL, duration, and a measure of inferred glottal aperture); consonant acoustics (voice-onset time, fricative and plosive spectra); and speech aerodynamics (speech initiation and termination lung volumes, volume of air per syllable, inferred subglottal pressure, oral

and nasal airflow, peak glottal airflow, minimum flow, AC flow, open quotient, and maximum flow declination rate). Generally, our findings continue to indicate significant deterioration of acoustic and articulatory speech parameters in postlingually deafened adults, and substantial long-term changes, generally in the direction of normalcy, following activation of the cochlear implant. During this last year, we have made the following specific findings.

Voice Onset Time (VOT)

The effects on VOT of turning subjects' speech processors off and on within an experimental session were generally consistent with, but smaller than, the long-term effects of deafening and restoration of some hearing. Data were analyzed for two implant users who participated in a longitudinal study of VOT described in last year's report. Only their voiced plosives and not the voiceless showed significant increases in VOT when comparing mean values in the first processor-off condition with the following processor-on condition. Changes in processor state produced widespread postural changes (in average SPL, *F0* and speaking rate). Those changes may be responsible for the observed changes in voiced VOT.

We have begun perceptual tests using synthetic speech stimuli in order to assess whether Ineraid implant users can discriminate VOT differences of the order of magnitude of the discrepancy between their VOT values and normative data. We tested a voicing contrast in syllable-initial plosives, created with the Klatt synthesizer. Onset frequencies of *F0* and *F1* varied across the series. Two psychometric functions were determined, one for voiced and one for voiceless syllables. In this initial experiment, one implant user yielded jnd values of 50 and 35 ms for voiceless and voiced plosives VOTs respectively. These jnds are more than twice as large as the discrepancies between her pre-implant VOTs and normative values, suggesting that, in her case, changes in VOT with the speech processor on are not due to auditory validation of the VOT contrast. A subject with normal hearing gave jnds of 15 and 5 ms, respectively.

Syllable-to-syllable Regulation of F0 and SPL

Syllable-to-syllable fluctuation of *F0* and SPL were measured in readings of a passage by four postlingually deafened adults, recorded before and after they received cochlear implants, and one adult with neurofibromatosis-2 (NF2—see section 1.2.3 below), who was initially profoundly deaf in one ear

and had a mild to moderate hearing loss in the other (aided). Three of the four cochlear implant users reliably reduced syllable-to-syllable fluctuation in $F0$ and SPL following the activation of their speech processors. The fourth speaker began with and maintained the regularity achieved by the others post-activation. In recordings made following the onset of bilateral profound deafness, the NF2 subject showed increased syllable-to-syllable fluctuation in $F0$ and SPL. Results from another experiment,[16] in which multiple repetitions of vowels in an /hVd/ context were produced by cochlear implant users with their processors turned off and on, suggest that some subjects showed less token-to-token variability in $F0$ and SPL with their processors turned on. The present results indicate that auditory feedback may also constrain the syllable-to-syllable variability of $F0$ and SPL contours.

1.2.2 Speech Perception by Cochlear Implant Users

The two long-term goals of this research are (1) to understand the acoustic cues that underlie speech reception by cochlear implant users and (2) to find the optimal stimulation strategies for different subjects and different types of cochlear implants.

Development of Software and Hardware Tools

The Nucleus device has an external part (including microphone, speech processor and transmitter) and an internal implanted receiver that receives power and stimulation information via a radio frequency link. In order to meet the goal of optimizing speech processing strategies, we have implemented an interface to encode computer-stored stimuli for presentation to users of the Nucleus cochlear implant. This interface bypasses the external part of the Nucleus device and can be used to stimulate subjects in an unrestricted fashion, constrained only by the limitations of the Nucleus device. Our system includes software running on two different CPUs: a PC and a Texas Instruments 320C25 fixed-point digital signal processor installed in the PC. The PC performs high-level functions (such as processing input speech signals according to different strategies); it downloads stimulation information to the 320C25 DSP chip and sends a signal to the DSP to start stimulation. The DSP sends appropriate electrical signals to a custom made

hardware driver that generates the radio frequency pulses that are sent to the implanted receiver.

We have written code to run a number of psychophysical and speech perception tests. Psychophysical tests include determination of threshold (T) and maximum comfortable stimulation level (C); loudness balancing of two channels; and pitch ranking of different channels. Speech processing is done off-line (not in real time), and it involves a number of steps. Files are either synthesized with KLSYN88 or digitized from a speech signal, and then converted to a standard binary format. Files are then processed according to a subject's T and C levels and to different stimulation strategies. We find the power density in a set of frequency bands that may or may not overlap, covering the speech range up to 6 kHz. Any arbitrary function may be used to map a specified acoustic range into the subject's electrical dynamic range. Finally, the program chooses what channels will be stimulated and in what sequence and outputs a subject-specific stimulus file to disk. We have already digitized and converted a large number of vowel and consonant tokens, and these files have been processed for presentation to specific subjects. Stimulus presentation and recording of subject responses will be done with a program developed at the Sensory Communications Group (see section 2).

Pitch Ranking

Three subjects were tested with a pitch ranking task to assess their channel discrimination. Two adjacent channels were stimulated in sequence and the subject was asked whether the first or the second stimulus was higher pitched. We then calculated d' values for all pairs of adjacent channels. Subjects 1 and 2 were tested with the standard clinical interface. Subject 1 had 19 active channels. Nine of the 18 channel pairs were perfectly discriminated and correctly pitch-ranked (i.e., more basal channels sounded higher pitched). Of the remaining nine channel pairs, seven were adequately discriminated (d' values between 1 and 2) and correctly pitch ranked; channels 13 and 14 were totally indiscriminable, and channels 10 and 11 were well discriminated (d'=1.9), but their pitch was reversed (that is, the more basal channel had a lower pitch). Subject 2 also had 18 pairs of adjacent channels, but only three of those pairs were perfectly discriminated and pitch ranked. Ten additional channel pairs were correctly pitch ranked. Five of those

---

16 M.A. Svirsky, H.L. Lane, J.S. Perkell, and J.F. Wozniak, J. Acoust. Soc. Am. 92(3): 1284-1300 (1992),

were adequately discriminated (1 < d' < 2), and the other five were poorly discriminated (0.4 < d' < 1). The four remaining pairs had reverse pitch ranking, three of them with low discriminability (0.4 < d' < 1) and one pair with perfect discrimination. The third subject was tested with our interface and software. Nineteen pairs of channels were tested. Six of them were perfectly discriminated and correctly pitch ranked. Four additional pairs were correctly pitch ranked, but only three of those showed adequate discrimination (1 < d' < 2). The remaining nine channel pairs had reverse pitch ranking, with three showing perfect discriminability and two more showing adequate discriminability (1 < d' < 2). In summary, all subjects showed marked differences in channel discriminability across the electrode array, and the nominally more basal channels did not always sound higher pitched than their nominally more apical neighbors. This may have important consequences in our thinking about speech processing strategies. For example, we may choose to implement strategies that only include stimulation channels that are well discriminated from their neighboring channels, and correctly pitch-ranked.

Mathematical Model of Vowel Perception with the Ineraid Cochlear Implant

The Ineraid cochlear implant has been used successfully to help deaf subjects understand speech. The incoming speech signal is filtered into four overlapping frequency bands, with crossover frequencies at approximately 700 Hz, 1.4 kHz and 2.3 kHz. The filters are broad, with slopes of 12 dB per octave. An analog representation of each filter output is delivered to an intracochlear electrode. Electrodes are placed at 4 mm intervals, spanning a range of 8 mm to 20 mm from the cochlear base. Electrodes closer to the base of the cochlea are associated with filters that have higher cutoff frequencies. Using this strategy, a typical patient can have a fluent one-on-one conversation (with the help of lipreading) and he also has significant speech reception scores without lipreading. However, there is still controversy about the cues employed by users of the Ineraid device to understand speech.

The aim of the current study was to determine the cues employed by users of the Ineraid cochlear implant to identify vowels. Our approach was to implement various mathematical decision models and determine which one gives the best fit to data on vowel perception reported by Dorman et al. (1992). In particular, we implemented models that used only temporal cues, only amplitude cues, or both kinds of cues at the same time. Our preliminary conclusion was that perceptual models with

only one type of cue (either temporal or amplitude cues) may be insufficient to explain important characteristics of the data. A simple model incorporating both kinds of cues appears to succeed where other models fail. The most successful model gave a good fit to actual speech reception data, for all vowels used in the study. This model used one temporal dimension (an estimate of F1 obtained from the temporal stimulation waveform delivered to channel 1) and three amplitude dimensions, namely the ratios of stimulation amplitude delivered to channels 2, 3 and 4 with respect to the stimulation amplitude delivered to channel 1.

Our main conclusion is that it is very difficult to explain vowel perception by Ineraid patients based only on the time waveforms delivered to each channel, or only on the distribution of energy among channels. User of the Ineraid device seem to integrate temporal and channel amplitude information in order to understand speech.

1.2.3 Degradation of Speech and Hearing with Bilateral Acoustic Neuromas

In this project, we are studying the relation between speech and hearing in people who become deaf from bilateral acoustic neuromas (auditory nerve tumors) associated with the genetic disorder neurofibromatosis type II (NF2). The primary goal of the study is to increase our understanding of the role of hearing in the control of adult speech production. The rationale and approach for this work are similar to those of our ongoing work on the speech production of cochlear implant patients. Speech acoustic and physiological parameters have been recorded and speech perception has been tested in a group of 40 (still hearing) NF2 patients. Then the same parameters are recorded at intervals for any patients who suffer further hearing loss.

Two of the enrolled patients have suffered further hearing loss, and we have obtained post-loss recordings on both of them. We have analyzed the pre- and post-loss recordings from one of the two, who lost her hearing due to tumor surgery and received a multichannel auditory brainstem implant (ABI), which restored some limited hearing. Recordings have been made longitudinally prior to and following hearing loss, covering a period of about two years. Perceptual testing following surgery showed that the subject (when using the ABI) had an awareness of sounds but poor (i.e., ineffective) speech discrimination ability. Informal observations indicate a slowly progressing, slight decrease in intelligibility. Longitudinal production changes appear to fall into three categories: changes in baseline values of the postural parame-

ters of SPL, speaking rate and *F0*, some changes in phonemic contrasts, and changes in the syllable-to-syllable regulation of *F0* and SPL (reported above in section 1.2.1). These changes are roughly consistent with an initial inability to make much use of the auditory feedback provided by the ABI, followed by a period of accommodation in which postural parameters appear to be returning to pre-loss values, but phonemic and syllable-to-syllable parameters continue to move away from pre-loss values. (The surgery also resulted in facial paralysis on one side, which is taken into account when interpreting speech production results.)

Several post-implant sessions have also included a recording made after 24 hours with the speech processor of the ABI turned off. Multiple repetitions of the speech material were recorded first with the processor off, then with it turned on for a number of minutes, then with it off again. Analysis of these recordings is almost complete.

1.2.4 Objective Assessment of Vocal Hyperfunction

This project is funded under a subcontract to the Massachusetts Eye and Ear Infirmary in collaboration with Dr. Robert E. Hillman.

The major goal of this project is to further develop and use quantitative measures of voice production to provide objective descriptions of conditions referred to as vocal hyperfunction. This work involves the use of noninvasive acoustic, aerodynamic and physiologic measures to study both organic (nodules, polyps, contact ulcers) and nonorganic (functional dysphonia/aphonia) manifestations of vocal hyperfunction. These measures are examined in relation to encoded descriptions of the pathophysiologic status of the larynx (from videolaryngoscopy/stroboscopy), pertinent medical and psychosocial factors, and judgments about perceptual characteristics of the voice. Interpretation of results will be aided through use of a simple model of vocal-fold vibration, as described in section 1.1.4.

Comparisons Among the Glottal Airflow Waveform, the Electroglottographic Signal, and the Acoustic Spectrum

The inverse filtering technique used in this work is not without problems: the potential for an air leak between the subject's face and the mask; cases of unsuccessful inverse filtering resulting in formant residuals (although some *F1* ripple may be due to subglottal coupling); and smoothing effects of low-pass filtering at around 1 kHz, resulting in a

rounding of an otherwise-abrupt change in the flow waveform that occurs at the moment of vocal-fold closure. Therefore, we added measures of the acoustic spectrum slope (at low and high frequencies) to our paradigm that have been found salient for vocal intensity as well as voice quality. We also calculated an indirect measure of vocal fold closure (adduction quotient), obtained from an electroglottographic (EGG) signal. The goal was to improve methods of evaluating voice disorders and their treatment by finding complements and/or substitutes for those airflow measures that are especially sensitive to recording and analysis difficulties. In addition, relationships among the aerodynamic, acoustic, and EGG measures were studied for two speech tasks; strings of repeated /pæ/ syllables and sustained phonation of /æ/. Each task is needed for different kinds of voice analyses.

The main results for group and individual data led to the following conclusions: (1) Adduction quotient, measured from the waveform, is sensitive enough to differentiate between waveforms with gradual versus abrupt vocal-fold closing movements. (2) Measurements of the spectrum slope (especially A1-A3, the difference in amplitudes of the spectrum prominences of *F1* and *F3*) may add important information to glottal waveform measures about voices with high vocal-fold closing velocities, often found for patients with hyperfunctionally-related voice disorders. (3) Some measures differ significantly depending upon whether energy in the *F3* region is comprised predominantly of harmonics or noise. (4) The EGG-based adduction quotient can serve as a complement to flow measures when there are problems with formant residuals that make the flow-based adduction quotient less reliable. (5) Large inter-speaker and session-to-session intra-speaker variation in DC flow may be at least partly due to air leakage not noticed during the recording. Thus, DC flow data must be interpreted with caution, and the value of DC flow as a clinical tool seems questionable. (6) Cross-task comparisons of measures obtained from /æ/ and /pæ/ can be made after adjustments for SPL differences.

Phonatory Function in Women with Vocal Nodules

In initial work we conducted a study that compared phonatory function of a group of 12 adult female bilateral vocal-fold nodules patients with a group of age matched normal female controls. After modifying our signal processing system, we reanalyzed the recordings from the initial study and combined the data with data for new subjects to yield a study comparing a group of 22 adult female nodules patients with 22 matched normal controls.

Statistical analyses of these data lead to the suggestion that bilateral vocal-fold nodules are associated with a pathophysiological phonation mechanism in which normal relationships between SPL and underlying vocal parameters are disrupted, i.e., there is reduction in the efficiency with which aerodynamic energy is converted into SPL, potentially contributing to ever-increasing vocal hyperfunction. Nodules patients appeared to expend higher than normal levels of aerodynamic energy (reflected in higher levels of subglottal pressure, peak flow, and AC flow) to attain a given SPL. These results also support our previous hypothesis about the potential existence of a "vicious cycle" in which the presence of lesions makes voice production more difficult (disrupts the normal vibratory mechanism), requiring higher levels of aerodynamic driving and myoelastic restoring forces. Such increases in the forces underlying phonation may lead to excessively stiff folds and a further increase in the collision forces and the further development of nodules. The tendency for subjects with nodules to speak louder than normal may exacerbate the problem and could conceivably accelerate the pathological process. The existence of this hypothesized pathophysiological mechanism should also be apparent in acoustic spectral measures (e.g., steeper spectral slope). We are currently examining acoustic spectra for such evidence of this mechanism.

1.2.5 Acoustic and Perceptual Effects of Endoscopic Nasal Surgery

This project is carried out in collaboration with Dr. Ralph Metson of the Massachusetts Eye and Ear Infirmary. The aim of this research is to determine the effects of surgical alterations of nasal anatomy on the spectral characteristics and perceptual attributes of speech. Five patients who underwent endoscopic nasal surgery were recorded prior to surgery, then one week and one month after surgery. The utterances were six vowel types adjacent to nasal consonants and stop consonants. In addition to endonasal surgery enlarging the sinus openings, some patients had a correction of deviated septum, resection of a turbinate, or removal of polyps. Acoustic analysis was carried out for nasalized vowels, nonnasalized vowels, and the nasal consonants /m/ and /n/. Significant differences in spectral properties were observed in the consonants and nasalized vowels recorded before and after surgery. For example, changes were found in the prominence of extra peaks due to nasalization and the first formant bandwidth was narrower for the nasal consonants and nasalized vowels. All except one patient perceived a change in their speech post-operatively. Attempts have

been made to account for the speech changes by incorporating the anatomical alterations caused by surgery into a speech production model. In particular, the removal of surfaces within the nasal cavity reduces the acoustic loss and leads to narrower bandwidths for some resonances.

1.2.6 Aiding Dysarthric Speakers

One practical aim of this research is to determine how best to use speech recognition techniques for augmenting the communication abilities of dysarthric speakers. As a first step toward this goal, we have performed the following kinds of analyses and tests on words produced by eight dysarthric speakers: a closed-set intelligibility test, critical listening and transcription, acoustic analysis of selected utterances, and an evaluation of the recognition of words by a commercial speech recognizer. An outcome of this evaluation is a description of attributes of vowels and consonants which are produced adequately and consistently and those which are produced in a distorted or inconsistent manner. The analysis and testing have led to several conclusions concerning the control of the articulators for these speakers: production of obstruent consonants was a particular problem, whereas sonorant consonants were less of a problem. Of the obstruent errors, most were place errors for alveolar consonants (particularly fricatives), but voicing errors were also high. These obstruent consonants were often produced inconsistently, as inferred from acoustic analysis and from low scores from the recognizer for words with these consonants. In comparison, vowel errors were less prevalent. The practical use of a speech recognizer for augmenting the communication abilities of the speakers requires the selection of an inventory of utterances that are produced consistently and consequently are recognized with reasonable accuracy. Based on the experimental data, word lists that lead to the enhancement of recognition accuracy for each speaker have been designed.

1.3 Speech Production Planning and Prosody

1.3.1 Interaction Between Prosody and Planning

Our work in speech production planning is focused on the interaction between prosody and other aspects of planning, such as segmental serial ordering, glottalization and error detection/correction. To facilitate these studies, we have spent

several years developing a range of speech databases for prosodic analysis. The largest database, consisting of seven hours of speech read "on air" by seven different professional FM radio news broadcasters, has been digitized, orthographically transcribed, automatically phonetically aligned and labeled for the occurrence of audible breath intakes. It will be made available to the research community via the Linguistics Data Consortium (LDC) at the University of Pennsylvania. In addition, we have helped to develop a new prosodic labeling system (ToBI), which was recently evaluated for reliability across more than a dozen transcribing sites.[17] Using this transcription system, we have marked the pitch accents, boundary tones and prosodic constituent boundaries of four stories that were read by six of the seven speakers, and more than 50 stories recorded in the broadcast studio by several of the speakers. These prosodic labels, which will also be made available via the LDC, are the basis for the analyses described below. Finally, we are prosodically labeling digitized versions of 200 spoken errors from the ATIS and Switchboard corpora of spontaneous speech, as well as 400 utterances read by nonprofessional speakers as part of a corpus developed for the study of lexical access. Training of new prosodic labelers and evaluation of the labeling conventions are ongoing in our lab.

In recent studies, the FM radio news database has provided evidence for the view that speakers signal the onset of a new intonational phrase using an early pitch accent where possible (as in e.g., "In MASsachusetts HOSpitals"), and prefer to avoid the pitch accent clash that results when two adjacent syllables in the same intonational phrase are pitch accented. These two factors together, phrase onset marking and pitch accent clash avoidance, provide an alternative account of what has been called "stress shift", and provide substantial support for the claim that constituents of the prosodic hierarchy play an active role in speech production planning.

Further evidence for this hypothesis has been derived from the study of glottalization of vowel-onset syllables. In earlier work, we demonstrated that speakers glottalize vowel onsets at prosodically significant locations, i.e., at the beginning of a new intonational phrase, or when the syllable is pitch-accented, in a subset of the BU FM news database containing more than 2000 candidate syllables in a range of prosodic and segmental contexts. More recently, we have shown that this pattern of glottalization of prosodically significant syllables holds for nonprofessional read speech as well. We are currently examining the glottalization of vowel-onset syllables in spontaneous (nonread) speech.

Examining the role of glottal gestures in speech production more extensivly, we are also testing the hypothesis that when speakers detect an error and interrupt an utterance to correct it, the interruption is accomplished via a relatively sudden glottal closure. While some of the errors in our recorded corpus show this pattern, particularly if the interruption occurs within a syllable, others appear to be characterized by completion of the word with a diminution in amplitude that is more characteristic of an utterance-final word. We are currently examining the correlation of this contrast with the type of error being corrected.

1.3.2 Some Phrase-level Influences on the Glottal Waveform

The prosody of an utterance is often described in terms of the locations of attributes such as pitch accents, boundary tones, and vowel reduction. These prosodic descriptions are usually represented in the sound by changes in fundamental frequency, duration, and other temporal and spectral attributes of the segments. We have collected data showing that the amplitude and spectrum of the glottal pulses, as well as their frequency, undergo modification depending on locations of pitch accents and boundary tones and depending on vowel reduction. The following results were obtained: (1) Reduced vowels tend to have significantly lower amplitude (7 to 13 dB) than nonreduced vowels, together with an increased first-formant bandwidth and an increased tilt in the source spectrum. (2) A nonreduced vowel has a somewhat reduced amplitude (by about 6 dB) when it follows a nuclear pitch accent. (3) A nonreduced vowel in utterance-final position in a declarative sentence shows a decreased amplitude of the first-formant peak, an increased open quotient, and an increased spectrum tilt. Production mechanisms that could account for these changes include modifications in subglottal pressure, adjustments of the glottal configuration, and influences of constrictions in the supraglottal airway.

---

17 J.F. Pitrelli, M.E. Beckman, and J. Kirschberg, *Proceedings of the International Conference on Spoken Language Processing* 3, Yokohama, Japan, 1994, pp. 123-126.

1.4 Studies of the Acoustics, Perception, and Modeling of Speech Sounds

1.4.1 Use of Transient Response to Determine the Natural Frequencies of the Nasal Cavity

Direct estimates of the natural frequencies of the nasal cavity are difficult to make, and these frequencies must often be inferred from calculations based on anatomical data or from measurements of the spectra of nasal consonants and vowels. In our present research, we have attempted to determine these natural frequencies based on measurements of the transient response of the nasal cavity. An abrupt excitation of the nasal cavity is obtained by first creating a continuous flow of air through the nasal passages with the velum lowered and the lips closed. The velum is then raised abruptly so that there is rapid closure of the velopharyngeal port, aided by the buildup of intraoral pressure near the instant of closure. This closure creates an acoustic transient that excites the nasal cavity. A spectral analysis of this transient response reveals peaks that represent the natural frequencies of the nasal cavity when it is closed at the posterior end. Some of these peaks are well-defined and consistent for a given subject, and others are less prominent. The latter are probably resonances of the sinuses, which appear as pole-zero pairs in the transfer function, whereas the former represent natural frequencies of the nasal cavity proper.

Measurements of the two lowest natural frequencies of the nasal cavity proper for six adult speakers give an average value of 540 Hz (range of 400-640 Hz) and 2160 Hz (range of 1880-2420 Hz). Additional resonances, presumably due to sinuses, appear for some subjects at frequencies of 210-250 Hz and 1150-1420 Hz. The values of the natural frequencies of the nasal cavity proper are consistent with sweep-tone measurements reported by others, and with estimates of these resonances based on anatomical data. There appears to be some variability in estimates of the natural frequencies of the sinuses.

1.4.2 Modification of /ð/ by a Preceding Nasal Consonant

Presumably as a result of coarticulation, /ð/ often assimilates to a preceding /n/ in phrases like "win those," but this assimilation is not complete for all features. With respect to the feature [nasal], the assimilation is often radical. The entire consonant region in the middle of the two-word sequence is nasalized. However, acoustic evidence suggests that contextually-nasalized /ð/ retains its dental place of articulation. Specifically, F2 is considerably lower at the release of a contextually-nasalized /ð/ than at the release of a true /n/, as would be expected for a dental consonant. Perception tests show that listeners can generally tell the difference between natural tokens of pairs like "win nos" and "win those," even when the /ð/ is completely nasalized. In addition, a synthetic stimulus continuum was constructed in which items differed only with respect to F2 frequency in the vicinity of the nasal consonant regions of phrases like "win nos". Listeners systematically reported hearing "those" more often when F2 was low at the release of the nasal consonant. These results are consistent with the claim of Krakow, Fowler, and others, that listeners can at least sometimes factor out coarticulatory effects.

1.4.3 Glottal Characteristics of Female Speakers—Acoustic, Physiological, and Perceptual Correlates

In a previous study of vowels produced by female speakers, we showed substantial individual differences in acoustic parameters related to glottal characteristics. Based on measurements taken on the speech signal or spectrum, inferences were made about glottal configuration during phonation and the nature of the glottal pulse. Subjects fell into two groups based on the acoustic measurements: Group 1, assumed to have abrupt glottal closure, and Group 2, assumed to have nonsimultaneous closure. Results of a listening test have shown that members of Group 2 are perceived to be breathier than members of Group 1. Our present research extends the earlier work in several ways. Physiological measurements have been made for four subjects, including inversed filtered airflow recordings and glottal images obtained by fiberscopy. Also, listening tests using synthesized speech have been carried out to study the role of glottal characteristics in speaker discrimination, and the usefulness of the acoustic measures for guiding the synthesis of natural sounding speech. The physiological and perceptual data are found to be consistent with the earlier interpretations of the acoustic measures of glottal characteristics.

1.4.4 The Effect of Emphatic Accent on Contextual Tonal Variation

Recent data show that the *F0* contour of a tone in Mandarin is perturbed differently by the tone preceding it than by the tone following it: the influence of the preceding tone (carryover effect) is assimilatory—the beginning of the *F0* contour of a tone becomes similar to the ending pitch value of the preceding tone; the influence of the following tone (anticipatory effect) is mostly dissimilatory—the *F0* maximum of a tone dissimilates from the *F0* minimum of the following tone. The present study examines the effect of emphatic accent on different syllables in disyllabic words in order to learn more about the nature of contextual tonal variation.

Native speakers of Mandarin produced /mama/ with different tone combinations. They were requested to produce those utterances with emphasis on the first or last syllable, or with no emphasis on either syllable. Various *F0* measurements were taken, including maximim and minimum *F0* values and *F0* values at different locations in the vowel and the nasal segments. Comparisons among the three accent conditions confirms that carryover assimilation is strongest when the first syllable is emphasized. As for the anticipatory effect, dissimilation is found in all three accent conditions, but the magnitude of the effect is strongest when the second syllable receives accent. In summary, for both carryover and anticipatory effects, the accented syllables are found to exert greater influence on the unaccented syllables than the other way around, regardless of the nature of the influence.

1.4.5 Time-variation in Mandarin Voiceless Sibilant Spectra

The three Mandarin sibilant fricatives, [s], [ʂ], and [ɕ], were studied using a combination of time-averaging and ensemble-averaging of the spectra. Ten native speakers of Beijing Mandarin (five male and five female) produced these fricatives followed by both unrounded and rounded high vowels. Time-averaged spectra were obtained by averaging thirteen consecutive 8-ms FFT windows over 20-ms intervals. Token-ensemble-averaged spectra were then obtained by averaging over the time-averaged spectra of ten repetitions of each sentence produced by each speaker at time-normalized locations in the utterances. Further ensemble-averaging was also performed over the male and female speakers separately to obtain group-ensemble-averaged spectra. Preliminary examination of both token-ensemble-averaged and group-ensembled-averaged spectra reveals that (a) Mandarin sibilants have substantial time-variation in their noise spectra, indicating continuous articulatory movement throughout the production of the frication; and (b) all three sibilants manifest substantial spectral variation due to anticipatory coarticulation with the following vowel. This kind of variation is evident from the very beginning of the frication noise, indicating that the extent of the anticipatory coarticulation may go beyond the beginning of the frication.

1.4.6 Acoustic and Perceptual Attributes of Fricative Consonants

Our studies of fricative consonants have been extended to include examination of the spectra of the fricative consonants in relation to adjacent vowels and the time variation of these spectra at the boundaries between fricatives and adjacent vowels. The following results emerged from the acoustic analyses:

1. For the strident fricatives /s š /, the maximum spectrum amplitude in the frequency range of the third formant and higher is 15-22 dB higher than the corresponding spectrum amplitude for the weak fricatives /f θ/. Most of this difference can be accounted for by the influence of the resonance of the front cavity (or tongue blade channel) on the transfer function from the turbulence noise source to the output.

2. The changes in the spectrum that occur in a time of 20-odd ms preceding voicing onset for the vowel generally consist of a reduction in high-frequency amplitude for strident consonants and the appearance of aspiration noise in the region of the second and third formants.

Perceptual experiments with synthesis of fricative-vowel syllables showed the importance of maintaining a strong high-frequency spectrum amplitude for the consonant /s/, but with limits on the range of amplitudes that was acceptable. Listeners accepted a wider range of amplitudes for the weak fricative /f/.

1.5 Models for Lexical Representation and Lexical Access

1.5.1 Theoretical Background

Our work on the development of models for lexical access stems from a view that the lexical items that underlie an utterance are stored in memory in terms of sequences of segments, each of which is classified by a set of distinctive features. The retrieval of these segments and features from the speech signal as produced by a speaker is a two-stage process. (1) The locations of a sequence of landmarks are determined, and these landmarks identify critical regions in the signal where acoustic information about individual segments is concentrated. The landmarks are the result of implementing articulator-free features, and each landmark is labeled with these features. (2) A directed search for acoustic properties is carried out in the vicinity of these landmarks, leading to estimates of the articulator-bound features of the segments.

1.5.2 Identification of Landmarks

This work is a component of a proposed knowledge-based speech recognition system which uses *landmarks* to guide the search for *distinctive features*. An algorithm has been developed for automatically detecting acoustically-abrupt landmarks. The algorithm is hierarchically-structured, and is rooted in linguistic and speech production theory. It uses several factors to detect landmarks: energy abruptness in six frequency bands and at two levels of temporal resolution, segmental duration, broad phonetic class constraints, and constraints imposed by articulation on landmark sequences. Tested on a database of continuous, clean speech of women and men, the landmark detector has detection rates 90 percent and above, substitution rates between 1 percent and 3 percent, and insertion rates 16 percent and below. The effects of landmark type, phonetic class, and prosody have been evaluated. These promising results indicate that landmarks are robustly-identifiable points in a speech waveform and that a landmark detector as a front-end in a speech recognition system could be feasible.

1.5.3 Automatic Context-sensitive Estimation of Distinctive Features at Landmarks

We have continued to develop procedures for estimation of distinctive feature values at articulatory landmarks. Toward this end, we are modeling each distinctive feature as a table containing phonetic contexts, a list of signal measurements (acoustic correlates) which provide information about the feature in each context, and, for each context, a statistical model for evaluating the feature given the measurements. The model of a distinctive feature may include several sets of acoustic correlates, each indexed by a different set of context features. Context features are typically lower-level features of the same segment; for example, manner features ([continuant, sonorant]) provide context for the identification of articulator-bound features ([lips, blade]). The acoustic correlates of a feature can be any static or kinematic spectral measurements defined relative to the time of the landmark. The statistical model is a simple N-dimensional Gaussian hypothesis test. A measurement program has been developed to test the usefulness of user-defined phonetic contexts. Measures of voice-onset time and formant locus classification are being developed as examples.

1.6 Development of Computer and Data Processing Facilities

A program has been written to display high-quality spectrograms on our engineering workstations. The spectrogram display includes a cursor which can be time-aligned with cursors in MITSYN-based programs for interactive data extraction or in the Klatt software used for acoustic analyses. This display will facilitate the identification of abrupt spectral transitions in acoustic signals.

The KLSYN88 speech synthesis facility has been augmented to include a transient source, consisting of a single sample, which can be filtered in the same way as the frication source. This new source is used to simulate the transients that occur at the release of stop or affricate consonants.

1.7 Publications

1.7.1 Papers Published

Holmberg, E., J. Perkell, R. Hillman, and C. Gress. "Individual Variation in Measures of Voice." *Phonetica* 51(1-3): 30-37 (1994).

Holmberg, E., R. Hillman, J. Perkell, and C. Gress. "Relationships between Intra-speaker Variation in Aerodynamic Measures of Voice Production and Variation in SPL Across Repeated Recordings." *J. Speech Hear. Res.* 37: 484-495 (1994).

Johnson, M.A. "Automatic Context-Sensitive Measurement of the Acoustic Correlates of Distinctive Features at Landmarks." *Proceedings of the International Conference on Spoken Language Processing* 3, Yokohama, Japan, 1994, pp. 1639-1642.

Lane, H., J. Wozniak, and J.S. Perkell. "Changes in Voice-Onset Time in Speakers with Cochlear Implants." *J. Acoust. Soc. Am.* 96(1): 56-64 (1994).

Matthies, M.L., M.A. Svirsky, H. Lane, and J.S. Perkell. "A Preliminary Study of the Effects of Cochlear Implants on the Production of Sibilants." *J. Acoust. Soc. Am.* 96 (3): 1367-1373 (1994).

Perkell, J., R. Hillman, and E. Holmberg. "Group Differences in Measures of Voice Production and Revised Values of Maximum Airflow Declination Rate." *J. Acoust. Soc. Am.* 96 (2): 695-698 (1994).

Shattuck-Hufnagel, S. "Stress Shift as Early Pitch Accent Placement: A Comment on Beckman and Edwards." In *Phonological Structure and Phonetic Form, Papers in Laboratory Phonology III.* Ed. P.A. Keating. Cambridge, Massachusetts: Cambridge University Press, 1994, pp. 34-42.

Shattuck-Hufnagel, S., M. Ostendorf, and K. Ross. "Stress Shift and Early Pitch Accent Placement in Lexical Items in American English." *J. Phonetics* 22: 357-388 (1994).

Sperry, E., R. Hillman, and J.S. Perkell. "Use of Inductance Plethysmography to Assess Respiratory Function in a Patient with Vocal Nodules." *J. Med. Speech-Language Path.* 2: 137-145 (1994).

Stevens, K.N. "Phonetic Evidence for Hierarchies of Features." In *Phonological Structure and Phonetic Form, Papers in Laboratory Phonology III.* Ed. P.A. Keating. Cambridge, Massachusetts: Cambridge University Press, 1994, pp. 242-258.

Stevens, K.N. "Scientific Substrates of Speech Production." In *Introduction to Communication Sciences and Disorders.* Ed. F. Minifie. San Diego: Singular, 1994, pp. 399-437.

Stevens, K.N., C.A. Bickley, and D.R. Williams. "Control of a Klatt Synthesizer by Articulatory Parameters." *Proc. Int. Conf. on Spoken Language Process.* 1: 183-186 (1994).

Stevens, K.N., and H.M. Hanson. "Classification of Glottal Vibration from Acoustic Measurements." In *Vocal Fold Physiology: Voice Quality Control.* Eds. O. Fujimura and M. Hirano. San Diego: Singular, 1994, pp. 147-170.

Wilhelms-Tricarico, R., and J.S. Perkell. "Biomechanical and Physiologically based Speech Modeling." *Proceedings of the 2nd ESCA/IEEE Workshop on Speech Synthesis,* New Paltz, New York, 1994, pp. 17-20.

1.7.2 Papers Accepted for Publication

Hillman, R.E., E.B. Holmberg, J.S. Perkell, J. Kobler, P. Guiod, C. Gress, and E.E. Sperry. "Speech Respiration in Adult Females with Vocal Nodules." *J. Speech Hear. Res.* Forthcoming.

Keyser, J.S., and K.N. Stevens. "Feature Geometry and the Vocal Tract." *Phonology.* Forthcoming.

Perkell, J.S. "Articulatory Processes." In *Handbook of Phonetic Science.* Eds. W. Hardcastle and J. Laver. Oxford: Blackwell. Forthcoming.

Perkell, J.S., M.L. Matthies, M.A. Svirsky, and M.I. Jordan. "Goal-based Speech Motor Control: A Theoretical Framework and Some Preliminary Data." *J. of Phonetics.* Forthcoming.

Shattuck-Hufnagel, S. "The Importance of Phonological Transcription in Empirical Approaches to 'Stress Shift' vs. Early Accent: a Commentary on Grabe and Warren Stress Shift: Do Speakers do it or do Listeners hear it? and Vogel, Bunnell and Hoskins The Phonology and Phonetics of the Rhythm Rule." In *Papers in Laboratory Phonology IV.* Eds. B.A. Connell and A. Arvaniti. Cambridge, Massachusetts: Cambridge University Press. Forthcoming.

Shattuck-Hufnagel, S. "Slips of the Tongue." In *Encyclopedia of Language and Linguistics*. Eds. R.E. Asher and J.M.Y. Simpson. Oxford: Pergamon Press. Forthcoming.

Stevens, K.N. "Articulatory-acoustic-auditory Relationships." In *Handbook of Phonetic Sciences*. Eds. W. Hardcastle and J. Laver. Blackwell: Oxford. Forthcoming.

Wilhelms-Tricarico, R. "Physiological Modeling of Speech Production: Methods for Modeling Soft-tissue Articulators." *J. Acoust. Soc. Am.* Forthcoming.

1.7.3 Papers Submitted for Publication

Chen, M. "Acoustic Parameters of Nasalized Vowels in Hearing-impaired and Normal-hearing Speakers." Submitted to *J. Acoust. Soc. Am.*

Dilley, L., S. Shattuck-Hufnagel, and M. Ostendorf. "Glottalization of Vowel-Initial Syllables as a Function of Prosodic Structure." Submitted to *J. Phonetics.*

Holmberg, E.B., R.E. Hillman, J.S. Perkell, and S. Goldman. "Comparisons Among Aerodynamic, Electroglottographic, and Acoustic Spectrum Measures of Female Voices." Submitted to *J. Speech Hear. Res.*

Lane, H., J. Wozniak, M.L. Matthies, M.A. Svirsky, and J.S. Perkell. "Phonemic Resetting vs. Postural Adjustments in the Speech of Cochlear Implant Users: an Exploration of Voice Onset Time." Submitted to *J. Acoust. Soc. Am.*

Manuel, S.Y. "Speakers Nasalize /ð/ After /n/, But Listeners Still Hear /ð/." Submitted to *J. Phonetics.*

Ostendorf, M., P. Price, and S. Shattuck-Hufnagel. "A Database of Prosodically Labelled Speech." In preparation.

Perkell, J.S., Matthies, M.L., and Svirsky, M.A. "Goal-based Speech Motor Control: A Theoretical Framework and some Preliminary Data." Submitted to *J. Phonetics.*

Perkell, J.S. "Properties of the Tongue Help to Define Vowel Categories: Hypotheses Based on Physiologically-oriented Modeling." Submitted to *J. Phonetics.*

Shattuck-Hufnagel, S., L. Dilley, and M. Ostendorf. "Prosodic Constraints on Glottalization of Vowel-initial Syllables in American English." Abstract submitted for Accoustical Society of America Meeting, June 1994.

Stevens, K.N. "Models of Speech Production." In *Handbook of Acoustics*. Ed. M. Crocker. New York: Wiley.

Wilhelms-Tricarico, R. "Biomechanical and Physiologically Based Speech Modeling." Submitted to *J. Phonetics.*

1.7.4 Thesis

Massey, N.S. *Transients at Stop Consonant Releases.* S.M. thesis. Dept. of Electr. Eng. and Comput. Sci., MIT, 1994.