# Chapter 1. Speech Communication

## Academic and Research Staff

Professor Kenneth N. Stevens, Professor Jonathan Allen, Professor Morris Halle, Professor Samuel J. Keyser, Dr. Krishna K. Govindarajan, Dr. Helen M. Hanson, Dr. Joseph S. Perkell, Dr. Stefanie Shattuck-Hufnagel, Dr. Reiner Wilhelms-Tricarico, Seth M. Hall, Jennell C. Vick, Majid Zandipour

## Visiting Scientists and Research Affiliates

Dr. Ashraf S. Alkhairy,[1] Dr. Corine A. Bickley, Dr. Suzanne E. Boyce,[2] Dr. Carol Y. Espy-Wilson,[3] Dr. David Gow,[4] Dr. Frank Guenther,[5] Dr. Robert E. Hillman,[6] Dr. Caroline Huang,[7] Dr. Harlan Lane,[8] Dr. John I. Makhoul,[9] Dr. Sharon Y. Manuel,[10] Dr. Melanie L. Matthies,[11] Dr. Richard S. McGowan,[12] Dr. Alice E. Turk,[13] Dr. Rajesh Verma,[14] Dr. Lorin F. Wilde,[15] Astrid Hagen,[16] Jane Wozniak[17]

## Graduate Students

Helen Chen, Harold A. Cheyne, Jeung-Yoon Choi, Erika Chuang, Michael P. Harms, Mark A. Hasegawa-Johnson, Lekisha S. Jackson, Hong-Kwang J. Kuo, Kelly L. Poort, Adrienne Prahler, Andrew I. Russell, Janet L. Slifka, Jason L. Smith, Yong Zhang

## Undergraduate Students

Laura C. Dilley, Emily J. Hanna, Dameon Harrell, Stefan H. Hurwitz, Mariya A. Ishutkina, Genevieve R. Lada, Teresa K. Lai, Adrian D. Perez, Dawn Perlner, Delsey M. Sherrill, Erik Strand, Jeremy Y. Vogelmann, Sophia C. Yuditskaya

## Technical and Support Staff

Katherine W. Kwong, Arlene E. Wint

---

1 KACST, Riyadh, Saudi Arabia.
2 Department of Communication Disorders, University of Cincinnati, Cincinnati, Ohio.
3 Department of Electrical Engineering, Boston University, Boston, Massachusetts.
4 Department of Psychology, Salem State College, Salem, Massachusetts.
5 Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts.
6 Massachusetts Eye and Ear Infirmary, Boston, Massachusetts.
7 Altech, Inc., Cambridge, and Boston University, Boston, Massachusetts.
8 Department of Psychology, Northeastern University, Boston, Massachusetts.
9 Bolt, Beranek and Newman, Inc., Cambridge, Massachusetts.
10 Department of Communication Sciences and Disorders, Emerson College, Boston, Massachusetts.
11 Department of Communication Disorders, Boston University, Boston, Massachusetts.
12 Sensimetrics Corporation, Cambridge, Massachusetts.
13 Department of Linguistics, University of Edinburgh, Edinburgh, Scotland.
14 CEERI Centre, CSIR Complex, New Delhi, India.
15 Lernout and Hauspie Speech Products, Burlington, Massachusetts.
16 University of Erlangen-Nürnberg, Erlangen, Germany.
17 Concord Area Special Education (CASE) Collaboratives, Concord, Massachusetts.

**Sponsors**

C.J. Lebel Fellowship
Dennis Klatt Memorial Fund
National Institutes of Health
- Grant R01-DC00075
- Grant R01-DC01291
- Grant R01-DC01925
- Grant R01-DC02125
- Grant R01-DC02978
- Grant R01-DC03007
- Grant R29-DC02525-01A1
- Grant F32-DC00194
- Grant F32-DC00205
- Grant T32-DC00038

National Science Foundation
- Grant IRI 93-14967[18]
- Grant INT 94-21146[19]

## 1.1 Studies of the Acoustics, Perception, Synthesis, and Modeling of Speech Sounds

### 1.1.1 Glottal Characteristics of Female and Male Speakers: Data from Acoustic and Physiological Recordings

The configuration of the vocal folds during the production of both normal and disordered speech has an influence on the voicing source waveform and, thus, affects perceived voice quality. Voice quality contains both linguistic and nonlinguistic information which listeners utilize in their efforts to understand spoken language and to recognize speakers. In clinical settings, voice quality also relays information about the health of the voice-production mechanism. The ability to glean voice quality from speech waveforms has implications for computer-based speech recognition, speaker recognition, and speech synthesis, and may be of value for diagnosis or treatment in clinical settings.

Theoretical models have been used to predict how changes in vocal-fold configuration are manifested in the output speech waveform. In previous work, we used acoustic-based methods to study variations in vocal-fold configuration among female speakers (nondisordered). We found a good correlation among the acoustic parameters and perceptions of breathy voice. Preliminary evidence gathered from fiberscopic images of the vocal folds during phonation suggest that these acoustic measures may be useful for categorizing the female speakers by vocal-fold configuration. That work has been extended in two ways.

First, we collected data from 21 male speakers (also nondisordered). Because male speakers are less likely than female speakers to have posterior glottal openings during phonation, we expect to find less parameter variation among male speakers, in addition to significant differences in mean values, when compared with the results from females. The data are consistent with those expectations. The average values of the acoustic parameters were significantly lower for males than for females. Although there are individual differences among the male speakers, the variation is smaller than that among female speakers. These differences in mean and variation are stronger for some parameters than for others. Several of the subjects displayed evidence of a secondary excitation, possibly occurring at glottal opening, which was not evident in the data from the female speakers. This observation is also in line with gender-based differences in glottal configuration.

In a further attempt to verify the correlation among the acoustic measures and actual glottal configurations, images of the vocal folds during phonation were collected in collaboration with the Research Institute for Logopedics and Phoniatrics at Tokyo University. The images were obtained using an endoscope and were recorded via a high-speed digital imaging system. As with the earlier fiberscopic data, preliminary analysis suggests that female speakers with relatively large posterior glottal openings also have relatively large degrees of spectral tilt and wider first-formant bandwidths.

### 1.1.2 Cues for Consonant Voicing for Postvocalic Consonants: The *Writer-Rider* Distinction

It is generally assumed that the voicing distinction of the alveolar consonant in *writer* versus *rider* is neutralized in the flap, but that the distinction is cued by the duration of the preceding vowel. As part of a

18 Under subcontract to Stanford Research Institute (SRI) Project ECU 5652.

19 U.S.-India Cooperative Science Program

larger study of the voicing distinction for consonants, we have been examining in detail the acoustic characteristics of the stressed vowel and the consonant in contrasting words like *writer*/*rider*, *doughty*/*dowdy*, *ricing*/*rising*, etc., produced by several talkers. In the case of *writer*/*rider*, the measured acoustic attributes include vowel length, time course of the first two formants in the vowel, and certain acoustic properties of the burst. An unexpected finding was that the trajectories of the formants of the diphthong were different for the two words. The offglide of the diphthong toward /i/ was more extreme for *writer* than for *rider*; that is, the first formant frequency was lower and the second formant frequency was higher immediately preceding the flap for *writer* than for *rider*. A perception experiment was prepared with synthesized versions of these words, in which the offglides of the formants were systematically varied. This experiment verified that listeners used these offglides to distinguish between *writer* and *rider*.

One interpretation of this result is that the more extreme vowel offglide for *writer* is a manifestation of a more extreme pharynx expansion, creating a condition that prevents further expansion during the following consonant. Since expansion of the vocal-tract volume is a prerequisite for a voiced obstruent consonant, then the more extreme offglide creates a condition that inhibits voicing during the consonant. For rider, the pharyngeal expansion in the offglide is less extreme, and there is room for further expansion in the consonant. Similar patterns for the vowel offglide are also observed for pairs like *ricing*/*rising* and *doughty*/*dowdy*. This method of enhancing the voicing contrast for postvocalic obstruent consonants can apparently be added to the inventory of cues for voicing in such consonants.

### 1.1.3 Burst and Formant Transition Cue Integration and Spectral Asynchrony Perception in Stop Consonant Perception

Two primary cues, formant transitions and burst cues, have been implicated in the perception of stop consonants. However, the exact nature of these cues, and how they are integrated, remains unknown. This study investigated the interaction and representation of these cues through two experiments.

In the first experiment, listeners identified synthetic syllable-initial consonant-vowel (CV) stimuli in which the second (*F*2) and third formant (*F*3) transitions, and the burst were independently varied in both a front vowel (/ɛ/) and back vowel (/ɑ/) context. The resulting identification surface shows similarities across the four subjects tested. When there was no burst cue, most listeners were reliably able to identify /b, d/ in the back vowel context and /b, g/ in the front vowel context; and, most listeners had a difficult time identifying /g/ in the back vowel context, and half of the listeners had difficulty identifying /d/ in the front vowel context. With the addition of the burst, listeners' percepts were more consistent and showed a trading relation between the burst cue and the formant transition cue, especially for /d, g/. In the back vowel context, the influence of the burst was to increase /gɑ/ responses for formant transitions corresponding to a /gɑ/ percept, and increase /gɑ/ responses for formant transitions corresponding to a /dɑ/ percept when the burst frequency is close to *F*2 (near 1500 Hz). In addition, for a high burst center frequency, most subjects tended to hear /dɑ/ even if the formant transitions corresponded to a /gɑ/ percept. In the front vowel context, two of the subjects showed little influence of the burst. The other two listeners showed a trading relationship between /d/ and /g/ when the burst was near 1800 Hz, i.e., near *F*2.

Based on the identification surfaces, it was hypothesized that the burst center frequency is less critical to identification than which formant the burst is near, e.g., *F*2. Thus, a second experiment was performed using burstless /Cɑ/ stimuli in which the *F*2 transition was varied. In addition, the stimuli either had one formant start prior to the other formants or had all formants start at the same time. The results show that when *F*2 started prior to the other formants, listeners identified all stimuli as /gɑ/ even though the formant transition cue corresponded to /bɑ/ or /dɑ/. However, if *F*3 started ahead of the other stimuli, then listeners based their identification primarily on the formant transition cue. Listeners appear to be interpreting the burst as a formant that starts prior to the other formants, i.e., a "leading" formant. Thus, the results suggest that listeners identify stop consonants based on spectral asynchrony in conjunction with formant transition cues.

### 1.1.4 A Longitudinal Study of Speech Production

Measurements have been made on a corpus of bisyllabic nonsense utterances spoken by three adult male talkers on two occasions separated by 30

years. The 1960 and 1990 recordings were processed similarly so that comparisons could be made directly. For the vowels, properties included measures of spectrum tilt, formant frequencies, fundamental frequency, and duration. Consonantal measures included durations, spectra of frication noise for stop bursts and fricatives, spectra of aspiration noise for voiceless stop consonants, and spectra of nasal consonants and of the adjacent nasalized vowels. In informal listening tests, subjects were presented with utterances from the two recording sessions and asked to determine which utterances were made by the older speakers.

For the most part, the acoustic properties remained remarkably stable over the 30-year interval. The stable properties included vowel and consonant durations and vowel formant frequencies. The fundamental frequency increased by a small but consistent amount for all talkers (2 to 8 Hz). Two talkers showed a significant decrease in high frequency amplitude for vowels, suggesting a reduction in the abruptness of the glottal source at the time of glottal closure during phonation. For all talkers there was a reduction in high-frequency amplitude of aspiration noise in voiceless aspirated stop consonants. One talker also showed a reduction (10 dB or more) in high frequency amplitude of frication noise in the tongue-blade consonants /d/, /s/, and /š/. Spectrum measurements in nasal consonants indicated that for one talker there were changes in the acoustic properties of the nasal cavity between the two recording times. Listeners were somewhat inconsistent in judging which utterances were produced at the later age, but were more successful in making this judgment for the talkers who showed an increased spectral tilt with age.

The number of talkers is, of course, too small to provide reliable data on changes in the acoustic characteristics of speech with age. The data do show, however, that large individual differences can be expected in the attributes that undergo change, and in the amount of change of these attributes. It should also be noted that for the three talkers attributes such as duration patterns, fundamental frequency, and formant frequencies were quite different, and these differences that were unique to the talkers did not change over the 30-year period.

### 1.1.5 MEG Studies of Vowel Processing in Auditory Cortex

In collaborative research between MIT, the University of Delaware, University of Maryland, and the University of California, San Francisco, the brain imaging technique magnetoencephalography (MEG) is being used to determine which attributes of vowels give rise to distinctive responses in auditory cortex. Previously, it was hypothesized that the M100 latency—the peak neuromagnetic activity that occurs at approximately 100 ms after the stimulus onset—was a function of the first formant frequency (*F*1) and not the fundamental frequency (*F*0). This hypothesis was derived from three-formant vowels, single-formant vowels, and two-tone complexes that matched *F*0 and *F*1 in the single formant vowels.

In 1996, Ragout and LePaul-Ercole presented two-formant vowels to subjects and found that the M100 latency tracked the fundamental frequency (*F*0) and not *F*1. However, their stimuli did not match normal speech: they used stimuli in which the amplitude of *F*2 was 20 dB greater than the amplitude of *F*1, leading to a "duck-like" sound. In order to reconcile Ragout and LePaul-Ercole's results with the aforementioned results, a set of two-formant vowels with varying formant amplitudes was presented to subjects. Preliminary results indicate that when the second formant amplitude becomes greater than the first formant amplitude by 6 dB, the M100 no longer tracks *F*1 but tracks *F*0. Thus, the M100 latency shifts from tracking *F*1 to *F*0 as the amplitude of *F*2 becomes greater than *F*1, i.e., as the stimuli becomes less speechlike. Future experiments are planned to determine if this result is providing evidence for a special speech processing module in the brain.

### 1.1.6 Synthesis of Hindi

A collaborative project on rule-generated synthesis of Hindi speech has been initiated with the Central Electronics Engineering Research Institute (CEERI) Center in New Delhi, under support from the Division of International Programs of the National Science Foundation. The project at CEERI has led to the development of an inventory of Hindi syllables synthesized with a Klatt formant synthesizer, and the formulation of procedures for concatenating the control parameters for these syllables to produce continuous utterances. Some collaboration in the fine-tuning of

the syllables and in the formulation of rules for intonation has been provided by the Speech Communication group.

### 1.1.7 Modeling and Synthesis of Lateral Consonant /l/

The lateral consonant in English is generally produced with a backed tongue body, a midline closure of the tongue blade at the alveolar ridge, and a path around one or both of the lateral edges of the tongue blade. In pre-vocalic lateral consonants, the release of the closure causes a discontinuity in the spectral characteristics of the sound. Past attempts to synthesize syllable-initial lateral consonants using formant changes alone to represent the discontinuity have not been entirely satisfactory. Data from prior research have shown rapid changes not only in the formant frequencies but also in the glottal source amplitude and spectrum and in the amplitudes of the formant peaks at the consonant release. New measurements of these parameters have been made from additional recordings of a number of speakers. The measurements have been guided by models of lateral-consonant production. Based on these data, new attempts of synthesis have incorporated changes in source amplitudes, formant bandwidths, and the location of a pole-zero pair. Including these additional parameters improves the naturalness of the synthesized lateral-vowel syllables in initial perception tests.

### 1.1.8 Modeling and Synthesis of Nasal Consonants

During the production of nasal consonants, the transfer function of the vocal tract contains zeros and poles with certain bandwidths and frequencies which change as a function of time. We are trying to refine the theory related to how these values change. An objective is to develop improved methods for synthesizing these consonants—methods which are consistent with the theory. Some analysis of the spectrum changes in utterances of intervocalic nasal consonants has been carried out, and these data have been used to refine the parameters for numerical simulation of models of the vocal and nasal tracts. Synthesis of some brief utterances was carried out, based on these simulations. The synthesized utterances sounded better than synthesis done using only the old methods. More work is being done to determine what acoustic characteristics of nasal consonants are perceptually important, so that focus can be directed to only the aspects of the theory that are most salient.

## 1.2 Studies of Normal Speech Production

### 1.2.1 Experimental Studies Relating to Speech Clarity, Rate, and Economy of Effort

#### Clarity Versus Economy of Effort in Speech Production I: A Preliminary Study of Inter-subject Differences and Modeling Issues

This study explores the idea that clear speech is produced with greater "articulatory effort" than normal speech. Kinematic and acoustic data were gathered from seven subjects as they pronounced multiple repetitions of utterances in different speaking conditions, including normal, fast, clear and slow. Data were analyzed within a framework based on a dynamical model of single-axis frictionless movements, in which peak movement speed is used as a relative measure of articulatory effort. There were differences in peak movement speed, distance and duration among the conditions and among the speakers. Three speakers produced "clear" utterances with movements that had larger distances and durations than those for "normal" utterances. Analyses of these data within a peak speed, distance, duration "performance space" indicated increased effort (reflected in greater peak speed) in the clear condition for the three speakers. The remaining speakers used other combinations of parameters to produce the clear condition. The validity of the simple dynamical model for analyzing these complex movements was considered by examining several additional parameters. Some movement characteristics departed from those required for the model-based analysis presumably because the articulators are structurally complicated and interact with one another mechanically. More refined tests of control strategies for different speaking styles will depend on future analyses of more complicated movements with more realistic models.

### Clarity Versus Economy of Effort in Speech Production II: Kinematic Performance Spaces for Cyclical and Speech Movements

This study was designed to test the hypothesis that the kinematic manipulations used by speakers to control clarity are influenced by kinematic performance limits. A range of kinematic parameter values was elicited by having the same seven subjects produce cyclical CV movements of lips, tongue blade and tongue dorsum (/bɑ/, /dɑ/, /gɑ/), at rates ranging from 1 to 6 Hz. The resulting measures were used to establish speaker- and articulator-specific kinematic performance spaces, defined by movement duration, displacement, and peak speed. These data were compared with speech movement data produced by the subjects in several different speaking conditions in the preceding study. The amount of overlap of the speech data and cyclical data varied across speakers from almost no overlap to complete overlap. Generally, speech movements were larger for a given movement duration than cyclical movements, indicating that the speech movements were faster and produced with greater effort, according to the performance space analysis. It was hypothesized that the cyclical movements of the tongue and lips were slower than the speech movements because they were more constrained by (coupled to) the relatively massive mandible. To test this hypothesis, a comparison was made of cyclical movements in maxillary versus mandibular frames of reference. The results indicate that the cyclical movements were not strongly coupled to mandible movements. The overall results indicate that the cyclical task did not succeed in defining the upper limits of kinematic performance spaces within which the speech data were confined for most speakers. Thus, the performance limits hypothesis could not be tested effectively. The differences between the speech and cyclical movements may be due to other factors, such as differences in speakers' "skill" with the two types of movement.

### Variations in Speech Movement Kinematics and Temporal Patterns of Coarticulation with Changes in Clarity and Rate

This study tests the hypothesis that the relative timing of articulatory movements at sound segment boundaries is conditioned by a compromise between economy of effort and a requirement for clarity. On the one hand, articulatory movements (such as lip rounding movements from /i/ to /u/ in /iC($C_n$)u/) are programmed to minimize effort (e.g., peak velocity); therefore, they cannot be too fast. On the other hand, if the movements are too slow, they begin too early or end too late (with respect to the /iC/ and /Cu/ boundaries) and produce less distinct vowels. Movement (EMMA) and acoustic data were collected from the same seven subjects. The speech materials were designed to investigate coarticulation in movements of the lips and of the tongue. They included $V_1C(C_n)V_2$ sequences embedded in carrier phrases, in which $V_1$ and $V_2$ were /i/ and /u/. For example: "Say leaked coot after it." (for lip movements), "Say he moo after it." (for tongue movements). They were spoken in three conditions, normal, clear and fast. Each subject recorded about 1100 tokens. The analyses focused on the amount of coarticulation (overlap) of the /i-u/ transition movement within the acoustic interval of the /i/, along with several other measures. Consonant string duration was longest for the clear condition and shortest for the fast condition. Peak velocities were higher in the fast and clear conditions than in the normal condition. The coarticulation effects were small and were observed more for the lip than the tongue-body movements. Generally, there was more overlap in the fast condition than the normal condition, but not less overlap in the clear condition than the normal condition. The effects of overlap on formant values were small. Thus, producing the clear condition involved increases of consonant string duration and peak velocity but not coarticulation differences. While there were some small increases in coarticulation in the fast condition, they did not seem to affect the spectral integrity of the /i/. Even though there was evidence of increased effort (as indexed by peak velocity) in the clear and fast conditions, the hypothesized effects of a tradeoff between clarity and economy of effort were minimally evident in formant values for /i/ and measures of coarticulation (overlap).

### Interarticulator Coordination in Achieving Acoustic-phonetic Goals: Motor Equivalence Studies

These studies are based on our preliminary findings of "motor equivalent" trading relations between lip rounding and tongue-body raising for the vowel /u/, which we have interpreted as supporting our hypothesis that speech motor control is based on acoustic goals. We have hypothesized that when two articulators contribute to producing an acoustic cue and the planned movement of one of the articulators might make the resultant acoustic trajectory miss the goal region for the cue, a compensatory adjustment is planned in the movement of the other articulator to

help keep the acoustic trajectory within the goal region. Furthermore, due to economy of effort, the compensation is limited to an amount that makes the acoustic trajectory just pass through the edge of the goal region on its way to the next goal. Thus, we expect to observe such compensatory covariation mainly among tokens near the edge of the goal region, i.e., less canonical tokens. We also hypothesize that the most canonical tokens (near the center of the acoustic goal region) are produced with "cooperative coordination." A canonical token of /u/ would be produced by cooperative lip rounding and tongue body raising movements. Overall, we expect to find positive correlations of lip protrusion and tongue body raising among tokens of /u/ that are acoustically most canonical and negative correlations among tokens that are least canonical.

We have tested this hypothesis for the sounds /u/, /r/ and /ʃ/ pronounced in carrier phrases by the seven speakers. These sounds were chosen because they are produced with independently controllable constrictions formed by the tongue and by the lips, making it possible to look for motor-equivalent behavior with correlation analysis. Each subject pronounced a total of about 650 tokens containing the sounds embedded in carrier phrases. To avoid biasing the correlations used to test the hypothesis through partitioning the data set into more and less canonical tokens, we attempted to create less canonical subsets *a priori*, with manipulations of phonetic context and speaking condition.

In one analysis approach, we have extracted and analyzed all of the articulatory data (mid-sound tongue and lip transducer positions) and the acoustic data for /u/ (formants) and /ʃ/ (spectral median and symmetry).

The acoustic measures indicate that we were only partly successful in eliciting subsets that were more and less canonical. There is evidence supporting motor equivalence for all three sounds. For /u/, the findings are generally related to how canonical the tokens were: in most cases there was compensatory coordination (motor equivalence) for less canonical tokens and cooperative coordination for more canonical tokens. For /ʃ/, there were significant correlations in 28% of the possible cases. All reflected motor equivalence: when the tongue blade was further forward, the lips compensated with more protrusion, presumably to maintain a front cavity that was large enough to achieve a low spectral center of gravity. When there was a difference between a subject's utterance subsets in how canonical the tokens were, the motor equivalent tokens were less canonical. No evidence of cooperative coordination was found for /ʃ/.

A study of acoustic variability during American English /r/ production also tests the hypothesis that speakers utilize an acoustic, rather than articulatory, planning space for speech production. Acoustic and articulatory recordings of the seven speakers reveal that speakers utilize systematic articulatory tradeoffs to maintain acoustic stability when producing the phoneme /r/. Distinct articulator configurations used to produce /r/ in various phonetic contexts show systematic tradeoffs between the cross-sectional areas of different vocal tract sections. Analysis of acoustic and articulatory variabilities reveals that these tradeoffs act to reduce acoustic variability, thus allowing large contextual variations in vocal tract shape; these contextual variations in turn apparently reduce the amount of articulatory movement required in keeping with the principle of economy of effort in speech production.

### 1.2.2 Physiological Modeling of Speech Production

#### Studies of Vocal-tract Anatomy

In cooperation with Dr. Chao-Min Wu at the University of Wisconsin, software was developed for the interactive visualization and mapping between two data sets: (1) anatomical sections from the Visible Human project and (2) a series of detailed drawings of histological sections of tongue. One part of the software compiles a set of images into a stack and computes spatial display sections through the data. Another component can be used to identify homologous points between two 3D image data sets. These point pairs calibrate a mapping between the data sets to partially integrate them. These methods were incorporated into the current work, described below.

#### Prototyping an Accurate Finite-element Model of the Vocal Tract

Progress was made in prototyping an accurate finite-element model of the tongue and floor of the mouth, based mainly on cryo-section data from the Visible Human project. Software was developed for various visualization and measurement tasks. For example, multiple arbitrarily oriented cross sections of a 3D stack of section images are combined into a 3D view, making it possible to capture measurements of vocal-

tract morphology from the cryo-section data. Commercial visualization and measurement programs are used for drafting a topologically accurate model, and locations of node points and spline control points are imported from the programs. These techniques will be used to incorporate data on individual vocal-tract morphology from MR images into vocal-tract simulations.

### 1.2.3 Theoretical Developments in Speech Production

In a theoretical paper on speech motor control, an overview and supporting data are presented about the control of the segmental component of speech production. Findings of "motor-equivalent" trading relations between the contributions of two constrictions to the same acoustic transfer function provide preliminary support for the idea that segmental control is based on acoustic or auditory-perceptual goals. The goals are determined partly by non-linear, quantal relations (called "saturation effects") between motor commands and articulatory movements and between articulation and sound. Since processing times would be too long to allow the use of auditory feedback for closed-loop error correction in achieving acoustic goals, the control mechanism must use a robust "internal model" of the relation between articulation and the sound output that is learned during speech acquisition.

Studies of the speech of cochlear implant and bilateral acoustic neuroma patients provide evidence supporting two roles for auditory feedback in adults: (1) maintenance of the internal model, and (2) monitoring the acoustic environment to help assure intelligibility by guiding relatively rapid adjustments in "postural" parameters underlying average sound level, speaking rate and amount of prosodically-based inflection of *F*0 and SPL.

### 1.2.4 Laryngeal Behavior at the Beginning of Breath Groups During Speech

Each person speaks with a particular timing that obviously depends on his or her linguistic intent but also must depend on the physical system generating the speech. Speech is created with an aerodynamic energy source that must be repeatedly replenished. This repeated activity imposes gross timing constraints on speech, and the mechanics of creating this source may impose particular timing effects at the beginning and end of each breath. We have initiated an experimental study of the influence of respiratory constraints on the temporal patterns of speech. First, we are examining how utterances are initiated at the beginnings of breath groups.

When a person draws in a breath, the air pressure inside the lungs is less than atmospheric pressure, creating a pressure gradient that causes air to move into the lungs. To produce speech, the speaker needs to compress the air in the lungs to increase the pressure while managing to retain a volume of air in the lungs during that compression. Multiple strategies could be used by the speaker to meet these requirements. Pressure may also be built up behind a complete blockage of the airway such as closure of the glottis or lips. Pressure may be built up behind an increased impedance of the vocal tract as its configuration is adjusted in anticipation of creating a speech segment. The particular method used may depend on the type of sound to be produced, the length of utterance, the location of emphasis in the utterance, and the relative timing of the muscles of respiration.

In an attempt to determine which method or methods the speaker uses, concurrent recordings of the acoustic signal, the signal from an electroglottograph (which provides an estimate of times of glottal openings and closings), a lung volume estimate (thorax and abdomen measures), and an unstrobed endoscopic videotape of the larynx were collected at the facilities of the Voice Laboratory at Massachusetts Eye and Ear Infirmary under the direction of Dr. Robert Hillman. Preliminary data show that at the initiation of some utterances, a speaker creates a glottal closure, which is then released when the articulators are in place for the initial segment of the utterance.

## 1.3 Speech Research Relating to Special Populations

### 1.3.1 Speech Production of Cochlear Implant (CI) and Bilateral Acoustic Neuroma (NF2) Patients

#### Longitudinal Studies

We have made three baseline pre-implant recordings on each of five additional CI subjects in the second year of this project. Additionally, all six research subjects have returned for post-implant recordings one week, one month, and three months following processor activation. Three of our six implant subjects have returned for their six-month visits and the remaining three subjects will complete six-month recordings by June. A total of 38 recordings have

been made this year with CI subjects. Over 75% of these recordings have been digitized an over 50% of the data from these recordings have been analyzed.

### Short-term Stimulus Modification Studies

One CI subject has participated in a "stimulus modification" experiment. During the recording of the subject's speech, an experimental speech processor was used to modify the subject's auditory feedback. Feedback alternated between the subject's regular program and another program that simulated no hearing. This recording has been digitized and the data are currently being analyzed.

### Perceptual Studies

All six CI subjects participated in this study at the time of each speech recording. Subjects are asked to discriminate eleven consonants and eight vowels from the natural speech of a same-gender speaker. We anticipate this test of perceptual ability will result in a diagnostic measure of perceptual benefit from the implant and subsequently support or disconfirm our hypotheses regarding relations between speech perception and production.

### Correlates of Posture

Four hypotheses inspired by our theory of the role of hearing in speech production were tested using acoustic and aerodynamic measures from seven CI subjects, three speakers who had severe reduction in hearing loss following surgery for Neurofibromatosis-2 (NF2), and one hard-of-hearing control speaker. These speakers made recordings of the Rainbow Passage before and after intervention. Evidence was found that supports the four hypotheses:

1. Deafened speakers who regain some hearing from cochlear prostheses will minimize effort with reductions in speech sound level.
2. To help reduce speech sound level, they will assume a more open glottal posture.
3. They will also minimize effort by terminating respiratory limbs closer to tidal-end respiratory level, FRC.
4. An effect of postural changes will be to change average values of air expenditure toward normative values.

### Coarticulation

Less coarticulation is presumably a feature of clear speech, and our theory predicts that deafened speakers will engage in clear speech. Therefore, we predict that deafened adults will show less coarticulation before and more coarticulation after their implant speech processors have been turned on. We have begun to test this prediction by extracting values of second formant frequency from readings of the vowel inventory in /bVt/ and /dVt/ contexts by seven implant users. Initial findings with one adult male reveal a statistically reliable overall increase in coarticulation following the activation of his implant speech processor. In particular the vowels /ɛ/, /æ/, and /u/ showed marked and reliable increases in coarticulation in sessions following the first session after activation of the processor.

## 1.3.2 Modeling and Analysis of Vowels Produced by Speakers with Vocal-fold Nodules

Speakers with vocal-fold nodules commonly use greater effort to speak. This increased effort is reflected in their use of higher than normal subglottal pressures to produce a particular sound pressure level (SPL). The SPL may be low because of decreased maximum flow declination rate (MFDR) of the glottal flow waveform, increased first formant bandwidth, or increased spectral tilt.

A parallel acoustic and aerodynamic study of selected patients with vocal nodules has been conducted. At comfortable voice, aerodynamic features are significantly different on average for the nodules group (at $p = 0.001$), whereas acoustic features are not significantly different (at $p = 0.001$) except for SPL and H1 – A1, the amplitude difference between the first harmonic (H1) and the first formant prominence (A1). Pn, defined as the percentage of data associated with speakers with nodules which are more than 1.96 standard deviations from the normal mean, is a measure of how well separated the two populations are for a particular feature. Even though SPL and H1 – A1 are significantly different statistically, the effect size is small, as indicated by the small Pn (21 and 2 percent, respectively). The difference in means for SPL is 1.8 dB and for H1 – A1 is 2.3 dB. In contrast, Pn for the aerodynamic features ranges from 8 to 21 percent. Ranked in order of Pn, the aerodynamic features are subglottal pressure, average flow, open quotient, minimum flow, AC flow, and MFDR. These observations show that it is easier

to differentiate the nodules group from the normal group using aerodynamic rather than acoustic features. By performing linear regression on the acoustic features with SPL as the independent variable, the two groups can be better separated. For example, the difference in group means for H1 – A1 increases to 3.9 dB from 2.3 dB, and Pn increases to 18 from 2 percent. The results of aerodynamic measures agree with previous findings.

The presence of a glottal chink can widen the first formant bandwidth. However, based on the glottal chink areas estimated from the minimum glottal flow, the bandwidth is increased on the average by only 1.1 times. It is not clear if the spectral tilt is increased, based on acoustic features. After regression with SPL, the mean H1 – A3 (where A3 = amplitude of third formant prominence) is 2.4 dB larger and mean A1 – A3 is 1.7 dB smaller for the nodules group compared with the normal group.

A modified two-mass model of vocal-fold vibration is proposed to simulate vocal folds with nodules. This model suggests that the MFDR can be decreased because of increased coupling stiffness between the masses and the presence of nodules which interfere with the normal closure activity of the vocal folds. The model also suggests that increasing the subglottal pressure can be used to compensate for the reduced MFDR, but energy dissipated due to collision is increased, implying greater potential for trauma to the vocal-fold tissues.

In summary, the greater effort used by speakers with vocal nodules results in differences in the aerodynamic features of their speech. However, the acoustic features show a smaller difference, reflecting the achievement of relatively good sound characteristics despite aberrant aerodynamics. A modified two-mass model is able to explain the need for higher subglottal pressures to produce a particular SPL, and it also demonstrates increased trauma potential to the vocal folds as a result of increased pressures.

### 1.3.3 Fricative Consonant Production by Some Dysarthric Speakers: Interpretations in Terms of Models for Fricatives

The aim of this research is to develop an inventory of noninvasive acoustic measures that can potentially provide a quantitative assessment of the production of /s/ by talkers with speech motor disorders. It is part of a broader study whose goal is to develop improved models of normal as well as dysarthric speakers' speech production and to use these models to interpret acoustic data obtained from the utterances of those speakers.

Utterances of eight dysarthric speakers and two normal speakers were used for this study. In particular, one repetition of each of nine words with initial /s/ sound was selected from a larger database, which was available from previous doctoral thesis work of Hwa-Ping Chang. The intelligibility of the words was determined previously by Chang.

Eight measurements were made from the fricative portions for each of the words. These included three measurements that assessed the spectrum shape of the fricative and its spectrum amplitude in relation to the vowel and five estimates of the degree to which the onset, offset, and time course of the fricative deviated from the normal pattern. These five estimates were made on a three-point scale and were based on observations of spectrograms of the utterances.

For purposes of analysis, the dysarthric speakers were divided into two groups according to their overall intelligibility: a low-intelligibility group (word intelligibility in the range 55-70 percent) and a high-intelligibility group (80-98 percent). The contributions of all eight measures to this classification were statistically significant. The measure that showed the highest correlation with intelligibility was a spectral measure (in the fricative) giving the difference (in dB) between the amplitude of the largest spectrum peak above 4 kHz (i.e., above the third-formant range) and the average spectrum peak corresponding to the second and third formants.

A task of future studies is to automate the measurements that are based on the scaling methods used in this research and to apply the measures to a larger population of speakers and utterances. The objective measures should also be related to assessments of clinicians along several different dimensions. The ultimate aim is to specify an inventory of quantitative acoustic measures that can be used to accurately assess the effectiveness of interventions as well as the amount of speech degeneration in speakers with neuromotor disorders.

### 1.3.4 Stop-consonant Production by Dysarthric Speakers: Use of Models to Interpret Acoustic Data

Acoustic measurements have been made on stop consonants produced by several normal and dysarthric speakers. The acoustic data were previously recorded by Hwa-Ping Chang and Helen Chen at MIT. In the present study, various aspects of production following release of the oral closure were quantified through the use of acoustic measures such as spectra and durations of noise bursts and aspiration noise, as well as shifts in frequencies of spectral prominences. Through comparison of these measurements from the normal and dysarthric speech, and based upon models of stop-consonant production, inferences were drawn regarding articulator placement (by examining burst spectra), rate of articulator release (from burst duration), tongue-body movements (from formant transitions), and vocal-fold state (from low-frequency spectra). The dysarthric speakers deviated from normal speakers particularly with respect to alveolar constriction location, rate of release, and tongue-body movement into the following vowel. For example, the lowest front-cavity resonance in the burst spectrum of an alveolar stop is normally in the range 3500-5500 Hz. For three of eight dysarthric speakers, the range was lowered to 1500-2800 Hz, indicating either placement of the tongue tip further back on the palate or formation of the constriction with the tongue body in a location similar to that of a velar stop.

## 1.4 Speech Production Planning and Prosody

### 1.4.1 Labeling Speech Databases

Speech labeling projects include the development and evaluation of several new transcription systems (for the perceived rhythm of spoken utterances and for the excitation source) which enrich existing labels for phonetic segments, part of speech, word boundaries and intonational phrases and prominences. Several new types of speech have been added, including samples from the CallHome database of extremely casual spontaneous telephone dialogues between family members and close friends, and a growing sample of digitized speech errors. In addition, a subset of the CallHome utterances are being labeled for distinctive features, following the methods described in section 1.5.1 below. These labeled databases provide the basis for evaluation of acoustic correlates of prosodic structure, as well as a resource for other laboratories to use (for example, the Media Lab at MIT made use of the ToBI-labeled radio news sample earlier this year.)

### 1.4.2 Acoustic-phonetic Correlates of Prosodic Structure

We evaluated the correlates of prosodic structure in several production experiments. In one, we examined durational cues to the direction of affiliation of a reduced syllable, as in "tuna choir" versus "tune acquire" versus "tune a choir." Results showed that rightward affiliation was reliably distinguished from leftward by patterns of syllable duration, but that the nature of the rightward affiliation (e.g., lexical, as in "acquire", versus phrasal, as in "a choir") was not. Moreover, the left-right distinction was most reliably observed for pitch accented words. In the other, we explored the extent of boundary-related duration lengthening in critical pairs of utterances such as "(John and Sue) or (Bill) will stay" versus "(John) and (Sue or Bill) will stay." Preliminary results for a single speaker show that preboundary lengthening is most pronounced on the final syllable of an intonational phrase, but significant lengthening is also found to the left up to and including the main stressed syllable of the final content word.

### 1.4.3 Evaluation and Application of a Transcription System for Regular Rhythm and Repeated *F*0 Contours

In an ongoing effort to investigate rhythm and intonation in running speech, a labeling system and on-line tutorial have been developed for indicating perceived rhythm and repeated *F*0 contours. The consistency of use of this labeling system was recently evaluated and a high degree of inter-labeler agreement was obtained. Furthermore, the system was applied to three minutes of continuous natural speech and a number of regions of speech were identified for which multiple listeners heard regular rhythms. For these regions, the interval between syllables heard as "beats" was found by measuring the time between successive vowel midpoints. Beat intervals were found to be in the range of 200-800 ms, which is consistent with recent findings of other investigators. Moreover, we found that for regions where five out of five labelers agreed, the speech contained a perceived regular rhythm; successive beat intervals were less variable in duration than when fewer than five out of five labelers agreed the speech was rhythmic. This observation is consistent with a hypothesis

that speech may in some cases be acoustically regularly timed and fits well with informal observations that speech frequently sounds regularly rhythmic. Furthermore, regions where three or more listeners heard a regular rhythm were more likely to have been also heard as containing repeated *F*0 contours than other regions of speech, and vice versa. In other words, regions perceived as containing regular rhythms tended to be heard as also bearing repeated intonation. This finding is consistent with observations from the literature on music and other non-speech auditory perceptual stimuli that rhythm and intonation may be interdependently perceived. We are currently working on a theory of prosody perception which unites observations from music, auditory stream analysis, and linguistics.

### 1.4.4 Initial Studies of the Perception of Prosodic Prominence

Initial studies of prominence perception have indicated that large *F*0 excursions between two adjacent full-vowel syllables (as in "transport") are perceptually ambiguous as to the location of the perceived prominence and that listeners can hear a sequence of *F*0 peaks and valleys on a sequence of full-vowel syllables (e.g., “We’re all right now”) as prominent either on the peaks or on the valleys. These results support an interpretive model of *F*0-governed prominence perception, in which listeners construct a pattern of syllable prominence using information of several kinds both from the signal and from their language knowledge. Such a view is also supported by pilot results showing a difference in which syllables are perceived as prominent for a normal production of a syllable string, where lexical stress is known, versus its reiterant imitation, where knowledge can provide no such constraints on the interpretation of the ambiguous *F*0 contour.

## 1.5 Models of Lexical Representation and Lexical Access

### 1.5.1 Labeling a Speech Database with Landmarks and Features

The process of lexical access requires that a sequence of words be retrieved from the acoustic signal that is radiated from a speaker’s mouth. In the lexical access process proposed here, an initial step is to transform an utterance into an intermediate representation in terms of a discrete sequence of sub-lexical or phonetic units. This representation consists of a sequence of landmarks at specific times in the utterance, together with a set of features associated with each landmark. The landmarks are of three kinds: acoustic discontinuities at consonantal closures and releases, locations of peaks in syllabic nuclei, and glide-generated minima in the signal. The inventory of features is the same as the distinctive features that are used in phonological descriptions of language.

Procedures have been developed for hand-labeling utterances in terms of these landmarks and features. The starting point in the labeling of a sentence is to generate (automatically) an idealized sequence of landmarks (without time labels) and features from the lexical representation of each individual word. This ideal labeling for each consonant would contain a sequence of two landmarks, each containing the features of the consonant. Each vowel and glide is assigned a single landmark with the lexically specified features attached. The labeling for the sentence consists of assigning times to the landmarks and modifying the landmarks (by deleting or adding landmarks) and the features based on observations of the signal. These procedures involve a combination of observation of displays of waveforms, spectra, and spectrograms, together with listening to segments of the utterance. Labeling has been completed for 50-odd sentences containing about 2500 landmarks.

There are several reasons for preparing this database:

1. Comparison of the hand-generated labels with the labels automatically generated from the sequence of lexical items provides quantitative information on the modifications that are introduced by a talker when running speech with various degrees of casualness is produced;
2. The landmarks and features for an utterance can be used as an input for testing and developing procedures for recovering the word sequence for the utterance; and
3. In the development of automatic procedures for extracting the landmarks and features for utterances, the hand-generated labels can be viewed as a goal against which potential automatically generated labels can be compared.

It is hoped that the database of utterances and labels can be publicly available when a sufficiently large number of utterances have been labeled.

### 1.5.2 Detection of Landmarks and Features in Continuous Speech: The Voicing Feature

A major task in implementing a model for lexical access from continuous speech is to develop well-defined procedures for detection and identification of the hierarchy of landmarks and features from analysis of the signal. For each landmark and feature there is a specific inventory of acoustic properties that contribute to this detection process. The inventory of properties that must be tapped to make the voiced/voiceless distinction for consonants is especially rich. Some initial progress has been made in developing signal processing methods that are relevant to this distinction.

In this initial work we have examined the rate of decrease of low-frequency spectrum amplitude in the closure interval for a number of intervocalic voiced and voiceless obstruent consonants produced by two speakers (one male, one female). The data show the expected more rapid decrease in low-frequency amplitude for the voiceless consonants than for the voiced. The amounts of decrease following consonantal closure are consistent with predictions based on estimates of vocal-tract wall expansion and of transglottal threshold pressures for phonation. There is, however, considerable variability in this measure of low-frequency amplitude, and it is clear that additional measures are needed to increase the reliability of voicing detection.

### 1.5.3 Deriving Word Sequences from a Landmark- and Feature-based Representation of Continuous Speech

Several steps are involved in designing and implementing a lexical access system based on landmarks and features. These include location of landmarks, determining features associated with these landmarks, converting this landmark-feature representation to a segment-feature representation, and matching to a lexicon that is specified using the same inventory of features. This project has examined the conversion and matching steps of this process. It uses as input landmarks and features obtained by hand-labeling a number of sentences produced by several talkers. These sentences contain words drawn from a small lexicon of about 250 items. The landmarks in the sentences identify discontinuities at consonantal closures and releases, vowel peaks, and glide-generated minima in the signal.

The conversion from temporal locations of landmarks, together with feature labels, to a lexically-consistent segment/feature representation requires that temporal information and the feature labels be used to collapse consonantal closures and releases into sequences of segments. For example, sequences of two consonants between vowels ($VC_1C_2V$) are often produced with just two acoustically-evident landmarks ($C_1$ closure and $C_2$ release), and duration information may be needed to determine whether the landmarks signal one consonant or two consonants.

Matching of the segment/feature representation for running speech to sequences of words from a stored lexicon requires knowledge of rules specifying modifications that can occur in the features of the lexical items depending on the context. As a first step in accounting for these modifications, some of the features that are known to be potentially influenced by context are marked in the lexicon as modifiable. The matching process also requires that criteria be defined for matching of features, since many features remain unspecified, both in the labeled representation and in the stored lexical representation. Several criteria for accepting a match are possible when a feature is unspecified or is marked as modifiable.

In this project, the performance of the matcher was evaluated using several different criteria for matching individual features. For example, one criterion was that a lexical item containing the feature [+nasal] was not accepted unless the labeled representation also contained [+nasal]; that is, there would be no match if the feature was unspecified for [nasal]. Experiments with different matching criteria led to a number of candidates for word sequences for most sentences and the correct sequence was almost always one of these candidates. The most effective matching criterion, which included a metric for the number of modifications that were needed to create the matches, led to word sequences that were in the top three candidates in 95 percent of the sentences.

This exercise is leading to the formulation of an improved model that separates the lexicon from the rules and proposes matching procedures that invoke the rules as an integral part of the lexical access process.

### 1.5.4 A Model for the Enhancement of Phonetic Distinctions

Models for the production and perception of speech generally assume that lexical items are stored in memory in terms of segments and features. In the case of speech production, the feature specifications can be viewed as instructions to an articulatory component which produces an acoustic output. The speaker's perceptual knowledge of the minimal distinctions in the language plays an important role in this production process. Coupled with the speaker's awareness of the perceptual/acoustic manifestations of the minimal distinctions of the language is knowledge that certain articulatory gestures over and above those specified by the phonological features can contribute to shaping the properties of the sound. Recruiting of these additional gestures can enhance the perception of these distinctions.

We have formulated speech production model with components that incorporate these enhancing processes. Four types of enhancement have been examined: voicing for obstruent consonants, nasalization for vowels, place distinctions for tongue blade consonants, and tongue body features for vowels. In each case, there is a gesture that is implemented in response to specifications from a phonological feature, and a separate articulatory gesture is recruited to enhance the perceptual contrast for that feature.

## 1.6 Laboratory Facilities for Speech Analysis and Experimentation

Facilities for performing direct-to-disk recordings, for transferring speech data from a digital audio tape (DAT) to a hard drive and for performing automated speech perception experiments were set up in a sound-attenuated room using a Macintosh computer with a Digidesign soundcard and Psyscope software.

In addition, development of an in-house speech analysis software program, xkl, is continuing. Xkl is an X11/Motif based program that performs real-time analysis of waveforms, plays and records sound files, creates spectrograms and other spectra, and synthesizes speech.

## 1.7 Publications

### 1.7.1 Journal Articles

Chen, M. "Acoustic Correlates of English and French Nasalized Vowels." *J. Acoust. Soc. Am.* 102: 2360-70 (1997).

Chen, M., and R. Metson. "Effects of Sinus Surgery on Speech." *Arch. Otolaryngol. Head Neck Surg.* 123: 845-52 (1997).

Guenther, F., C. Espy-Wilson, S. Boyce, M. Matthies, M. Zandipour, and J. Perkell. "Articulatory Tradeoffs Reduce Acoustic Variability during American English /r/ Production." Submitted to *J. Acoust. Soc. Am.*

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.

Lane, H., J. Perkell, M. Matthies, J. Wozniak, J. Manzella, P. Guiod, M. MacCollin, and J. Vick. "The Effect of Changes in Hearing Status on Speech Level and Speech Breathing: A Study with Cochlear Implant Users and NF-2 Patients." *J. Acoust. Soc. Am.* Forthcoming.

Lane, H., J. Wozniak, M. Matthies, M. Svirsky, J. Perkell, M. O'Connell, and J. Manzella. "Changes in Sound Pressure and Fundamental Frequency Contours Following Changes in Hearing Status." *J. Acoust. Soc. Am.* 101: 2244-52 (1997).

Matthies, M., P. Perrier, J. Perkell, and M. Zandipour. "Variation in Speech Movement Kinematics and Temporal Patterns of Coarticulation with Changes in Clarity and Rate." Submitted to *J. Speech, Lang. Hear. Res.*

Perkell, J., and M. Zandipour. "Clarity Versus Economy of Effort in Speech Production: Kinematic Performance Spaces for Cyclical and Speech Movements." Submitted to *J. Acoust. Soc. Am.*

Perkell, J., M. Zandipour, M. Matthies, and H. Lane. "Clarity Versus Economy of Effort in Speech Production: A Preliminary Study of Inter-Subject Differences and Modeling Issues." Submitted to *J. Acoust. Soc. Am.*

Perkell, J.S., M.L. Matthies, H. Lane, F. Guenther, R. Wilhelms-Tricarico, J. Wozniak, and P. Guiod. "Speech Motor Control: Acoustic Goals, Saturation Effects, Auditory Feedback and Internal Models." *Speech Commun.* 22: 227-50 (1997).

Svirsky, M.A., K.N. Stevens, M.L. Matthies, J. Manzella, J.S. Perkell, and R. Wilhelms-Tricarico. "Tongue Surface Displacement During Obstruent

Stop Consonants." *J. Acoust. Soc. Am.* 102: 562-71 (1997).

Wu, C., R. Wilhelms-Tricarico, and J.A. Negulesco. "Landmark Selection for Cross-Mapping Muscle Anatomy and Volumetric Images of the Human Tongue." Submitted to *Clin. Anat.*

### 1.7.2 Conference Papers

Dilley, L.C., and S. Shattuck-Hufnagel. "Ambiguity in Prominence Perception in Spoken Utterances in American English." *Proceedings of the Joint Meeting of the International Congress on Acoustics and the Acoustical Society of America*, Seattle, Washington, June 1998. Forthcoming.

Govindarajan, K. "Latency of MEG M100 Response Indexes First Formant Frequency." *Proceedings of the Joint Meeting of the International Congress on Acoustics and the Acoustical Society of America*, Seattle, Washington, June 1998. Forthcoming.

Perkell, J., M. Matthies, and M. Zandipour. "Motor Equivalence in the Production of /ʃ/." *Proceedings of the Joint Meeting of the International Congress on Acoustics and the Acoustical Society of America*, Seattle, Washington, June 1998. Forthcoming.

Poort, K. "Stop-Consonant Production by Dysarthric Speakers: Use of Models to Interpret Acoustic Data." *Proceedings of the Joint Meeting of the International Congress on Acoustics and the Acoustical Society of America*, Seattle, Washington, June 1998. Forthcoming.

Prahler, A.M. "Modeling and Synthesis of Lateral Consonant /l/." *Proceedings of the Joint Meeting of the International Congress on Acoustics and the Acoustical Society of America*, Seattle, Washington, June 1998. Forthcoming.

Shattuck-Hufnagel, S. and A. Turk. "The Domain of Phrase-Final Lengthening in English." *Proceedings of the Joint Meeting of the International Congress on Acoustics and the Acoustical Society of America*, Seattle, Washington, June 1998. Forthcoming.

Stevens, K.N. "Toward Models for Human Production and Perception of Speech." *Proceedings of the Joint Meeting of the International Congress on Acoustics and the Acoustical Society of America*, Seattle, Washington, June 1998. Forthcoming.

Turk, A.E., and S. Shattuck-Hufnagel. "Duration as a Cue to Syllable Affiliation." *Proceedings of the Conference on the Phonological Word*, Berlin, Germany, October 1997. Forthcoming.

Wilhelms-Tricarico, R., and C.-M. Wu. "A Biomechanical Model of the Tongue." *Proceedings of the 1997 Bioengineering Conference*, BED-Vol. 35. Eds. K.B. Chandran, R. Vanderby, and M.S. Hefzy. New York: AMSE, (1997), pp. 69-70.

### 1.7.3 Chapter in a Book

Shattuck-Hufnagel, S. "Phrase-level Phonology in Speech Production Planning: Evidence for the Role of Prosodic Structure." In *Prosody: Theory and Experiment: Studies Presented to Gosta Bruce.* Ed. M. Horne. Stockholm, Sweden: Kluwer. Forthcoming.

### 1.7.4 Thesis

Hagen, A. *Linguistic Functions of Glottalizations and their Language Specific use in English and German.* Diplomarbeit (M.) Comput. Sci., University of Erlangen-Nürnberg, Germany.