Speech Communication

Sponsors
C.J. LeBel Fellowship
Dennis Klatt Memorial Fund
Donald North Memorial Fund
National Institutes of Health (Grants R01-DC00075,
R01-DC01925, R01-DC02978, R01-DC03007, T32-DC00038.
National Science Foundation (BCS - 0418205)

Academic and Research Staff
Professor Kenneth N. Stevens, Professor Morris Halle, Professor Samuel J. Keyser, Dr. Joseph S. Perkell, Dr. Stefanie Shattuck-Hufnagel, Dr. Helen Hanson, Dr. Janet Slifka, Dr. Margaret Denny, Dr. Satrajit Ghosh, Dr. Majid Zandipour, Mark Tiede, Dr. Miwako Hisagi, Seth Hall.

Visiting Scientists and Research Affiliates
Dr. Corine A. Bickley, Department of Hearing, Speech and Language Sciences, Gallaudet University, Washington, District of Columbia.
Dr. Suzanne E. Boyce, Department of Communication Disorders, University of Cincinnati, Cincinnati, Ohio.
Dr. Krishna Govindarajan, Nuance Communications Inc., Burlington, Massachusetts.
Dr. David Gow, Department of Psychology, Salem State College, Salem, Massachusetts, and Department of Neuropsychology, Massachusetts General Hospital, Boston, Massachusetts.
Dr. Frank Guenther, Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts.
Dr. Andrew Howitt, Otolith, visit site at: http://www.otolith.com.
Dr. Robert E. Hillman, Department of Voice Surgery and Rehabilitation, Massachusetts General Hospital, Boston, Massachusetts.
Dr. Harlan Lane, Department of Psychology, Northeastern University, Boston, Massachusetts.
Dr. Steven Lulich, Harvard School of Public Health, Boston, Massachusetts.
Dr. Melanie Matthies, Department of Communication Disorders, Boston University, Boston, Massachusetts.
Dr. Richard McGowan, CReSS LLC, Lexington, Massachusetts.
Dr. Lucie Menard, Department of Linguistics and Language Education, University of Quebec, Montreal, Canada.
Dr. Anthony Okobi, Brown Medical School, Providence, Rhode Island.
Dr. Rupal Patel, Department of Speech Language Pathology and Audiology, Northeastern University, Boston, Massachusetts.
Dr. Alice Turk, Department of Linguistics, University of Edinburgh, Edinburgh,
United Kingdom.
Dr. Nanette Veilleux, Department of Computer Science, Simmons College, Boston, Massachusetts.
Dr. Lorin Wilde, STAR, Massachusetts.

Graduate Students
Marc Boucek, Lan Chen, Nancy Chen, Xuemin Chi, Elisabeth Hon, Youngsook Jung, Chi-youn Park, Yoko Saikachi, Sherry Zhao

Undergraduate Students
Alicia Patterson, Rebecca Pomerantz (Harvard University)

Technical and Support Staff
Arlene E. Wint

1. Constraints and Strategies in Speech Production

Introduction

The objective of this research is to refine and test a theoretical framework in which words in the lexicon are represented as sequences of segments and syllables and these units are represented as complexes of auditory/acoustic and somatosensory goals. The motor programming to produce sequences of sensory goals utilizes an internal neural model of relations between articulatory motor commands and their acoustic and somatosensory consequences. The relations between articulatory motor commands and the movements they generate are influenced by biomechanical constraints, which include characteristics of individual speakers' anatomies and more general dynamical properties of the production mechanism. To produce an intelligible sound sequence while accounting for biomechanical constraints, speech movements are planned so that sufficient perceptual contrast is achieved with minimal effort. There are individual differences in planning movements toward sensory goals that may be due to relations between production and perception mechanisms in individual speakers.

In a current project, funded by the NIDCD, the internal model is implemented as a neurocomputational model that is used to control a vocal-tract model (an articulatory synthesizer). The combined models provide the bases of hypotheses about the planning of speech movements. To test these hypotheses, we are conducting experiments with speakers and listeners in which we measure articulatory movements, speech acoustics, perception, and brain activation. We are manipulating speaking condition, phonemic context and speech sound category and we introduce transient and sustained perturbations. We are also performing modeling and simulation experiments, in which we adapt the vocal-tract model to the morphologies of individual speakers. We are testing properties of the neurocomputational model by using it to control the individualized vocal tract models in efforts to replicate those speakers' production data.
During this last year, we have made further progress on several major studies.

1.1 Variation in vowel production

We completed analyses of recordings of articulatory movements and acoustics of vowels pronounced in different contexts and speaking conditions by 10 male and 10 female subjects (reported in preliminary form last year), with the following results. 1) Measures of contrast decreased with speaking condition, from clear to normal to fast. 2) Because of the presumed influence of economy of effort, when changing from clear to fast, point vowel distributions moved toward the center and non-point vowel distributions expanded toward the periphery, making clear and fast distributions overlap more for non-point than for point vowels. Work in progress includes dimension reduction of the articulatory movement data so they can be compared more directly with acoustic data, and trajectory analysis measurements for use in a quantitative assessment of a hypothesized trade-off between clarity and economy of articulatory effort.

1.2 Relation between perceptual acuity and production contrasts

In previously reported work, we found preliminary support of a relation between speaker auditory acuity and produced sound contrast. In an expansion of this research, we completed the development of a Matlab facility for generating speech sound continua for perceptual testing using the Klatt synthesizer. Gender-neutral continua were generated for two vowel pairs (*pap* to *pep* and *pep* to *pip*), the sibilant contrast (*said* to *shed*), voicing onset time continuum (*doe* to *toe*) and vowel duration (using the utterance *pup*). Additional Matlab software was developed for carrying out the required perceptual testing. We have determined the perceptual acuity of 18 subjects who participated in a corresponding speech production experiment (1.1), and we are now analyzing the relationship between the perceptual test results and the corresponding production data.

1.3 Development and application of an algorithm for perturbation of vowel-to-vowel formant transitions

We have significantly improved our existing algorithm for real-time perturbation of vowel formant frequencies for use in sensorimotor adaptation experiments. A previously developed algorithm has been extended to reliably track and shift multiple formants in steady state as well as changing formant trajectories. This processing can now be performed on a personal computer with a sound card instead of separate digital signal processing hardware. We have also implemented a Matlab interface to this algorithm so that we can run our experiments from within the Matlab environment. This offers additional flexibility and significantly reduces the time needed for paradigm development. In a preliminary study, subjects were asked to pronounce utterances containing diphthongs with formant movement from /Θ/ to /i/, such as *bike* and *tight*. The transduced speech signal was processed by this algorithm and fed back to the subject with a delay of 9ms. The signal processing uses LPC analysis and resynthesis to detect and shift the F1 and F2 frequencies during the transition; the resynthesis uses the LPC residual as the source, so the voiced sounds fed back to the subject sound reasonably natural. The introduced formant shift was in a direction normal to a straight-line F1, F2 trajectory connecting the starting and ending points of the transition; the shift was maximal at the transition mid-point and zero at the starting and ending points. We have now run seven female and four male subjects using this perturbation in an adaptation paradigm. We are presently analyzing the data.

1.4 Development of an algorithm for sibilant perturbation

We have also developed an algorithm for shifting sibilants (e.g. "/□/" in *shed*) in nearly real-time. Similar to the vowel perturbation algorithm, this algorithm runs on a computer and no longer requires external digital signal processing hardware. Subjects pronounce words like *shed* into a microphone and the transduced signal is processed by the algorithm. The algorithm shifts the spectral mean during the sibilant portion of the utterance so that *shed* is transformed to sound like *said*. The output is fed back to the subject over headphones with a 10ms delay. We have run 8 pilot subjects in an adaptation paradigm to determine if the subjects will compensate for the perturbation, presumably by changing their articulations. Thus far, some subjects have compensated in response to the shift and others have not. We are in the process of refining the algorithm.

1.5 Trajectory planning in the concatenation of larger units

As reported last year, we are analyzing data bearing on the interplay of sufficient auditory contrast and economy of production effort, in the context of overlapping articulatory gestures for adjacent consonants. /kt/ clusters formed across word boundaries (*pack top*) compared with those originating in the lexicon (*pact op*) show distinct patterns of behavior: Compared to lexical sequences, those formed across word boundaries show in general greater separation in timing between points of maximum constriction at normal speech rates, and greater reduction in these separations under fast production rates. Lexical sequences tend to preserve the rate-scaled relative phasing between the two constrictions. Regardless of utterance type, distinct release bursts are not produced consistently, especially as speaking rate increases. Currently, data have been acquired and analyzed for 15 subjects. In general, tautosyllabic /kt/ clusters (*pact op*) are significantly less variable in timing than heterosyllabic (*pack top*) sequences, relative to the labial gestures of the carrier context.

1.6 Control of tongue movements in acoustic and articulatory spaces

A previously described project was completed, in which we implemented a biomechanical/articulatory vocal-tract model and neurocomputational control models, and acquired acoustic, kinematic and EMG data from two speakers to investigate the planning of vowel-to-vowel tongue movements. Two different control regimes were investigated: planning in acoustic space using feedback mechanisms, and motor planning through feedforward control of muscle activation. Simulation results showed that: a) modulations of the motor commands accounted for the effects of speaking rate on EMG, kinematic, and acoustic outputs; b) the movement and acoustic trajectories were influenced by vocal tract biomechanics; and c) both planning schemes produced similar articulatory movement, EMG, muscle length, force, and

acoustic trajectories, which were also comparable to the subject's data under normal speaking conditions. In addition, the effects of a bite-block and auditory masking on measured EMG signals, articulatory movements and formants were measured in the subjects and simulated by the models. Acoustic planning produced successful simulations but motor planning did not. These results lead to the inference that with somatosensory feedback available but in the absence of auditory feedback, feedforward commands are relied upon initially and are modified with practice by corrections based on the difference between expected and produced somatosensory feedback. Overall these simulation and experimental results provide support for the use of an acoustic space rather than a motor space for articulatory movement planning.

2. Modeling of Human Lexical Access

2.1 Modeling of landmark detection

Acoustically- and lexically-significant time-points (i.e. landmarks) in the speech signal can provide useful information in a knowledge-based speech recognition system. Due to physiological constraints on landmarks, only 44% of the possible landmark pairs are "legal". In this project, we propose to use bigrams to quantitatively model the grammar of consonant landmarks, which can be used to predict landmark sequences. Our work consists of two steps --- (1) landmark candidate detection and (2) landmark sequence determination. The landmark candidate detection algorithm candidate detection algorithm expands Liu's work to a probabilistic framework (ref. Liu's paper). In the next step, the bigram was applied to determine the most likely landmark sequences. Results tested on the TIMIT database are consistent across gender and dialects of American English. Even though lower thresholds than Liu's were used to allow more insertions in the initial landmark candidate detection step, we still managed to halve the insertion errors by applying the bigram model in the following step. Many errors are systematic landmark variants such as flaps or syllabic nasals; they can be used to improve landmark detection methods and help us further understand acoustic variations in the speech signal.

2.2 Approximants

This study addresses the classification of approximant consonants (liquids and glides), as part of the Lexical Access project in the Speech Communication Group at MIT. In a departure from conventional analyses of formant frequency positions, this study considers the empty space left between the formant peaks during their extreme movements at approximant landmarks. The center frequencies of these spectral valleys (Fvalley) were analyzed as potential acoustic classifiers for liquid (/l/ and /r/) and glide (/w/ and /y/) consonants in syllable-initial positions in American English. A male and a female speaker produced recordings of nonsense words containing the four approximant consonants in eleven following vowel contexts for measurement. Statistical analysis showed no vowel-consonant interactions or main effects of vowel context on the acoustic measure Fvalley. Highly significant main effects of consonant identity were found, with Fvalley [/r/] > Fvalley [/l/] > Fvalley [/w/]> Fvalley [/y/]. All Tukey pairwise comparisons for consonant identity were also highly significant. It is concluded that the spectral valley center frequency measure reveals sharp distinctions between the four American English approximant consonants for the two speakers studied, and is a good candidate for an acoustic classifier of liquids and glides. Future work should expand the study to include more speakers and prosodic contexts so that the results can be generalized in the model. Theoretical studies will be carried out to determine the reasons for these effects for the different approximants.

2.3 The Consonant /ð/

Phonetic variation is pervasive in everyday speech. Studying these variations is essential for building acoustic models and lexical representations that effectively capture the variability of speech. This thesis examines one of the commonly-occurring phonetic variations in English: the stop-like modification of the dental fricative /ð/. This variant exhibits a drastic change from the canonical /ð/; the manner of production is changed from one that is fricative to one that is stop-

like. Furthermore, the place of articulation of stop-like /ð/ has been a point of uncertainty, leading to the confusion between stop-like /ð/ and /d/. The aims of this project are to uncover the segmental context of stop-like /ð/, possible causes of the modification, whether the dental place of articulation is preserved despite modification, and if there are salient acoustic cues that distinguish between stop-like /ð/ and /d/.

Word-initial /ð/ in the read speech of the TIMIT Database, the task-oriented spontaneous speech of the AEMT Corpus, and the non-task-oriented spontaneous speech of the Buckeye Corpus are examined acoustically. It is found that stop-like /ð/ occurs most often when it is preceded by silence or when preceded by a stop consonant. The occurrence is less frequent when /ð/ is preceded by a fricative or an affricate. This modification rarely occurs when /ð/ is preceded by a vowel or liquid consonant. These findings suggest that possible factors that may contribute to the stop-like modification of /ð/ include physiological mechanisms of speech production, prosody, and/or other aspects of speaking style and manner.

Acoustic analysis indicates that stop-like /ð/ is significantly different from /d/ in burst amplitude, burst spectrum shape, burst peak frequency, and second formant at following-vowel onset. Moreover, the acoustic differences indicate that the dental place of articulation is preserved for stop-like /ð/. Automatic classification experiments involving these acoustic measures suggest that they are robust in distinguishing stop-like /ð/ from /d/. Applications of these findings may lie in areas of automatic speech recognition, speech transcription, and development of acoustic measures for speech disorder diagnosis.

2.4 Acoustic Evidence for Word Boundaries

An important component in the human speech recognition task is the segmentation problem of parsing a continuous stream of speech into individual words for comprehension. Word onsets are hypothesized to be stronger and less prone to modification to help the listeners in the segmentation task. CVC syllables, with the same stop consonants at both word-initial and -final positions, were analyzed for comparison. The maximum in the stop burst transient, closure duration, and VOT (for voiceless stops) show trends that word-initial stops are articulatorily stronger than word-final stops. The same set of measurements was taken on real words from the LAFF (lexical access from features) database for position-dependent differences. On average, the transient peak magnitude and VOT values (for 450 occurrences of /k/ for each of the 4 speakers) for word-initial /k/'s were found to be greater than word-medial /k/'s, indicating that the burst is "stronger" at initial positions. Closure durations for 150 /k/'s across the 4 speakers were found, on average, to be longer for word-initial than medial positions. There are large regions of overlap in these measurements, which were made with consonants occurring in randomly selected sentences and, as a consequence there are large regions of overlap in the data, presumably due to lack of control of prosodic context. A new database involving real words, with controlled lexical access within a controlled intonation contour, is being recorded to investigate the word-position effects on English stops and affricates. More spectral measurements are also being investigated as word-segmentation cues.

3. Quantal Theory, Enhancement, and Speech Production Models

3.1 Physical Mechanisms Underlying Quantal Relations

When a particular articulatory dimension in speech production is manipulated through a range of values, there is often a nonlinear relation between the dimension and its acoustic consequence. The acoustic parameter is relatively insensitive to the change in the articulatory parameter over one portion of its range and shows a rapid change with articulation over another part of its range. It has been proposed that regions of insensitivity of acoustic attributes to changes in articulation could provide a quantitative basis for defining distinctive features.

During the past year or two we have been examining articulatory/acoustic relations for a number of distinctive features, and we have attempted to develop a way of characterizing the different

types of quantal relations. Based on our present data, there appear to be two sources of quantal properties. One of these arises from aerodynamic forces and mechanical properties of vocal-tract surfaces. An example is the feature [stiff vocal folds], which specifies a contrast *pat-bat*. A second source of quantal relations is related to the acoustic filtering by vocal-tract resonators. An example is the place contrast of *pot-dot*. Almost all quantal relations can be put into one of these two categories, which are sometimes called articulator-free features and articulator-bound features.

3.2 Role of F1sub in Defining the Feature [low]

The goal of this study is to determine whether acoustic coupling between the first subglottal resonance F1sub and the F1 frequency for vowels creates a region near 600 Hz in which the F1 prominence shows an irregularity. Such a finding would provide evidence for a defining quantal articulatory-acoustic relation for the distinctive feature [low].

The time course of F1 in relation to F1sub was examined for certain diphthongs and for several monophthongs produced by a number of speakers using Chi's data [X. Chi and M. Sonderegger, J. Acoust. Soc. Am., (In press)]. For the diphthongs, a discontinuity in F1 or a dip in amplitude of the F1 prominence was observed as it passed through F1, while for the monophthongs F1 was above F1sub for [+low] vowels and below F1sub for [-low] vowels. In a preliminary further study, data from the literature on F1 for vowels from various languages showed that the boundary between F1 values of [+low] vowels and those of [-low] vowels agrees with the average value of F1sub obtained in the laboratory study with English.

3.3 Vowel Nasalization in American English

This study quantifies acoustic variation of vowel nasalization arising from phonetic context in American English, with an emphasis on carryover contexts. While qualitative articulatory trajectories and phonetic descriptions suggest that a vowel is nasalized in carryover contexts, few acoustic studies have examined this issue. Our acoustic analyses have examined nasalization in vowels through measurement of the amplitude of a nasal resonance (at about 1 kHz) relative to the amplitude of the F1 prominence. Data for the vowel /i/ in NV, N, and NVN contexts show that: (1) a vowel can be nasalized with at least one adjacent nasal consonant, even if the nasal consonant is pre-vocalic; (2) vowels with nasal consonants on both sides (NVN) do not guarantee more vowel nasalization.

4. Studies of Speech Development and Speech Disorders

4.1 Improving Naturalness of Electrolarynx Speech

Lack of F0 variation in electrolarynx (EL) speech has been identified as a major contributor to its unnatural quality. This study investigated whether the naturalness of EL speech could be significantly improved by co-varying its fundamental frequency (F0) in synchrony with naturally occurring variations in the RMS amplitude of the EL speech signal. Experiments using synthesized EL speech with declarative sentences were carried out with constant F0 contours, with F0 contours derived from measurements of amplitude modulation, and with F0 contours extracted from the pre-laryngectomy speech of two speakers. A group of listeners evaluated the naturalness of the three types of stimuli. Amplitude-based F0 modulation resulted in EL speech that was judged to be more natural sounding than EL speech having constant F0, but less natural sounding than EL utterances with F0 contours extracted from pre-laryngectomy speech.

5. Effects of Hearing Status on Adult Speech Production

Introduction

An application to NIDCD for five years of continuation support was approved and funding has been awarded. The goals of this research are: to deepen significantly our understanding of how a speaker's auditory acuity influences his or her speech motor planning; to describe the speech perception and production of hearing-impaired adults and the effects of cochlear prostheses; and to evaluate and help refine a quantitative model of the role of hearing in speech. We are conducting experiments with normal-hearing speakers, and with postlingually deafened adults who receive cochlear implants. The experiments measure the effects of the implants on speech in recordings made before implantation and up to two years after, as speakers' auditory acuity evolves. According to our model of the role of hearing in adult speech motor control, many of the goals of speech movements are in the auditory domain. Consequently, a central theme of this research is the role of auditory perception in the feedback and feedforward control systems that are used to achieve auditory goals during speech. Feedforward control is almost entirely responsible for generating articulatory movements in adults. However, when there is a mismatch between the speaker's intention and the resulting auditory feedback during production of a speech sound, that error leads to corrective motor commands that serve to update feedforward commands for subsequent movements. The speaker's ability to detect such a mismatch depends on his or her auditory acuity.

The proposed research examines the role of auditory acuity when producing phoneme and lexical stress contrasts; when compensating for feedback perturbation of vowel formants; and when imitating synthesized vowels. To measure acuity we present synthetic speech continua for discrimination testing. To assess relations between production and acuity, we measure the degree of separation of produced contrastive phonemes; dispersion of productions around their phoneme means; compensation for vowel formant shift; and imitation accuracy for vowels. In most of these experiments we also block auditory feedback temporarily in order to reveal the state of feedforward commands. Analyses take demographic variables such as age at hearing loss and duration of implant use into account.
Since the start of funding in December, 2006, we have hired and trained a new research specialist and have developed new algorithms and paradigms and conducted pilot studies for the proposed research.

5.1 Development of facilities

A version of the Klatt synthesizer has been developed that can be called directly from Matlab in order to support the online adjustment of synthesis parameters as a function of subject responses during perceptual experiments. We have also upgraded our facilities for psychophysical testing in one of our sound-attenuating rooms.

5.2 Mapping the perceptual space

A new interactive testing procedure has been developed for eliciting judgments of vowel quality, in order to map subjects' vowel perceptual spaces. Using this procedure, subjects are instructed to interactively find the best match to a displayed prototype by 'exploring' a two dimensional vowel space in which one axis maps to F1 and the other to F2 (with both the direction and choice of axes are randomly permuted). Clicking at some offset within the displayed vowel space causes online synthesis of a vowel token having those formant values (with additional parameters specified by rules). The procedure terminates when the Euclidean distance of all subject formant choices from the median value falls below a predetermined threshold. Output measures include perceptual vowel category loci and sizes (quantified as dispersion around the means and curvature of best-fitting paraboloids to the responses).

6. Speech Prosody

6.1 Effect of Tone Distribution on Subglottal Pressure

The current work is part of a project to characterize the subglottal pressure (Ps) contour in terms of the distribution of pitch accents and of phrase and boundary tones. Declination of the working phase, and the transition from the working phase to the termination phase are studied. It is found that the nuclear pitch accent does not define the start of the termination phase; the utterance offset is a better marker. Declination rate of the working phase and its relation to the phrase and boundary tones at utterance offset are found to vary among speakers. These differences could result in variations in SPL and F0 that contribute to a speaker's individuality. The results have implications for models of speech production, and for applications such as computer speech synthesis and recognition.

6.2 Robustness of Acoustic Landmarks in Spontaneous Speech

Acoustic landmarks (abrupt changes associated with consonant closures and releases, vowels and glides) play an important role in some models of lexical access (e.g. Stevens 1998, 2002), so it is important to determine how often they actually survive the rigors of articulatory overlap and weakening in spontaneous speech production. A corpus of spontaneous American English speech was collected from 8 adult female speakers and hand labeled for the occurrence of landmarks. Preliminary results for one conversation (240 secs., 610 words, analysis completed for 1003 of 2750 predicted landmarks) show that 86% of landmarks were realized overall, with a sharply lower rate for coronal stops /t/ and /d/. These results suggest that the majority of landmarks are available for detection both by human listeners and automatic recognition algorithms. Ongoing analyses are comparing the rate of automatic detection of these acoustic events with the hand labels, and tabulating the relatively limited set of contexts in which predicted landmarks are lost or changed.

6.3 Irregular Vocal-Fold Vibration as a Cue to Speaker Identity

This study tests the hypothesis that, among the other paralinguistic information conveyed by voice qualities, irregular vocal fold vibration (glottalization) can serve as a cue to speaker identity. Earlier reports of systematic differences across speakers in the rate and type of intermittent glottalization support this idea. Still, it remains an open question whether human listeners use this speaker-specific information when recognizing familiar voices. Results of a perceptual experiment, in which listeners first learned the voices of the speakers they were later tested on, suggest that irregular pitch periods in utterance-final regions can be a cue in the recognition of individual speaker voices.

6.4 Acoustic Correlates of Lexical Prosody

During the past year, we completed a series of experiments to determine the acoustic parameters that differentiate between primary stress and non-primary full vowels using two-syllable real words and novel words with identical syllable compositions. The location of the high focal pitch accent within a declarative carrier phrase was varied using an innovative object naming task that allowed for a natural and spontaneous manipulation of phrase-level accentuation. Results from male native speakers of American English show that when the high focal pitch accent was on the target word, vowel differences in pitch, intensity prominence, and amplitude of the first harmonic (H1*) (corrected for the effect of the vocal tract filter), accurately distinguished full vowel syllables carrying primary stress vs. non-primary stress. Acoustic parameters that correlated to word stress under all conditions tested were syllable duration, H1*-A3*, as a measurement of spectral tilt, and noise at high frequencies, determined by band-pass filtering the F3 region of the spectrum. Furthermore, the results indicate that word stress cues are augmented when the high focal pitch accent is on the target word. This became apparent after a formula was devised to correct for the masking effect of phrase-level accentuation on the spectral tilt measurement, H1*-A3*. Perceptual experiments also show that male native speakers of American English utilized differences in syllable duration and spectral tilt, as controlled by the KLSYN88 parameters DU and TL, to assign prominence status to the syllables of a novel word embedded in a carrier phrase. Results from this study suggest that some correlates to word stress are produced in the

laryngeal region and due to vocal fold configuration. The model of word stress that emerges from this study has aspects that differ from other widely accepted models of prosody at the word level. The model can also be applied to improve the prosody of synthesized speech, as well as to improve machine recognition of speech.

Publications

Journal Articles, Published

H. Lane, M. Denny, F.H. Guenther, H. Hanson, N. Marrone, M.L. Matthies, J.S. Perkell, E. Stockmann, M. Tiede, J. Vick, and M. Zandipour, "On the Structure of Phoneme Categories in Listeners with Cochlear Implants," *J. Speech, Language, and Hearing Research 50*: 2-14 (2007).

H. Lane, M. Matthies, M. Denny, F. Guenther, J. Perkell, E. Stockmann, M. Tiede, J. Vick, and M. Zandipour, "Effects of Short- and Long-Term Changes in Auditory Feedback on Vowel and Sibilant Contrasts," *J. Speech, Language and Hearing Research 50*: 913-927 (2007).

L. Menard, M. Polak, M. Denny, E. Burton, H. Lane, M.L. Matthies, N. Marrone, J.S. Perkell, M. Tiede, and J. Vick, "Interactions of Speaking Condition and Auditory Feedback on Vowel Production in Postlingually Deaf Adults with Cochlear Implants," *J. Acoustical Society of America, 121*: 3790-3801 (2007).

J.S. Perkell, M. Denny, H. Lane, F.H. Guenther, M.L. Matthies, M. Tiede, J. Vick, M. Zandipour, and E. Burton, "Effects of Masking Noise on Vowel and Sibilant Contrasts in Normal-Hearing Speakers and Postlingually Deafened Cochlear Implant Users," *J. Acoustical Society of America, 121*: 505-518 (2007).

J.S. Perkell, H. Lane, M. Denny, M.L. Matthies, M. Tiede, M. Zandipour, J. Vick, and E. Burton, "Time Course of Speech Changes in Response to Unanticipated Short-Term Changes in Hearing State," *J. Acoustical Society of America, 121*: 2296-2311 (2007).

M. Kenstowicz and C. Park, "Laryngeal Features and Tone in Kyungsang Korean: A Phonetic Study." *The Phonology-Morphology Circle of Korea 12 (2)*: 247-264 (2006).

Journal Articles, Accepted for Publication

V. Villacorta, J. Perkell, and F. Guenther, "Sensorimotor Adaptation to Feedback Perturbations of Vowel Acoustics and its Relation to Perception," *J. Acoustical Society of America*, forthcoming.

K. Stevens and S. Keyser, "Quantal Theory, Enhancement and Overlap," *J. Phonetics*, forthcoming.

X. Chi and M. Sonderegger, "Subglottal Coupling and Its Influence on Vowel Formants," *J. Acoustical Society of America*, forthcoming.

S. Lulich, A. Bachrach, and N. Malyska, "A Role for the Second Subglottal Resonance in Lexical Access," *J. Acoustical Society of America*, forthcoming.

Journal Articles, Submitted for Publication

S. Ghosh, F. Guenther, and J. Tourville, "A Neuroimaging Study of the Production of Elementary Speech Units," submitted to *J. Speech, Language and Hearing Research*.

C.Y. Lee and K.N. Stevens, "Strident Fricatives in Mandarin Chinese: From Acoustics to Articulation and Features," submitted to *Phonetica.*

Book/Chapters in Books

S. Shattuck-Hufnagel, Y. Yasinnik, N. Veilleux, and M. Renwick, "A Method for Studying the Time Alignment of Gestures and Prosody in American English: 'Hits' and Pitch Accents in Academic-Lecture-Style Speech." In *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, eds. A. Esposito, M. Bratanic, E. Keller, and M. Marinaro, (Amsterdam: IOS Press, pp. 34-44, 2007).

A. Esposito, D. Esposito, M. Refice, M. Savino, and S. Shattuck-Hufnagel, "A Preliminary Investigation of the Relationship Between Gestures and Prosody in Italian." In *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, eds. A. Esposito, M. Bratanic, E. Keller, and M. Marinaro, (Amsterdam: IOS Press, pp. 65-74, 2007).

Meeting Papers, Presented

T. Bohm, and S. Shattuck-Hufnagel, "Listeners Recognize Speakers' Habitual Utterance-Final Voice Quality," Presented at the Paralinguistic Speech – Between Models and Data Satellite Meeting, ICPhS 07, Saarbrucken, Germany, August 3, 2007.

J.S. Perkell, M. Zandipour, S. Ghosh, L. Menard, H. Lane, M. Tiede, and F.H. Guenther, "Variation in Vowel Production," Presented at the Fourth Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Honolulu, HI, Nov. 38-Dec. 2, 2006.

M. Zandipour, J.S. Perkell, M. Tiede, F.H. Guenther, K. Honda, and E.Z. Murano, "Different Motor Strategies for Increasing Speaking Rate: Data and Modeling", Presented at the Fourth Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Honolulu, HI, Nov. 38-Dec. 2, 2006.

J.S. Perkell, "Sensorimotor Control of Speech Production: Models and Data", Keynote talk, 11th Australasian International Conference on Speech Science and Technology, Auckland, New Zealand, December 6-8, 2006.

Meeting Papers, Published

N. Chen, J. Slifka, and K. Stevens, "Vowel Nasalization in American English: Acoustic Variability Due to Phonetic Context," *Proceedings 16th International Congress of Phonetic Sciences ICPhS-07,* Saarbrucken, Germany, 905-908, 2007.

H. Hanson, J. Slifka, S. Shattuck-Hufnagel, and J. Kobler, "Tone Distribution and Its Effect on Subglottal Pressure during Speech," *Proceedings 16th International Congress of Phonetic Science ICPhS-07,* Saarbrucken, Germany, 545-548, 2007.

J. Perkell, "Sensory Goals and Control Mechanisms for Phonemic Articulations," *Proceedings 16th International Congress of Phonetic Sciences ICPhS-07,* Saarbrucken, Germany, 169-174, 2007.

S. Shattuck-Hufnagel and N. Veilleux, "Robustness of Acoustic Landmarks in Spontaneously-Spoken American English" *Proceedings 16th International Congress of Phonetic Sciences ICPhS-07,* Saarbrucken, Germany, 925-928, 2007.

J. Slifka, "Irregular Phonation and Its Preferred Role as Cue to Silence in Phonological Systems," *Proceedings 16th International Congress of Phonetic Sciences ICPhS-07,* Saarbrucken, Germany, 229-232, 2007.

M. Tiede, S. Shattuck-Hufnagel, B. Johnson, S. Ghosh, M. Matthies, M. Zandipour and J. Perkell, "Gestural Phrasing in /KT/ Sequences Contrasting Within- and Cross Word Contexts," *Proceedings 16th International Congress of Phonetic Sciences ICPhS-07,* Saarbrucken, Germany, 521-524, 2007.

N. Veilleux, A. Brugos, S. Shattuck-Hufnagel, A. Patterson, "Distribution of Disfluencies and Errors in English Discourse," *Proceedings 16th International Congress of Phonetic Sciences ICPhS-07,* Saarbrucken, Germany, 1349-1352, 2007.

T. Bohm, S. Shattuck-Hufnagel, and G. Nemeth, "A Simple Method to Convert Modal into Irregular Voice, *Interspeech 07*, Antwerp, 2007.

J.S. Perkell, M. Zandipour, S. Ghosh, L. Menard, H. Lane, M. Tiede, and F.H. Guenther, "Variation in Vowel Production," J. Acoust. Soc. Am. 120, 3293 (A); Fourth Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Honolulu, HI, Nov. 38-Dec. 2, 2006.

M. Zandipour, J.S. Perkell, M. Tiede, F.H. Guenther, K. Honda, and E.Z. Murano, "Different Motor Strategies for Increasing Speaking Rate: Data and Modeling" J. Acoust. Soc. Am. 120, 3293 (A), Fourth Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Honolulu, HI, Nov. 38-Dec. 2, 2006.

J.S. Perkell, "Sensorimotor Control of Speech Production: Models and Data," *Proceedings of the International Speech Communication Association Workshop on Experimental Linguistics*, Athens, Greece, Aug. 28-30, 2006.

J.S. Perkell, "Achieving Speech Motor Goals: Feedback and Feedforward Control," *Proceedings of the Seventh International Seminar on Speech Production*, Ubatuba, Brazil, Dec. 13-15, 2006.

M. Zandipour, J.S. Perkell, F.H. Guenther, M. Tiede, K. Honda, and E. Murano, "Speaking With a Bite Block: Data and Modeling," *Proceedings of the Seventh International Seminar on Speech Production*, Ubatuba, Brazil, Dec. 13-15, 2006.

Theses

S. Lulich, *The Role of Lower Airway Resonances in Defining Vowel Feature Contrasts*, PhD thesis, Graduate Program in Speech and Hearing Biosciences and Technology, Harvard-MIT Division of Health Sciences and Technology, MIT, 2006.

A. Okobi, *Acoustic Correlates of Word Stress in American English*, PhD thesis, Graduate Program in Speech and Hearing Biosciences and Technology, Harvard-MIT Division of Health Sciences and Technology, MIT, 2006.

J. Yoo. *FMRI Studies of Effects of Hearing Status on Audio-Visual Speech Perception,* PhD thesis, Graduate Program in Speech and Hearing Biosciences and Technology, Harvard-MIT Division of Health Sciences and Technology, MIT, 2006.