Part V  Language, Speech and Hearing

Section 1  Speech Communication
Section 2  Sensory Communication
Section 3  Auditory Physiology
Section 4  Linguistics
Section 1  Speech Communication

Chapter 1  Speech Communication
Chapter 1. Speech Communication

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

Visiting Scientists and Research Affiliates
Dr. Vladimir M. Barsukov, Dr. Corine A. Bickley, Dr. Suzanne E. Boyce, Dr. Carol Y. Espy-Wilson, Dr. Richard S. Goldhor, Dr. Robert E. Hillman, Dr. Eva B. Holmberg, Dr. Caroline Huang, Dr. Harlan Lane, Dr. John Locke, Dr. John I. Makhoul, Dr. Sharon Y. Manuel, Dr. Takahide Matsuoka, Dr. Jorgen Pind, Dr. David Williams, Anna Esposito, Giulia Arman Nassi, Roeland P. Schaefter, Jane Wozniak

Graduate Students
Hwa-Ping Chang, Marilyn Y. Chen, Helen M. Hanson, David M. Horowitz, Mark A. Johnson, Hong-Kwang Kuo, Sharlene A. Liu, Noel S. Massey, Kelly L. Poort, Lorin F. Wilde

Undergraduate Students
Howard Cheng, Leanne Clarke, Laura C. Dilley, Helena Geng, Brett A. Geoffroy, John Hardy, Marnie Harker, Peggy Li, Henry Y. Lu, Rebecca Mallin, Rachel Molinaar

Technical and Support Staff
Michelle Halverson, D. Keith North, Arlene E. Wint

---

1 Audiofile, Inc., Lexington, Massachusetts, and Boston University, Boston, Massachusetts.
2 Boston University, Boston, Massachusetts.
3 Massachusetts Eye and Ear Infirmary, Boston, Massachusetts.
4 MIT and Massachusetts Eye and Ear Infirmary, Boston, Massachusetts.
5 Dragon Systems, Inc., Newton, Massachusetts.
6 Department of Psychology, Northeastern University, Boston, Massachusetts.
7 Massachusetts General Hospital, Boston, Massachusetts.
8 Bolt, Beranek and Newman, Cambridge, Massachusetts.
9 Wayne State University, Detroit, Michigan.
10 Utsunomiya University, Utsunomiya City, Japan.
11 University of Iceland, Reykjavik, Iceland.
12 Sensimetrics, Inc., Cambridge, Massachusetts.
13 Department of Physics, Salerno University, Salerno, Italy.
14 University of Utrecht, The Netherlands.
1.1 Introduction

The overall objective of our research in speech communication is to gain an understanding of the processes whereby (1) a speaker transforms a discrete linguistic representation of an utterance into an acoustic signal, and (2) a listener decodes the acoustic signal to retrieve the linguistic representation. The research includes development of models for speech production, speech perception, and lexical access, as well as studies of impaired speech communication.

Sponsors

C.J. Lebel Fellowship
Dennis Klatt Memorial Fund
National Institutes of Health
Grant R01-DC00075
Grant P01-DC00361-06A115
Grant R03-DC01721
Grant R01-DC01291
Grant R01-DC0026115
Contract R01-DC0077616
National Science Foundation
Grant IRI 89-0524917
Grant IRI 89-10561
Grant INT 90-2471318

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

1.2.1 Nasal Vowels and Consonants

Some further experimental data on the acoustics of nasalization have been collected, and these data have contributed to a refinement of the acoustic theory of nasalization. An experimental approach has been developed for measuring the natural frequencies of the nasal cavity when it is closed at the posterior end. The natural frequencies are determined by measuring the response of the nasal cavity when a transient excitation is produced by rapidly closing the velopharyngeal port. In accordance with previous findings, two of the principal natural frequencies are in the range 490-640 Hz and 1880-2400 Hz for different individuals. Additional natural frequencies, presumably due to sinuses, are observed at 210-250 Hz and at frequencies between the principal resonances.

Acoustic analysis of nasalized vowels in English and French shows that the lowest sinus resonance contributes to an enhancement of the spectrum amplitude at low frequencies. This increased low-frequency amplitude has been observed to be as great as 10 dB for French nasal vowels. This enhanced low-frequency amplitude is an additional attribute that contributes to the decreased first-formant prominence that occurs when a vowel is nasalized.

1.2.2 Lateral Consonants

In earlier studies of lateral consonants, acoustic events at the release and closure of these consonants were examined. These acoustic data showed the nature of the discontinuity in amplitudes and frequencies of spectral peaks, particularly the abrupt change in second-formant amplitude due to bandwidth changes associated with acoustic losses during the lateral. This study has been extended to include syllabic laterals, which contrast with the unstressed vowel [o] in word pairs like bucko-buckle. Measurements show similar values for the first two formant frequencies $F_1$ and $F_2$ for the two syllabic nuclei, with some spectral differences in the $F_3-F_4$ region which do not appear to be consistent. The most consistent difference between the lateral and the vowel is the $F_2$ bandwidth. Informal listening with synthesized utterances confirms that the increased $F_2$ bandwidth shifts the identification from the vowel to the lateral. More extensive analyses and listening tests are in progress.

1.2.3 Formant Transitions for Fricative Consonants

Measurements have been made of the transitions of the second and third formant frequencies ($F_2$ and $F_3$) in vowels following the release of different voiceless and voiced fricative consonants produced by four speakers. This is part of a larger study of the acoustics, perception, and modeling of fricative consonants. The data show that the range of $F_2$ starting frequencies across different following

---

15 Under subcontract to Massachusetts Eye and Ear Infirmary.
16 Under subcontract to Massachusetts General Hospital.
17 Under subcontract to Boston University.
18 U.S.-Sweden Cooperative Science Program.
vowels is much greater for labiodental fricatives than for palatoalveolar fricatives, with the ranges for dentals and alveolars being between these extremes. These data support a view that the tone body is relatively free to anticipate the following vowel for labials, leading to a wide range of F2 values at the release. The production of palatoalveolars, on the other hand, places constraints on the position of the tongue blade and the anterior portion of the tongue body, thereby constraining the degree to which the tongue can be positioned to anticipate the following vowel. If the range of F2 starting frequencies across different consonants is examined for each vowel, this range is least for the vowel /i/ and most for /u/ with other vowels in between. These range differences reflect differences in the extent of tongue-body movements that are required as movements are made between the consonants and the different vowels.

1.2.4 Transients in Stop Consonant Releases

Theoretical analysis, coupled with acoustic data, indicates that an acoustic transient occurs at the release of a stop or affricate consonant, immediately preceding the noise burst. This acoustic transient is generated by the abrupt pressure release that occurs when the cross-sectional area of the consonantal constriction increases rapidly. The Klatt speech synthesizer has been modified to include the capability of generating such a transient source, with a spectrum that is in accord with theoretical predictions. Listening tests with synthetic consonants produced with different levels of this transient have shown that listeners usually judge the stop consonants with a transient having an appropriately adjusted amplitude to sound more natural than consonants produced without the transient.

1.2.5 Acoustic Study of Italian Vowels

A study of the acoustic characteristics of Italian vowels has been initiated. Formant frequencies, durations, and fundamental frequencies were examined for the vowels produced in a standard consonant-vowel frame with different stop consonants preceding the vowel. The effects of consonant voicing and vowel height on fundamental frequency follow patterns similar to those observed in English. The influence of the consonant on the following vowel is greatest for the second formant, and also follows a pattern similar to that for English. Italian has four vowels with intermediate height—two front and two back vowels, with one member of each pair being more open than the other. The more open vowel of the pair is longer than its more closed counterpart, in contrast to the opposite duration pattern for the corresponding tense-lax pairs in English. Different models are being developed to account for the data.

1.2.6 Preaspirated and Postaspirated Stops in Icelandic

We have examined the perception of two speech cues in Icelandic, Voice Onset Time (VOT) as a cue for word-initial aspiration of stops, and Voice Offset Time (VOFT) as a cue for preaspiration. These speech cues contrast in an interesting manner in perception though signaled by acoustically identical cues. The studies replicate the well-known finding that place of articulation has an effect on the placement of the VOT boundaries separating unaspirated from unaspirated stops. No such effect was found for the perception of VOFT. Both cues are sensitive to the duration of the vowel carrying the relevant segments, VOFT much more so than VOT. While the effect of vowel duration on the perception of VOT has usually been attributed to the effect of rate normalization, in the case of VOFT it emerges that the ratio of VOFT at the phoneme boundary to the duration of the vowel acts as a higher-order invariant, thus obviating the need to posit a mechanism of rate normalization. The differences in the behavior of these speech cues is related to their different positions in the syllable.

1.2.7 Acoustic and Articulatory Studies of Stop Consonant Production

The production of a stop consonant such as [p] requires that the articulator that makes the consonant closure be released rapidly, so that an abrupt onset of acoustic energy is achieved. Data on the movement of the jaw and the lips have been obtained using an electromagnetic midsagittal articulometer (EMMA) to track movements of points on these structures during the production of stop consonants in various phonetic environments. Simultaneous recordings of the sound output were made. Data for the consonant [p] in utterances like [espV] or [epV] (where V = [i] or [a]) show that the downward movements of the jaw and lower lip appear to be coordinated to achieve maximum rate of lip opening at the time of release, in spite of the requirement that the jaw remain high during a preceding [s] in order to enhance turbulence noise generation. In general, the total time interval over which the stop consonant has an influence on the movement patterns is in the range 150-200 ms.
1.2.8 Individual Differences in Laryngeal Waveforms for Female Speakers

The aim of this project is to develop procedures for characterizing the laryngeal behavior of individual speakers based on measurements on the speech waveform. A number of measurements were made in several vowels produced by several different speakers. The measurements are intended to provide information about the glottal area waveform and hence about the configuration and state of the glottis during modal vibration. The parameters include various measurements of spectral tilt, first-formant bandwidth, and degree of acoustic coupling through the glottis to the trachea. Substantial individual differences in these parameters have been observed, and some partial interpretations in terms of average glottal opening have been made. Further studies are needed if a compact set of descriptors for individual speakers is to be developed.

1.3 Studies of Normal Speech Production

1.3.1 Goal-based Speech Motor Control: A Theoretical Framework and Some Preliminary Data

A theoretical framework for the segmental component of speech production has been outlined and some preliminary supporting data have been presented. According to the framework, articulatory movements are programmed to achieve sequences of articulatory and acoustic goals. The goals are defined partly by characteristics of speakers that enable them to produce and perceive speech sounds that have distinctive acoustic properties; these characteristics are correlates of distinctive features. Some feature correlates are determined by quantal (non-linear) relations between articulation and sound. Goal definitions also may be partly language-specific, according to other principles such as a compromise between sufficient perceptual contrast and economy of articulatory effort. When utterances are produced, goal specifications are modified by prosodic influences and reduction. Based on the sequence of modified goal specifications, articulatory movements are timed appropriately and smoothed by the control mechanism and the biomechanical properties of the articulators. To help keep acoustic variability within perceptually-acceptable limits, speech production mechanisms include a strategy of "motor equivalence", which takes advantage of the fact that for some sounds, a similar acoustic transfer function can be achieved with somewhat different area functions. This hypothesis is supported tentatively by data on motor equivalence at the level of the transformation between the vocal-tract area function and the acoustic transfer function.

1.3.2 Factors Underlying Findings of Motor Equivalence in the Transformation Between the Vocal-tract Area Function and the Acoustic Transfer-function

Using measurements made with an electromagnetic midsagittal articulometer (EMMA), we have found evidence of trading relations between lip rounding and tongue-body raising for multiple repetitions of the vowel /u/ in seven male speakers of American English. In other words, when the tongue body is less raised, the lips are more rounded and vice versa. (An eighth speaker did not show such a trading relation.) As mentioned above, these findings support the hypothesis that there is a "motor equivalence" strategy which functions at the level of the transformation between the vocal-tract area function and the acoustic transformation. For all of the subjects, examination of subsets of data defined by a measure of acoustic distance from a normative mean has shown that there is an increasing likelihood of negative correlations for tokens that are less /u/-like. For a majority of the data sets, tokens that are less /u/-like are produced at a faster rate. Thus, this strategy could help to keep acoustic variation within perceptually-acceptable limits, and it seems to come into play more when it might be needed, to prevent production of a sound that could lead to the wrong percept.

1.3.3 Speaking Rate and Associated Kinematic Patterns of Tongue Movements

Kinematically-based measurements of effort on tongue movements were examined for a speaker with a cochlear implant. Turning the speech processor of the implant on and off induced changes in the subject's speaking rate. Kinematic measurements were made of EMMA data on tongue-body movements toward and away from the steady-state target for the vowel /u/ in the utterance "... who hid ....". Although the movements toward the target were quite small—on the order of a millimeter—the kinematic measurements still exhibited relationships that were strikingly similar to previous results of others: there was a strong positive relation between movement distance and duration, and values of an index of velocity profile shape (the ratio of peak velocity to average velocity) were positively related to values of movement duration.
According to measurements of peak velocity and effort, the most effort was exerted in the condition when the cochlear implant had been turned off for the longest time.

### 1.4 Speech Research Relating to Special Populations

#### 1.4.1 Speech Production of Cochlear Implant Patients

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

This research aims to characterize the speech production of postlingually deafened adults before and after they receive cochlear implants. One goal of the work is to contribute to the development of models of the role of hearing in speech. The findings can also contribute to evaluating and improving prostheses and to focusing speech therapy for implant users. Toward these goals, we have been making measurements of (1) vowel acoustics (vowel formants, F0, SPL, duration, and inferred glottal aperture); (2) consonant acoustics (voice-onset time, fricative and plosive spectra); (3) speech aerodynamics (speech initiation and termination lung volumes, volume of air per syllable, inferred subglottal pressure, oral and nasal airflow, peak glottal airflow, minimum flow, AC flow, quotient, and maximum flow declination rate); and (4) tongue, lip and jaw trajectories. A major concern has been to separate "phonemic" changes in these parameters from changes due to "postural" adjustments in overall speaking rate, SPL, and F0.

#### Production of the Plosive Place Contrast

To date, we have examined the spectra of plosive bursts of four implant users before and after activation of the speech processors of their cochlear implants. The spectral slopes of the bursts (dB/octave) in all three places of articulation did not change reliably with activation. Comparison with normative data suggests that these speakers had normal spectral slopes for bilabial and alveolar positions, but more compact spectra for the velar place of articulation.

#### Production of Fricative Place Contrasts

Prior to activation of their cochlear implants, two deafened adult speakers produced the alveolar and palatal voiceless fricatives with spectral parameters more similar to one another than in normal hearing individuals. With the restoration of some hearing, these speakers differentiated these productions more normally. When their implant speech processors were turned off and on for short periods, they reduced and increased spectral differentiation correspondingly.

We have conducted an initial experiment to explore the articulatory basis of changes in fricative spectra with prosthetic hearing and to assess whether they are due to selective control by prosthetic hearing or a postural change in rate or both. With another implant user, an electro-magnetic midsagittal articulometer system was used to record movements of the tongue, lips and jaw. Preliminary analysis showed that tongue blade positions during the palatoalveolar fricative /sh/ changed with access to hearing; moreover, the shift in place of articulation directly correlated with the observed shift in the fricative spectrum between processor on and off conditions, a shift that sharpened the fricative place contrast with the processor on.

#### Voicing Onset Time

Implant users lengthen voice-onset time (VOT) toward normative values with processor activation. However, VOT changes may reflect postural adjustments. VOT and syllable duration have been measured for the plosives spoken by five patients pre- and post-activation of their implant speech processors over a period of several years. Pre-implant, all five speakers characteristically uttered voiced plosives with too-short VOT, compared to normative data. VOTs of voiceless plosives were also abnormally short for three of the speakers, and close to normal for the remaining two. With some hearing restored, implant users made relatively few errors with respect to voicing when identifying plosives in listening tests. Three of the five speakers lengthened VOT, and three increased the contrast. The findings are consistent with the hypothesis that deafened speakers reset articulatory routines when some hearing is restored and they can hear the relevant voicing features. However, this study controlled for the indirect contribution to VOT changes of changes in only one postural variable, speaking rate. In the case of the voiced plosives, multiple correlations between changes in the postural variables of rate, a measure of breathiness and SPL and changes in VOT were large and significant, suggesting that much, if not all, of the VOT changes could be attributed to the mediation of
postural changes. In the case of the voiceless plosives, however, multiple correlations were small and could account for only part of the observed increases in VOT, suggesting that the speaker increased VOT partly under the control of his or her renewed ability to hear VOT due to the implant.

Our findings to this point continue to indicate significant deterioration of acoustic and articulatory speech parameters in adults after becoming deaf, and substantial long-term changes, generally in the direction of normalcy, following activation of the cochlear implant.

1.4.2 Degradation of Speech and Hearing with Bilateral Acoustic Neuromas

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

Two of the enrolled patients have suffered further hearing loss, and we have obtained post-loss recordings on both of them. We have analyzed the pre- and post-loss recordings for one of the two. This patient had a mild-to-moderate hearing loss in one ear (aided) and was deaf in the other due to bilateral acoustic neuromas. She became deaf bilaterally when the second of these neuromas was surgically removed. Following this surgery, an auditory brainstem implant provided minimal speech reception benefit. Three speech production recordings made prior to deafening were compared to recordings made 11 and 35 weeks post-deafening. Results indicated significant changes in voice onset time (VOT), fundamental frequency (F0), F0 range, and vowel duration. VOTs, adjusted for changes in syllable duration, were significantly shorter for the voiceless plosives post-deafening. F0 was at the mean of a normative range prior to deafening, and rose significantly above this range post-deafening. An analysis of prosody indicated that the subject placed pitch accents on the same syllables post-deafening, but F0 mean and s.d. were significantly higher. Vowel formants, /s-sh/ contrast, an indirect measure of breathiness, and overall SPL, showed no significant changes by 35 weeks post-deafening.

1.4.3 Objective Assessment of Vocal Hyperfunction

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

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

Expansion and Correction of the Normative Database

We have gathered and completely analyzed data for 30 additional normal speakers (15 males and 15 females) using updated analysis methods. Statistical tests comparing previous (1988) and new (1993) normative data indicated overall group differences. For most individual parameters, differences between the two groups were relatively small (less than a 10 percent change) and not statistically significant. Such variation could be attributed to minor methodological differences and sampling effects, i.e., most of the new average values fell well within one standard deviation of the previous averages. However, for the parameter maximum flow declination rate (MFDR), there were instances (mostly for loud voice) of significant differences between the 1988 and 1993 data. These differences could be attributed largely to changes (corrections) in the signal processing algorithms that are used to obtain glottal waveform measurements. The new MFDR values more closely match predictions based on speech acoustic theory and are therefore believed to be more valid.
Respiratory Function Associated with Vocal Nodules in Females

We have completed the collection of respiratory data for 10 females with vocal nodules and 10 matched (sex, age, body type) normal controls. Preliminary results indicate that, compared to matched controls, nodules subjects display reduced efficiency of respiratory function for speech production by operating at lung volume levels which require greater muscular effort. Some of the nodules subjects also tended to stop at inappropriate (non-juncture) points within sentences to inhale and demonstrated oppositional chest wall displacements after the initiation of exhalations.

Modeling of Vocal Fold Vibration

A mathematical model of vocal fold vibration has been developed during this last year. We started with the well known Ishizaka-Flanagan two-mass model of the vocal folds. The physical system is described mathematically by a set of nonlinear differential equations with singular coefficients which represent the self-oscillating voice source composed of two stiffness-coupled masses. Also included are lumped approximations for supra- and sub-glottal vocal-tract loading and lip radiation characteristics. The fourth-order-Runge-Kutta numerical method for solving differential equations is applied because it gives more precise results than the Euler method used in Ishizaka and Flanagan's original model.

Treatment Efficacy Studies of Voice Therapy Protocols

Voice therapy protocols for use in our proposed treatment efficacy studies have been developed. The protocols include: (1) detailed step-by-step procedures for each behavioral approach being used (vocal hygiene, direct facilitation, relaxation, respiration, carryover, home practice), (2) daily behavioral charts for documenting voice use outside of the therapy sessions, (3) home practice records (verified by spot checking cassette recordings of home practice sessions), (4) response charts for documenting patient performance in each therapy session, (5) scales for clinician's ratings of patients' motivation and vocal status at each therapy session, and (6) a scale for patients' self evaluation of their vocal status at each therapy session. We have completed making repeated longitudinal recordings under this protocol with one patient and are progressing with several others. Data are currently being analyzed.

Development of a Multi-user Relational Database

Significant progress has been made in the development of a multi-user relational database application ("VoiceBase") which will provide improved capabilities for the storage, retrieval, and reporting of subject background (history), perceptual, acoustic, aerodynamic, electroglossographic, and videostroboscopic project data.

1.4.4 Aiding Dysarthric Speakers

We have been collecting data on the speech patterns of several dysarthric speakers. One objective of this research is to develop procedures for using speech recognition devices to augment the communication abilities of these individuals. The speech of each speaker is evaluated through (1) listening tests, (2) tests with a speech recognizer, and (3) acoustic analysis of a series of words produced by the speaker. An outcome of this evaluation is a description of attributes of vowels and consonants which are produced adequately and consistently and those which are produced in a distorted or inconsistent manner. Each of the speakers who have been evaluated show somewhat different patterns of deficiencies, but there are some common problems, particularly with regard to fricative and lateral consonants that require shaping of the tongue blade. These kinds of evaluations provide a basis for selecting words that are likely to be recognized consistently by a speech recognizer.

1.5 Speech Production Planning and Prosody

Our work in speech production focuses on both segmental and prosodic aspects of the planning process and their interaction. To make these studies possible, we are continuing the development of a prosodic database with our colleagues Mari Ostendorf of Boston University and Patti Price of Stanford Research Institute. With the help of students in the Undergraduate Research Opportunities Program at MIT, using the ToBI transcription system recently proposed, we have labeled the pitch accents, constituent boundaries and boundary tone markers for almost one hour of FM radio news style speech produced by two professional speakers. The resulting database, which also includes part-of-speech labels and automatic phonetic alignments (10 percent hand corrected), will be made available to the speech research community through the Linguistics Data Consortium at the University of Pennsylvania. We are participating in the creation of an online training hand-
book for the ToBI transcription system, being prepared at Ohio State University, which includes audible labeled examples. We are also involved in evaluating the reliability of the system across 30 transcribers. In addition, we are training several new transcribers, who will then participate in both the ongoing creation of the FM news database and the labeling of utterances containing speech errors in the new MIT Digitized Speech Error Database (see below).

Using the FM news database, we have carried out three studies of the role of prosody in speech planning. First, we found that the apparent stress shift (in which the first syllable of late-main-stress words like "Massachusetts" is perceptually more prominent than the main stress syllable, in phrases like "the Massachusetts senator") often results from the placement of an early prenuclear pitch accent on the first accentable syllable of a new intonational phrase and the loss of pitch accent from the main-stress syllable. These findings suggest that speakers and hearers may be able to use the occurrence of early accent within the word as a cue to the onset of a new intonational phrase. Second, we found that alternating-stress words (like "Massachusetts" and "university") are pitch-accented differently from adjacent-stress words (like "campaign" and "Hongkong"): alternating-stress words are more likely to be double-accented and are more reliably early-accented in phrase-initial contexts. In addition, both intuitions and dictionary markings of the placement of lexical stress are less consistent for adjacent-stress words. These results suggest that speakers are reluctant to place pitch accents on adjacent syllables within a word, just as they are for syllables across word boundaries.

Then we analyzed the occurrence of glottal onset in more than 400 vowel-initial syllables and found that it is most likely to occur when the syllable begins an intonational phrase and/or is pitch accented (61 percent, versus 16 percent for non-phrase-initial, non-accented syllables.) The pattern of phrase-initial glottal onsets was maintained even for syllables which begin with reduced vowels, which cannot be pitch accented. This finding suggests that glottalized onset of these syllables may provide a useful cue to the beginning of an intonational phrase. Studies are currently under way to evaluate whether phrase-initial glottal onsets are associated with low F0, post-pausal position or certain specific segment types.

Finally, we are developing and prosodically labeling a second database of utterances containing segmental and word-level speech errors, in order to investigate prosodic constraints on disfluencies. Error utterances are being harvested from a number of existing digitized speech databases, and prosodic transcribers are being trained in our laboratory. We anticipate being able to label and analyze 300-500 error utterances over the next year. We asked the questions (1) what kinds of error corrections speakers mark prosodically and (2) how prosodic factors such as pitch-accent prominence and constituent boundaries constrain the pattern of segmental and morphemic errors. Results can be expected to shed light on the question of what prosodic representations speakers employ during the production planning process and how those representations guide segmental planning.

### 1.6 Models for Lexical Representation and Lexical Access

We are developing a model for lexical access in which items are stored in the lexicon in terms of arrays of features, and procedures for estimating these features from the acoustic signal are specified. The estimation of feature arrays from measurements on the signal proceeds in three steps: (1) detection of landmarks in the signal where a major articulator forms or releases a narrow constriction in the vocal tract, or forms a minimum or maximum opening for a consonant or vowel; (2) making acoustic measurements in the vicinity of the landmarks to estimate the articulator-free and articulator-bound features being implemented; and (3) organizing the feature estimates into a form that can be used to match against a feature-based lexicon. During the past year, our efforts have been directed towards detection of landmarks, developing a framework for estimating landmarks, and hand-labeling a database of sentences in terms of landmarks and features, for testing the adequacy of our analysis procedures.

#### 1.6.1 Identification of Landmarks

Two types of landmarks are created by the formation of a narrow consonantal constriction in the vocal tract. For one type, the articulator forming the constriction makes a complete closure in the midline of the vocal tract, whereas for the other type, the closure is only partial. The first of these is further subdivided into landmarks where there is no increase or decrease in intraoral pressure across the landmark (i.e., nasals and laterals), and where there is a change in intraoral pressure.

The detection of these landmarks is accomplished by examining the rate of change of energy in different frequency bands. The bands that are selected, the averaging time for determining energy in a band, and the time increment for obtaining a
first difference are all adjusted to best suit the type of landmark being sought. The general procedure is to first scan the signal with fixed filters, averaging times, and increments in order to make an initial selection of candidate landmarks, and then to examine each of these landmarks with a more finely tuned set of measurements. The algorithms that have been developed miss a few landmarks, and steps are being taken to refine the algorithms, including the measurement of time variation of first-formant prominences. Procedures for detecting regions of minimum vocal-tract opening for glides are examining changes in low-frequency amplitude and looking for minima in first formant frequency coupled with extrema in other formant trajectories.

1.6.2 Estimation of Features Around Landmarks

A framework is being developed for estimating the features involved in generating a given landmark. The acoustic properties that must be measured to provide cues for some features are relatively context-independent, whereas for other features the properties are dependent on the context of features that co-occur with the feature or that are associated with adjacent segments. A program is being developed to guide the set of context-dependent measurements that are required to estimate a feature. The program is given a query feature (a feature to be determined) and a set of context features, and it specifies a list of measurements to be made to determine the value of the feature. The set of measurements that is required for a given feature is determined in part from theoretical models of speech production and in part through examination of a database of utterances. They include measurements of formant trajectories near landmarks, gross spectrum shape and its variation with time, and degree of spectrum prominence. The program is being tested for a limited set of features using some databases, including one under development in our laboratory.

1.6.3 Feature-labeled Database

Portions of database of 100 sentences spoken by four speakers have been hand-labeled to indicate the locations of landmarks, and the features being implemented at each landmark are listed. The feature estimates are based on careful acoustic analysis of the sound in the vicinity of each landmark and on listening to segments of the signal. This labeled database is to be used for two purposes: (1) it can serve as a standard against which various algorithms for estimating landmarks and features can be evaluated; and (2) it can provide data indicating which features in the spoken words in context differ from features in the lexicon. The second of these uses of the database can help to guide the refining of rules for describing modifications of feature patterns that occur in casual speech.

1.7 Publications

1.7.1 Papers Published


1.7.2 Papers Accepted for Publication


Perkell, J., R. Hillman, and E. Holmberg. "Group Differences in Measures of Voice Production and Revised Values of Maximum Airflow


1.7.3 Papers Submitted for Publication


