# Part IV 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. Corine A. Bickley, Dr. Suzanne E. Boyce, Dr. Carol Y. Espy-Wilson, Seth M. Hall, Dr. Marie K. Huffman, Dr. Sharon Y. Manuel, Dr. Melanie L. Matthies, Dr. Joseph S. Perkell, Dr. Mark A. Randolph, Dr. Carol Chapin Ringo, Dr. Stefanie R. Shattuck-Hufnagel, Dr. Mario A. Svirsky, Dr. Victor W. Zue

### Visiting Scientists and Research Affiliates

Giulia Arman-Nassi,[1] Dr. Harvey R. Gilbert,[2] Dr. Richard S. Goldhor,[3] Dr. Robert E. Hillman,[4] Dr. Jeannette D. Hoit,[5] Eva B. Holmberg,[6] Jacques Koreman,[7] Dr. Harlan L. Lane,[8] Dr. John L. Locke,[9] Dr. John I. Makhoul,[10] Aniruddha Sen,[11] Dr. Victor N. Sorokin,[12] Dr. Arend Sulter,[13] Dr. Noriko Suzuki,[14] Jane W. Webster[15]

### Graduate Students

Abeer A. Alwan, Marilyn Y. Chen, A. William Howitt, Caroline B. Huang, Lorin F. Wilde

### Undergraduate Students

Anita Rajan, Lorraine Sandford, Veena Trehan

### Technical and Support Staff

Ann F. Forestell, Laura B. Glicksman, Sandra I. Lugo, D. Keith North

---

1 Milan Research Consortium, Milan, Italy.

2 Pennsylvania State University.

3 Sensimetrics Corporation.

4 Boston University.

5 Department of Speech and Hearing Sciences, University of Arizona.

6 MIT and Department of Speech Disorders, Boston University.

7 University of Nijmegen, The Netherlands.

8 Department of Psychology, Northeastern University.

9 Massachusetts General Hospital.

10 Bolt Beranek and Newman, Inc.

11 Tata Institute of Fundamental Research, Bombay.

12 Institute of Information Transmission Problems, Moscow.

13 University of Groningen, The Netherlands.

14 First Department of Oral Surgery, School of Dentistry, Showa University, Tokyo.

15 Massachusetts Eye and Ear Infirmary.

## 1.1 Introduction

**Sponsors**

C.J. Lebel Fellowship
Dennis Klatt Memorial Fund
Digital Equipment Corporation
National Institutes of Health
Grants T32 DC00005, R01 DC00075, F32 DC00015, S15 NS28048, R01 NS21183,[16] P01 NS23734[17], T32 NS 07040
National Science Foundation
Grant IRI 88-05680[16]

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.

## 1.2 Studies of the Acoustics, Production and Perception of Speech Sounds

### 1.2.1 Vowels

Although vowels often show a great deal of variability, humans can identify them well enough to be able to communicate through speech. The present study continues work (performed here and at other laboratories) to improve the understanding of (1) factors that cause the variability of vowels and (2) the perceptual effects of this variability.

The effects of consonant context, lexical stress, and speech style (e.g., carrier phrase, continuous read, or spontaneous) on vowels have been investigated in a number of previous acoustic studies. However, those studies of consonant context and lexical stress involved only isolated words or words in a carrier phrase, and the studies of continuous speech did not consider each consonant context separately. In the present study, formant frequencies are measured in vowels taken from a read story which is well-controlled with respect to consonant context and lexical stress. The vowel set was chosen to illustrate the feature distinctions high/non-high (/i/-/e/, /I/-/ɛ/), front-back (/ɛ/-/ʌ/), and tense-lax (/i/-/I/, /e/-/ɛ/). The consonant contexts were chosen to include liquids and glides (/w/, /r/, and /l/) and stops (/b/, /d/, /g/). Liquids and glides have been found to have a stronger effect on adjacent vowels than stops, presumably because liquids and glides constrain the tongue body more than stops (with the possible exception of /g/). The vowels in the corpus carry primary or secondary lexical stress. The same vowels and contexts are elicited in spontaneous speech and in nonsense words spoken in a carrier phrase. In total, the database consists of approximately 1700 vowel tokens from four speakers. Vowels from each speaker are analyzed separately.

Preliminary results indicate that consonant context has a greater effect on vowel formant frequencies at the midpoint than lexical stress or speech style. For example, the tokens /I/ from the syllables /bIt/ and /wIt/, both taken from a carrier phrase, tend to differ more than two tokens of /wIt/, one taken from a carrier phrase and one from spontaneous speech. Of the factors considered in this study, consonant context is the most important for understanding the variation found in non-reduced vowels. This result is especially relevant now for speech synthesis and speech recognition systems, which are moving from read speech to more spontaneous styles.

Future work will follow several directions. The interaction of the factors consonant

---

[16] Under subcontract to Boston University.

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

context, lexical stress, and speech style will be quantified. The formant trajectory throughout the duration of the vowel, not only the midpoint value, will be examined. Also, the vowels will be presented to listeners for identification to determine if the measured differences in the vowel tokens are perceptually relevant. The listeners' ability to identify these tokens will be compared to the performance of statistical classifiers on the same tokens.

### 1.2.2 Nasal Consonants and Vowel Nasalization

A nasal sound is produced when the nasal cavities are coupled to the oral cavity by movement of the velum (soft palate). For some speech sounds (e.g., /m/, /n/), this coupling is a requirement; for others, coupling occurs due to phonetic context — the vowel in *bean* is nasalized because nasal coupling occurs on the vowel in anticipation of the nasal sound *n*. We are carrying out a detailed analysis of this contextual, anticipatory nasalization.

In one study, the central task involves tracking changes in nasalization over time, by comparing spectral properties of nasalized vowels with those of oral counterparts (e.g., the vowels in *bean* and *bead*). Spectral analysis is used to identify differences in harmonic amplitudes between the nasalized vowels and the oral vowels. Additional poles due to nasal coupling will increase harmonic amplitudes of nasal vowels relative to oral vowels; additional zeroes have the opposite effect.

Our results indicate that in such nasalized vowels, additional energy appears fairly consistently in the range of 600–1000 Hz; the additional pole which is thus indicated moves up through this frequency range as the velopharyngeal port opens and the mouth closes for the approaching nasal consonant. Since acoustic theory predicts addition of a pole-zero pair(s) to the spectrum of the vowel, we also want to identify the presence and behavior of the zero. Because this is not easily done with spectral data, analysis-by-synthesis is used to clarify the contribution of the nasal zero to the spectral characteristics of the nasalized vowels.

One goal of this work is to provide measures of nasalization which give reliable information about changes in velopharyngeal opening, without the invasiveness characteristic of physiological studies. The methods used here are particularly useful for determining changes in nasalization over time. Understanding when velum movement begins for production of nasal speech sounds, and the time course it follows, is an essential part of characterizing motor planning for speech. An important additional benefit of this work should be improved synthesis of contextually nasalized vowels. Studies in progress will use synthesized speech to investigate the role of various acoustic factors in the perception of nasalization.

A second study of nasalization has examined the acoustics, synthesis and perception of vowels in a nasal consonant context (e.g., /mam/) and in a nonnasal context (e.g., /bab/), with emphasis being placed on the acoustic attributes around the midpoints of the vowels. There was considerable variability in the acoustic data, but one reasonably consistent result was that the spectrum amplitude of the first formant for the vowel in relation to the amplitude of the first harmonic was less for the nasal context than for the nonnasal context. This property is presumably due, at least in part, to the increased acoustic losses with nasal coupling, causing a widening of F1. In one of the perception experiments, judgments of nasality were obtained with synthetic vowels whose spectra were manipulated to match those of naturally spoken vowels. The data showed that use of a pole-zero pair to synthesize a matching vowel yielded reasonably consistent nasality judgments, but that attempts to match vowel spectra without using a pole-zero pair failed to produce vowels that were judged to be nasal.

In addition to the experimental studies of vowel nasalization in the context of nasal consonants in English, we have attempted to refine existing theoretical treatments of the acoustics of vowels produced with coupling to the nasal cavity. Particular attention was given to the effect of the combined output from the nose and the mouth on the locations of the principal low-frequency

pole-zero pair that is contributed by the nasal coupling. For reasonable areas of the velopharyngeal opening, the predicted frequency of the additional pole is in the range of 750–900 Hz, and of this pole causes a peak with an amplitude that is up to 13 dB above the spectrum amplitude for the corresponding nonnasal vowel.

### 1.2.3 Analysis, Modeling, and Synthesis of Fricative Consonants

Data on airflow and intraoral and subglottal pressure have been collected for voiced and voiceless fricatives in intervocalic position. From these data, estimates have been made of the time course of the turbulence noise sources at the supraglottal and glottal constrictions. Reasonable agreement was obtained between the time-varying acoustic spectra of the fricatives calculated from these sources and the spectra actually observed in the sound. Based on these analyses, new strategies for the synthesis of fricative consonants are being developed. These strategies are focusing particularly on the time variation of the frication, aspiration, and voicing sources at the boundaries between the fricative consonant and the adjacent vowels.

### 1.2.4 Influence of Selected Acoustic Cues on the Perception of /l/ and /w/

Although the sounds /l/ and /w/ are articulated quite differently, they are acoustically quite similar. In a recognition system we have developed, /l/ and /w/ were frequently confused, especially when they occurred intervocalically. In the present study, we are attempting to refine some of the acoustic properties used to distinguish between these sounds by investigating the perceptual importance of some of the cues used in the recognition system, as well as some others which appear to be salient.

An [ala]-[awa] continuum was synthesized. The starting point was an easily identifiable [ala] stimulus. Three factors were varied orthogonally to shift the percept towards [awa]. First, the rate of change in the formant transitions between the semivowel and following vowel was varied in five steps from 20 msec to 60 msec. Second, the rate of change in the amplitudes of F3, F4 and F5 between the semivowel and following vowel was varied in five steps so that the amplitude of the spectral peaks in the higher frequencies could change as slowly as 10 dB in 60 ms or as fast as 10 dB in 20 ms. Finally, the spectral shape of the consonant could contain either prominent peaks above F2 (coronal shape) or little energy in the higher frequencies (labial shape). All combinations of the parameters were synthesized, yielding a total of 50 stimuli, and stimuli were presented in a test to listeners who gave a forced choice response of /l/ or /w/.

Results of the identification tests show that the rate of change of the formant transitions is an important cue. Stimuli with formant transitions of 30 ms or less are heard as [ala] . As the durations of formant transitions are increased above 30 ms, the perception moves towards [awa]. Spectral shape also greatly influences which consonant is heard. In general, for each formant transition duration, more of the stimuli synthesized with a labial spectral shape are heard as [awa]. However, there were two types of responses. Regardless of the rate of formant transitions between the consonant and following vowel, one set of listeners could not hear [awa] unless the consonant was synthesized with a labial spectral shape. On the other hand, regardless of the spectral shape, the response of another set of listeners moved from [ala] to [awa] as the formant transitions became more gradual.

The effect on the listeners' responses of the rate of change in the amplitude of F3, F4 and F5 between the consonant and following vowel was negligible. However, since this abrupt change in amplitude of some high frequency spectral prominence between an /l/ and a following vowel (as much as 20 dB in 10 ms) has been observed systematically in natural speech, we plan to investigate the perceptual importance of this cue more carefully.

### 1.2.5 Acoustic Theory of Liquid Consonants

As a part of our systematic study of different classes of speech sounds, we have been attempting to develop a theory of sound production for the liquid consonants /r/ and /l/. A distinguishing attribute of these consonants appears to be that the acoustic path from the glottis to the mouth opening consists of at least two channels of different lengths, and possibly includes a side branch. The side branch is under the tongue blade for /r/ and between the tongue blade and the palate for /l/. A consequence of these configurations is a transfer function that includes a zero and an additional pole in the frequency range below about 4 kHz. When the consonant is released into a vowel, the zero and the pole ultimately cancel. Acoustic data appear to support this analysis.

### 1.2.6 The Consonant /h/ and Aspiration Noise

The consonant /h/ in English is unique in that its primary articulation is at the glottis, and it apparently lacks an invariant configuration (the oral gestures accompanying /h/ seem to be due almost entirely to surrounding phones). We have been studying acoustic phenomena related to /h/ to get a better understanding of how it behaves and to improve synthesis of it. We have used a common measure of breathy voice, that is, the difference in amplitudes of the first two harmonics (H1-H2), to characterize changes in the glottal spectrum as speakers move into and out of an /h/. Across speakers, we find consistent H1-H2 changes associated with /h/ of about 10 dB, in spite of large individual differences in absolute H1-H2. We have also examined the characteristics of the turbulence noise that is generated during /h/. In addition to a source of noise near the glottis, we find evidence for noise generation at points in the upper vocal tract, particularly adjacent to high vowels when the oral tract is relatively constricted. The measured spectrum of the radiated sound for /h/ is in reasonable agreement with the spectrum calculated from theories of turbulence noise generation.

### 1.2.7 Modeling of Vocal-Fold Vibration

A computer model of the interaction between vocal-fold vibration and vocal-tract configuration has been implemented; this model is useful in analyzing interactions between the vocal tract and the glottal source. The model includes two-mass representations of the vocal folds and an open or closed vocal tract. In many ways, the behavior of the model is consistent with measurements of vocal-fold vibration and inferences of vibratory characteristics which are based on acoustic observations. The model generates periodic glottal pulses when the vocal tract is unconstricted, if a sufficient pressure drop across the glottis exists, and if the vocal folds are appropriately positioned. For the cases of modal and "breathy" voicing, there is good agreement between many characteristics of the behavior of the model and measurements of speech. The results of the model also seem reasonable in cases in which voicing is inhibited, such as a reduction in the transglottal pressure and changes in the glottal configuration. When the vocal-tract load is changed, the model also produces the expected behavior. In the case of a closed vocal tract, the model shows a decay in vocal-fold vibration within a time interval of 2–3 glottal pulses, and the vocal folds come to rest with the upper fold more abducted than the lower.

### 1.2.8 Perception of Some Constant Contrasts in Noise

The goals of this study are: first, to examine perceptual confusions of place of articulation of some consonants when these consonants are heard in noise; and second, to use the results of the masking theory of the human auditory system in predicting when these confusions occur.

In order to determine whether the results of masking theory (which were established primarily for simple tones) could be applied to formant frequencies of vowels, two psychophysical experiments were conducted. In the first one, a 1-kHz tone was used; in the second, a synthetic vowel with only one formant frequency. The formant frequency of

the vowel was the same as that of the tone, i.e., 1 kHz. The levels of both the tone and the one-formant sound were the same, as was the duration of both the tone and the vowel. The masker in these experiments was white noise. Results of the experiment show that both the tone and the synthetic vowel were masked at about the same noise level.

The second series of experiments consisted of perceptual experiments using natural /Ca/ syllables spoken by one male speaker where the consonant C was one of the four consonants /b/, /d/, /m/, or /n/. The utterances were then degraded by adding various levels of white noise and presented to subjects in identification experiments. Preliminary results show that the thresholds where confusions in place of articulation occur can, indeed, be estimated with a high degree of accuracy from masking theory. Future work includes conducting more psychophysical experiments using different vowels and extending the stimuli to include synthetic consonant-vowel syllables for better control of the different parameters of the speech signal.

## 1.3 Studies and Models of the Perception and Production of Syllables, Words, and Sentences

### 1.3.1 Distinctive Features and Lexical Access

A model of lexical access has been proposed in which acoustic correlates of distinctive features are identified in the speech signal prior to lexical access. These acoustic properties are identified by a two-stage process: first, the locations of acoustic landmarks or events are determined; and second, additional acoustic properties are extracted by examining the signal in the vicinity of these landmarks. As a first step in implementing automatic procedures for extracting these properties, we are hand-labeling some speech data according to a specified inventory of landmarks and properties that bear a rather direct relation to distinctive features. A system for facilitating this hand-labeling has been developed to permit the observer to examine, with any degree of detail, the waveform and successive spectra in the vicinity of specified landmarks in the signal. These labeled data will be used as a basis for evaluating automatic algorithms for extracting the relevant properties.

### 1.3.2 Syllable-Based Constraints on Properties of English Sounds

The objective of this research is to develop a phonological representation and corresponding rule framework for modeling constraints on an utterance's acoustic-phonetic pattern. The primitives of our descriptive framework are distinctive features, which characterize an utterance's underlying surface-phonemic representation. An associated set of acoustic properties comprise an utterance's acoustic description. At present, rules for relating these two domains of representation are based on the syllable. Specifically, constraints on patterns that comprise the acoustic representation of an utterance are stated as conditions on the realization of well-formed syllables; an immediate constituent grammar is assumed for representing syllable structure at the surface-phonemic level. Thus far, our efforts have been directed towards two lines of investigation: (1) providing empirical justification for the use of the syllable in the current feature-based representational framework, and (2) developing a formal model of lexical representation and lexical access incorporating syllable-based constraints.

It has been argued by linguists that the syllable provides the basis for effectively describing the phonotactics of English. The syllable is also viewed as providing an appropriate domain of application for a wide range of rules of segmental phonology. We have developed statistical methods for testing these two claims. In addition, we have performed experiments for evaluating the possible role of syllabic constraints during word recognition. For example, binary regression trees have been used to quantify the extent to which knowledge of a segment's position within the syllable, in conjunction with other contextual factors, aids in the prediction of its acoustic realization. The principle of Maximum Mutual Information is used in con-

structing trees. The principle of mutual information is also used in quantifying collocational constraints within the syllable. We have performed a hierarchical cluster analysis in order to determine an appropriate immediate constituent structure for describing syllable internal organization. Finally, a model of lexical access has been proposed based on the notion of constraint satisfaction. Constraints, stated primarily in terms of the syllable, are applied to an utterance's acoustic description in order to derive a partial phonemic specification of an utterance in which features are arranged in columns and assigned to the terminal positions of the syllable's internal structure. In this partial phonemic specification, selected constituents of the syllable are left unspecified for their feature content. Partitions of a 5500-syllable lexicon have been constructed and evaluated to determine which constituents within the syllable are most informative in identifying word candidates.

In the study examining the role of syllable structure in predicting allophonic variation, we evaluated a database of approximately 12,000 stop consonants obtained from the TIMIT database. We found that a stop's syllable position is the single most important factor in explaining the variation in its acoustic realization. Furthermore, we have found interesting interactions between acoustic-phonetic constraints and constraints on an utterance's phonemic representation. For example, a stop consonant is almost certain to be released when placed in the syllable-onset position, whereas in the syllable coda, a stop tends to be unreleased. A syllable-coda stop may also exhibit a number of other phonetic variants (e.g., released, glottalized, deleted, etc.). Given that place-of-articulation and voicing features are well represented for released stops, the latter result suggests the syllable coda to be a less reliable source of phonetic information than the syllable onset. In agreement with this finding, results from lexical partitioning experiments suggest that the coda is the least informative of the syllable's constituents in terms of providing information regarding the lexical identity of the syllable.

The results of our experiments suggest that linguistic information is conveyed in the signal in a highly constrained and redundant manner. We are currently developing a formal model of lexical representation capable of exploiting this redundancy. Concurrent with the development of this model, we are also extending the above experiments, as well as exploring parsing methods capable of implementing the emerging framework for word recognition.

### 1.3.3 Temporal Spreading of the Feature Retroflex

We are investigating some of the domains in which we expect a significant influence of /r/ on neighboring sounds. In particular, we are looking at the effects of speaker, speaking rate and phonetic context on the spreading of the feature retroflex.

In an earlier acoustic study, we observed that postvocalic /r/s are often merged with preceding vowels, especially in words where the /r/ is followed by a syllable-final consonants such as in "cartwheel." In comparison with spectrograms of the word "car" where there are two distinct time periods for the /a/ and /r/, spectrograms of "cartwheel" show one vocalic region for these sounds which appears to be an r-colored /a/. In addition, we observed that in words like "everyday" ([ɛvrIdeʸ]), retroflexion from the /r/ often spreads across the preceding labial consonant (/v/) to the end of the vowel (/ɛ/). The spreading of retroflexion is evidenced by the lowering of the third formant F3.

In the present study, we recorded several speakers saying the minimal pair words "car," "cart," "card," "carve," and "carp" at a slow and a more casual speaking rate. We also recorded the speakers saying minimal pair sentences which contained various combinations of a vowel, one or more /r/s, and one or more labial consonants.

Preliminary results suggest that spreading of the feature retroflex occurs mainly when the talkers are speaking casually. Merging of postvocalic /r/s and preceding vowels does not always occur when the /r/ is followed by a syllable-final consonant, even when the utterance is spoken at a casual speaking rate. The feature retroflex can spread across one or two labial consonants into the preceding

vowel, although most of the vowel is unaffected, and F3 is lowered only in the last two or three pitch periods. The labial consonants, on the other hand, can be considerably affected by retroflexion. For example, in a sentence which contains the sequence /vwɝ/, the /v/ and /w/ are merged into one consonant and they have a lower F3 than the following /ɝ/.

### 1.3.4 Speech Production Planning

Work has continued on experiments eliciting error evidence to evaluate a frame-and-insert model of serial ordering for speech production. A task involving sentence generation from target words is being used to extend results from read-aloud tongue twisters, showing that (1) word onsets are more susceptible to segmental interaction errors than are stressed-syllable onsets located in word-medial position, suggesting a word-based representation at the point where such errors occur; and (2) word-final segments are protected against interaction errors in phrases but not in lists of words, suggesting a different planning framework for grammatically- and prosodically-structured utterances.

### 1.3.5 Prosodic Prominence and Stress Shift

Work has continued on acoustic and perceptual measures of the change in prosodic prominence called "stress shift" (e.g., in the word "thirteen," the syllable "thir-" is perceived as more prominent than the main-stress syllable "-teen" when the word is combined in the phrase "thirteen men"). Perceptual judgments by both phonetically sophisticated and unsophisticated listeners show that stress is perceived to shift; acoustic measures suggest little or no increase in fundamental frequency (F0) or in duration. Both of these results are consistent with a model in which perceived stress shift results from the disappearance of the pitch accent from the main stress syllable of the target word.

### 1.3.6 Analysis and Synthesis of Prosody

During the first year of this new project we have collected most of our speech database, and have demonstrated that prosodic patterns are effective at disambiguating certain types of structural ambiguity.

#### *Prosody Database*

We have collected an extensive database of speech from six speakers using the FM radio newscaster style, with the cooperation of WBUR radio at Boston University. Both in-studio newscasts and laboratory recordings of experimental utterances have been obtained. Several hours of speech have been transcribed orthographically, and much of it digitized. A prosodic transcription system has been developed for labeling the parts of speech, word stress patterns, and intonation units of the utterances; we are working on a labeling system for phrase-level prominences.

#### *Prosodic Disambiguation*

Many pairs of sentences in English consisting of the same string of words and segments differ strikingly in their structure and, thus, in their interpretation. For example, "The men won over their enemies" might mean that the men persuaded their opponents to their point of view, or that the men vanquished their foes. By providing contrasting preceding contexts, we obtained two spoken versions for each of 30 such sentences, ascertained by listening that the prosody of each was appropriate to the preceding context, and then asked listeners which preceding context was the appropriate one for each version. Results demonstrate conclusively that prosodic differences can lead to appropriate structural interpretations of phonologically identical sentences. We are currently developing a transformation algorithm that will allow us to impose the durations and F0 patterns of one version onto the phonetic shape of the other, providing a stronger test of the hypothesis that prosodic differences alone are responsible for the difference in interpretation by listeners.

## 1.4 Basic Speech Physiology

### 1.4.1 Timing of Upper Lip Protrusion Gestures for the Vowel /u/

Timing of upper lip protrusion gestures and accompanying acoustic events was examined for multiple repetitions of word pairs such as "lee coot" and "leaked coot" for four speakers of American English. The duration of the intervocalic consonant string was manipulated by using various combinations of /s,t,k,h,#/. Data from one of the speakers formed a bimodal distribution which made them difficult to analyze. For the other three subjects, pairwise comparisons and other analyses were made of times of: acoustic /i/ offset to acoustic /u/ onset (consonant string duration), protrusion onset to acoustic /u/ onset (onset interval), maximum acceleration to acoustic /u/ onset (acceleration interval), and acoustic /u/ onset to protrusion offset (offset interval). In spite of considerable token-to-token and cross-speaker variation, several general observations were made: There were some consonant-specific effects, primarily for /s/. The non-s subset evidenced two patterns: (1) The lip protrusion gesture for /u/ had a relatively invariant duration, but its timing varied with respect to the oral consonant gesture complex: the longer the consonant string, the earlier the lip protrusion gesture, or (2) The protrusion gesture duration correlated positively with consonant duration. In the predominating pattern 1, the slope of the timing relationship between oral and labial gestures differed across subjects.

### 1.4.2 Kinematics of Upper Lip Protrusion Gestures for the Vowel /u/ at Normal and Fast Speaking Rates

Acoustic events were identified manually and kinematic events were identified algorithmically on the lip-protrusion versus time signal for one of the above four subjects, using a program developed for this purpose. In addition to the parameters examined in the previous analysis, peak velocity and peak acceleration for the lip protrusion gestures were included in pairwise correlations and comparisons. As anticipated, peak velocity and peak acceleration of the lip protrusion gesture for /u/ correlated with one another and with movement distance. Consistent with the timing results (above), peak velocity and acceleration were not correlated with duration of the consonant string preceding the /u/. This result, in combination with overall gesture timing that correlated with consonant string duration (pattern 1 from the previous analysis), was found in the fast speech of this subject as well as in his normal rate speech. Most parameters differed significantly between the two rate conditions in expected ways. Higher average values of peak velocity and peak acceleration in the fast condition agreed with results of others which suggest that movement kinematics are adjusted as part of the mechanism for speaking at different rates.

### 1.4.3 Articulatory Movement Transduction: Hardware and Software Development

Our alternating magnetic field system for transducing articulatory movements has been recalibrated after making several changes in the electronics to reduce field strength and improve electrical safety. The performance of the system with the revised electronics was initially somewhat degraded; considerable effort was expended in restoring performance to an acceptable level. New single-axis transducers have been received, and tests are being conducted on the system's performance in simultaneously tracking multiple receivers while recording an acoustic signal. An interactive, menu-driven computer program has been implemented for the display and initial analysis of data from the movement transducer. The program simultaneously displays time-synchronized acoustic and multi-channel displacement data, and generates x-y plots of the displacement data over selected time intervals. It also allows for audition of the acoustic signal and a number of other useful data-analysis functions.

## 1.5 Speech Production of Cochlear Implant Patients

### 1.5.1 Speech Breathing in Cochlear Implant Patients

A study has been completed of the effects on speech breathing of postlingual profound-to-total deafness and of the reintroduction of auditory stimulation, including some self-hearing. Three postlingually deafened adults read passages before and after receiving stimulation from a cochlear prosthesis while changes in their respiratory volumes were transduced and recorded. (The schedule of eight to ten recordings of each subject covers a period two years; it is complete for two of the three subjects.) All subjects read initially with abnormal average airflow and volume of air expended per syllable. These two parameters changed significantly in the direction of normalcy following the onset of stimulation.

### 1.5.2 Work in Progress

We have begun an experiment on the short-term changes in speech breathing and speech acoustics caused by brief periods of auditory stimulation and its interruption within a single session. The subjects are two of the three from the previous experiment.

A program has been completed for the efficient measurement of fundamental frequency, vowel formants and harmonic amplitudes from the acoustic signal, and open quotient from the electroglottographic signal. The resulting data will allow us to explore several areas of potential acoustic phonetic abnormalities and to correlate those data with the associated changes in speech breathing.

Longitudinal recordings of speech samples pre- and post-implant and signal processing of those recordings continue. In a later stage, measures from a pneumotachometer and from a nasal accelerometer will be integrated with the respiratory and acoustic measures with the aim of clarifying the role of audition and the effects of its loss on adult speech elaboration.

## 1.6 Phonatory Function Associated with Misuse of the Vocal Mechanism

### 1.6.1 Studies in Progress

Complete sets of subject recordings have been obtained, and data extraction is nearly complete for the following studies:

(a) Intra-subject variation in glottal airflow, transglottal pressure, and acoustic and electroglottographic measurements for normal male and female speakers. We have made three recordings each of three normal male and female speakers (with at least one week between each repeated recording). Results from this study will begin to provide important information concerning the reliability of our measurements. Such information is critical to evaluating the utility (sensitivity) of these measurements for the clinical assessment of vocal pathology.

(b) Relationships between glottal airflow and electroglottographic measurements for female speakers in soft, normal and loud voice. Recordings have been obtained for twelve normal females. Results from this study will represent the first group-based information concerning relationships between glottal airflow and electroglottographic measures of vocal function.

(c) Phonatory function associated with vocal nodules in females. Recordings have been obtained for a group of twelve females with vocal nodules, and for twelve sex- and age-matched normal subjects serving as a control group. These data will enable the first group-based statistical tests for significant differences in our measures between normal vocal function and hyperfunctionally-related vocal pathology. The results of this study will be used to identify those measures which differentiate hyperfunctional (adducted) from normal voice. Five of these subjects have also been recorded following therapy. Comparisons of their pre- versus post-therapy data will provide information about the efficacy of the therapeutic techniques.

(d) Changes in phonatory function associated with spontaneous recovery from functional dysphonia: A case study. We have

obtained repeated recordings of a female subject who initially displayed functional dysphonia and then (approximately one year later) appeared to spontaneously recover normal voice. Results of this study will enable us to identify those measures which differentiate hyperfunctional (non-adducted) from normal voice.

### 1.6.2 Development of Facilities and Methodological Refinements

Signal processing has been refined, and an interactive, menu-driven analysis program has been developed for extraction of aerodynamic, electroglottographic and acoustic measures on a new engineering workstation that was installed last year. With the new software, we can measure additional parameters (relative harmonic amplitudes, adduction quotients from electroglottographic and flow waveforms). Extensive use of command procedures and algorithmic data extraction has greatly increased the ease and efficiency with which signals are processed and analyzed.

We have decided to eliminate the measure of vocal-fold adduction that is obtained from the first derivative of the electroglottographic (EGG) waveform. The first derivative of the EGG is often too weak and noisy (particularly for female voices and for males in the soft voice condition) to reliably locate the points on the signal that are needed to obtain the adduction measure. We still obtain two more reliable estimates of vocal-fold adduction, one from the undifferentiated EGG signal and another from the inverse filtered flow signal (i.e., the glottal airflow waveform).

## 1.7 Computer Facilities

Our Vax 11/750, which was used for analysis and synthesis of acoustic signals, has been replaced with a Local Area VaxCluster consisting of five engineering workstations received through a grant from the Digital Equipment Corporation. Four of the workstations have hardware for real-time A/D and D/A; the analysis and synthesis software developed by Dennis Klatt has been ported to run on these machines (with the help of colleagues in the Linguistics Department, University of Texas). The new "speech" cluster is on the same DECNet network as our previously-described physiology cluster, allowing for the effortless interchange of data and shared printing and backup functions. An 8-mm tape backup system has been added, with dual porting for access by both clusters. An erasable optical disk subsystem has been added to the physiology cluster, providing for storage and rapid access of large signal files.