16.0 Speech Communication

Academic and Research Staff

Prof. K.N. Stevens, Prof. J. Allen, Prof. M. Halle, Prof. S.J. Keyser, Dr. C. Bickley, Dr. S. Boyce, Dr. C. Chapin Ringo, Dr. C. Espy-Wilson, Dr. D.H. Klatt, Dr. S. Manuel, Dr. J.S. Perkell, Dr. S. Seneff, Dr. S. Shattuck-Hufnagel, Dr. V.W. Zue, M. Cohen, D.H. Kaufman, N. Lauritzen, M. Phillips

Collaborating Scientists

Dr. A. ni Chasaide,[1] Dr. R. Goldhor,[2] Dr. M. Gosy,[3] Dr. R.E. Hillman,[4] Dr. T. Hirahara,[5] Dr. K. Hirose,[6] E.B. Holmberg,[7] Dr. H. Kawasaki,[8] Dr. H. Lane,[9] Dr. L.S. Larkey,[10] Dr. J. Locke,[11] Dr. J.I. Makhoul,[12] Dr. N. Suzuki,[13] J. Webster,[14] Dr. L. Wheeler,[15] Dr. K. Yoshida[16]

Graduate Students

A. Alwan, M. Anderson, K.K. Key, N. Daly, S. Dubois, J.R. Glass, A.W. Howitt, C. Huang, R. Kassel, L. Lamel, H. Leung, J.N. Marcus, L. Pastel, J. Pitrelli, M. Randolph, T. Wilson

---

1 Visiting Scientist, Trinity College, Dublin

2 Kurtzweil Applied Intelligence

3 Visiting Scientist, Hungarian Academy of Sciences, Budapest

4 Boston University

5 Visiting Scientist, ATR, Osaka

6 Visiting Scientist, University of Tokyo, Tokyo

7 Boston University

8 Voice Processing Corporation

9 Northeastern University

10 Kurtzweil Applied Intelligence

11 Massachusetts General Hospital

12 Bolt, Beranek and Newman, Inc.

13 Visiting Scientist, Showa University, Tokyo

14 Massachusetts Eye and Ear Infirmary

15 Visiting Scientist

16 Visiting Scientist, NEC, Kawasaki

Undergraduate Students

M. Blush, G. Hopkins, C. Jankowski, J. Landry, A. Lim, H. Mitra, C. Pao, A. Shaw, S. Tierney, D. Whitney, A. Wong

Support Staff

A. Forestell, K. Kline, K. North, A. Wint

Part-Time Assistants/Special Projects

K. Abramson, N. Bateman, L. Glicksman, K. Isaacs, M. McCandless, L. Volaitis, D. Wong

C.J. Lebel Fellowship
National Institutes of Health (Grants 5 T32 NS07040, 5 R01 NS04332, 5 R01 NS21183, 5 P01 NS 13126, and 1 P01-NS23734)
National Science Foundation (Grant BNS 8418733)
U.S. Navy - Naval Electronic Systems Command (Contracts N00039-85-C-0254, N00039-85-C-0341, N00039-85-C-0290)

16.1 Acoustic Correlates of Breathiness: First Harmonic Amplitude, Turbulence Noise, and Tracheal Coupling

A selected sample of reiterant speech has been collected from ten female speakers and six male controls in order to quantify acoustic correlates of perceived breathiness of the female voice, and to contrast these measures with comparable data from males. Two sentences with differing stress patterns were spoken by replacing each syllable by [hV] and by [ʔV], where V = [a, i, æ, o, ɝ]. Detailed analysis of the [a] data reveals: 1) wide variation in the strength of the first harmonic (relative to first formant amplitude), with an average increase of about 6 dB for females relative to males; 2) a greater tendency for the third formant to be excited by noise rather than voicing harmonics in the female population; and 3) indications of tracheal poles and zeros in the spectra of vowels adjacent to voiceless consonants in utterances produced by both genders. These three measures of breathiness tend to be greatest in unstressed syllables and toward the end of an utterance.

The acoustic data have been used to design a synthesis experiment in order to determine which dimensions of breathiness are perceptually most salient. Preliminary results of a listening test involving a female voice producing the vowel [a] under 16 different conditions indicate that noise in the third formant region is the most effective cue for most listeners, although a few listeners are more responsive to an increase in first harmonic amplitude and/or spectral tilt. Some changes (first formant bandwidth increase or increase to the first harmonic amplitude) induced the perception of nasality if done alone, but were interpreted as a highly breathy non-nasal vowel if accompanied by aspiration noise and spectral tilt.

16.2 Speech Recognition

The overall objectives of our research in machine recognition of speech are:

1. to carry out research aimed at collecting, quantifying, and organizing acoustic-phonetic knowledge, and
2. to develop techniques for incorporating such knowledge, as well as other relevant linguistic knowledge, into speech recognition systems.

During the past year, progress has been made on several projects related to these broad objectives.

16.2.1 Acoustic Segmentation and Classification

As part of our goal to better understand the relationship between the speech signal and the underlying phonemic representation, we have developed a procedure that segments the speech signal into an acoustic structure, and have determined an acoustically motivated set of broad classes. The segmentation algorithm makes use of the output of an auditory model. Every 5 ms, the algorithm measures the similarity of a given frame of data to its near neighbors. A frame is associated with either its past or future, depending on whether the backward or forward similarity is greater. We have initially biased the algorithm towards over-segmentation, since mechanisms exist for us to combine segments at a later stage.

Since there is no single level of segmentation that can adequately describe all the acoustic events of interest, we adopted a multi-level representation. In this representation, each initial "seed region" is associated with either its left or right region, again using a distance metric. When two regions are associated with each other, they are merged into one. The procedure is repeated until the entire utterance is described by a single acoustic event. By keeping track of the distance at which two regions merge into one, the multi-level description can be displayed in a tree-like fashion as a dendrogram. The acoustic description varies from fine at the bottom of the dendrogram, to coarse at the top. Thus, for example, the release of a stop consonant may be considered to be a single acoustic event or a combination of two events (release plus aspiration), depending on the level of detail desired.

In order to evaluate the effectiveness of this representation, we first developed an algorithm to automatically find the path through the dendrogram which best matched a time-aligned phonetic transcription. We then tabulated the insertion and deletion errors of these paths. An analysis of the acoustic structure, using 500 utterances from 100 different talkers, shows that it captures over 96% of the acoustic-phonetic events of interest with a 5% insertion rate.

Our objective with the acoustic classification algorithm is to group similar sounds into the same category, and to separate sounds that are widely different. While we did not know how many classes would be appropriate, we suspected that the number of classes would be small in order for the results to be robust against contextual and extra-linguistic variations.

To classify the acoustic segments, we first determined a set of prototype spectral templates based on training data. This was accomplished by using a stepwise-optimal agglomerative hierarchical clustering procedure, resulting in many possible codebooks of varying size and content. We then evaluated the effectiveness of various codebooks in several ways, using 500 sentences from 100 speakers. A comparison of the phonetic content of the resulting clusters over several databases indicates that the top three or four levels are quite stable, suggesting that the total number of clusters should not exceed twenty. We also measured the decrease in mean square distortion error as a function of the cluster size and found that the rate of decrease is less than 1% after the cluster size exceeds ten. In addition, we judged the relative merit of a set of clusters by examining the distribution of phonetic information within each set. This was done by performing hierarchical clustering of all phones using their distribution across the set of clusters as a feature vector. This procedure is very helpful in facilitating visualization of the data structure captured by a set of clusters. A qualitative analysis of these structures showed that after ten to twelve clusters the hierarchical organization did not change significantly.

16.2.2 Recognition of Semivowels in American English

A set of procedures has been developed for the detection and classification of the semivowels /w, y, r, l/ in American English. The detection process marks those acoustic events which may signal the occurrence of a semivowel. Once marked, the acoustic events are used by the classification process in two ways, based on their times of occurrence and their relative strengths. First, a small region from which to extract the values of particular acoustic properties is determined. Second, the number of possible classifications of the detected sound is reduced. Almost all of the acoustic properties are based on relative measures, and hence they tend to be independent of speaker, speaking rate and speaking level.

Fairly consistent overall recognition results in the range of 78.5% to 95% were obtained across different contexts. (Higher performance was obtained when /w/ and /l/ were allowed to be confusable.) These results are for corpora which include polysyllabic words and sentences produced by many speakers, both males and females, of several dialects. Thus the recognition data show that much of the across-speaker variability is overcome by using a feature-based approach to recognition where relative measures are used to extract the acoustic properties.

Several issues were brought forth by this research. First, an acoustic study revealed several instances of feature assimilation. Some of the domains over which feature spreading occurred are limited to syllables whereas others spread across syllable boundaries. Second, an analysis of the sounds misclassified as semivowels showed that, due to contextual influences, the misclassified vowels and consonants had patterns of features similar to those of the assigned semivowels. This result suggests that there may be an advantage in representing lexical items in terms of matrices of binary features as opposed to, or in addition to, phonetic labels. Finally, the system's ability to recognize semivowels which were in the underlying transcription of the utterances, but were not included in the hand transcription, suggests that caution should be exercised in using hand-transcribed data to evaluate recognition systems, without appropriate interpretation of the results.

Our experience with the recognition of the semivowels has helped us to evolve a framework for a feature-based approach to speech recognition. This approach is based on three assumptions: that phonetic segments are represented as bundles of binary features; that the abstract features have acoustic correlates which, due to contextual influences, have varying degrees of strength; and that the acoustic manifestation of a change in the value of a feature or a group of features is marked by specific events in the sound. These acoustic events correspond to maxima or minima in particular acoustic parameters.

16.2.3 Recognition of Continuously-Spoken Letters by Listeners and Spectrogram Readers

Because of acoustic similarities between the pronunciation of some letters of the alphabet, automatic recognition of continuously-spoken letters is a difficult task. The goal of this study is to determine and compare how well listeners and spectrogram readers can recognize continuously-spoken letter strings from multiple speakers. Our interest in spectrogram reading results is motivated by the belief that this procedure may help us identify acoustic attributes and decision strategies that are useful for system implementation. Listening and spectrogram reading tests involving eight listeners and six spectrogram readers, respectively, were conducted using a corpus of one thousand word-like strings designed to minimize the use of lexical knowledge. Results show that listeners' performance was better than readers' (98.4% vs 91.0%). In both experiments, string lengths were determined very accurately (98.1% and 96.2%), presumably due to the large number of glottal stops inserted at letter boundaries to facilitate segmentation. Most of the errors were due to substitution of one letter for another (68% and 92%), and they generally fall into two categories. Asymmetric errors can often be attributed to subjects' disregard for contextual influence, whereas symmetric errors are largely due to acoustic similarities between certain letter pairs. Our subsequent acoustic study of four of the most confusable letter pairs has resulted in the identification of a number of distinguishing acoustic attributes. Using these attributes, we achieved overall recognition performance better than that of the readers.

16.2.4 Vowel Recognition Using Artificial Neural Nets

This study is concerned with the application of artificial neural nets to phonetic recognition. Our work is motivated by the observation that our improved knowledge of acoustic-phonetic feature extraction is often overshadowed by our relative ignorance on how to combine them into a robust decision. We suspect that artificial neural nets may provide a natural self-organizing mechanism for different acoustic cues to simultaneously interact, cooperate and compete. Our goal is to investigate how the mathematically well-defined framework of artificial neural nets can be exploited in phonetic recognition when they are augmented with acoustic-phonetic knowledge. Specifically, we explored issues such as the selection of an appropriate network, the choice of the error metric, the use of contextual information, and the determination of the training procedure. Our investigation is couched in a set of experiments that attempts to recognize the 16 vowels in American English independent of speaker.

Our experimental results were based on some 10,000 vowel tokens that appear in all phonetic contexts. They were extracted from 1,000 sentences spoken by 200

speakers, 140 male and 60 female. We found that by replacing the mean-squared error metric with a weighted one to train a multi-layer perceptron, we consistently obtained better recognition accuracy and rank order statistics. Using the two-layer perceptron in a *context-independent* manner, we were able to achieve a top-choice vowel recognition accuracy of 54%, which compares favorably with results reported in the literature. Our *context-dependent* experiments reveal that heterogeneous sources of information can be integrated to improve recognition performance. Our top-choice accuracy of 67% is comparable to the average agreement on vowel labels provided by listeners when they are given the immediate phonetic context. Finally, we found that the rate of improvement of recognition accuracy may be used as a terminating criterion for training, and that reasonable performance can be achieved using as few as 800 training tokens.

We are presently investigating ways to improve recognition performance by incorporating additional acoustic attributes, including those that preserve temporal aspects of the signal. In addition, we want to understand how different sources of information are being used in the network. Finally, we plan to move on to the recognition of other classes of phonemes.

16.3 Speech Planning

We are pursuing the development of a speech production planning model in three areas, and are beginning to explore interactions among them. The first is word structure, the second is lexical stress, and the third is phrasal prosody.

1. Word Structure. Earlier experimental results showed the importance of word structure in production planning, by demonstrating that segmental speech errors are more likely to occur between word-onset consonants than between non-word-onset consonants. This is true even when the word-onset consonants occur before vowels with mixed stress levels, showing that the special susceptibility of word-onset segments is not due to the fact that in English word onsets are usually in stressed syllables. We have now extended these results from the original set of list stimuli ("parade fad foot parole") to a second set of list stimuli with a different metrical structure ("fad parade parole foot") and to phrasal stimuli, ("It's a fad to parade with parole on your foot") with nearly identical findings. Since the preferential occurrence of interaction errors among word-onset consonants is not diminished by changing the metrical structure of the stimuli or by introducing syntactic and prosodic structure, we conclude that it is a robust and reliable effect. Interestingly, the effect is even more pronounced for non-word stimuli like "pareed fid fet perile," suggesting that word structure imposes this effect not during the process of lexical access, but during the integration of the accessed segments into the planning frame.

2. Lexical Stress. Although word-onset consonants appear to be the strongest candidates for segmental interaction errors, prestressed consonants also participate in errors, although at a reduced rate. For example, in the pair of stimuli "parade fad foot parole" and "repeat fad foot repaid," /p/ and /f/ interact about twice as often in the first (word-onset) twister as in the second (prestressed), but both error rates are significantly higher than that for "ripple fad foot rapid," where the /p/ is neither in the word onset nor in prestressed position. This result suggests that lexical stress may also play a role in the planning frame.

3. Phrasal Prosody. Although list versus phrasal context did not affect the error pattern in the Word Structure experiment described above, an earlier pilot experiment had suggested that phrasal stimuli do have one significantly different effect: word-final consonants are protected against errors when the elicitation stimuli are phrases but not when they are lists. That is, when the four words of the list twister "peal tone pan tool" are placed in a phrasal context of "From the peal of the tone to the pan of the tool," the number of /l/-/n/ (final-position) errors is reduced significantly from that of the number of /p/-/t/ (initial position) errors. However, these pilot results were obtained in a between-subject experimental design, raising the possibility that the two sets of speakers were differentially susceptible to the effects of list versus phrasal contexts. A replication of the experiment, using within-subject design, has now shown the same effect. We are currently pursuing the question of whether this final-position protection effect in phrases is associated: 1) with the computation of syntactic structure for these phrases; or 2) with the computation of metrical structure to support their more complex prosody.

4. Interactions: Lexical vs. Phrasal Prominence. In a separate experimental domain, we are following up on the hypothesis that the pitch gestures traditionally associated with lexical stress are actually associated with the pitch accents of phrase-level intonation patterns. In measurements made on a single speaker, when the pitch accent is moved off of a target word (by any one of a number of means), the lexically-stressed syllable of that word exhibits no residual pitch gesture. If confirmed in a number of speakers, this finding would support the view that lexical stress or prominence is a rhythmic or durational phenomenon.

16.4 Studies of the Acoustics and Perception of Speech Sounds

16.4.1 Stops, Fricative, and Affricate Consonants

As one component of our study of the acoustics and perception of a variety of speech sounds, we are examining the mechanisms of sound generation during the constricted interval and at the release for stop, affricate, and fricative consonants. This research includes a theoretical component in which the events at the implosion, during the constricted interval, and at the release of these consonants are calculated through the analysis of models of the processes. The analysis includes: 1) the effects of intraoral pressure on the release of the closure for stop consonants and on the shaping of the constriction for fricative consonants; 2) the role of the yielding walls of the vocal tract in determining airflows during the constricted interval and following the release; 3) the effects of obstacles in the airflow on the generation of turbulence noise; 4) determination of the amplitude and spectral characteristics of the initial transient at the release of stop consonants; and 5) calculation of formant trajectories near the consonantal implosion and release. Comparison of acoustic data for the consonants with predictions of the model show reasonable agreement. Analysis of the model is hampered, however, by the lack of quantitative data on the impedance of the vocal-tract walls at low frequencies and on the rates of movement of articulatory structures at the instant of release of stop consonants.

One outcome of the theoretical and acoustical study of the release of stop and affricate consonants is the finding that the initial transient has an amplitude that is often comparable to or exceeds that of the following noise burst. This transient is expected, therefore, to contribute significantly to the perception of the place and manner of articulation for these consonants. A perceptual experiment examining the contrast between fricative and affricate consonants in English has been carried out, and has indeed shown that the initial transient is an important contributor to the perception of the affricate-fricative distinction.

16.4.2 Vowel Perception

We have been continuing our studies of the perception of vowels in consonant contexts in which the formants undergo rising-falling-rising trajectories (i.e., concave downwards) or falling-rising-falling trajectories (concave upwards). Subjects adjusted the formants of a steady-state vowel so that this vowel matched the quality of various vowels with time-varying trajectories. The subjects did not always show perceptual compensation for undershoot in the second-formant trajectory, as might be expected based on the results of earlier experiments. That is, the subjects did not always adjust the second formant of the matching stimulus to be beyond the extreme maximum or minimum point in the trajectory of the time-varying second formant. This lack of perceptual overshoot was evident in concave downward trajectories. Studies of these effects are continuing in an effort to separate out effects that might be auditorily based from effects that are based on linguistic experience.

16.4.3 Voicing for Fricatives

The maintenance of vocal-fold vibration during voiced fricatives is expected to be subject to considerable variability, since careful adjustment of the glottal and supraglottal constrictions is required to provide an intraoral pressure in the appropriate range. We have been examining the acoustic characteristics that distinguish voiced from voiceless fricative consonants when they occur in a variety of phonetic contexts, including sequences of two fricatives that may or may not agree in voicing. The observations verify that it is quite common for voiced fricatives to be produced with vocal-fold vibration over only a portion of the time interval in which the vocal tract is constricted. The distinction between voiced and voiceless fricatives is usually carried, however, by the presence or absence of significant vocal-fold vibration over a 20-30 millisecond time interval immediately adjacent to the preceding or following sonorant interval. Preliminary listening tests with synthetic fricatives in intervocalic position have verified that glottal vibration with these temporal characteristics leads to appropriate listener identification of the fricatives as voiced or voiceless.

16.4.4 Cross-Language Study of Nasalization

In collaboration with the Linguistics Center of the University of Lisbon, we have been conducting a comparative study of the implementation of nasalization in the three languages: English, French, and Portuguese. In particular, we have examined nasalization in utterances where the nasal interval is followed by a stop consonant (as in the words *banter* in English, *tante* in French, or *tinto* in Portuguese). In a series of perceptual experiments with synthetic utterances of the form *tante*, the following parameters were

systematically manipulated: 1) the duration of nasalization in the vowel; 2) the amount of nasalization in the vowel; and 3) the duration of the nasal murmur following the vowel.

The stimuli were presented to native speakers of Portuguese, English and French, which differ with respect to the occurence of nasal vowels in their phonological systems. The listeners were asked to judge, for each stimulus: 1) the presence or absence of nasalization; and 2) the adequacy of the stimulus as a natural utterance with respect to its nasalization.

The different language groups gave similar responses with regard to the presence or absence of nasalization. However, judgments of the naturalness of the stimuli in the different languages depend on the temporal characteristics. French listeners preferred a longer duration of nasalization in the vowel and were indifferent to the presence of murmur, English listeners preferred some murmur and accepted briefer nasalization in the vowel, and Portuguese responses were intermediate. Preliminary acoustic data from utterances in the three languages are in accord with these perceptual findings.

16.5 Physiology of Speech Production

16.5.1 Articulatory Movement Transduction

Work has continued on refinement of techniques for transduction of articulatory movements with the use of alternating magnetic fields. This effort has concentrated on exploration of a system which uses three magnetic-field transmitters as an alternative to the already-completed two-transmitter system (see RLE Technical Report No. 512). A successful three-transmitter system is potentially more desirable because: 1) it would simplify the complicated and time-consuming protocol for running experiments; 2) the data might be more reliable; 3) its transducer/receivers are much less expensive; and 4) it uses field strengths which are several times lower than the two-transmitter system. The last factor has potential implications for the use of humans as experimental subjects, because a very recent epidemiological study has found an increased incidence of cancer in children who live in proximity to magnetic fields generated by high-tension power lines.

The function of both the two- and three-transmitter systems has been simulated. The simulations predict: the characteristics of fields generated by solenoidal transmitters, the signal strengths induced in the transducer/receivers, and the resulting transduced displacements. An important requirement of the device is that it transduce displacements with sufficient accuracy in the face of a moderate amount of receiver tilt, and with receivers mounted slightly lateral to the midsagittal plane (because it is impossible assure absolutely midsagittal transducer placement). Use of the simulation suggests that a three-transmitter system could perform adequately only if the transducers are mounted within about 2 mm of the midline of the transmitter assembly. The simulation predicts that the two-transmitter system allows for more off-midline displacement error, but at the expense of having to calibrate each transducer separately and mount it on the subject with the same orientation in which it was calibrated. The simulations have also been used to investigate and specify the transmitter characteristics

and spacing for new versions of both designs. These new versions have been built, along with additional testing and calibration devices, and final testing is in progress.

16.5.2 A Quantitative Study of Anticipatory Coarticulation

Analysis is almost complete on lip-protrusion movement data for the vowel /u/ from four speakers of American English. Test utterances consisted of pairs of words such as "leak hoot" and "leaked coot," each pair embedded in a carrier phrase. Word pairs were chosen so that different numbers of consonants intervene between the unrounded vowel /i/ and the rounded /u/. Utterances were repeated 15-20 times in random order. Protrusion movements of the upper lip and the acoustic signal were recorded, digitized and processed. Times of end of voicing of the vowel /i/ and onset of voicing of the vowel /u/ were marked interactively in the acoustic signal stream. Times of protrusion onset (onset of positive velocity) and of maximum acceleration were identified algorithmically in the movement signal stream.

Consistent with previous qualitative observations on one subject, there was a lot of variation in the relative timing of the identified events. Lip protrusion movements were partitioned into two components: an initial slow one, with onset of positive velocity occurring around the end of the preceding unrounded /i/ (possibly corresponding to relaxation of spreading for the /i/), and a later faster one, with its onset marked by the largest acceleration peak, occurring after the end of the /i/.

Quantitative analysis shows that as the duration of the acoustic interval between the end of the /i/ and the onset of the /u/ increased (with increasing numbers of intervocalic consonants), the interval between the onset of the slow movement component and onset of the /u/ also increased, but at a lesser rate. Thus, for the shortest intervocalic intervals the protrusion onset occurred slightly earlier than the end of the /i/, and for the longest intervocalic intervals, the initial protrusion onset occurred slightly later than the end of the /i/. This finding suggests that the time of the initial protrusion onset is somewhat constrained by the end of the acoustic requirement of lip spreading for the /i/, but that constraint interacts with a competing tendency to produce the protrusion gesture at a relatively invariant, (biomechanically) optimal velocity. The analysis also suggests that when the consonant string contains a /k/, the movement onset is slightly delayed in comparison to utterances not containing /k/. It is speculated that this effect is due to the coordination of the lip and tongue body gestures for the vowel /u/, in which lip protrusion onset "waits" for completion of the tongue body gesture for the /k/. Further experimentation is required to explore these ideas.

16.5.3 Glottal Airflow and Pressure Measurements for Male and Female Speakers with Normal Voices

National Institutes of Health (Grant 5 R01-NS21183), subcontract with Boston University

Measurements on the inverse filtered airflow waveform (the "glottal waveform") and of estimated average transglottal air pressure and glottal airflow were made from non-invasive recordings of productions of syllable sequences in soft, normal and loud voice and in normal, low and high pitch for 25 male and 20 female speakers. Analysis of the

results for intensity manipulation is complete and shows a large amount of inter-subject variation in most parameters. There was an almost universal DC offset in the glottal airflow waveform, indicating incomplete closure of the inter-arytenoid portion of the glottis. With change from normal to loud voice, both males and females produced loud voice with increased pressure, accompanied by increased AC flow and increased rate of change of airflow during the glottal closing phase (called "closing velocity"). Soft voice was produced with decreased pressure, decreased AC flow and closing velocity and increased DC and average flow. Within the loudness conditions, there was no significant male-female difference in pressure. Several glottal waveform parameters separated males from females in normal and loud voice. The waveforms evidenced higher AC flow and higher closing velocity for males. In soft voice, the male and female glottal waveforms were more alike, and there was no significant difference in closing velocity. The DC flow did not differ significantly between males and females in all three loudness conditions. Because of the DC flow component, previously-employed measures of glottal resistance and vocal efficiency may be less useful indicators of vocal-fold function than other measures. Most of the findings may be related to biomechanical differences and differences in voice source acoustic characteristics between males and females and across loudness conditions.

Analysis of the same measures for normal, low and high F0 is almost complete. It shows increased pressure for change both from normal to high and normal to low pitch. The unexpected finding for low pitch may be due to the nature of the task, in which low pitch is below the normal speaking range. A lack of correlations of the measured parameters with F0 change and strong correlations with change in sound pressure level (SPL) suggest that the measured aerodynamic parameters are more closely related to mechanisms which determine SPL than F0.

16.5.4 Objective Assesment of Vocal Hyperfunction: An Experimental Framework and Preliminary Results

National Institutes of Health (Grant R01-NS21183), subcontract with Boston University

This study is part of the initial phase of a project which focuses on the development and use of quantitative measures to provide objective descriptions of conditions called "vocal hyperfunction." Some of these conditions can be manifested in the form of vocal-fold edema, nodules, polyps, and contact ulcers. They are accompanied by acoustic abnormalities and are believed to be due to abuse and/or misuse of the vocal apparatus. More advanced cases may require surgical treatment. Experimental design for the project is based on a descriptive theoretical framework which maintains that there are different types and stages of hyperfunctionally-related voice disorders. Data consist of indirect measures made from non-invasive aerodynamic and acoustic recordings, including: 1) parameters derived from inverse filtered approximations of the glottal volume velocity waveform; 2) estimates of transglottal air pressure, average glottal airflow, glottal resistance and vocal efficiency; and 3) measures of vocal intensity and F0. Preliminary results (based on comparisons among 15 voice patients and the above-mentioned 45 normal speakers) support major aspects of the theoretical framework and indicate that the measurement approach is capable of differentiating hyperfunctional from normal voices and different hyperfunctional conditions from one another. Organic manifestations of vocal hyperfunction (nodules, polyps, contact

ulcers) are accompanied by abnormally high values of AC flow and maximum closing velocity, suggesting an increased potential for vocal fold trauma due to high collision forces. In contrast, non-organic manifestations of vocal hyperfunction (called "functional disorders") tend to be associated with abnormally high levels of unmodulated DC flow, without high values of AC flow and maximum closing velocity, suggesting reduced potential for vocal fold trauma. The measures also suggest different underlying mechanisms for nodules and polyps vs. contact ulcers. The results are interpreted with respect to predictions based on the theoretical framework. These interpretations have led to refinement of the experimental approach for use in the next phase of the work. Currently, completed measures from an additional 15 voice patients are being examined.

16.5.5 The Speech of Cochlear Implant Recipients

National Institutes of Health (Grant 1 P01-NS23734), subcontract with the Massachusetts Eye and Ear Infirmary

Work has begun on studying the speech of post-lingually-deafened cochlear implant recipients as part of an NIH Program Project entitled "Basic and Applied Research on Cochlear Prostheses" (J.B. Nadol, MD, Principal Investigator). The aims of this component project are to: 1) characterize the speech of post-lingually deafened adults before and after they receive cochlear implants to help evaluate and improve prostheses; and 2) understand the role of hearing (auditory feedback) in speech production. These aims will be met by making: 1) perceptually-based measurements of sound segment deletions and substitutions and ratings of prosodic variables and overall intelligibility; 2) acoustically based measures of segmental productions (temporal and spectral) and prosody (F0 and temporal variables); and 3) physiological measures of respiration and air flow regulation and articulatory coordination. The measures will be compared within each implanted subject across time and between aided and unaided conditions, as well as between implanted subjects and normal-hearing controls. The measures are derived from recordings of the acoustic signal, oral and nasal air flow, nasal and throat vibrations (for a nasality index), an electroglottographic signal and, chest and abdominal cross-sectional areas (for a measurement of lung volume). Thus far, experimental protocols have been established, longitudinal recordings have been made of four implant patients, individual recordings have been made of several controls, initial signal processing and data extraction procedures have been devised, and preliminary measures of respiration have been examined longitudinally for two implant recipients. Analysis will be resumed and expanded when a new data processing facility (currently being completed) comes on line.