

XXIII. SPEECH COMMUNICATION

Academic and Research Staff

Prof. K.N. Stevens	Dr. A.W.F. Huggins‡‡	Dr. R.N. Ohde
Prof. M. Halle	Dr. D. Isenberg	Dr. J.S. Perkill
Prof. S.J. Keyser	Dr. T.E. Jacobs	Dr. E. Putney
Dr. S.E. Blumstein*	Dr. P.A. Keating	Dr. L. Santerre
Dr. W.E. Cooper+	Dr. D.H. Klatt	Dr. S. Shattuck-Hufnagle****
Dr. S. Ehrlich	Dr. M. Laferriere***	Dr. J.R. Westbury
Dr. F. Grosjean‡	Dr. J.I. Makhoul‡‡	Dr. V.W. Zue
Dr. J.A. Gurlekian**	Dr. L. Menn+++	S. Boyce
Dr. C.S. Hawkins++	Dr. P. Menyuk‡‡‡	M. Danly
Dr. W.L. Henke	Dr. J. Miller‡	J.A. Jakimik

Graduate Students

T.T. Allard	B. Delgutte	D.K. Oka
C.A. Bickley	U.G. Goldstein	S. Seneff
M.A. Bush	H.L. Golub	C.H. Shadle
R.C. Concia	S.K. Holford	M.G. Stella
	B.M. Kinney	

* Associate Professor, Department of Linguistics, Brown University.

+ Associate Professor, Department of Psychology and Social Relations, Harvard University.

‡ Assistant Professor, Department of Psychology, Northeastern University.

** Assistant Professor, Consejo Nacional de Investigaciones Cientificas y Tecnicas, Buenos Aires, Argentina.

++ Assistant Professor of Communicative Disorders, Emerson College.

‡‡ Staff Member, Bolt Beranek and Newman, Inc.

*** Staff Member, Mitre Corporation, Bedford, Massachusetts.

+++ Research Associate, Aphasia Research Center, Boston University.

‡‡‡ Professor of Special Education, Boston University.

**** Assistant Professor, Department of Psychology, Cornell University.

(XXIII. SPEECH COMMUNICATION)

1. STUDIES OF SPEECH PRODUCTION AND PERCEPTION

National Institutes of Health (Grant 2 R01 NS04332 and
Training Grant 5 T32 NS07040)

C.J. LeBel Fellowships

Sheila E. Blumstein, Rodolfo Concia, Bertrand Delgutte, Jorge A. Gurlekian, Morris Halle, William L. Henke, David Isenberg, Samuel J. Keyser, Patricia A. Keating, Dennis H. Klatt, Ralph N. Ohde, Joseph S. Perkell, Stephanie Seneff, Christine H. Shadle, Stefanie Shattuck-Hufnagel, Kenneth N. Stevens, John R. Westbury, Victor W. Zue

a. Segmental Aspects of Speech

As part of our continuing series of studies on the perceptual, acoustic, and physiological bases for the phonetic features, we have been examining cross-linguistically the acoustic and articulatory attributes of stop and fricative consonants with different places of articulation, particularly those produced by raising the tongue blade -- the so-called coronal consonants. These studies during the past year have included: an investigation of the acoustic and articulatory correlates of the opposition between /s/ and /ʃ/ in English, and similar oppositions for fricatives in several other languages; a study of the acoustic differences that account for the contrast between dental and alveolar consonants that occurs in some languages; observations of the distinctive acoustic attributes for retroflex consonants; a perceptual study of certain acoustic cues for the identification of the Argentine Spanish fricatives /s/ and /f/; a detailed acoustic study of the characteristics of the frication burst, the aspiration noise, and the onset of voicing for voiceless stop consonants in English; and development of a mechano-acoustic model of the aerodynamic events immediately following release of stop consonants, in order to explain in detail the nature of the acoustic events at the release for different places of articulation.

b. Peripheral Constraints on Speech Timing for Vowels

In this study we have examined the temporal and amplitude characteristics of electromyographic activity in the jaw-lowering musculature relative to the durations of vowels whose articulations differ in displacement of that structure. The

principal concern of this work – involving recording of vertical movements of the mandible via a strain gauge, and electromyographic recordings from the anterior belly of digastric via intramuscular hooked wires – is to discover the interrelationships between strength and duration of muscle activity, duration and magnitude of articulatory displacement, and the acoustic duration of speech segments. It is anticipated that this research will provide insight into principles which govern the control of time in the speech signal.

c. Motor Control During Speech: Implications from a Study of Respiration

In collaboration with Dr. P. Schonle (University Hospital, Tübingen), we have conducted an investigation of the temporal and amplitude characteristics of electromyographic activity of certain of the respiratory muscles relative to abrupt speaker-generated variations in upper airway resistance. The focus of this work is to determine whether the inspiratory and/or expiratory muscles exhibit either anticipatory or short-latency (compensatory) activity relative to aerodynamic loadings and unloadings which accompany the articulation of individual speech sounds.

d. A Reaction Time Study of the Production of /s/ and /š/

This study is designed to reveal dynamics of preparation and movement in the articulation of fricative consonants following vowels. Measures were obtained of the time to initiate and to complete articulation of /s/ and /š/ following a go-signal during a steady vowel. Systematic differences in reaction time were obtained for different fricatives and vowels, and these differences appear to be linked to "activation time of the motor programs" for the two fricatives rather than to movement difficulty per se.

e. Phonological Structure of the Mental Lexicon

Previously published data on spontaneous speech errors have been examined in order to speculate on the representational form of words in the mental lexicon that is accessed during speech production. The analysis suggests that words are represented in terms of morphemes, which are, in turn, represented in terms of phonemes.

(XXIII. SPEECH COMMUNICATION)

There is little evidence in the error corpus for independently moveable distinctive features or syllables as phonologically real processing units, although the concept of a syllable does play a useful role in specifying positional constraints on phonetic movement in an error. Words appear to be processed as if they were composed of linear sequences of phonemes.

f. Toward a Measure of Phonetic Similarity

As part of our effort to develop a new model of speech perception and lexical access, we are developing a metric for comparing an input spectrum with a spectral template, independently of the nature of the speech spectra to be compared. A set of 66 vowels acoustically similar to /æ/ were synthesized by adding together sinusoidal harmonics, and these vowels were presented to subjects in separate tests of psychophysical distance and phonetic distance. This work is continuing with additional types of stimuli, and preliminary results suggest that a metric can be developed without explicitly computing formant-frequency values.

g. Intrinsic Pitch of Vowels in Sentence Context

As a first step toward an understanding of how the intrinsic fundamental frequency (F_0) of vowels interacts with intonation in running speech, the fundamental frequency of the vowels /i,a,u/ was examined in a controlled sentence context. Results were obtained for four speakers at four locations in sentences matched for length and stress pattern. Vowel height was observed to influence fundamental frequency at all locations in the sentence except sentence-final position. Various hypotheses have been proposed and additional acoustic measurements have been made in order to explain these results in terms of acoustic and/or physiological interactions between supralaryngeal and laryngeal activity.

h. Study of the Phonological Processes in American English

We are continuing our research that seeks to obtain quantitative information on the variation of the properties of speech sounds in context. As a part of this program, we have been examining various palatalization effects that occur across

word boundaries, including the palatalization of word-final /s/ or /z/ followed by /y/. These and other sentence-level effects are being measured from a large corpus of data, and are being organized into a set of rules that describe the effects in a systematic way. One source of new insights into these rules is our continuing project on spectrogram reading, in which members of our research group are improving the speed and accuracy with which they can interpret phonetic information in spectrograms.

i. Representation of Speechlike Sounds in the Discharge Patterns of Auditory Nerve Fibers

We are collaborating in this work with the Eaton-Peabody Laboratory of Auditory Physiology, Boston, Mass. In order to study how acoustic characteristics important for speech discrimination can be coded at the level of the auditory nerve, responses of auditory-nerve fibers to speechlike stimuli were recorded in anesthetized cats. Results demonstrate that a knowledge of response properties for simple stimuli is useful to interpret responses to speechlike sounds.

Response patterns of single fibers show a peak in discharge rate just after the onset of a stimulus with a rapid onset. This peak is more prominent for a /č̣/-like noise burst (with an abrupt onset) than for a /ṣ̌/-like burst (with a gradual onset). Following the offset of a stimulus to which a unit responds, there is a recovery period during which the unit responds at a lower discharge rate to test stimuli. This effect (analogous to forward masking) was demonstrated for a synthetic /ma/ stimulus: During the formant transitions following the low-frequency nasal murmur, fibers with low characteristic frequencies (i.e., those that responded to /m/) discharged at a lower rate than they did for a /ba/ stimulus with identical formant transitions. The coding of fundamental frequency was investigated using single-formant synthetic stimuli. At sound levels typical of speech, information about fundamental frequency is present as periodic fluctuations in the short-time average discharge rate of units over a wide range of characteristic frequencies (CF). Experiments with /s/ and /ṣ̌/-like noise bursts suggest that there is information in the profile of average discharge rate vs CF that could be used to distinguish between the spectra of these sounds. The results of pilot

(XXIII. SPEECH COMMUNICATION)

experiments on responses to speechlike sounds presented in a continuous noise background suggest that the noise can affect more or less strongly the manner in which particular acoustic characteristics of speechlike sounds are represented.

j. A Speech Analysis-Synthesis System with Separable Models for Spectrum Envelope and Excitation

A speech analysis-synthesis system has been developed which separates the speech waveform into a model for the vocal-tract transfer function and an excitation model. Both components can then be modified independently, and speech can be reconstructed with a transformed spectrum and/or transformed pitch. The system does not make use of a simple excitation model that would require the extraction of a sequence of pulses spaced by the fundamental period, and therefore the reconstructed speech is more natural-sounding than typical vocoder speech. The system has potential applications in a number of different areas, such as speaker normalization for computer speech recognition, studies of female versus male voice quality differences, modification of speech rate, spectral modification of speech (as a potential aid to persons with residual low-frequency hearing), and implementation of a baseband-excited vocoder, in which the upper portions of the reconstructed excitation spectrum are generated from the lower portion.

k. Software for a Laboratory Speech Synthesizer

The software formant synthesizer that has been used in our laboratory for a number of years to generate stimuli for perceptual experiments has been fully documented and published as a set of Fortran programs. The theory of formant synthesis is presented and strategies for the imitation of a number of English speech sounds are presented in tabular form.

l. Refinement of Videofluorographic Techniques for Studying Articulation

As a continuation of work on reducing radiographic dosages for x-ray studies of speech production, an experiment was run to improve the resolution of pellet-tracking measurements using frame-by-frame analysis of videotaped fluoroscopy. A paraffin wedge was used to produce more uniform exposure density and a relatively

large lead pellet was affixed to the tongue. Accuracy of resulting plots of pellet movement was found to approach the accuracy obtained from cineradiographs with a dosage reduction of a factor of 40 as compared to conventional 35-mm cineradiography.

m. A Magnetic-Field Ranging Device for Studying Midsagittal Plane Movements of Points inside the Vocal Tract

A 3-channel magnetic-field ranging device has been constructed incorporating previously used principles. Two large transmitting coils, each excited by a sinusoidal signal at a different frequency, are mounted on a helmet above the forehead and below the occiput. Extremely small transducer-receivers consist of inductors mounted on a base with lead wires attached. The transducers are attached to the structures of interest and leads are brought out of the corner of the mouth. The magnetically induced signal in each transducer is demodulated to produce voltages which are related to distances between transmitters and transducer, allowing for spatial localization by triangulation. In spite of technical problems which remain to be resolved, the potential usefulness of such a device has been demonstrated in a test on 3 subjects of a hypothesis about the nature of articulatory targets for point vowels.

n. Speech Synthesis by Rule

The MITalk-79 text-to-speech system that has been under development at M.I.T. for a number of years under the direction of Professor Jonathan Allen has been fully documented as a monograph. The monograph includes chapters by Jonathan Allen describing an overview of the system and its potential future as a practical device; chapters by Sheri Hunnicutt on morphemic decomposition, letter-to-sound conversion, partial syntactic analysis, and fundamental-frequency contour generation; chapters by Dennis H. Klatt on prediction of segmental durations, phonological recoding, phonetic-to-parametric conversion, and the theory of formant synthesis; and a chapter by David B. Pisoni on a detailed evaluation of the intelligibility and comprehensibility of words, sentences and paragraphs produced fully automatically from English text. The MITalk-79 system has been frozen as

(XXIII. SPEECH COMMUNICATION)

a set of BCL programs that can be obtained by license agreement with the M.I.T. Patent Office.

2. STUDIES OF SPEECH PRODUCTION BY CHILDREN AND DISORDERS
OF SPEECH PRODUCTION

National Institutes of Health (Grant 2 R01 NS04332 and
Training Grant 5 T32 NS07040)

National Science Foundation (Grant BNS77-26871)

C.J. LeBel Fellowship

Suzanne Boyce, Marcia A. Bush, Ursula G. Goldstein, Howard L. Golub,
William L. Henke, Lise Menn, Kenneth N. Stevens

a. Modeling the Effects of Growth on Vowel Production

A static articulatory model incorporating growth parameters has been developed to study the effects of anatomy on women's and children's production of vowels. The model is characterized by articulatory variables specifying a number of parameters influencing vocal-tract shape, including the position of a circular tongue body relative to the angle of the jaw. The sizes of parts of the vocal tract are specified by growth curves which express anatomical distances as a function of age. Vocal-tract length and cross sections in the midsagittal plane are computed and used as the basis for determination of area functions and formant frequencies. The model is used to compute formant frequencies when age and sex of the speaker are given, along with positions of the articulators appropriate for particular vowels. The next step in this research is to examine the formant frequencies for vowels produced by the model for children, women, and men, and to compare these model-generated data with data from real speech, in an effort to explain the non-uniform scaling that has been observed in the formant frequencies for different vowels.

b. Diagnostic Use of the Infant Cry

We have attempted to advance our understanding of the diagnostic value of the infant cry by extending past work in two directions. First, we have developed a

model of cry production that will enable the observer to relate more closely the acoustic properties of the cry to the anatomical and physiological characteristics of the infant producing the cry. Second, we have refined the procedures for extracting parameters from the sound, for deriving features from these parameters, and for statistical analysis of the data by making use of the computer-based signal processing and data-manipulation capabilities that have been developed in recent years by speech researchers. Recently, we have completed a pilot study analyzing the cry of about 50 normal infants and about 50 infants with several different abnormalities. Based on the model of cry production and on statistical analysis of the cry data, eight diagnostic tests were designed by combining a number of acoustic features extracted from the cry. The cry for infants with medical abnormalities shows certain characteristics that deviate from those for normal infants, depending on the type of abnormality. These deviant characteristics include acoustic evidence that indicates respiratory problems, instability of laryngeal control, and constricted vocal tract. Although the number of infants in the different abnormal groups is small, there are indications that, with judicious selection of features and diagnostic tests, infants with particular abnormalities can be identified with a reasonable probability of success, and a relatively small false alarm rate.

c. Pitch and Marked Voice Quality in Parent-Child Discourse: Acoustics and Semantics

We have completed the acoustic analysis of recordings from 16 parent-child conversations. The children were aged 2 to 5, and the conversations took place in a semistructured laboratory playroom situation. We are proceeding with an analysis and interpretation of the data. In particular, we are examining the pitch contours during these interactions to determine the extent to which regularities in pitch ranges occur in successive clauses (within and across speakers) as the semantic focus of the discourse is maintained or as it shifts.

d. Control of Fundamental Frequency by Profoundly Deaf Speakers

The general goal of this study is to determine by an acoustic analysis the relationship between segmental articulation and laryngeal control by profoundly

(XXIII. SPEECH COMMUNICATION)

deaf boys and girls. In particular, we are trying to discover whether some of the exaggerated variations in voice fundamental frequency (F0) produced by deaf speakers are a consequence of inappropriate articulatory maneuvers and/or laryngeal postures used in vowel and consonant production. The results of preliminary analyses indicate that an excessively high F0 is often associated with the deaf speaker's production of high vowels and of vowels following voiceless consonants. For some deaf boys and girls, these deviations in F0 appear to be correlated with the use of "extreme" articulatory strategies (e.g., excessive aspiration, exaggerated formant (tongue/jaw) movement). Implications of these findings for the development of speech-training procedures are also being considered.

e. Acoustic Study of the Effects of Intubation on Laryngeal Function

In collaboration with the Anesthesiology Department at Beth Israel Hospital, we have conducted a study to determine the effects of intubation on laryngeal function. Recordings of vocalizations of patients were made before intubation (preoperative) and postoperatively at various times after intubation. Acoustical analyses of these recordings were performed to seek evidence for changes in the vocal-fold vibration pattern that might be the result of temporary or permanent influences on the vocal folds as a consequence of intubation. In most patients who have been intubated for several hours, there was evidence for changes in the vocal-fold vibration pattern, and for gradual recovery to the preoperative condition within a few days.