XX. SPEECH COMMUNICATION

Academic and Research Staff

Prof. Kenneth N. Stevens
Prof. Morris Halle
Prof. Samuel J. Keyser
Dr. Jared Bernstein
Dr. Sheila Blumstein*
Dr. Margaret Bullowa
Dr. William E. Cooper†
Dr. Louis M. Goldstein
Dr. Francois Grosjean‡
Dr. Jorge A. Gurlekian**
Dr. William L. Henke
Dr. A. W. F. Huggins††
Dr. Margaret Kahn
Dr. Dennis H. Klatt
Dr. Martha Laferriere‡‡
Dr. John Makhoul††
Dr. Lise Menn***
Dr. Paula Menyuk†††
Dr. Joanne Miller‡
Dr. Barbara J. Moslin
Dr. Ralph N. Ohde
Dr. Colin Painter‡‡‡
Dr. Joseph S. Perkell
Dr. David B. Pisoni****
Dr. Stefanie Shattuck-Hufnagel††††
Dr. John R. Westbury
Dr. Katherine Lee Williams‡‡‡‡
Dr. Victor W. Zue
Jola A. Jakimik
John M. Sorensen

Graduate Students

Marcia A. Bush
David W. Chidakel
Rodolfo C. Concia
Bertrand Delgutte
Ursula G. Goldstein
Howard L. Golub
Stephen K. Holford
Brian M. Kinney
Peter V. LaMaster
David K. Oka
Afarin Ordubadi
Stephanie Seneff
Christine H. Shadle
Michael G. Stella

---

*Associate Professor, Department of Linguistics, Brown University.

†Assistant 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.

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

‡‡Assistant Professor of Linguistics, Southeastern Massachusetts University.

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

†††Professor of Special Education, Boston University.

‡‡‡Associate Professor of Communicative Disorders, Emerson College.

****Professor, Department of Psychology, Indiana University.

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

‡‡‡‡Research Associate, Department of Psychology, University of Denver.

1. STUDIES OF SPEECH PRODUCTION AND PERCEPTION

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

C. J. LeBel Fellowships

Sheila E. Blumstein, Bertrand Delgutte, Morris Halle, William L. Henke, Samuel J. Keyser, Dennis H. Klatt, Ralph N. Ohde, Colin Painter, Joseph S. Perkell, David B. Pisoni, Kenneth N. Stevens, Victor W. Zue

a. Segmental Aspects of Speech

Our research on the segmental aspects of speech is examining the acoustic, perceptual, and articulatory correlates of the phonetic features that classify speech sounds in language. The objectives of the work are to determine how the human perceptual and articulatory systems place constraints on the selection of an inventory of phonetic features, and to utilize evidence from speech acoustics, auditory perception, speech production, and phonology to work toward a revised inventory of features.

We have completed several studies of the acoustical properties and of the perception of place of articulation for stop consonants in English, and these studies have suggested that the listener samples the speech signal in the vicinity of points in time when there is a rapid change in the spectrum, and classifies the sound in these regions in terms of certain gross characteristics of the short-time spectrum. Further work in this area is investigating the acoustic correlates and the phonological evidence for features that differentiate among coronal consonants (i. e., consonants produced with the tongue blade) that have been traditionally classified in terms of different places of articulation in different languages. We have also begun to examine the acoustic bases for the categorization of consonants as voiced or voiceless, with the aim of finding an integrated acoustic property that applies over a variety of phonetic contexts and consonantal manners of articulation.

Research on the laryngeal features is continuing with a laryngeal fiberscope study of what the vocal folds are doing during the production of the following set of consonants:

$\underset{\circ}{b}$, p^h, p, b, $p^?$, ph, b^{η}, ?b, ɓ, kp, gb, p'.

Each of these consonants is recorded in carrier sentences before the eight primary cardinal vowels. This is a follow-up on an acoustic study using the same data set.

As a continuation of our study of the articulatory correlates of certain vowel features, we have run two preliminary palatographic experiments to explore the notion that patterns of tongue-to-maxilla contact are invariant correlates of "vowel height" features. Results from six speakers with differently-shaped palatal vaults suggest that such invariant patterns may not exist on the palatal surfaces and that further study is necessary

to determine whether there is some other basis for postulating invariant articulatory correlates for the vowel height features.

b. Acoustic Study of the F_0 Contours of Cantonese Tone

As part of our research directed toward a better understanding of the behavior of the larynx and the features underlying it, we have examined the fundamental-frequency (F_0) contours of the nine Cantonese tones. The data were collected from two native speakers of the Canton dialect. The corpus was designed such that the effects of sentence intonation were minimized and modifications of tone contours by adjacent consonants were counterbalanced. In addition to the F_0 contours of the nine tones, durational data were also obtained. From the production data that we have collected, we were able to obtain average F_0 contours for all the tones. We are currently in the process of synthesizing simple consonant-vowel syllables with the averaged tone contours for perceptual experiments. One aim of the study is to gain some insight into whether these contours are perceived and produced in a quantal manner.

c. Study of the Phonological Processes in American English

The goal of this part of our research is to provide, through acoustic studies, quantitative information on the variation of the properties of speech sounds in context. Whenever the variations appear to be systematic, either for all speakers or for a subset of the speakers, rules are proposed to describe such phonological variations. Over the past year we have conducted a study of the acoustic effect of a phonological process commonly known as palatalization. In particular, we investigated the effect of palatal consonants (/$\overset{\vee}{s}$, $\overset{\vee}{z}$, y/) on the adjacent alveolar fricatives (/s, z/). Our results indicate that the palatalization of alveolar fricatives occurs much more readily and completely when the palatal consonant precedes the alveolar fricative (e.g., this ship, gas shortage) than when it follows the alveolar fricative (e.g., Irish setter, tunafish sandwich). The observed difference in acoustic data can be accounted for by several hypotheses, ranging from explanations that are based on the relationship between anticipatory and perseveratory articulation to those that are more motivated by considerations of the underlying articulatory constraints. We are currently exploring these hypotheses by examining additional acoustic as well as physiological data.

d. Experiments on Spectrogram Reading

Several spectrogram reading experiments were conducted over the past year. The experiments were designed to determine the amount of phonetic information that is contained in the speech signal or, more specifically, in a spectrographic representation of the speech signal. The task involved identifying the phonetic content of an unknown utterance only from a visual examination of the speech spectrogram. In the first experiment,

the subject attempted to phonetically label spectrograms of normal and anomalous English utterances as well as words in a known carrier phrase. The results, when compared with the transcriptions of three phoneticians who listened to the utterances, indicated an overall agreement of better than 85% on the sentences, and 93% for words in a carrier phrase. Subsequent experiments were designed to see how fast spectrogram reading can be accomplished and to what extent such a procedure can be taught. The general conclusions from these experiments were that there exists a great deal of phonetic information in the acoustic signal, and that a spectrographic display captures a substantial amount of such information. Furthermore, spectrogram reading is often based on the application of explicit rules, and thus can be taught efficiently. Preliminary data from these experiments also suggest that spectrogram reading can probably be done in 30-40 times real-time.

e. A New Model of Lexical Access

During speech perception, how does the perceptual apparatus generate lexical hypotheses "bottom-up" from direct analysis of the input acoustic waveform? The conventional point of view is that the process proceeds logically in stages, where the peripheral auditory system first performs a spectral analysis of the input, and then a set of property detectors extracts relevant properties from this neural spectrogram and generates a phonetic transcription (in the form of a feature matrix in which the columns denote segments and the rows denote distinctive features). The phonetic representation may be errorful and incomplete, but it forms the basis for a search of the lexicon for candidate word hypotheses through an analysis-by-synthesis procedure.

Examination of the strategies employed in the construction of several computer-based speech-understanding systems has suggested that the analysis-by-synthesis model is seriously suboptimal in a computational sense. The knowledge contained in the verification component should be applied earlier in the recognition process so that phonological and lexical constraints of the language can be used to reduce the alternatives and thus reduce errors. All of the knowledge that is embodied in the generative rules of the verification component can be precompiled into a representation that is ideal for direct bottom-up lexical hypothesis formation without post-verification. Phonetic segmentation and labeling decisions need not be made during lexical search, since the decoding network can be made to represent directly the acoustic manifestations of words and word sequences. We intend to pursue these theoretical arguments and ask whether the speech perception apparatus might have evolved in such a way as to take advantage of these strategies. The first objective will be to program a computer simulation of these algorithms in order to establish their potential for accurate decoding of acoustic data. These efforts should lead to refinements in the theory and perhaps lead to new kinds of perceptual tests of competing theories of speech perception.

f. Prediction of Segmental Duration in English Sentences

A set of rules has been developed for the prediction of segmental durations in any English sentence. The input representation for a sentence is a string of symbols drawn from an inventory of 52 phonemes, three alternative stress markers, morpheme boundary, word boundary, and eight syntactic structure distinctions. Eleven durational rules are applied to predict acoustic durations (in ms) of phonetic segments derived from this abstract representation. The rules are intended to quantify many of the larger rule-governed changes in duration that are associated with syntactic environment, segmental position within a word, stress, and phonetic context. The effects of different rules are combined multiplicatively (subject to an incompressibility constraint). The rule system is offered as a first start toward more sophisticated and powerful algorithms.

An objective evaluation of the rules has been performed in which durational predictions have been compared with durations measured in new paragraphs read by the author. Results indicate that the rules account for 84% of the variance in measured segmental durations. Perceptual evaluation of speech synthesized using rule-governed durations indicates that both naturalness ratings and intelligibility of sentences synthesized using these rules are comparable to results using sentences synthesized with durations obtained from a natural recording.

g. Perceptual Interpretation of Durational Cues

The concept of a lexically-based perceptual strategy has been extended to the interpretation of durational cues. Many factors influence the duration of phonetic segments in an utterance. How is it possible to determine whether a particular segment has been lengthened due to syntactic, stress, or phonetic factors? We argue that the answer lies in the specification of expected durations for segments in the representation for each word of the lexicon. Then the lexical search strategy can include durational criteria to select among lexical candidates without unraveling phonetic causality, and, simultaneously, it can look for certain kinds of segmental lengthening and shortening patterns (relative to the word under consideration) that indicate particular syntactic structures. We plan to look in detail at the kinds of rule-durational systems proposed for English to determine the relative advantages of this viewpoint.

h. Analysis of Speech Error Data

A corpus of over 8000 speech errors collected by Merrill Garrett and Stefanie Shattuck-Hufnagel has been examined for evidence concerning the active use of distinctive features and markedness concepts during early stages of the speech production process. The results of our analyses support the view that, when segmental speech errors occur, individual distinctive feature values of segments rarely, if ever, move

about independently. It is entire phonetic segments that move.

Our analyses also indicate that markedness plays no role in the determination of which intrusion consonants are selected in an error. There is no measurable response bias toward a favored set of intrusion consonants. The relative frequency with which a consonant functions as an intrusion in a segmental speech error is statistically indistinguishable from the relative frequency with which it participates as the intended segment. In addition, error frequencies for different consonants are highly correlated with frequency of occurrence in content words in English, suggesting that errors are based on confusions with segments in similar positions in other words which are being manipulated during sentence planning.

Finally, there are more palatalization errors involving alveolar obstruents, particularly [s] → [š], [t] → [č], than one might expect by chance. These errors appear to be caused by the misapplication of a familiar palatalization rule of English, because normal application of the rule (to, for example, the [s] of "this shirt") would not be counted as an error.

i. Electrophysiological Investigation of Peripheral Processes in Speech Perception

Electrophysiological studies of the coding of speechlike sounds in the auditory nerve have been conducted in collaboration with the Eaton-Peabody Laboratory of Auditory Physiology. Single-unit recordings from the auditory nerve of anesthetized cats were obtained using the methods described in 1965 by Kiang et al.

Tone-burst stimuli were used to study the characteristics of short-term adaptation. Over a wide range of stimulus conditions, the time course of firing rate after the onset of a tone burst can be described as a superposition of a rapidly and slowly decaying component with time constants of a few msec and a few tens of msec, respectively. These results are likely to be relevant to the coding of certain characteristics of the onsets that occur in speech (e.g., abruptness, voice-onset time).

Fibers which have been adapted by a tone burst of adequate frequency and level respond with a decreased firing rate to the onset of a second tone burst occurring up to 100-250 msec after the offset of the adapting stimulus. This "forward-masking" effect suggests that the spectral content of previous stimulation may affect the way the spectrum of an onset is coded in the distribution of firing rate across fibers.

Single-formant synthetic stimuli were used to study the coding of certain vowel characteristics. Information about formant frequency is present in the discharge pattern of fibers with a characteristic frequency close to the formant frequency. At levels typical of speech, information about fundamental frequency is present in the firing pattern of fibers over a wide range of characteristic frequencies.

j. Physiology of Speech Production

A preliminary experiment has been performed to test the idea that the articulators assume a "speech posture" which is different from rest and from which articulatory movements may be made most efficiently. Initial electromyographic (EMG) and cine-radiographic data suggest that while there may be such a speech posture, its nature (in terms of whether it is expressed as the position of a structure, its configuration, or the tension of certain muscles) may depend on the particular articulatory structure as well as a number of other factors. A great deal of additional analysis is needed.

Two EMG and movement studies aimed at gaining an understanding of coarticulation strategies have been run, and these data are currently being analyzed.

Studies on reducing dosages in radiographic pellet-tracking experiments suggest that dosages may be reduced by using: the lowest possible cine-frame rate, 16-mm film size, high-speed recording film (Kodak 2474), or videotape. While videotape requires the lowest possible dose at 60 frames per second, frame-by-frame analysis thus far seems to be less accurate than with cine.

We are continuing to build up our physiological data-gathering and analysis facilities, and we expect to be close to completion by the end of this year.

2. SYNTACTIC-TO-PHONETIC CODING IN SPEECH PRODUCTION

National Institutes of Health (Grant 5 ROl NS13028)

William E. Cooper, John M. Sorensen

The speech wave contains a number of characteristics that reflect the speaker's structural representation and processing of syntactic constituents. A theory of syntactic-to-phonetic coding has been developed in order to account for a variety of experimental results obtained in the recent years of this project. The structural component of the theory contains a metric of syntactic boundary strengths to account for the presence and rank magnitude of syntactic influences on speech timing, fundamental frequency (F_o), and the application of cross-word phonetic conditioning rules. The processing component of the theory includes constraints on the speaker's on-line planning and execution. In addition to theory development, experimental studies have been conducted on fundamental-frequency patterns to obtain more information about the following F_o attributes: (a) the form and domain of F_o declination in declarative utterances, (b) the influence of syntactic boundaries on the application of cross-word F_o conditioning effects, and (c) the influence of speaking rate, utterance and constituent length, and parenthetical expressions on the form and domain of F_o declination. These experimental studies have been conducted primarily with native speakers of English. Two related studies have been completed for

Japanese in collaboration with Kazuhiko Yorifuji.

3. STUDIES OF SPEECH PRODUCTION AND SPEECH DISCRIMINATION BY CHILDREN AND BY THE HEARING-IMPAIRED

National Institutes of Health (Training Grant 5 T32 NS07040)
National Science Foundation (Grants BNS76-80278 and BNS77-26871)
C. J. LeBel Fellowship

Jared Bernstein, Suzanne Boyce, Marcia A. Bush, Ursula G. Goldstein, Howard L. Golub, William L. Henke, Lise Menn

a. Speech and Sound Production by Infants and Children

Measurements of various parts of the vocal tract of children at different ages have been stored in a computer data base and fitted with growth curves. These curves are used to specify the dimensions of a static model to simulate the midsagittal outline of the vocal tract at various stages of development. The model will be further developed to include generation of area functions, formant frequencies, and speech waveforms for use in the exploration of the phonetic capabilities of children and the relationship between children's articulatory configurations for different vowels and the configurations used by adults.

b. Acoustic Analysis of Infant Cries

The aim of this project is to describe statistically the properties of cries elicited from infants who are a few days old, and to specify the way in which the cries of infants known to have particular pathologies differ from the cries of normal infants. Twenty-odd parameters of the cries of a large number of infants have been measured and assembled into a data base. These parameters include temporal properties of the cries, modes of vocal-fold vibration, formant frequencies, presence of nasalization, fundamental frequency, etc. Statistical analysis of the data for normal infants and for subgroups characterized by particular pathologies is proceeding. A model of infant cry production has been developed based on the acoustic theory of speech normally used for adult sound production, but modified by some physiological and anatomical hypotheses for neonates. The most important hypothesis deals with the control strategies involved in cry production. It is assumed that neonates tend to control their muscles (especially in the larynx) in a quantal fashion, thereby helping to explain most of the observed unique acoustic phenomena found in the cry.

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

We have been analyzing speech from recordings of 16 parent-child conversations; the children were aged 2 to 5, and the conversations took place in a semistructured laboratory playroom setting. For the parents we have found some semantically determined regularities in the pitch ranges used in successive clauses (within and across speakers) and some regularities in the use of certain marked voice qualities (falsetto, creaky voice, singing voice).

d. Speech Production by the Deaf

This area of research involves the application of acoustic theory and analysis to the study of problems commonly encountered in the speech of the profoundly deaf. One ongoing project is concerned with comparing the effects of segmental variables (such as vowel height, consonantal context, and vowel nasalization) on fundamental-frequency control in the speech of deaf and normal-hearing children and adolescents. Preliminary results suggest that certain inadequate modes of vocal-fold vibration may be maximally sensitive to such segmental variables and, furthermore, may be associated with the erratic pitch and breathy or falsetto voice quality characteristics of many deaf speakers.

In another project, a systematic study has been made of the kinds of anomalies that are present in the speech of deaf children when they concatenate words. These anomalies include pausing (with or without inspiration), glottalization, and errors in articulation of word-initial and word-final consonants. The data suggest a lack of awareness of many deaf children of how to produce phrasal units that encompass sequences of several words.

An attempt has also been made to describe the speech-production capabilities of ten adventitiously deafened adults. The most common segmental errors made by these speakers involved the production of the sibilant consonants /s/ and /$\check{s}$/. Inadequate velopharyngeal control at both the segmental and suprasegmental levels was also frequent. The best predictors of speech errors among the ten subjects appeared to be lack of hearing-aid use which, in turn, seemed related to the type and severity of a speaker's hearing loss.