XIX. SPEECH COMMUNICATION

Academic and Research Staff

Prof. Kenneth N. Stevens
Prof. Morris Halle
Dr. Jared Bernstein
Dr. Sheila Blumstein*
Dr. Margaret Bullowa
Dr. William E. Cooper
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. Colin Painter††
Dr. Joseph S. Perkell
Dr. Ernest A. Scatton‡‡
Dr. Stefanie Shattuck-Hufnagel***
Dr. Katherine Lee Williams
Dr. Catherine D. Wolf
Dr. Victor W. Zue
Barbara J. Moslin
John M. Sorensen

Graduate Students

Marcia A. Bush
Bertrand Delgutte
William F. Ganong III
Ursula Goldstein
Howard L. Golub
Stephen K. Holford
George P-N. Hui

1. STUDIES OF SPEECH PRODUCTION AND PERCEPTION

National Institutes of Health (Grants 5 ROl NS04332-14 and 5 T32 NS07040-02, and Fellowship 1 F22 NS00796-01)

Sheila E. Blumstein, Marcia A. Bush, Bertrand Delgutte, William L. Henke, Dennis H. Klatt, Colin Painter, Joseph S. Perkell, John M. Sorensen, Kenneth N. Stevens, Victor W. Zue

a. Perception of Stop Consonants

We are engaged in several studies relating to the perception of stop consonants. In one of these studies we are reexamining the cues for place of articulation of consonants through experiments in which responses are obtained from stimuli whose acoustic characteristics of bursts and transitions are systematically manipulated. The experiments suggest that listeners process the abrupt acoustic events at the release of stop consonants by utilizing cues that result from integration of onset characteristics of bursts and formant transitions. These integrated cues appear to be more independent of context than the acoustic properties of the individual components that contribute to the cues. In other experiments we are investigating the range of conditions of burst amplitude

---

*Associate Professor, Department of Linguistics, Brown University.

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

‡Assistant Professor of Linguistics, Southeastern Massachusetts University.

**Professor of Special Education, Boston University.

††Associate Professor of Communicative Disorders, Emerson College.

‡‡Assistant Professor, Department of Slavic Languages and Literatures, State University of New York at Albany.

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

and time location under which this cue integration occurs. The relationship between consonant cues in syllable-initial and syllable-final position is also being examined through experiments on the identification of these consonants and on selective adaptation.

In a collaborative study with the Eaton-Peabody Laboratory of the Massachusetts Eye and Ear Infirmary, the possible physiological correlates (in the auditory nerve) of abrupt stimulus onsets and offsets of the type that occur in stop consonants are being examined. Initially, we are using as stimuli sinusoids with abrupt onsets and offsets.

b. Study of Lexical Tones in Tone Languages

The aim of this study is to gain a better understanding of larynx behavior and of the features underlying it. We are interested in examining the various aspects of fundamental frequency (F_0) contours, such as rate of movement, direction, and timing of frequency changes, as they appear in various tone languages. Through a cross-language comparison of the use of F_0 contours for lexical tones, we hope to gain insight into the mechanism and the limitations on F_0 production and perception. In a series of perceptual experiments on Mandarin tones, syllables were synthesized by using "averaged" F_0 contours obtained from measured data. We found that these averaged contours possess some essential features for absolute identification. We also found that the frequency range spanned by these contours can be reduced by an order of magnitude without seriously impairing performance. The results of these experiments suggest a perceptual decoding mechanism based on simple property detectors. In ongoing perceptual experiments simple F_0 contours of various shapes are being used to determine those aspects of the contours that are distinctive. Acoustic measurements of F_0 contours for other tone languages are also being made and compared.

c. Acoustic-Phonetic Characteristics of English

We have begun a project whose ultimate goal is to produce a handbook describing the acoustic-phonetic characteristics of English. Speech data will be recorded from three men and three women in order to formulate a set of rules that describe the general acoustic characteristics of spoken English sentences. It is intended that this handbook will ultimately contain semicomplete normative data on the acoustic characteristics of a single dialect. The analyses should prove useful in the development of programs for speech synthesis by rule and for automatic speech recognition. In conjunction with background material on the physiology of speech production and the perception of speech, the acoustic analyses can also be used in programs for teaching spoken English (e.g., to teach second-language learners, the deaf, or the speech-handicapped), and to define objective criteria for the attainment of acceptable speaking performance.

Analysis techniques that will be employed include use of the traditional sound

spectrograph and several newer computer methods of spectral analysis. These newer methods include the discrete Fourier transform, linear prediction spectral analysis, automatic formant frequency tracking, and use of a digital spectral analyzer[1] that simulates processing which occurs in the human auditory system (based on psychophysical evidence).

d. Acoustic-Phonetic Studies in Languages Other than English

In an effort to develop further understanding of the acoustic mechanism of speech production and of the phonetic categories that it is capable of producing, several projects have been initiated. These include an acoustic study and interpretation of the various click sounds of Xhosa, and an acoustic and physiological investigation of stop consonants produced with various laryngeal configurations in several different languages.

e. Speech Synthesis by Rule

In a continuing effort to improve the quality of a speech synthesis-by-rule program,[2] a new programming language has been developed and the entire program has been rewritten in this higher level language.[3] Rules can now be stated in highly readable notation that is then automatically converted to Fortran code. The role of a phonological component in synthesis by rule has been defined and implemented.[3] The phonological component accepts as input a linear string of symbols produced by the (hypothetical) semantics component, syntactic component, and lexical component of a grammar of English. This abstract representation of an utterance is transformed by the phonological component into a narrow phonetic transcription and a specification of stress levels, segmental durations, and aspects of the fundamental frequency contour.

f. Segmental Duration as Part of the Speech Code

A review article has been prepared on the linguistic uses of segmental duration in English.[4] The pattern of durations of individual phonetic segments and pauses conveys information about the linguistic content of an utterance. Acoustic measures of segmental timing have been used by many investigators to determine the variables that influence the durational structure of a sentence. The literature on segmental duration is reviewed and related to perceptual data on the discrimination of duration and to psychophysical data on the ability of listeners to make linguistic decisions solely on the basis of durational cues. We conclude that, in English, duration often serves as a primary perceptual cue in making distinctions among (i) inherently long vowels and short vowels, (ii) voiced and voiceless fricatives, (iii) phrase-final and nonfinal syllables, (iv) voiced and voiceless postvocalic consonants, as indicated by changes in the duration of the preceding vowel in phase-final position, (v) stressed and unstressed or reduced vowels, and (vi) the presence or absence of emphasis.

g. Physiological Studies

Experiments are being performed on short-latency feedback effects of unexpected reductions in intraoral air pressure during the production of bilabial stops. We are looking at pressure-related changes in the timing of EMG signals from muscles of the larynx, lips, and mandible under conditions of normal lip sensation and labial anesthesia.

References

1. D. H. Klatt, "A Digital Filter Bank for Spectral Matching," Conference Record 1976 IEEE International Conference on Acoustics, Speech and Signal Processing, Philadelphia, April 12-14, 1976 (IEEE Catalog Number 76 CH1067-8 ASSP, pp. 573-576, 1976).
2. D. H. Klatt, "Acoustic Theory of Terminal Analog Speech Synthesis," Conference Record 1972 International Conference on Speech Communication and Processing, Boston, Massachusetts (IEEE Catalog Number 72 CHO 596-7, AE, pp. 131-135, 1972).
3. D. H. Klatt, "Structure of a Phonological Rule Component for a Synthesis-by-Rule Program," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-24, No. 5, pp. 391-398, October 1976.
4. D. H. Klatt, "Linguistic Uses of Segmental Duration: Acoustic and Perceptual Evidence," J. Acoust. Soc. Am. 59, 1208-1221 (1976).

2. SYNTACTIC-TO-PHONETIC CODING IN SPEECH PRODUCTION

National Institutes of Health (Grant 1 RO1 NS13028-01)

William E. Cooper, John M. Sorensen, Kenneth N. Stevens

a. Theory

A theory of syntactic-to-phonetic coding is being developed that focuses on two related issues: (i) What domains of syntactic coding influence aspects of phonetic coding, including influences on speech rhythm, intonation, and the blocking of phonological rules normally operating across word boundaries? and (ii) What details of syntactic coding are there in cases that are controversial on purely linguistic grounds? One specific objective in the theory concerns attaining a formal description of phrase and clause boundary strengths that utilizes the two structural properties of branching complexity and node height. This part of the theory is designed to account for differences in the magnitude of phrase-final lengthening and clause-final lengthening, as well as for differences in whether particular phrase and clause boundaries block the operation of phonological rules.

b. Studies of Speech Timing

Experimental studies have been conducted to test aspects of the theory. These include experiments on the blocking of phonological rules at syntactic boundaries, phrase-final lengthening, clause-conditioned lengthening for a segment just prior to the beginning of an embedded clause, and correlations of durations for nonadjacent word segments that lie within the same clause or span a clause boundary. Experiments have also been conducted on temporal disambiguities of structurally ambiguous sentences. The results of these studies have provided support for some of the major features of the theory and have stimulated revision and further development. We plan to continue conducting experimental studies on speech timing as a means of testing and developing the theory. The results of recent work in this area will also be used in designing new experiments on speech perception, to test the extent to which listeners detect syntactic influences on speech rhythm, and, more important, whether listeners actually utilize this information in decoding the syntactic representation of utterances.

c. Studies of Fundamental Frequency Contours

In addition to studies on speech timing, we are extending our work to fundamental frequency (F_0) contours, since these contours also reflect the influence of syntactic coding. We plan to obtain F_0 measurements for many of the experimental utterances already analyzed for syntactic influences on timing. Thus far, we have completed one such study involving complement clause structures.[1] Additional experiments are planned to test hypotheses about clause-conditioned resetting of F_0 declination and other aspects of syntactic influences on F_0.

References

1. W. E. Cooper, "Syntactic Control of Timing in Speech Production: A Study of Complement Clauses," J. Phonet. 4, 151-171 (1976).

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

National Institutes of Health (Grant 5 T32 NS07040-02 and Fellowship 1 F22 MH58258-02)

Jared Bernstein, Marcia A. Bush, Ursula G. Goldstein, Howard L. Golub, Lise Menn, Katherine Lee Williams

a. Speech and Sound Production by Infants and Children

A literature search has been conducted in an effort to establish the dimensions and growth of children's vocal tracts. The data from this search will be entered into a

computer data base for use with a vocal-tract model. The model will be used in an analysis-by-synthesis manner to estimate vocal-tract area functions and articulatory positions for various speech gestures by children of various ages.

Measurements of vowel formants of one two-year-old child are being made from spectrograms and by means of a linear prediction program. These measurements will be related to the average vocal-tract anatomy of a two-year-old child, through use of the vocal-tract model. An experiment on the adult perception of these vowels is planned. Segment and syllable durations for this child are also being examined from utterances taken at 18, 21, and 24 months.

In another project that is just beginning, we are attempting to apply modern signal-processing techniques to an analysis of infant cries, with a view toward developing improved methods of classifying the attributes of cries and individual differences in these attributes.

b. Speech Discrimination by Infants

Two studies of stop-consonant place discrimination by infants 6-12 weeks old have been carried out by using a high-amplitude sucking procedure. In one experiment, two synthetic consonant-vowel pairs were used. Both have five-formant patterns with formant transitions appropriate for either an initial [d] or [g], followed by a steady-state portion appropriate for the vowel [a]. One pair also had a 5-ms noise burst prior to the transition onsets which modeled the acoustic characteristics of the natural speech versions of [d] or [g]. An analysis of the results, in terms of levels of significance for discrimination, indicates that there is a greater probability of infant discrimination of a [d/g] contrast with an initial release burst than without it. The results suggest that the strong integrated cue resulting from both burst and appropriate onsets of formant transitions is required for infant discrimination, whereas adult listeners are able to extract information from the weaker cue provided only by transitions, possibly as a consequence of experience with the stimuli.

The purpose of the second study was to investigate infants' ability to discriminate between two speech sounds that contrast in syllable-initial, syllable-medial and syllable-final position and for which the contrasting element in medial position may or may not be stressed. A comparison was made of infant discrimination of contrasts: [da] vs [ga], [ad] vs [ag], [adá] vs [agá], and [áda] vs [ága]. All stimuli were composed of 5 formant patterns characteristic of unreleased stop consonants and steady-state vowel portions, with stress conveyed by a sudden rise and fall in F_0. Preliminary analysis of the results indicates that infants can discriminate syllable-initial but not syllable-final contrasts. Discrimination of intervocalic contrasts depends on stress placement. The exact nature of this dependency is now being investigated.

Future studies are planned to study infant discrimination of segmental contrasts

in duration and fundamental frequency for various phonetic contexts and language environments of infants.

c. Speech Production and Perception by the Deaf

The goal of this research is to understand the nature and causes of phonetic differences between speech of the deaf and speech of those with hearing. Detailed descriptions of the speech of deaf children are being prepared, based on reported work and on new measurements. These descriptions include analysis of articulatory, phonatory, and rhythmic problems of deaf children. By synthesizing deaflike speech with selected combinations of these problems at various levels of severity, we hope to develop an analytic predictor of intelligibility based on measurable aspects of a speech sample. We are also trying to develop signal-processing techniques for classifying the phonatory behavior of children and to use these techniques to describe the distribution of phonation qualities in a large population of deaf children.

4. ACOUSTIC STUDIES OF SPEECH SOUNDS: INVARIANT ATTRIBUTES, CONTEXT EFFECTS, AND SPEAKER DIFFERENCES

U. S. Army – Maryland Procurement Office (Contract MDA904-76-C-0331)

William L. Henke, Dennis H. Klatt, Kenneth N. Stevens, Victor W. Zue

a. Acoustic Study of Stop Consonants

The primary aim in the acoustic studies of stop consonants is to investigate the acoustic characteristics of these consonants under a controlled phonetic environment.[1] By controlling the environment and limiting the influence of the higher level sources of knowledge, we hope to isolate those acoustic attributes that are presumably invariant from those that are context- and speaker-dependent. The corpus of data used in this study is some 1700 nonsense [hə'CVC] utterances spoken by three males. Fifteen vowels and 27 singletons and clusters containing the English stops /p, t, k, b, d, g/ are included in the data base. Signal processing and statistical analysis programs are used to facilitate data analysis. Various aspects of the temporal and spectral characteristics of the stops have been quantified. The results suggest in general the presence of more context-independent attributes for the stop consonants than has been believed. This study is being extended to include other speech sounds and more complicated linguistic environments. Data are also being collected from more speakers to investigate inter-speaker differences.

b. Statistics of Fundamental-Frequency (F_0) Contours in Sentences

We have extracted F_0 contours from several minutes of read and spontaneous speech produced by different speakers. The material has been divided into speech bursts or "breath groups," separated by pauses, and the statistics of the contours within breath groups are being examined. These include the distribution of maximum and minimum F_0 values at different points in time within each breath group. The data show the contributions to the overall distribution of F_0 for a given speaker that are due to the baseline fall of F_0 within a breath group and also the fluctuations within an utterance that indicate stress and demarcate phrases.

c. Study of the Phonological Processes in English

The purpose of this study is to describe the phonological processes of English within a unified framework. Although most of the phonological processes have been known for some time, their treatment has been fragmental and has been based primarily on the introspection of the researcher working on his own speech, or judging the speech of an informant. Inherent in this approach is the danger of faulty judgment. As a consequence, description of these processes may be inaccurate and lead to incorrect formulation of rules. In our study we hope to avoid controversy over the phonetic corpus by documenting each process with acoustic and articulatory measurements, when these do not already exist. By making measurements of processes that have not been studied, we hope to improve on fragmental data, and develop a comprehensive and unified corpus of processes that will serve as the data base not only for this study but also for anyone interested in investigating phonological processes of natural language. Ongoing research includes classification, annotation, and cross-referencing of phonological processes. Acoustic data are also being collected from quite a few speakers. Measurements will be made on spectrograms to provide us with quantitative data on the nature of such phonological processes as nasalization, palatalization, and dental flapping.

References

1. V. W. Zue, "Acoustic Characteristics of Stop Consonants: A Controlled Study," Sc.D. Thesis, Department of Electrical Engineering and Computer Science, M. I. T., May 1976.