25. Communications Biophysics

A. SIGNAL TRANSMISSION IN THE AUDITORY SYSTEM

Academic and Research Staff

*Prof. W.T. Peake, Prof. W.M. Siebert, Prof. T.F. Weiss,
Dr. T.R. Bourk, Dr. J.J. Guinan, Jr., Dr. E.M. Keithley,
Dr. N.Y.S. Kiang, Dr. W.M. Rabinowitz, Dr. J.J. Rosowski,
Dr. W.P. Sewell, R.M. Brown, A.H. Crist, V.B. Kava,
J.W. Larrabee, F.J. Stefanov, D.A. Steffens*

Graduate Students

*D.W. Carley, B. Delgutte, M.L. Gifford, T. Holton
R.S. Leong, T.J. Lynch III, R.S. Waissman, J.R. White*

25.1 Basic and Clinical Studies of the Auditory System

National Institutes of Health (Grant 5 P01 NS13126 and Training Grant 5 T32 NS0704)

National Science Foundation (Grant BNS80-06369)

Nelson Y.S. Kiang, William T. Peake, William M. Siebert, Thomas F. Weiss

Our study of signal processing in the auditory system has continued in cooperation with the Eaton-Peabody Laboratory at the Massachusetts Eye and Ear Infirmary.

25.2 Middle Ear

Measurements of acoustic impedance at the tympanic membrane at low frequencies (200 Hz) have been useful clinically for detection of middle ear abnormalities. In addition, measurements of changes in the impedance which occur during contractions of the middle ear muscles (e.g., as a reflex response to high intensity sound), have been useful in diagnosing abnormalities of the inner ear and the neural pathways of the acoustic reflex in the brain stem.

Measurements of impedance in humans have often been limited to low frequencies (below 1 kHz). One of the difficulties is accounting for the effect of the space in the ear canal between the measurement location and the tympanic membrane. Recently published results have demonstrated the importance of a precise method for assessing this effect and the resulting measurements in human subjects extend to a higher frequency (4 kHz) and are significantly different from most earlier measurements.[1]

In experimental animals it is possible to minimize this problem and impedance measurements in cats have now been obtained up to 20 kHz in both normal animals and those with various modifications of middle ear structures.[2] These measurements lead toward a model of the cat middle ear in which the model elements can be identified with particular entities.[3] By applying the same methods to the ear of the alligator lizard, which has a rather different structure, we can test ideas about the importance of particular structures to signal transmission properties.[4] Also it is possible to interpret the effects of middle-ear muscle contractions (in man and other mammalian species) in terms of models of the middle ear.[5] We plan to broaden our work on the acoustic reflex so that the signal processing properties that involve inputs from both ears, and the anatomical substrate can be better defined. This reflex is of interest both because it is a relatively simple neural system requiring only the more primitive parts of the brain, and also because the knowledge can be clinically useful.

25.3 Inner Ear Transduction

Mechanical inputs are transduced to electrical responses by the receptor cells of the inner ear. Only in recent years has it been possible to record the receptor potentials inside the receptor cells so that specific models of the generation process can be tested.[6] Our experimental work has concentrated on the alligator lizard inner ear, where systematic measurements of DC and AC components of responses to tones have been completed.[7] The acoustic, mechanical, and mechano-electric system from external ear to receptor potential of the alligator lizard has been represented in a comprehensive model.[8] In this model the important nonlinearities are in the mechano-electric transducer. Unlike most earlier conceptions of inner ear processes, the frequency selectivity of the receptor potential is thought to result from the mechanical properties of the hairs (stereocilia) on the receptor cells. Both the experimental and the theoretical results will be prepared for publication in the coming year.

25.4 Coding in the Cochlear Nerve

Study of responses from the nerve fibers which carry signals from the inner ear to the brain continues to be important both for learning about inner ear mechanisms and for understanding how stimuli are coded at this level in the system. A reexamination of some experimental behavioral results reported in 1947 in the light of more recent physiological and anatomical information has led to the conclusion that behavioral thresholds to tones can be interpreted in terms of a simple picture in which those neurons which are most sensitive to a particular stimulus frequency determine the behavioral threshold at that frequency.[9] A study of cochlear-nerve responses to speech-like stimuli has led to descriptions of response measures and indications of the utility of these measures for detecting acoustic features of the stimuli.[10] A general conclusion of this study is that the different measures may be useful in different situations, so that systems for extracting information about speech stimuli (such as the central nervous system) might use multiple processing schemes.[11]

25.5 Central Nervous System

One focus of our work on the central nervous system has been the group of cells where the cochlear nerve fibers terminate (the cochlear nucleus). The first of a series of papers dealing with the anterior ventral cochlear nucleus has been published.[12] The results include a precise description for the cells of this region of the relationship of location within the nucleus to frequency selective properties of the cells. In this work computer manipulation of the anatomical information and computer generated displays played a key role in reaching conclusions about spatial organization of neural response properties. It is expected that further developments in the use of computer representation and processing of anatomical information will be an important effort in the next few years.

References

1. W.M. Rabinowitz, "Measurement of the Acoustic Input Immitance of the Human Ear," J. Acoust. Soc. Am. 70, 1025-1035 (1981).
2. T.J. Lynch III, "Signal Processing by the Cat Middle Ear: Admittance and Transmission, Measurement and Models," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., February, 1981.
3. T.J. Lynch III and W.T. Peake, "Measurements and Models of Cat Middle-ear Mechanics," J. Acoust. Soc. Am. 69, S14 (1981).
4. J.J. Rosowski, T.J. Lynch III, and W.T. Peake, "Middle-ear Mechanics: Comparisons of Cat and Alligator Lizard," J. Acoust. Soc. Am. 69, S14 (1981).
5. W.M. Rabinowitz, "Acoustic-reflex Effects on Middle-ear Performance," J. Acoust. Soc. Am. 69, S44 (1981).
6. T.F. Weiss, C.L. Searle, and K. Baden-Kristensen, "Proportionality of Intracellular Resistance Changes and Receptor Potentials in Hair-Cell Models," Hear. Res. 4, 243-250 (1981).
7. T. Holton, "Mechanoelectric Transduction by Cochlear Receptor Cells," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., September, 1981.
8. T.F. Weiss, W.T. Peake, R. Leong, T. Holton, J.J. Rosowski, and J.R. White, "Mechanical and Electrical Mechanisms in the Ear: Alligator Lizard Tales," J. Acoust. Soc. Am. 70, S50 (1981).
9. N.Y.S. Kiang, "A Reexamination of the Effects of Partial Section of the Auditory Nerve," Perspectives in Biol. and Med. 254-269, Winter 1981.
10. B. Delgutte, "Representation of Speech-like Sounds in the Discharge Patterns of Auditory-nerve Fibers," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., August 1981.
11. B. Delgutte, "Coding of Some Consonantal Features in the Auditory Nerve," J. Acoust. Soc. Am. 69, S54 (1981).
12. T.R. Bourk, P.P. Mielcarz, and B.E. Norris, "Tonotopic Organization of the Anteroventral Cochlear Nucleus of the Cat," Hear. Res. 4, 215-241 (1981).

B. AUDITORY PSYCHOPHYSICS AND AIDS FOR THE DEAF

Academic and Research Staff

L.D. Braida, R. Boduch, C. Chomsky, J. Coker,
L.C. Dowdy, N.I. Durlach, C.L. Farrar, M.S. Florentine,
D.M. Freeman, M.F. Garrett, A.J.M. Houtsma, W.M. Rabinowitz,
C.M. Reed, M.C. Schultz, C.L. Searle, W.M. Siebert,
E. Villchur, G.P. Widin, P.M. Zurek

Graduate Students

T.T. Allard, J.C. Anderson, D.K. Bustamante,
M.A. Clements, M.F. Davis, S.V. DeGennaro, S.G. Feldman,
K.J. Gabriel, D.M. Horowitz, Y. Ito, J.R. Moser,
P.M. Peterson, M.A. Picheny, R.J. Rohlicek, R.A. Siegel,
G.M. Skarda, J.P.Stautner, R.M. Uchanski, L. Zhang

25.6 Intensity Perception and Loudness

National Institutes of Health (Grant 5 R01 NS11153)
National Science Foundation (Grant BNS77-16861)
Louis D. Braida, Nathaniel I. Durlach, Catherine L. Farrar, Mary S. Florentine, William M. Rabinowitz, Charlotte M. Reed, Rosalie M. Uchanski

This research is concerned with the development of a unified, quantitative theory of intensity perception and loudness, and involves the construction and integration of models of sensory processes, short-term memory, perceptual context effects, and decision making, as well as extensive psychophysical experimentation.

A critical review of the SISI (Short Increment Sensitivity Index) Test[1] of intensity discrimination used in clinical audiological diagnoses to distinguish between various kinds of hearing loss has been completed.[2,3]

Two different versions of a model[4,5] for intensity discrimination based on excitation patterns derived from masking experiments have been analyzed and compared to published data.[6] In the single-band version it is assumed that discrimination is based on that critical band in which the excitation grows most rapidly with increasing intensity. In the multi-band version, discrimination is based on an optimum combination of information from all critical bands. Predictions from the two versions of the model were compared to data vfor intensity discrimination of tones as a function of level and frequency, for partially-masked tones, and for white noise. In general, the single-band version yields predictions in qualitative agreement, but not in quantitative agreement with the data for pulsed tones. The multi-band version yields predictions in better agreement with the data, except for high frequency-tones.

Data obtained in loudness discrimination experiments[7] between pure and partially masked tones of the same frequency have been analyzed in terms of the Preliminary Theory of Intensity Resolution.[8,9] To avoid problems of masking and the fatigue caused by continuous exposure to the masking noise, the pure tones were presented to one ear and the masked tones to the opposite ear. Five roving-level, two-interval, two-alternative symmetric forced-choice experiments have been conducted. Two experiments involved monaural comparisons between the same stimulus types, two involved interaural comparisons between the same stimulus types, and one involved interaural comparison between different stimulus types. Comparisons between the loudness of stimuli of the same types were considerably more accurate when made monaurally than when made interaurally.

The data were analyzed by assuming that the trace mode[9] was used for loudness comparisons and that accuracy was limited by four independent noise components: sensation noise, which limits resolution when no memory noise is present; trace noise, associated with memory decay between the stimulus presentation intervals; interaural comparison noise, which accounts for the degradation in performance seen when interaural comparisons are made for the same stimulus type; and a noise component which accounts for the additional degradation seen when pure tones are compared to partially masked tones. The size of various noise components were estimated by compensating for the level dependence of basic resolution, averaging across observers and levels, and using previously published estimates of the size of the trace memory noise. Estimates of the sensation noise for the stimuli used are roughly consistent with previous measurements[10] and indicate that it was the smallest component present. The trace noise was slightly larger than the sensation noise for the 200-msec interstimulus interval used. The interaural comparison noise was roughly independent of stimulus type and was found to be much larger than either the sensation or trace noises. Roughly speaking the size of this component is the same as that observed when comparing the loudnesses of 800 and 1000 Hz tones in the same ear.[9] The noise associated with the comparison of different stimulus types was large relative to the sensation and trace noises, but significantly smaller than the interaural comparison noise. The same pattern of relative sizes of noises was found for all four observers who participated in the study.

References

1. J. Jerger, J. Shedd, and E. Harford, "On the Detection of Extremely Small Changes in Sound Intensity," Arch. Otolaryngol. 69, 200-211 (1959).
2. S. Buus, M. Florentine, and R.B. Redden, "The SISI Test: A Review. Part I," accepted for publication in Audiology (1982).
3. S. Buus, M. Florentine, and R.B. Redden, "The SISI Test: A Review. Part II," accepted for publication in Audiology (1982).
4. E. Zwicker, "Die Elementaren Grundlagen zur Bestimmung der Informationskapazitat des Gehors," Acustica 6, 365-381 (1956).
5. D. Maiwald, "Berechnung von Modulationsschwellen mit Hilfe eine Functionsschemas," Acustica 18, 193-207 (1967).
6. M. Florentine and S. Buus, "An Excitation-Pattern Model for Intensity Discrimination," J. Acoust. Soc. Am. 70, 1646-1654 (1982).

7. R.M. Uchanski, "Variability of Loudness Comparisons for Stimuli with Different Dynamic Ranges," S.M. Thesis, Dept. of Electrical Engineering and Computer Science, M.I.T., 1981.
8. N.I. Durlach and L.D. Braida, "Intensity Perception I. Preliminary Model of Intensity Resolution," J. Acoust. Soc. Am. 46, 372-383 (1969).
9. J.S. Lim, W.M. Rabinowitz, L.D. Braida, and N.I. Durlach, "Intensity Perception VIII. Loudness Comparisons between Different Types of Stimuli," J. Acoust. Soc. Am. 62, 1256-1267 (1977).
10. A.J.M. Houtsma, N.I. Durlach, and L.D. Braida, "Intensity Perception XI. Experimental Results on the Relation of Intensity Resolution to Loudness Matching," J. Acoust. Soc. Am. 68, 807-813 (1968).

25.7 Hearing Aid Research

National Institutes of Health (Grant 5 R01 NS12846)

Louis D. Braida, Diane K. Bustamante, Steven V. DeGennaro, Nathaniel I. Durlach, Dennis M. Freeman, Paul Milner, Patrick M. Peterson, Michael A. Picheny, Charlotte M. Reed, Robin Rohlicek, Rosalie M. Uchanski, Edgar Villchur, Victor W. Zue, Patrick M. Zurek

This research is directed toward improved aids for people with sensorineural hearing impairments. We intend both to develop improved aids and to obtain a fundamental understanding of the limitations on such aids. The proposed work includes study of i) linear amplification, ii) amplitude compression, iii) frequency lowering, and iv) clear speech.

i) Research on linear amplification is concerned with modelling the dependence of speech reception performance on the linear amplification system, the speech materials, the background interference, and the impairment. Current work is focused on the use of Articulation Theory for predicting results in quiet and on the effects of additive noise on performance.

An extensive study[1] of the perception of filtered nonsense syllables by listeners with sensorineural hearing impairments, normal listeners, and normal listeners with simulated hearing losses has recently been completed and analyzed in terms of the predictions of Articulation Theory.[2,3] For the normal hearing listeners, Performance-Intensity (PI) functions generally increase with level until a plateau is reached, and at any given level scores decrease when spectral information is removed. However, maximum performance for the highpass 700 Hz condition was roughly equal to maximum performance for the unfiltered condition. In order of diminishing maximum scores, the remaining conditions were ordered as follows: lowpass 2800 Hz, bandpass 700-2800 Hz, highpass 1400 Hz, bandpass 1400-2800 Hz, lowpass 1400 Hz, bandpass 700-1400 Hz, lowpass 700 Hz and highpass 2800 Hz. Relative to previously reported results,[2] these data indicate superior performance for lowpass 700 Hz, lowpass 1400 Hz, and highpass 2800 Hz; slightly inferior performance for highpass 700 Hz and highpass 1400 Hz; and roughly equivalent performance for the wideband and lowpass 2800 Hz conditions.

The listeners with sensorineural hearing losses generally exhibited a dependence of performance on filter condition that was similar to the normal hearing listeners in that removing spectral information generally reduced intelligibility at a given presentation level. Maximum scores in each condition, including the unfiltered speech, were lower than for normal hearing listeners, and were also dependent on the degree of hearing loss. The specific contribution of the highest frequency band (highpass 2800 Hz) did not appear to be as large for these listeners since scores did not increase greatly when this band was added to conditions containing information below this cutoff frequency. However, at the higher presentation levels, removing information below 700 Hz resulted in maximum scores equal to or better than the scores for those conditions where this information was present. Finally the PI functions for these listeners exhibited significant rollover at the highest levels tested much more frequently than for the normals.

Listeners with simulated losses exhibited higher performance than listeners with real losses, although the dependence on filtering condition was roughly the same. Generally, these listeners exhibited rollover for the same test conditions (700 Hz lowpass, 700-1400 and 1400-2800 Hz bandpass) as the listeners with sensorineural hearing loss.

Although there is considerable internal consistency in the data obtained, these data are not in good agreement with the quantitative predictions of Articulation Theory. In general, the data for the highly filtered conditions (700 Hz lowpass, 700-1400 Hz and 1400-2800 Hz bandpass, and 2800 Hz highpass) are relatively better, and those for the broadband conditions (unfiltered and 700 Hz highpass) relatively poorer than predicted for all listeners. Checks of the additivity assumption of Articulation Theory were generally positive for low and moderate presentation levels, but negative at the highest presentation levels (where spread of masking and discomfort effects are significant) for nearly all listeners tested.

ii) Research on amplitude compression, directed towards listeners with reduced dynamic range, involves the development of an advanced, digital, multiband compression system and further study of syllabic compression, limiting, and automatic volume control.

A new study of multiband syllabic compression for severely impaired listeners has been conducted using an improved 16-channel compression system.[4] Speech reception experiments were conducted on three listeners with severe congenital sensorineural hearing impairments characterized by flat audiograms and dynamic ranges of 20-30 dB. Three compression systems which placed progressively larger amounts of speech energy within the listener's residual dynamic range by raising to audibility and compressing 25%, 50%, and 90% of the short-term input amplitude distribution in each of 16 frequency bands were tested. For comparison, a linear system designed to raise six octave-wide bands of speech to comfortable levels was also tested. System performance was evaluated with nonsense CVC syllables presented at a constant input level and spoken by two talkers. Extensive training was provided to ensure stable performance. The results were notably speaker dependent, with compression consistently providing better performance for one, linear amplification for the other. Averaged over speakers, however, there was no net advantage for any of the compression systems for any listener. The use of high compression ratios and large input ranges tended to degrade perception of initial consonants and vowels. Under some conditions, however, final consonant scores were higher with compression than with linear amplification. Detailed analyses of performance are now being prepared.

iii) Research on frequency lowering, directed towards listeners with poor hearing at high frequencies, includes continued analysis of our pitch invariant frequency lowering scheme and exploration of low-frequency representations of speech achieved by means of artificial codes.

In one study[5] the discriminability of 32 pairs of CV syllables frequency-lowered by two different pitch-invariant techniques, the warping/dilation scheme used in much of our previous work[6] and an LPC-based technique,[7] has been compared using listeners with normal hearing. Although the signal processing techniques differ computationally, they were used to achieve the same spectral mapping: nonuniform lowering to a 1250 Hz bandwidth, with greater compression of the high frequencies. For reference purposes, the materials were also processed by lowpass filtering to the same bandwidth. Consonant discrimination for the warping-dilation scheme was slightly inferior to lowpass filtering but substantially better than for the LPC scheme. The relative performance of the two frequency-lowering schemes was roughly independent of the contrast tested. (The results for the warping/dilation scheme and lowpass filtering were similar to those reported previously for identical conditions.[8]) Although poor performance of the LPC scheme may be attributed to both temporal smearing (associated with the use of a 25 msec window) and spectral smearing (resulting from the limitations of an 8-pole vocal tract model) not present in the warping/dilation scheme, the similarities in the performance patterns for the two schemes suggest that most of the discriminations were adversely affected by the spectral-transformation itself rather than by artifacts introduced by the specific characteristics of the implementation.

In a second study[9] four listeners with high-frequency sensorineural hearing loss were tested on consonant discrimination and/or identification under conditions of linear amplification and frequency lowering. Linearly amplified stimuli were presented with either uniform amplification or high-frequency emphasis. Frequency lowering was accomplished by a technique involving pitch-invariant non-uniform compression of the short-term envelope,[6] and included lowering to bandwidths of 2500 Hz and 1250 Hz. Of three subjects tested on pairwise discriminability of CV syllables, two who listened to frequency lowering to a 1250-Hz bandwidth performed worse (by roughly one unit of d') on lowering than on linear amplification. For a third subject, who was tested with the less severe lowering factor (lowering to 2500 Hz), discrimination performance was similar for the two conditions. Three subjects received training on CV identification for conditions of linear amplification and lowering to a bandwidth of 2500 Hz; in addition, one subject also received training for lowering to 1250 Hz. With one exception (that of one subject for lowering to 2500 Hz), post-training performance on linear amplification exceeded that on lowering. The frequency-lowering scores of two of the subjects were roughly 20 percentage points below those of normal subjects tested on a similar lowering condition, while the scores of the third subject for two lowering factors were consistent with those obtained by normals. The results indicate some variability in performance among subjects with similar audiometric loss.

iv) Research on clear speech is concerned with the creation of speech materials that are highly intelligible to impaired listeners through the use of techniques that focus on the speech source rather than on signal processing, and with the determination of the acoustical properties of these materials that are responsible for the high intelligibility.

One study, consisting of three components,[10] measured intelligibility differences for impaired listeners associated with variations in speaking mode; evaluated the acoustic differences associated with these variations; and probed the effects on intelligibility of variation in one of the parameters, speaking rate, associated with variation in speaking mode. The first component consisted of measurements of word intelligibility in nonsense sentences for five impaired listeners. The impaired listeners ranged in age from 24 to 64 and had primarily cochlear losses (4 bilateral, 1 unilateral; 3 flat, 2 sloping; 1 congenital, 2 noise exposure, 2 of unknown origin; speech reception thresholds 0-70 dB; speech discrimination scores 28-92%). The speech materials were recorded in an anechoic chamber using three male speakers and two speaking modes, "conversational" and "clear", which were achieved by instructions and coaching. All materials (conversational as well as clear) were highly intelligible to normal hearing listeners (scores exceeded 94%). These materials were presented to each impaired listener using two frequency-gain characteristics and three intensity levels. The results of these tests showed a clear and robust intelligibility advantage for clear speech. Substantial improvements were obtained for essentially all listeners, all speakers, all frequency-gain characteristics, and all levels. In addition, the improvements appeared to be roughly independent of word position in the sentence and independent of phoneme class. Overall, the scores with clear speech exceeded those with conversational speech by 17 percentage points.

The second component consisted of an analysis of prosodic, phonological, and phonetic differences between conversational and clear speech. Prosodic characteristics showed dramatic changes in temporal factors. In clear speech, the speaking rate was reduced by roughly a factor of two; the number and duration of pauses was substantially increased; and, unlike the results obtained when a speaker is instructed merely to speak slowly, the duration of most speech sounds was substantially increased, although not uniformly. Changes were also observed in fundamental frequency contours: in clear speech, the range (and to a slight extent, the mean) of fundamental frequencies was increased. Analysis of phonological properties showed that clear speech was associated with less vowel reduction, fewer burst eliminations, decreased degemination, decreased alveolar flapping, more sound insertions, and fewer sound deletions. Analysis of phonetic characteristics showed no substantial change in long-term RMS spectra, consistent with the notion that the intelligibility improvement in clear speech is roughly orthogonal to the improvement that can be achieved by manipulation of the frequency-gain characteristic. This analysis also showed that clear speech (unlike loud speech) is associated with an increase (non-uniform) in consonant-to-vowel amplitude ratio and with a variety of changes in individual consonants and vowels (e.g., a more pronounced high-frequency peak in the short-term spectrum for /t/ and an expanded and shifted formant frequency triangle for the lax vowels).

The third component probed the extent to which the change in speaking rate was responsible for the change in intelligibility between the two speaking modes. A subset of the nonsense sentences was processed using a high-quality algorithm[11] to speed up clear speech to the rate of conversational speech and to slow down conversational speech to that of clear speech. In this algorithm, all elements of speech are slowed down or sped up by the same amount and fundamental frequency is unaltered. Both casual listening and tests performed on the impaired listeners with "twice-processed" speech (sped up and then slowed down clear speech and slowed down and then sped up conversational speech) indicated that the processing was indeed of high quality. However, for both the sped-up clear speech and the slowed down conversational speech, impaired listeners achieved sharply reduced intelligibility scores: compared to unprocessed scores of 53 (conversational) and 75 (clear) and twice-processed scores of 48 (conversational) and 65 (clear), the processed scores were 40 (slowed-down conversational) and 45 (sped-up clear). These results indicate that a uniform decrease in rate by no means transforms conversational speech into clear speech and also that a uniform increase in rate for clear speech does not produce results that are more intelligible than conversational speech at the same rate.

References

1. P. Milner, "Perception of Filtered Speech by Hearing Impaired Listeners and by Normal Listeners with Simulated Hearing Loss," Ph.D. Thesis, Department of Speech and Hearing Sciences, The City University of New York, 1981.
2. N.R. French and J.C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," J. Acoust. Soc. Am. 19 (1947).
3. R.L. Dugal, L.D. Braida, and N.I. Durlach, "Implications of Previous Research for the Selection of Frequency-Gain Characteristics," in G.A. Studebaker and I. Hochberg (Eds.), Acoustical Factors Affecting Hearing Aid Performance and Measurement (University Park Press, 1980).
4. M.C.W. Coln, "A Computer Controlled Multiband Amplitude Compressor," S.M. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1979.
5. J.R. Rohlicek, "Analysis of Frequency Lowering Systems," S.B. Thesis. Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, Massachusetts (1979).
6. B.L. Hicks, L.D. Braida, and N.I. Durlach, "Pitch Invariant Frequency Lowering with Nonuniform Spectral Compression," Proc. of 1981 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1981.
7. O. Sotomayor-Diaz, "Frequency Lowering Systems Based on Linear Prediction of Speech," S.M. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1980.
8. C.M. Reed, B.L. Hicks, L.D. Braida, and N.I. Durlach, "Discrimination of CV Syllables Processed by Lowpass Filtering and Pitch-Invariant Frequency Lowering," accepted by J. Acoust. Soc. Am. (1982).
9. C.M. Reed, L.D. Braida, Durlach, and K.I. Schultz, "Discrimination and Identification of Frequency--Lowered CV Syllables by Hearing-Impaired Listeners," manuscript in preparation, 1981.

10. M.A. Picheny, "Speaking Clearly for the Hard of Hearing," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1981.
11. D. Malah, "Time Domain Algorithms for Harmonic Bandwidth Reduction and Time Scaling of Speech Signals," IEEE Trans. Acoust. Speech and Signal Processing, ASSP-27 121 (1979).

25.8 Tactile Communication of Speech

National Science Foundation (Grant BNS77-21751)
National Institutes of Health (Grant 1 R01 NS14092)
Karmazin Foundation through the Council for the Arts at MIT

Louis D. Braida, Raymond Boduch, Carol Chomsky, Mark A. Clements, Jackie Coker, Leonard C. Dowdy, Nathaniel I. Durlach, Michael A. Epstein, Merrill F. Garrett, Adrian J.M. Houtsma, James Moser, Joseph S. Perkell, William M. Rabinowitz, Charlotte M. Reed, Jean P. Reid, Martin C. Schultz, Gregory M. Skarda

The goal of this research is to develop tactile speech communication aids for the deaf and deaf-blind. Our research during the past year can be divided into four major areas: (i) Study of Tadoma, (ii) Development of a Synthetic Tadoma System, (iii) Studies with a Vibrotactile Stimulator Array, and (iv) Basic Study of Encoding and Display Schemes. In addition, we have completed work on a comparative review of recent studies of the tactile communication of speech, including work on Tadoma, spectral displays, and tactile supplements to lipreading.[1]

i) One aspect of our work in this area focuses on the study of the speech and language communication abilities of deaf-blind subjects who are highly trained in the use of Tadoma. A survey of Tadoma use in the United States and Canada[2] has been completed and a set of screening tests on nine experienced Tadoma users has been performed.[3] Additional, in-depth testing of three of these subjects has been conducted in the areas of speech reception,[4,5] linguistic competence,[6] and speech production. A second aspect of our work in this area is a study of the acquisition of Tadoma skills by normal-hearing and sighted subjects having simulated deafness and blindness. Results for discrimination and identification of short speech segments[7,8] have been obtained as well as preliminary results on perception of sentences formed from a small, closed vocabulary.[8] A group of four normal-hearing and sighted subjects is now involved in a long-term training program in Tadoma. These subjects have completed the discrimination and identification components of the study and are currently concentrating on word, sentence, and paragraph stimuli.

ii) The initial objective of this work is to develop a system composed of a sensor array to be placed on a talker's face and an artificial face (driven by the signals from the sensor array) to be sensed by a Tadoma reader's hand that will provide speech reception performance equivalent to that achieved with real Tadoma. The facial actions that are included in our initial version of the system (based on our understanding of real Tadoma) are laryngeal vibration, oral air flow, and jaw and lip movement. Simultaneous multichannel recordings of these actions, together with the acoustic speech signal, have been made with the aid of the Speech Communication Group. A variety of speech materials drawn primarily from the set used in tests of Tadoma subjects (including isolated syllables, words, and sentences) have been recorded by a male talker. The multichannel recordings have been digitized and transferred to disk storage and software is currently being developed to facilitate their display and editing. A preliminary study of the effects of phonetic context on various articulatory actions has also been performed. In the construction of the artificial face, a model skull is being adapted to incorporate the articulatory actions specified above. Both a jaw-motion and lip-movement system have been incorporated into the skull using position-programmable DC servo-motors and appropriate mechanical linkages. The lower lip is capable of both protrusion (in-out) and aperture (up-down) movements while the upper lip is constrained to protrusion movements only. Lower lip movements are accomplished relative to the jaw. The next stage of development includes construction of an air-flow system to convey aspiration characteristics of consonants, addition of a vibrator in the region of the throat to represent laryngeal vibration, the addition of a moldable rubber skin surface, and development of an interface to control the artificial face from the digital recordings. Initial evaluation of the synthetic Tadoma system will focus on confirming the physical operation of the system by comparing the actions of the artificial face to those of a real face producing the same utterances. Perceptual tests will also be conducted on experienced Tadoma users to compare speech-reception results for the artificial face with those of a real face. After adequate perceptual equivalence has been demonstrated, the synthetic system will be used to explore transformations in Tadoma that cannot be achieved naturally.

iii) Studies of tactile speech communication are being conducted using a vibrotactile stimulator array recently developed for our laboratory.[9] In these studies, the acoustic speech waveform is processed in some manner (e.g., spectral decomposition), and the processed signal is then displayed on the vibratory array using some arbitrary encoding scheme (e.g., a frequency to place transformation). In one study two types of artificial tactile displays, both based on different descriptions of the same acoustic speech waveform, were investigated. The vibatory array used for the study consisted of a 12x12 element square array of stimulators that was 13.3 cm on a side and was placed on the observer's thigh. The array was positioned such that the tangential motion of the vibrators was along the longitudinal axis. One display ("area-function" display) involved presenting the shape of a model of the vocal tract derived by LPC analysis, while the other ("spectral" display) involved presenting the shape of the short-term spectrum derived from the envelopes of the outputs of a bank of filters. In the area-function display, each of the twelve rows of vibrators (situated parallel to the length of the leg) corresponded to a different acoustic tube section whose area was indicated by the pattern of vibrators active in the columns (transverse to the leg) corresponding to that particular row. The overall vibratory amplitude was varied such that the perceived tactile intensity was proportional to the loudness of the corresponding acoustic segment. Voiced and unvoiced segments of the stimuli were represented by separate excitation waveforms (a 250-Hz sinewave and a wideband noise). In the spectral display, each of the twelve rows of the array (again arranged parallel to the length of the leg) corresponded to a different frequency band and each of the twelve columns (transverse to the subject) represented a time interval. The time interval of each analysis frame was 9 msec, corresponding to 108 msec of speech displayed at any instant. Overall loudness and the voiced-voiceless distinction were encoded in the manner described for the area-function display. Experiments have been carried out on two subjects to compare the discriminability of pairs of short speech segments through the two types of displays. The results of these experiments are currently being analyzed and will be compared to similar studies performed with Tadoma and with other types of artificial tactile displays.

iv) Our objective in this research is to gain fundamental knowledge for the design of effective tactile systems through experiments that are sufficiently simple to permit systematic analysis and yet are relevant to the encoding and display of complex signals. We intend to apply the results of these experiments not only to the development of tactile speech communication systems, but also to the development of tactile music systems. In an ongoing experiment,[10] whose goal is to study perceptual dimensionality at a fixed locus, we are investigating a number of physical variables to determine the subsets that are most potent for generating high-dimensionality perceptual spaces. A device has been constructed for these experiments in which a mechanical stimulator centered around a fixed surround (0.375 inch diameter) is brought into contact with the subject's finger tip (middle or index finger). The physical variables studied were frequency, intensity, and surface area of stimulator contact. The experiments employed an absolute-identification procedure with feedback, separate testing of each physical variable, stimulus durations of 500 msec with mechanical contact before and after stimulation lasting 400 msec, and aural masking noise and visual obstruction of the device. Ten frequency values ranging from 50 to 530 Hz were tested at each of three loudness levels; ten intensity values ranging from 3 to 30 dB SL were tested at each of three frequencies; and eight contact sizes ranging from 0.28 to 0.0l-inch diameter were tested for one combination of frequency and intensity. Roughly 500 trials per condition were obtained for each of four subjects and a confusion matrix was generated for each subject under each condition. The results, which were similar across subjects, may be summarized by information transfer in bits for each condition. For each of the three physical variables studied, information transfer averaged roughly 1.5 bits, corresponding to roughly 50% of the available stimulus information transferred. Further experiments are planned in which the variables will be presented in all possible combinations to determine if added dimensions contribute to an increase in information transfer. Other physical variables (including direction of vibration, waveshape of the signal, and force, texture, and temperature of the contactor) will be included in future experiments.

References

1. C.M. Reed, N.I. Durlach, and L.D. Braida, "Research on Tactile Communication of Speech: A Review," ASHA Monograph Number 20 (1982).
2. M.C. Schultz, S.J. Norton, S. Conway-Fithian, and C.M. Reed, "A Survey of the Tadoma Method in the United States and Canada," submitted to Volta Review, 1982.

3. C.M. Reed, S. Conway-Fithian, L.D. Braida, N.I. Durlach, and M.C. Schultz, "Further Results on the Tadoma Method of Communication," J. Acoust. Soc. Am. 67, Suppl. 1, S79, (1980).
4. S.J. Norton, M.C. Schultz, C.M. Reed, L.D. Braida, N.I. Durlach, W.M. Rabinowitz, and C. Chomsky, "Analytic Study of the Tadoma Method: Background and Preliminary Results," J. Speech Hearing Res. 20, 574-595 (1977).
5. C.M. Reed, N.I. Durlach, L.D. Braida, and M.C. Schultz, "Analytic Study of the Tadoma Method: Identification of Consonants and Vowels by an Experienced Tadoma User," J. Speech Hearing Res., 25, 108-116 (1982).
6. N. Maragioglio, and M.F. Garrett, "Sentence Processing in a Tactile Modality: Evidence from a Tadoma User," Manuscript in preparation.
7. C.M. Reed, S.I. Rubin, L.D. Braida, and N.I. Durlach, "Analytic Study of the Tadoma Method: Discrimination Ability of Untrained Observers," J. Speech Hearing Res. 21, 625-637 (1978).
8. C.M. Reed, M.J. Doherty, L.D. Braida, and N.I. Durlach,). "Analytic Study of the Tadoma Method: Further Experiments with Inexperienced Observers," J. Speech Hearing Res., 25, 216-223, (1982).
9. S.I. Rubin, "A Programmable Tactile Array," S.M. Thesis, M.I.T., 1979.
10. A.J.M. Houtsma, "Preliminary Data on Tactile Intensity Identification, Frequency Identification, and Contact Size Identification," Summary of work in progress, 1981.

25.9 Acoustic Emissions from Human Ears

National Institutes of Health (Fellowship 5 F32 NS06386)
National Science Foundation (Fellowship SPI79-14913)
William M. Rabinowitz, Gregory P. Widin, Patrick M. Zurek

Recent studies[1,2] indicate that human ears emit low-level sounds (< 20 dB SPL) spontaneously and in response to externally applied stimuli. Since these sounds appear to arise in the cochlea, but can be measured in the ear canal, they provide a noninvasive objective means to probe peripheral auditory function.

Our initial work[3] has focused on spontaneous acoustic emissions (SAEs), extending previous measurements[4,5] of their suppression by external tones, and relating suppression to the frequency and level of the suppressor. We have also observed the effect of one external tone on the suppression caused by another. In these measurements, the first tone is presented at a level which substantially suppresses the SAE. The SAE level can then be increased by introducing a low-level second tone of appropriate frequency. This is interpreted as an indication of suppression of the first tone by the second. As the level of the second tone is raised, the SAE can again be reduced by the suppressive action of the second tone. In one ear with a particularly intense SAE, 10 dB of stimulus-tone suppression has been observed in this way. While the SAE is typically suppressed by the first tone at a high rate (5 dB/dB), the first tone was suppressed by the second at a rate (1 dB/dB) similar to that observed neurophysiologically.[6,7]

In future work we plan to examine the time course of suppression for the SAE and stimulus tone, and its relation to forward masking as measured psychophysically. We will also explore possible relation of SAEs to certain types of tinnitus, and to noise exposure, as suggested by studies with chinchilla.[8] In addition, we will study emissions stimulated by external sounds to determine whether stimulated and spontaneous emissions have common origins. We plan to determine whether the measurement of acoustic emissions associated with aural distortion products can serve as a basis for objective audiometric tests.

References

1. D.T. Kemp, "Stimulated Acoustic Emissions From Within the Human Auditory System," J. Acoust. Soc. Am. 64, 1386-1391 (1978).
2. D.T. Kemp, "Evidence of Mechanical Nonlinearity and Frequency Selective Wave Amplification in the Cochlea," Arch. Otorhinolaryngol. 224, 37-45 (1979).
3. W.M. Rabinowitz and G.P. Widin, "Interaction of Spontaneous Acoustic Emissions with External Tones," J. Acoust. Soc. Am. Suppl. 1 70, S7 (1981).
4. J.P. Wilson, "Evidence for a Cochlear Origin for Acoustic Re-emissions, Threshold Fine-Structure and Tonal Tinnitus," Hear. Res. 2, 233-252 (1980).
5. P.M. Zurek, "Spontaneous Narrow-band Acoustic Signals Emitted by Human Ears," J. Acoust. Soc. Am. 69, 514-523 (1981).
6. M.B. Sachs, "Stimulus-response Relation for Auditory-Nerve Fibers: Two-tone Stimuli," J. Acoust. Soc. Am. 45, 1025-1036 (1969).
7. E. Javel, "Suppression of Auditory Nerv Responses I: Temporal Analysis, Intensity Effects, and Suppression Contours," J. Acoust. Soc. Am. 69, 1735-1745 (1981).
8. P.M. Zurek and W.W. Clark, "Narrow-band Acoustic Signals Emitted by Chinchilla Ears After Noise Exposure," J. Acoust. Soc. Am. 70, 446-450 (1981).

C. TRANSDUCTION MECHANISMS IN HAIR CELL ORGANS

Academic and Research Staff

Prof. L.S. Frishkopf, Prof. T.F. Weiss, Dr. C.M. Oman

The overall objective of this project is to study the sequence of steps by which mechanical stimuli excite receptor organs in the phylogenetically related auditory, vestibular, and lateral-line organs. The receptor cells in these organs are ciliated hair cells. Specific goals include the characterization of the motion of the structures involved, particularly the hair cell cilia; study of the nature and origin of the electrical responses to mechanical stimuli in hair cells; and investigation of the role of these responses in synaptic and neural excitation.

During the past year our goal has been to measure the mechanical response characteristics of a simple auditory organ to sound, with particular attention to the way in which the spatial pattern of response depends upon stimulus frequency.

25.10 Investigation of the Mechanical Basis of Frequency Selectivity and Tonotopic Organization in the Cochlea of the Alligator Lizard

National Institutes of Health (Grant 5 R01 NS11080)

Lawrence S. Frishkopf

Background

In the mammalian cochlea the receptor cells (hair cells) are located in the organ of Corti on the basilar membrane; this membrane (35 mm long in man) forms one wall of the coiled cochlear partition which vibrates when the ear is stimulated acoustically, thereby stimulating the hair cells. Bekesy (1960) and others have shown that, in mammals, mechanical frequency analysis occurs along this partition. This relationship of frequency to place - tonotopic organization - is maintained in the auditory nerve and at higher levels of the auditory system.

In the homologous but structurally simpler auditory organ of the alligator lizard, Gerrhonotus multicarinatus, studies of the responses from auditory nerve fibers indicate that spatial frequency analysis also takes place, i.e., a correlation exists between the frequency at which a primary auditory neuron is maximally sensitive (the characteristic frequency (CF) of the fiber) and the location along the basilar membrane at which that fiber's response is recorded (Weiss et al., 1978). In this species the basilar membrane is very short (0.5 mm); as in the mammal, the hair cells are located within an elongated ridge (the basilar papilla) which sits on the basilar membrane. Peake and Ling (1980) have measured the velocity of motion at several locations along the basilar membrane as a function of stimulus frequency, using the Mossbauer method. Their results indicate that in this species a single broad frequency response characterizes all measured locations and largely reflects the frequency response of the middle ear.

These results from mechanical and neural studies suggest that frequency analysis in the lizard ear occurs somewhere between the mechanical response of the basilar membrane and the neural output. In addition there is evidence from intracellular recordings that the hair cells are themselves frequency selective, exhibiting narrow frequency responses and a range of CF's like those seen in the nerve fibers of this organ (Weiss et al., 1978). Thus it seems reasonable to hypothesize that frequency analysis in this organ depends on the mechanical or electrical properties of the hair cells.

One feature of the hair cells that varies systematically along the papilla is the maximum length of the cilia; over a part of the papilla in which the neural CF varies monotonically from 1 to 4 kHz, the maximum cilia length varies from 30 μm to 12 μm. The suggestion has been made by Weiss et al (1978) that frequency analysis in this region results from the frequency selective mechanical responses of free-standing cilia of different lengths. Recently Weiss has calculated how the resonant frequencies of cilia-like structures in a boundary layer vary with cilia length; a model based on these results predicts tuning curves that fit the neural data remarkably well (Leong, 1982).

Methods and Results

The cochlear duct of the alligator lizard was excised and the basilar papilla with its supporting structures was dissected and mounted in a drop of lizard Ringer's solution across an opening in an air-filled chamber, where it could be stimulated with tones (>110 dB SPL) over the sensitive range of the organ (0.2 - 4 kHz). The resulting motion of the basilar papilla and hair cell cilia was observed using stroboscopic illumination at 400 - 800X under a water immersion objective (N.A. 0.75).

Some aspects of the motion of the basilar papilla have already been described (Frishkopf, 1981); the entire papilla rocks in phase about its long axis, in the direction of hair cell sensitivity, over the frequency range of the organ. Above 3 kHz the basal (high frequency) region moves somewhat more than the rest of the papilla. Phase differences between the motion of the hair cell cilia and underlying basilar papilla have been observed, but the dependence of these phase differences on frequency, location, and cilia length remains to be determined. Measurements of such relative phase (and amplitude) dependencies are planned in order to decide whether ciliary motion can provide a basis for observed frequency selectivity and tonotopic organization, and whether the characteristics of this motion are consistent with length-dependent ciliary mechanics.

References

1. G. von Bekesy, Experiments in Hearing, edited by E.G. Wever (McGraw-Hill, New York, 1960).
2. L.S. Frishkopf, "Mechanical Response Properties of the Basilar Papilla in Alligator Lizard: Failure to Find a Basis for Tonotopic Organization," M.I.T. Research Laboratory of Electronics Progress Report No. 123, pp. 206-207 (1981).

3. R. Leong, "Model for Signal Transmission in the Alligator Lizard Ear: From Drum Pressure to Receptor Potential of Hair Cells with Free-Standing Stereocilia," S.M. Thesis, M.I.T. 1982.
4. W.T. Peake and A. Ling, Jr., "Basilar-Membrane Motion in the Alligator Lizard: Its Relation to Tonotopic Organization and Frequency Selectivity," J. Acoust. Soc. Am. 67, 1736-1745 (1980).
5. T.F. Weiss, W.T. Peake, A. Ling, Jr., and T. Holton, "Which Structures Determine Frequency Selectivity and Tonotopic Organization of Vertebrate Cochlear Nerve Fibers? Evidence from the Alligator Lizard," in R. Naunton and C. Fernandez (Eds.), Evoked Electrical Activity in the Auditory Nervous System, (Academic Press, New York, 1978), pp. 91-112.