26. Communications Biophysics

A. Signal Transmission in The Auditory System

Academic and Research Staff

Prof. N.Y.S. Kiang, Prof. W.T. Peake, Prof. W.M. Siebert, Prof. T.F. Weiss, Dr. M.C. Brown, Dr. R.A. Eatock, Dr. D.K. Eddington, Dr. J.J. Guinan, Jr., Dr. E.M. Keithley, Dr. J.B. Kobler, Dr. W.M. Rabinowitz, Dr. J.J. Rosowski, J.W. Larrabee, F.J. Stefanov-Wagner, D.A. Steffens

Graduate Students

D.M. Freeman, M.L. Gifford, A.D. Krystal, M.P. McCue, X-D Pang, C. Rose

26.1 Basic and Clinical Studies of the Auditory System

National Institutes of Health (Grants 5 PO1 NS 13126, 1 RO1 NS 18682, and Training Grant 5 T32 NS 07047)

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

Studies of signal transmission in the auditory system continue in cooperation with the Eaton-Peabody Laboratory for Auditory Physiology at the Massachusetts Eye and Ear Infirmary. The goals of these studies are to determine the anatomical structures and physiological mechanisms that underlie vertebrate hearing and to apply that knowledge where possible to clinical problems. We report projects that were completed during 1983.

Mechanisms of mechanoelectric transduction in hair cells have been investigated in the alligator lizard.[1,4,5] Comparison of responses of hair cells and cochlear neurons to acoustic clicks and tone bursts have indicated that: adaptation of the response is a property of synaptic transmission at the receptor-neuron junction; frequency selectivity of the steady-state response to tones is a property of the mechanical input of the hair cell and/or a property intrinsic to each hair cell; important nonlinear phenomena accompany mechanoelectric transduction; the production of the receptor potential involves a low-pass filtering stage that follows mechanoelectric transduction.

Electric responses to sounds which are recorded outside the cochlea are widely used in auditory experiments to monitor the state of the cochlea as well as to assess its mechanical input. These responses are mixtures of potentials generated by different sources in the cochlea. To investigate their composition, electric responses to clicks and tones were recorded at the round windows of anesthetized alligator lizards before and after the neurotoxin "tetrodotoxin" (TTX) was added to scala tympani.[5] By combining click responses obtained in the presence and absence of TTX and at high and low click repetition rates, click responses were trisected into three components: (1) a rate-insensitive, TTX-insensitive component (that was identified as the cochlear microphonic potential or CM and assumed to be generated by hair cells); (2) a rate-sensitive, TTX-sensitive component (that was identified as the neural component); (3) a rate-sensitive, TTX-resistant component (which has not been identified previously and which was called component X). Component X is generated in the inner ear and has a latency between that of the CM and neural component. Several origins for component X are possible of which the most likely is that component X represents the compound post-synaptic potential of the nerve terminals. Measurements of responses to tones in the presence and absence of TTX demonstrate that the contribution of the neural component to the round-window response is appreciable below 1.5 kHz and negligible above this frequency.

The olivocochlear bundle (OCB) is a major efferent pathway from the central auditory system to the periphery. Both the anatomy[3] and physiology[2] of this pathway have been investigated in cats. The main anatomical finding is that the projections from both the medial superior olivary complex (MSOC) and from the lateral superior olivary complex (LSOC) to the cochlea are topographic but that while the LSOC mapping appears to connect locations with the same best frequencies, the MSOC mapping does not appear to do so. Electrical stimulation of the OCB at the midline of the brainstem is known to suppress sound-evoked responses of cochlear nerve fibers. A study was completed[2] in which the synchronized responses of fibers to tones were examined. The results are consisent with a model in which two processes that excite cochlear nerve fibers summate, and only one is affected by the electrical stimulation of the OCB.

The afferent neurons innervating hair cells in cats are of two types. Type I neurons innervate inner hair cells and Type II neurons innervate outer hair cells. A great deal is known about the physiological activity of the former and virtually nothing about the activity of the latter. An attempt was made to record from the central axons of Type II neurons with metal electrodes[7] (which have a low electrical noise level) and to look for long-latency events in response to electric shocks delivered to the cochlea. The rationale is based on the relatively small diameters of the axons of these neurons and their consequent small conduction velocities. It was possible to obtain long-latency spikes, but the amplitudes were barely above the noise levels of the electrodes. When sound was introduced, the electrode recorded gross responses of the cochlea that obscured these spikes. Thus it was concluded that other approaches to recording the activity of Type II neurons would need to be developed.

References

1. K. Baden-Kristenson and T.F. Weiss, "Receptor Potentials of Lizard Hair Cells with Free-Standing Stereocilia: Responses to Acoustic Clicks," J. Physiol. 335, 699-721 (1983).
2. M.L. Gifford and J.J. Guinan, Jr., "Effects of Crossed-Olivocochlear-Bundle Stimulation on Cat Auditory Nerve Fiber Responses to Tones," J. Acoust. Soc. Am. 74, 115-123 (1983).
3. J.J. Guinan, Jr., W.B. Warr, and B.E. Norris, "Differential Olivocochlear Projections from Lateral Versus Medial Zones of the Superior Olivary Complex," J. Comp. Neurol. 221, 358-370 (1983).

4. T. Holton and T.F. Weiss, "Receptor Potentials of Lizard Cochlear Hair Cells with Free-Standing Stereocilia in Response to Tones," J. Physiol. 345, 205–240 (1983).
5. T. Holton and T.F. Weiss, "Frequency Selectivity of Hair Cells and Nerve Fibers in the Alligator Lizard Cochlea," J. Physiol. 345, 241–260 (1983).
6. M.S. Kaplan, B.G. Szaro, and T.F. Weiss, "Components of Cochlear Electric Responses in the Alligator Lizard," Hearing Res. 12, 323–351 (1983).
7. N.Y.S. Kiang, E.M. Keithley, and M.C. Liberman, "The Impact of Auditory-Nerve Experiments on Cochlear Implant Design," Ann. N.Y. Acad. Sci. 405, 114–121 (1983).
8. E.A. Richter, B.E. Norris, B.C. Fullerton, R.A. Levine, and N.Y.S. Kiang, "Is There a Medial Nucleus of the Trapezoid Body in Humans?" Am. J. Anat. 168, 157–166 (1983).

B. Auditory Psychophysics and Aids for the Deaf

Academic and Research Staff

L.D. Braida, R. Boduch, H.S. Colburn, J. Coker, L.A. Delhorne, L.C. Dowdy, N.I. Durlach, C.L. Farrar, L.J. Ferrier, M.S. Florentine, D.M. Freeman, M. Furst, K.J. Gabriel, W. Jesteadt, J. Koehnke, N. Macmillan, K. Morey, W.M. Rabinowitz, C.M. Reed, J. Reid, R.P. Russell, M.C. Schultz, P.M. Zurek

Graduate Students

T.T. Allard, C.A. Bickley, D.K. Bustamante, M.M. Downs, K.K. Foss, E. Gilbert, D.M. Horowitz, Y. Ito, M. Jain, S.J. Leivy, D.F. Leotta, D. Opalsky, R.C. Pearsall II, J.C. Pemberton, P.M. Peterson, R.D. Reohr, R.J. Rohlicek, G.M. Skarda, M.J. Tsuk, R.M. Uchanski, R.G. Waissman

26.2 Intensity Perception and Loudness

National Science Foundation (Grant BNS77-16861)
National Institutes of Health (Grant 1 F 33 NS07202-01)
Louis D. Braida, Nathaniel I. Durlach, Neil Macmillan, William M. Rabinowitz, Michael J. Tsuk, Roberto G. Waissman

This research involves theoretical and experimental studies directed at a unified quantitative theory of intensity perception and loudness. It involves the development and integration of models of sensory processes, short-term memory, perceptual-context effects, and decision making, as well as psychophysical experimentation. During this period our work has consisted of i) development of a new model for context coding, ii) development of maximum-likelihood techniques for estimating sensitivity parameters, and iii) model-based analysis of categorical perception experiments.

i) The new model for context coding accounts for the increase in sensitivity, observed near the extremes of the range when unidimensional stimuli are identified, by assuming that sensations are estimated relative to noisy perceptual anchors with a stochastic ruler. In this model, the mean and variance of the anchor locations, and the mean number of ruler steps used to cover the sensation range, are the only free parameters. Further, these parameters are assumed to be independent of the stimulus parameters for the experiment. By fitting data (Braida and Durlach, 1972) from sets of intensity identification experiments in which the stimulus range was varied systematically, we have estimated that the mean anchor locations are roughly 2.0 jnd's outside the sensation range, the anchor variance is roughly 3.0 times the sensation variance, and roughly 35 steps are used to measure the sensation range. Independent estimates of the anchor position to sensation variance ratio have been derived based on the relation between sensitivity in fixed-level two-interval discrimination experiments and small-range one-interval identification experiments (1.2–4.0), and from data on the dependence of fixed-level discrimination on interstimulus interval (0.5–3.0). We plan to develop a model for the formation and maintenance of anchors in order to determine whether these parameters can be determined from more fundamental considerations, such as the composition of the stimulus set. A manuscript (Braida et al., 1984) describing this model has been accepted for publication.

ii) Estimates of sensitivity in identification experiments are often derived crudely merely by fitting straight lines to ROC curves by eye. Lippmann (1974) developed a procedure for computing maximum likelihood estimates under the assumption that the underlying densities were Gaussian and of equal variance. We are attempting to improve the computational efficiency of this procedure (which must solve sets of simultaneous nonlinear equations) and to relax the equal-variance assumption. Some preliminary work is reported on by Waissman (1983).

ii) We have begun to apply the Preliminary Theory of Intensity Resolution (Durlach and Braida 1969) to data obtained in categorical perception experiments. According to the theory, two memory modes are used in processing perceptual continua. In the trace mode, observers compare stimuli with the memory traces of other stimuli, and performance is limited by the inter-stimulus interval. In the context mode, observers compare stimuli to perceptual anchors, and performance is limited by the stimulus range. The theory predicts that sensitivity in discrimination experiments should be proportional to that in identification. For many continua used in categorical perception research, this prediction is upheld; for some (the fricative-affricate "continuum") it is not, suggesting that stimuli actually differ multidimensionally. Different stimulus domains were found to differ in (a) the amount of trace variance, (b) the amount of context-coding variance, and (c) the existence and location of anchors; but no single parameter captured the categorical/continuous distinction. Memory variances and anchor locations can be estimated from experiments in which fixed-level discrimination, as well as identification and roving-level discrimination, is measured. Among the few experiments with categorically-perceived continua that have used the critical fixed-level condition are some in which discrimination peaks arise from anchors, and others in which they reflect regions of high basic sensitivity. A manuscript reporting on this work has been prepared for publication (Macmillan, 1984).

References

1. L.D. Braida and N.I. Durlach, "Intensity Perception II: Resolution in One-Interval Paradigms," J. Acoust. Soc. Am. 51, 483–502 (1972).
2. L.D. Braida, N.I. Durlach, J.S. Lim, J.E. Berliner, W.M. Rabinowitz, and S.R. Purks, "Intensity Perception XIII: Perceptual Anchor Model of Context Coding," J. Acoust. Soc. Am., accepted for publication (1984).
3. N.I. Durlach and L.D. Braida, "Intensity Perception I: Preliminary Theory of Intensity Resolution," J. Acoust. Soc. Am. 46, 372–383 (1969).
4. R.P. Lippmann, "Effect of Payoff Matrix on Absolute Identification," S.M. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1974.
5. N.A. Macmillan, "Beyond the Categorical/Continuous Distinction: A Psychophysical Approach to

Processing Modes," in S. Harnad (Ed.) Categorical Perception, submitted for publication.
6. R.S. Waissman, "Maximum Likelihood Estimation of Decision Model Parameters," S.M. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 1984.

26.3 Binaural Hearing

National Institutes of Health (Grant 5 RO1 NS10916)

H. Steven Colburn, Nathaniel I. Durlach, Kaigham Gabriel, Manoj Jain, Janet Koehnke, Sander J. Leivy, David Opalsky, Patrick M. Zurek

Research on binaural hearing has consisted of i) experimental work, ii) theoretical work, and iii) development of facilities.

i) We have measured the effect of masker bandwidth and frequency on binaural detection for masker bandwidths from 6.4 Hz to 10 kHz and center frequencies of 250 and 4000 Hz. Specifically, a complete set of N0S0 and N0SPi thresholds were measured for two subjects using an adjustment procedure and for a third subject using a forced-choice, adaptive procedure. There were no significant differences between results from the two procedures. Both the N0S0 and the N0SPi data are being used in connection with modeling studies.

Also, we have continued with measurements of sensitivity to interaural correlation of narrowband noises in the presence of a spectral fringe of interaurally correlated noise. Preliminary results indicate that the presence of a spectral fringe interferes with detection of a narrowband stimulus more than one would expect from the critical bandwidth notion as it is applied in detection.

We are continuing to study the relation of detection to correlation. The outcomes of in-depth measurements from a small number of normal hearing subjects are testing our ability to predict detection from correlation and visa versa. Specifically, we are measuring psychometric functions for both tests. Early results, when plotted on common axes according to the calculated equivalent correlation coefficient or signal to noise ratio, indicate that the functions appear to be steeper for the detection experiment than for the correlation experiment. Further data are needed. These results will also be compared to those obtained from hearing-impaired subjects (Gabriel, 1983).

ii) In our theoretical work we continue to pursue both physiologically based models (e.g., Colburn, 1983) as well as stimulus-response models (e.g., Siegel and Colburn, 1983; Zurek, 1984). Although our analysis of physiological models is proceeding, most of our work this year has been focused on models that are easily described and analyzed without the complexity of the description of the peripheral transduction from pressure to neural firing patterns. We use available physiology as a guide, but believe that there are many models that can be made compatible with the physiology and that progress at this time is most likely to come from the black-box approach.

Predictions were made from the energy-detection model of Green and Swets for the N0S0 thresholds measured in the MLD-Bandwidth study. The fits are good with some choice of critical bandwidth (similar to values in the literature) and integration time (also similar to previous estimates). These quantities are of interest because the critical bandwidth estimate is used in predicting the N0SPi thresholds. Further, if it is assumed that one and the same integrator is effective in both homophasic and antiphasic detection tasks, then fitting the N0S0 data can provide an estimate of what we have been calling 'binaural sluggishness'. Indeed, an interesting prediction of this 'single-integrator' model is that as T (integration time) varies (from subject to subject), larger T's produce lower N0S0 thresholds and higher N0SPi thresholds. (The MLD thus varies inversely with T.) This co-variation is seen, roughly, in the data of our three subjects.

A simple model of binaural interaction incorporating a temporal averager for interaural differences has been developed and tested on results from both normal and hearing impaired subjects. This model has been successful at predicting NOSPi detection results and correlation discrimination results from observed interaural time and interaural intensity jnds.

Also, we have calculated the cross-correlation between the envelopes of band-pass noises that are identical except that they have an either an antiphasic tone or uncorrelated noise added. These quantities are needed to analyze detection and correlation discrimination at high frequencies where the envelope is assumed to be the effective binaural stimulus.

iii) Considerable efforts were applied to the development of two related but distinct experimental facilities, a VAX-based facility in the main laboratory and a stand-alone, 11/23-based, easily transportable facility that is designed for use in our anechoic chamber (Leivy, 1983; Opalsky, 1983). For the VAX-11/750 system, we have been developing a sophisticated stereo interface with microsecond interaural delay capabilities: the hardware is complete, but work still remains on the software development. For the 11/23-based system, a stand-alone experimental station designed for binaural experiments has been developed and is currently being used to perform a variety of different psychophysical experiments. A complete hardware/software package has been implemented which allows the manipulation and presentation of digitally stored or computed stimulus waveforms as well as the recording, storage and retrieval of subject responses. The experimental data recording routines have been developed around a data-base management system which enables the storage and retrieval of experimental results along with experimental and stimulus parameters. All of the software has been written the language FORTH. Because of the interpretive nature of this language, the user-system interaction is relatively simple compared with other experimental systems, and hence, novice users are more easily able to develop experimental programs.

References

1. H.S. Colburn, "Recent Developments in Binaural Modeling," in R. Klinke (Ed.), <u>Hearing-Physiological Bases and Psychophysics</u> 1983.
2. K.J. Gabriel, "Binaural Interaction in Hearing-Impaired Listeners," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., August 1983.

3. Sander J. Leivy, "A FORTH Program for Psychophysical Experiments," B.S. Thesis, Department of Biomedical Engineering, Boston University, 1983.
4. David Opalsky, "Binaural Psychophysical Experimental Facility," B.S. Thesis, Department of Biomedical Engineering, Boston University, 1983.
5. R.A. Siegel and H.S. Colburn, "Internal and External Noise in Binaural Detection," Hearing Res. 11, 117–123 (1983).
6. P.M. Zurek, "A Predictive Model for Binaural Advantages and Directional Effects in Speech Intelligibility," J. Acoust. Soc. Am., accepted for publication.

26.4 Hearing Aid Research

National Institutes of Health (Grant 5 RO1 NS12846)

Louis D. Braida, Diane K. Bustamante, Lorraine A. Delhorne, Nathaniel I. Durlach, Kristin K. Foss, Dennis M. Freeman, Eric Gilbert, Patrick M. Peterson, Miles P. Posen, Charlotte M. Reed, Roy P. Russell, Karen Silletto, Rosalie M. Uchanski, Victor W. Zue, Patrick M. Zurek.

This research is directed toward improving hearing aids for persons with sensorineural hearing impairments. We intend to develop improved aids and to obtain fundamental understanding of the limitations on such aids. The work includes studies of i) effects of noise on intelligibility, ii) amplitude compression, iii) frequency lowering, and iv) clear speech.

i) Although many persons with impaired hearing experience difficulty understanding speech in noise backgrounds which are not particularly deleterious for persons with normal hearing, the extent to which this reflects aspects of the impairments beyond loss of sensitivity is unknown. To obtain insight into this problem we have measured monaural speech intelligibility for CV syllables presented in a background of Gaussian noise with the spectral shape of babble. Speech and noise are added prior to spectral shaping with either a flat or a rising frequency–gain characteristic. Data has been obtained from five hearing-impaired subjects with roughly 40–60 dB losses and wide variety of audiometric configurations.

The results have been analyzed in terms of a model which assumes that the hearing loss is caused by an equivalent additive noise, (e.g., Dugal et al., 1980), and Articulation Theory is used to predict intelligibility. Plots of consonant identification scores versus Articulation Index for the conditions tested, including a few conditions with normal listeners, show a very strong degree of overlap. That is, these results indicate that, once the materials are 'calibrated' with normal listeners, consonant reception is highly predictable (within limits of test variability and between-subject variability in normal listeners) from knowledge of the speech and noise spectra and the listener's absolute thresholds.

ii) Many sensorineural hearing impairments are characterized by reduced dynamic range and abnormally rapid growth of loudness. Multiband amplitude compression has been suggested to improve speech reception for listeners with such impairments, but the intelligibility advantages associated with multiband compression are limited by distortion of speech cues associated with the short term spectrum (e.g., Braida et al., 1983, De Gennaro et al., 1984). To overcome this problem we are studying 'Principal Component Compression,' a means of inter-band control of compressor action that seems capable of achieving significant reductions in level variation with minimal distortion of the short term spectrum.

The basic structure of the principal component compression system is as follows. First, the speech signal is analyzed into bandpass components using a 16-channel critical-band filter bank. The bandpass signals are modulated to yield baseband signals. Estimates of the short-term RMS band level functions (envelopes) or band energies are derived by smoothing the squared baseband signals. The logarithms of the band energy signals provide estimates of the log-coded short-term speech amplitude spectrum. The principal component representation of the amplitude spectrum is obtained via the principal component transform; that is, the vector of band energies is multiplied by a matrix of the principal component basis vectors to yield a vector of principal component coefficients. The principal component coefficients are processed, which generally entails compression of one or more low-order coefficients. The inverse principal component transformation is applied to the compressed coefficients, another vector-matrix multiplication, yielding a set of modified band energies. The difference between the processed and unprocessed log band energies yields the band gain functions in dB. The original filter bank output signals are processed accordingly and summed to yield the processed speech signal.

Promising compression algorithms have been evaluated on the basis of three criteria; 1) effectiveness in reducing short-term speech level variations, 2) preservation of the shape of the short-term spectrum of various speech sounds and 3) intelligibility for normal hearing listeners in quiet and in noise. Results of these evaluations indicate that system configurations which substantially compress (e.g., by a factor of 50) only the first principal component or the first two principal components are most successful in reducing level variation across frequency while preserving important spectral cues. Compression of the first PC significantly compresses overall level and also provides some compression of spectral tilt. Compression of the second PC compresses spectral tilt, which effectively can provide high frequency emphasis for falling spectra and low frequency emphasis for rising spectra. However, the compression systems have been selected to avoid compressing the spectral tilt of rising spectra, which generally correspond to fricatives (i.e., /s/ and /sh/), as this would diminish high frequency energy in these spectra, the major cue for these sounds. Perceptual evaluations of two PC compression systems, one which compresses PC 1 and one which compresses both PC 1 and PC 2 will be performed on the basis of speech intelligibility experiments with severely impaired sensorineural hearing loss listeners.

iii) Frequency-lowering is a form of signal-processing intended to make high-frequency speech cues available to those who cannot hear high-frequency sounds. We have evaluated a frequency-lowering technique studied by Lippmann (1980). In this system speech levels in high-frequency bands modulate 1/3 octave bands of noise at low frequencies, which are then added to unprocessed speech. We found, in agreement with Lippmann, that processing improved the recognition of stop and fricative consonants when the listening bandwidth is restricted to 800 Hz. However, we also found that processing degrades the perception of nasals and (particularly) semivowels, consonants not included in Lippmann's study. We modified Lippmann's signal processing by reducing the level of the modulated noise when low-frequency components dominate the speech signal. Preliminary results indicate that the modified system does not degrade nasals and semivowels, but maintains the processing advantage for stops and fricatives. Further details of this work are available in Posen (1984).

A preliminary study of artificially coded consonants and vowels has been completed (Foss, 1983). In this study low-frequency (under 500 Hz) sounds have been synthesized to represent CV syllables. After training, identification scores were 90% correct, roughly 20 percentage points higher than those obtained on single tokens of low-pass filtered natural speech. These results suggest that, at present, the limitations to an effective frequency-lowering system depend upon signal processing constraints rather than auditory factors. Furthermore, the cues used in identifying the coded sounds appear to be relatively robust: the error rates and confusions observed for the full-size set were generally consistent with those obvserved on smaller training sets. For the filtered natural speech, performance was found to depend on the number of tokens used to represent each stimulus type: scores decreased from 70% to 35% when the type/token ratio increased from 1 to 3. This indicates that the cues used to identify small sets of filtered natural speech may prove unreliable when the set size is increased substantially. Comparison of these results with those of Miller and Nicely (1953), suggests that increases in the type-token ratio beyond 3 should have minimal effects on performance.

iv) Research on clear speech is directed toward understanding the improvements in intelligibility that can be achieved through techniques which focus on the speech source rather than signal processing. Previous research (e.g., Picheny, 1981; Chen, 1980) has established substantial intelligibility gains for "clear" speech relative to "conversational" speech. During the past year, work has focused on using principal components (PC) analysis to characterize the short term acoustic spectral differences between these two types of materials. PC basis vectors have been computed from the smoothed band energies of sixteen frequency bands with bandwidths that approximate critical bands.

To a first approximation, each of the first few PC basis vectors has a shape roughly independent of speaker and mode. This is consistent with the absence of long-term spectral differences (over speakers or modes) for these materials. In addition, the cumulative percent of the total variance in the first four basis vectors is constant, 91–93%, for all speakers and both speaking modes. However, we have also found that the sum of the variances of the smoothed energy bands increases from conversational to clear speech, i.e. there is a greater total variability in band levels for clear speech. In addition, PC1 (roughly corresponding to overall level) accounts for less of this variability in clear than in conversational speech for each speaker. By contrast, PC2 (roughly corresponding to spectral tilt) accounts for a larger portion of this variability in the clear speech of the talkers who achieved the greatest increase in intelligibility when speaking clearly. In an absolute sense, variation in spectral tilt increases from conversational to clear for each speaker studied, and is correlated with intelligibility, both within and across speaking modes. Equivalently, the proportion of total variability accounted for by PC1 is negatively correlated with intelligibility.

References

1. L.D. Braida, "Speech Processing for the Hearing Impaired," Proc. 2nd Int. Conf. on Rehab. Engr., Ottawa, Canada, 1984.
2. F.R. Chen, "Acoustic Characteristics and Intelligibility of Clear and Conversational Speech at the Conversational Level," S.M. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1980.
3. S.V. De Gennaro, L.D. Braida, and N.I. Durlach, "Multichannel Syllabic Compression for Severely Impaired Listeners," Proc. 2nd Int. Conf. on Rehab. Engr., Ottawa, Canada, 1984.
4. 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).
5. K.K. Foss, "Identification Experimentation on Low-Frequency Artificial Codes as Representation of Speech," S.B. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1983.
6. R.P. Lippmann, "Perception of Frequency Lowered Consonants," J. Acoust. Soc. Am. 67, S78 (1980).
7. G.A. Miller and P.E. Nicely, "An Analysis of Perceptual Confusions among Some English Consonants," J. Acoust. Soc. Am. 27, 338-352 (1953).
8. M.A. Picheny, "Speaking Clearly for the Hard of Hearing," Ph.D. Thesis, Dept. of Electrical Engineering and Computer Science, M.I.T., 1981.
9. M.P. Posen, "Intelligibility of Frequency Lowered Speech Produced by a Channel Vocoder," S.M. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1981.

26.5 Discrimination of Spectral Shape

National Institutes of Health (Grant 1 RO1 NS16917)

Louis D. Braida, Lorraine A. Delhorne, Nathaniel I. Durlach, Catherine L. Farrar, Mary S. Florentine, Yoshiko Ito, Charlotte M. Reed, Patrick M. Zurek

This research is concerned with determining the ability of listeners with normal and impaired hearing to discriminate stimuli with broadband, continuous, speechlike spectra, and to relate these measurements to underlying auditory abilities. The stimuli were generated by filtering Gaussian noise through a parallel synthesizer whose resonance parameters were adjusted to correspond to the spectra of steady-state unvoiced fricatives /f,s,sh/ and the burst portion of the unvoiced plosives /p,t,k/). In some tests, a masker having the spectral characteristics of cafeteria-noise babble was introduced. To prevent listeners from basing judgments on loudness cues, overall levels were roved. Discrimination data have been obtained for three experimental conditions (/p-t/ with 30-msec duration, /p-t/ with 300-msec duration, and /f-sh/ with 300-msec duration), each with a different group of five normal-hearing listeners. Roughly speaking, the shape of the psychometric function, d' vs. S/B (averaged over level pairs), is the same for all stimulus configurations and all subjects tested; this function is very shallow and has a slope of approximately 0.13 (change in d' per dB change in signal to babble ratio). Preliminary data have been obtained from two listeners with sensorineural loss indicate that they require a 20-30 dB increase in S/B to achieve discrimination performance comparable to that of listeners with normal hearing. The extent to which this shift in S/B should be assigned to effects of the impairment rather than to other differences between the subjects has not yet been determined.

A second set of experiments measured pure-tone thresholds in the background of a masker composed of a synthetic speech stimuli (30 msec /p/ and /t/ bursts) in quiet (where the /p-t/ discrimination was easy) or at S/B = −4 dB (where the discrimination was difficult). For all four masker complexes, the fixed-level masked thresholds generally follow the spectral shape of the masker at the three levels tested (45, 75, and 95 dB SPL). The most notable deviation from linearity of masking is observed at 125 Hz for the lower levels, where the threshold increase is smaller than the increase in masking level. Masking difference patterns, obtained by subtracting thresholds in the /p/-shaped masker complex at a given S/B from those obtained in the /t/-shaped masker complex at the same S/B, show (as expected) much larger differences in quiet than in babble. In quiet, these differences, which are several dB larger for the 75- and 95-dB SPL maskers than for the 45-dB masker, were highest in the frequency region 500-2000 Hz (averaging roughly −15 dB for the highest masker level) and smallest at 4000 Hz (3 dB). For S/B = −4 dB, the masking differences fluctuate around zero except at 3500 Hz for the 45-dB level, where the /t/-shaped masker complex provides roughly 7 dB more masking than the /p/-shaped masker.

In order to interpret the relation between spectral shape discrimination and basic auditory abilities, we are developing black-box models of both peripheral processing, which derives an internal spectrum from the physical input, and central processing, which relates the hypothesized internal spectrum to psychophysical performance. Our general approach to this modelling task is to start with models that are as simple as possible, apply them to an increasingly wide range of data, and elaborate them as dictated by these applications. Initial work on the peripheral processor has used a conventional filter-bank model and has assumed ideal central processing. Each filter output level is assumed to be estimated by rectification, integration, and a logarithmic transformation. The processing is corrupted by adding Gaussian internal noise after the level estimate. This model has been applied to our own data on discrimination and detection and also to results on frequency discrimination of tone bursts in broadband noise. For a given choice of filters and integration times, the model has one free parameter, K, corresponding to the level of noise added after the logarithmic transformation. Results show that (1) except for our own data on the detection of 30-msec tones in /t/-shaped masking noise (with no babble), all empirical curves can be reasonably well fit by some choice of model parameters; and (2) in the exceptional case just mentioned, the best fitting theoretical curve is too high at high frequencies and too low at low frequencies. The values of K required to fit the data vary by a unacceptable amount. However this variation can be substantially reduced if the effects of external noise fluctuations are taken into account. Our work on ideal central processing has assumed, for simplicity, that the internal spectrum (with the roving level included) can be characterized as an N-dimensional random vector whose components are Gaussian random variables. We have computed the optimal performance of such a receiver as a function of the variance of the roving overall level, and shown that this performance is equivalent to that for a receiver which processes only level-differences between adjacent bands optimally. In general, the roving of overall level is predicted to have only small effects on the discriminability of stimulus pairs equated in terms of overall energy or loudness.

26.6 Tactile Perception of Speech

National Institutes of Health (Grant 1 RO1 NS 14092-05)
National Science Foundation (Grant BNS 77 21751)
Louis D. Braida, Raymond Boduch, Jackie Coker, Lorraine A. Delhorne, Maralene M. Downs, Leonard C. Dowdy, Nathaniel I. Durlach, Daniel F. Leotta, Joseph C. Pemberton, William M. Rabinowitz, Charlotte M. Reed, Jean Reid, Richard D. Reohr, J. Robin Rohlicek, Roy P. Russell, Martin C. Schultz, Gregory M. Skarda

The ultimate goal of our research program is to develop tactile aids for the deaf and deaf-blind that will enable the tactile sense to serve as a substitute for hearing. Among the various components of our research in this area are (i) study of tactile communication methods employed by the deaf-blind, (ii) development of an augmented Tadoma system, (iii) development of a synthetic Tadoma system, and (iv) design and evaluation of a wearable, portable aid.

i) Experimental work is being conducted on two methods of communication (tactile fingerspelling and tactile signing) that are in common use among members of the deaf-blind population. One goal of our research is to examine the rates of communication attainable through these two methods and to compare these rates to those obtained through the Tadoma method. For tactile fingerspelling, five deaf-blind subjects have been tested on several sentence-repetition tasks and on tracking of continuous text. Sentences were presented by an experienced fingerspeller at rates varying from roughly 2 to 6 letters/sec. For "conversational" sentences, experienced deaf-blind receivers of fingerspelling achieved perfect scores for rates at and below 5 letters/sec (the equivalent of 1.5 syllables/sec, which is roughly three times slower than normal speaking rates). These subjects were able to track text at an average rate of roughly 30 words/min (a result similar to that obtained by experienced Tadoma users and roughly one-third of that obtained under normal auditory conditions). For tactile signing, the sentence materials used in the Tadoma and fingerspelling research have been translated into ASL and videotaped by a native signer who then serves as the "sender" in tests with experienced deaf-blind users of ASL. Results (currently available for three subjects) are being analyzed by glossing the subjects' responses and determining their semantic and syntactic accuracy as a function of rate of presentation in signs/sec.

ii) Work is underway to evaluate an "augmented" Tadoma system in which the normal Tadoma signal (received by placing a hand on the talker's face and neck) is supplemented by information on tongue position (derived by sensing the contact pattern of the tongue with the upper palate and displayed to a finger of the opposite hand through a tactile transducer array). A programmable interface has been developed between a Palatograph and an Optacon transducer and software allows tabular specification of vibrator action on a 24x6 array corresponding to 63 palate contact points. The mapping currently in use preserves the general shape of the palate on the vibratory array. Current experiments are concerned with the ability of laboratory subjects using augmented Tadoma to discriminate pairs of stimuli that contrast different tongue positions and are difficult to discriminate using normal Tadoma. Experiments will then be extended to include other subjects (deaf-blind Tadoma users) and other types of tasks and test materials (e.g., identification of speech segments and reception of continuous speech).

iii) To better understand the success of the Tadoma method and to explore transformations of the method that cannot be achieved directly, we are developing a synthetic Tadoma system in which an artificial talking face is driven by signals recorded from the facial actions of a real talking face. During the past year work has continued on the processing (digitizing, editing, filtering, etc.) of the tape-recorded facial-action signals into computer files appropriate for driving the artificial face (Skarda, 1983). The library has been substantially increased to include additional speech materials from the original male talker and materials from a second male talker and two female talkers.

On the artificial face, a four-channel position-control system has been constructed to drive the DC servomotors which effect the lip and jaw movements. Each motor operates closed-loop with feedback of shaft velocity (tachometer) and position (optical encoder). The system response (of shaft position over input voltage) is set to approximate a third-order Butterworth lowpass characteristic with a 3-dB cutoff of 32 Hz, which allows for (smoothly) driving the articulators at normal and higher rates. An interface (of specialized hardware and software) for outputing the facial-action signal files from our laboratory computer system to the artificial-face electronics has also been completed. Artificial-face representations for the two remaining facial actions, laryngeal vibration and oral airflow, are currently being installed and experiments will soon begin with laboratory-trained normal subjects and deaf-blind Tadoma users to assess the adequacy of the system as a simulation of Tadoma.

iv) A further project in this area concerns the development of a multi-channel, wearable/portable, tactile aid. (This project is being conducted under the auspices of the Rehabilitation Engineering Center at Gallaudet College.) The aid incorporates a linear array of vibrators and displays short-term spectral information to the abdomen. Efforts during the past year have focused on laboratory-based experiments to guide various design choices (Rohlicek, 1983). These experiments have provided rough estimates of absolute thresholds, intensity and frequency discrimination thresholds, modulation thresholds, and spatial and temporal masking thresholds. In addition, preliminary discrimination experiments were performed using a computer-controlled eight-channel tactile vocoder on speech segments to explore the effects of various signal processing schemes, e.g., amplitude compression (to combat small dynamic range), temporal truncation (to combat masking of consonants by vowels), and spectral peak picking (to combat spatial masking). None of the processing schemes tested improved the discrimination of speech segments significantly. Work has now begun on hardware development of the wearable aid.

References

1. J.R. Rohlicek, "Some Perceptual Constraints on the Tactile Coding of Speech," S.M. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1983.
2. G.M. Skarda, "An Analysis of Facial Action Signals during Speech Production Appropriate for a Synthetic Tadoma System," S.M. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1983.

C. Transduction Mechanisms in Hair Cell Organs

Academic and Research Staff

Prof. L.S. Frishkopf, Prof. T.F. Weiss, Dr. D.J. DeRosier[21], 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 stereocilia; 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.

26.7 Length-Dependent Mechanical Tuning of Free-Standing Stereociliary Bundles in the Alligator Lizard Cochlea is the Basis of Neural Tuning

National Institutes of Health (Grants 5 R01 NS 11080 and GM-21189)

Lawrence S. Frishkopf, David J. De Rosier

As reported earlier,[1] we have studied the motion in response to sound of free-standing stereociliary bundles of hair cells in the alligator lizard basilar papilla as a function of stimulus frequency, hair cell location, and bundle length. We have established that such motion provides a basis for frequency selectivity and tonotopic organization observed in nerve fibers to the organ. Our results indicate (1) that stereociliary bundles behave like damped mechanical resonators; (2) that resonant frequency varies inversely with bundle length raised to a power between 1.5 and 2; and (3) that bundle resonant frequency and neural CF are close in value in corresponding regions of the papilla and nerve. These findings are consistent with an explanation of frequency analysis in the papilla based on length-dependent mechanical tuning of stereociliary bundles.

We have reported these results fully in a recent paper;[2] other findings that corroborate this interpretation have been recently published as well.[3]

References

1. L.S. Frishkopf and D.J. DeRosier, "Evidence of Length-Dependent Mechanical Tuning of Hair Cell Stereociliary Bundles in the Alligator Lizard Cochlea: Relation to Frequency Analysis," M.I.T. Research Laboratory of Electronics Progress Report No. 125, pp. 190-192 (1983).
2. L.S. Frishkopf and D.J. DeRosier, "Mechanical Tuning of Free-Standing Stereociliary Bundles and Frequency Analysis in the Alligator Lizard Cochlea," Hearing Res. 12, 393-404 (1983).

[21]Professor, Department of Biology, Brandeis University

3. T. Holton and A.J. Hudspeth, "A Micromechanical Contribution to Cochlear Tuning and Tonotopic Organization," Science 222, 508–510 (1983).