

18.0 Communication Biophysics

Academic and Research Staff

Prof. L.D. Braid, Prof. L.S. Frishkopf, Prof. N.Y.S. Kiang, Prof. W.T. Peake, Prof. W.M. Siebert, Prof. T.F. Weiss, Prof. V.W. Zue, Dr. M.C. Brown, Dr. D.K. Bustamante, Dr. H.S. Colburn, Dr. B. Delgutte, Dr. R.A. Eatock, Dr. L.A. Delhorne, Dr. N.I. Durlach, Dr. D.K. Eddington, Dr. J.J. Guinan, Jr., Dr. D.R. Ketten, Dr. J.B. Kobler, Dr. W.M. Rabinowitz, Dr. J.J. Rosowski, Dr. P.M. Zurek, I.A. Boardman, R.M. Brown, M.L. Curby, J.A. Frisbie, K.W. Grant, J.D. Koehnke, N.A. Macmillan, E.M. Luongo, K. Payton, C.M. Reed, B. Schneider, F.J. Stefanov-Wagner, D.A. Steffens

Graduate Students

C.A. Bickley, C.R. Corbett, S. Dynes, D.M. Freeman, D.A. Frost, G. Girzon, R.P. Goldberg, Y. Ito, R.C. Kidd, G. Kline, D.F. Leotta, M.E. Machado, J.R. Melcher, X-D Pang, C. Passero, S.L. Phillips, M.H. Power, P.M. Peterson, L. Stephens, R.M. Uchanski

18.1 Basic and Clinical Studies of the Auditory System

National Institutes of Health (Grants 5 PO1 NS13126, 5 RO1 NS18682, 5 RO1 NS20322, 5 RO1 NS20269, and 5 T32NS 07047)

Symbion, Inc.

Lawrence S. Frishkopf, Nelson Y.S. Kiang, William T. Peake, William M. Siebert, Thomas F. Weiss, M. Christian Brown, Bertrand Delgutte, Ruth A. Eatock, Donald K. Eddington, John J. Guinan, Jr., Darlene R. Ketten, J.B. Kobler, William M. Rabinowitz, John J. Rosowski, Ian A. Boardman, Robert M. Brown, Mark L. Curby, Frank J. Stefanov-Wagner, D.A. Steffens, Scott B.C. Dynes, Dennis M. Freeman, Gary Girzon, R.C. Kidd, J.R. Melcher, Xiao-Dong Pang, Susan L. Phillips

Investigations of signal transmission in the auditory system are being carried out in cooperation with the Eaton-Peabody Laboratory for Auditory Physiology at the Massachusetts Eye and Ear Infirmary. The long-term objective is to determine the anatomical structures and physiological mechanisms that underlie vertebrate hearing and to apply that knowledge to clinical problems. Studies of cochlear implants in humans continue in the Cochlear Implant Laboratory in a joint program with the Massachusetts Eye and Ear Infirmary. The ultimate goal of these devices is to provide speech communication for the deaf by using electric stimulation of intracochlear electrodes to elicit patterns of auditory nerve fiber activity that the brain can learn to interpret.

18.1.1 External and Middle-Ear Transmission

Darleen R. Ketten, Xiao-Dong Pang, William T. Peake, Susan L. Phillips, John J. Rosowski

To determine the interaction between the transformations of acoustic signals performed by the external ear and by the middle ear, it is important to know not only the input impedance of the middle ear but also the output (or radiation) impedance of the

external ear. We measured the radiation impedance of the external ear of cat and evaluated the ear's performance as a coupler of acoustic power.¹ The results are important to our understanding of middle-ear function in that they allow us to calculate the impedance matching capabilities of the middle ear and also allow computations of power flow through the external and middle ears. Our measurements suggest that the pinna-flange acts to damp high-frequency resonances in the radiation impedance; the low-frequency impedance is primarily determined by the concha and external canal. Whereas the cat external and middle ear act as a nearly perfect (i.e., matched) acoustic-power coupler at 4 kHz, they are far from perfect at other frequencies in the audible range. The measure of acoustic performance that can be determined from these measurements, diffuse-field pressure ratio and diffuse-field absorption cross-section, will be important in comparing the performance of the ears of different species. For instance, it may be that some species have evolved highly directional external ears with decreased ability to absorb acoustic power.

Two earlier studies have been published.^{2,3}

18.1.2 Cochlear Mechanisms

Ruth A. Eatock, Dennis M. Freeman, R.C. Kidd, Thomas F. Weiss, Lawrence S. Frishkopf

A theoretical study of the hydrodynamic stimulation of the hair bundles of hair cells has been completed⁴ and some of the results have been published⁵ and presented at scientific meetings.⁶

A theoretical study of the signal processing properties of calcium channels was completed⁷. Models of calcium channel gating and calcium ion accumulation were investigated. It was found that these two mechanisms can contribute substantially to the loss of synchronized responses of cochlear nerve fibers in the ear. Thus there are three important lowpass filter mechanisms in the cochlea. They result from the membrane capacitance of the hair cell, the kinetics of gating of calcium channels, and the kinetics of intracellular accumulation of calcium channels.

18.1.3 Stimulus Coding in the Auditory Nerve

Bertrand Delgutte

A fundamental problem in auditory theory is to reconcile the wide range of stimulus level over which listeners make fine spectral discriminations with the relatively narrow dynamic range of auditory-nerve fibers. That is, the discharge rate of auditory-nerve fibers grows with stimulus level only over a 30-dB range between threshold and saturation. It has been suggested that such saturation might limit psychophysical performance at high sound levels, and that certain important features for discrimination might be encoded in temporal patterns of auditory-nerve activity ("phase-locking"). In the past year,^{8,9} we used a detection theoretic approach to examine quantitatively how much information for intensity discrimination is available in the discharge patterns of auditory-nerve fibers. This line of research is based on the principle that the stochastic behavior of auditory neurons imposes fundamental limitations on the performance that is achievable in any detection or discrimination task (Siebert, *Kybernetik* 2, 206-215, 1965). Specifically, we measured intensity difference limens (DL's) of single

auditory-nerve fibers for both tone and noise stimuli using paradigms and detectability measures that explicitly mimic those of psychophysics. These physiological data were then incorporated into a model that combines intensity information from many auditory-nerve fibers. This model is the first to include both a physiologically realistic distribution of fiber thresholds and multiple, frequency-selective channels.

Responses of auditory-nerve fibers to pairs of tone or broadband noise stimuli with different levels were recorded in anesthetized cats. Intensity DL's of single fibers were measured by adjusting the level of difference until the most intense stimulus evoked more spikes than the weakest stimulus for about 75% of the presentations. For both tones and noise, variations in physiological DL's with intensity had basically a U shape, with a minimum (meaning best performance) in the range where discharge rate grows rapidly with level. Minimum DL's ranged between 1 and 7 dB, implying that a small number of fibers could account for psychophysical performance (about 1 dB). However, this good correspondence between physiological and psychophysical DL's occurs only over a narrow (20-40 dB) range of levels, so that, in order to account for psychophysical performance over a broad range of levels, information from fibers with different thresholds must be combined.

Our model for combining intensity information from an array of 30,000 fibers was based on the optimum combination rule of signal-detection theory, under the assumption that the spike counts of different fibers are statistically independent Gaussian random variables. For both tones and noise, model predictions well exceeded psychophysical performance in intensity discrimination over at least a 90-dB range of stimulus levels. This implies that psychophysical performance is not limited by saturation of auditory-nerve fibers, but by the efficiency with which the central auditory system processes the information available at the level of the auditory nerve.

This conclusion is reinforced by the result that the dependence of model predictions on intensity clearly differed from that of psychophysical DL's; model predictions show a degradation at high stimulus levels, whereas psychophysical performance is either constant (for broadband noise) or improves with level (for tones). More realistic level dependence of predicted DL's was obtained by assuming that the central processor gives more "weight" to information from high-threshold fibers than to that from low-threshold fibers. This modified model also provided a robust representation of the spectra of speech sounds with respect to variations in stimulus level.

18.1.4 Afferent and Efferent Systems in Mammalian Cochlea

M. Christian Brown

Our goal has been to understand the functional significance of the several types of afferent and efferent innervations of the mammalian cochlea. We have used techniques involving horseradish peroxidase labeling of neurons to resolve several anatomical issues which have functional significance.^{10,11,12} One specific goal is to determine the central projections of type-II spiral ganglion cells. Peripherally in the cochlea, these ganglion cells provide the afferent innervation of the outer hair cells. In material from the guinea pig cochlea, central axons of labeled type-II ganglion cells could be traced from the spiral ganglion, through the auditory nerve, and into the cochlea nucleus, which is the first auditory nucleus in the central nervous system. Although these axons

followed the course taken by the thick, myelinated axons from type-I ganglion cells, the type-II axons were much thinner (about $0.5 \mu\text{m}$ diameter) and appeared to be unmyelinated. Conduction times which were computed on the basis of these diameters indicate that information from afferent terminals on hair cells arrives at the cochlear nucleus in 0.3 msec (type-I) or 6.1 msec (type-II).¹⁰ The functional significance of these observations is that the type-II neurons play a role in auditory processes which involve slow time courses. Ongoing studies in mice and gerbils indicate that the central projections of type-I and type-II ganglion cells are topographically similar, although some type-II central axons do not innervate the dorsal cochlear nucleus.

We have used similar techniques to reconstruct the cochlear projects of olivocochlear efferent neurons. For efferents, the unmyelinated fibers innervate near the inner hair cells whereas the myelinated fibers innervate the outer hair cells, which is the converse of the afferent innervation pattern. Unmyelinated efferent fibers innervating near inner hair cells have been shown to fall into two categories: those with very discrete terminations which might influence discrete regions of afferent fibers, and those with very broad terminations which presumably can influence afferent fibers conveying information from a large range of frequency regions.¹¹ Myelinated efferents innervating outer hair cells have also been reconstructed but it has not yet been possible to establish categories based on the terminal patterns of these fibers. Ongoing labeling studies of these myelinated fibers at the single-unit level in cat¹² and guinea pig may succeed in establishing categories for their peripheral innervation patterns.

18.1.5 Middle-Ear Muscle Reflex

John J. Guinan, Jr., James B. Kobler, Xiao-Dong Pang, Michael P. McCue

We aim to determine the structural and functional bases of middle-ear reflexes. We have continued our work recording from individual stapedius motoneurons in Ketamine-anesthetized or decerebrate cats. We have now distinguished five groups of motoneurons. In addition to the four groups previously distinguished on the basis of the laterality of their responses to sound, we have now distinguished a fifth category, motoneurons with spontaneous activity. These neurons typically respond to sound in either ear and have axons with conduction velocities that are approximately half as large as all of the other groups (based on the spike latency from electrical stimulation in the floor of the fourth ventricle). Manuscripts describing this work and related work are in preparation.

The cell bodies of stapedius motoneurons are located in four loose groups around the facial nucleus and near the descending facial nerve. We have attempted to find correlations between these four anatomical groups and our five physiological groups by injecting single, physiologically characterized stapedius motoneurons with HRP and indentifying their cells of origin in the brain. To date, over twenty stapedius motoneurons have been characterized and identified. Motoneurons in all classes, except spontaneously active motoneurons, have been labeled. Stapedius motoneurons which responded only to sound in the contralateral ear were found only in the ventromedial perifacial nucleus and no other type was found there. The other neuron types were located throughout the remaining stapedius motoneuron cell groups with no strong pattern emerging. These results are consistent with, and greatly extend, our earlier results based on gross lesions.

We have recently begun two projects, one aimed at determining the amplitudes and time courses of the middle-ear transmission changes produced by individual stapedius motor units and muscle fibers, and the second aimed at determining the effects of stapedius-muscle contractions on the responses of single auditory-nerve fibers. During the past year we completed a manuscript on the asymmetries in the acoustic reflexes of the cat stapedius muscle.¹³ Based on measurements of electromyographic activity (EMG), we found that the crossed stapedius reflex had a maximum amplitude which was about 1/3 of the maximum amplitude of the uncrossed reflex. This indicates that the crossed reflex probably produces correspondingly smaller changes in middle-ear sound transmission than the uncrossed reflex. Since most measurements of the effects of middle-ear reflexes have used only crossed sound, it seems likely that the effects of middle-ear reflexes due to uncrossed or binaural sound have been underestimated.

18.1.6 Cochlear Efferent System

John J. Guinan, Jr.

Our aim is to understand the physiological effects produced by medial olivocochlear (MOC) efferents which terminate on outer hair cells. To do this we have measured the sound pressure in the ear canal generated by a sound source, with and without electrical activation of MOC fibers by an electrode in the fourth ventricle (OCB stimulation). The efferent-induced change in the ear-canal sound pressure, P_{oc} , can be thought of as a change in a sound of cochlear origin. Measurements of P_{oc} are of interest because P_{oc} gives a measure (albeit indirect) of the MOC-induced change in cochlear mechanics. Using continuous tones, we found that: (1) the phase delay of P_{oc} was approximately a monotonic increasing function of sound frequency¹⁴ and (2) group delays calculated from P_{oc} phase vs. sound frequency were generally longer at lower frequencies. During the past year we have implemented a paradigm in which similar experiments are done using short tone 'pips' instead of continuous tones. These experiments show that for low frequencies, the sound in the ear canal outlasts by a few ms the sound produced by the sound source (i.e., there appears to be a 'cochlear echo') when there is no OCB stimulation. With OCB stimulation, however, this 'cochlear echo' is greatly reduced. Thus, OCB stimulation appears to suppress cochlear echoes. These data provide the interpretation that P_{oc} measured with continuous tones is the suppression of a cochlear echo.

During the past year we have analyzed and prepared for publication work done with M.L. Gifford. A paper has been submitted¹⁵ on the effects of electrical stimulation of efferent neurons with an electrode either at the origin of the MOC efferents (MOC stimulation) or at the floor of the fourth ventricle (OCB stimulation). Our results indicate that the effects produced by both kinds of stimulation are attributable only to MOC efferents and not to lateral olivocochlear (LOC) efferents. Furthermore, OCB stimulation in the cat probably stimulates both crossed and uncrossed MOC efferents. These experiments provide a new interpretation of the neural elements responsible for the empirical findings reported in previous experiments in which auditory efferents were electrically stimulated.

During the past year, two previously submitted papers were published. The anatomy and organization of the auditory efferents was reviewed in a joint paper with W.B. Warr and J.S. White of Boys Town National Institute.¹⁶ In a joint paper with other researchers

in the Eaton-Peabody Laboratory, data from changes induced in the firing patterns of auditory-nerve fibers were used to consider the interrelationships among the effects produced by the efferent system, acoustic trauma, ototoxic lesions, and pharmacological manipulation of endocochlear potential.¹⁷

18.1.7 Cochlear Implants

Donald K. Eddington, Gary Girzon

This year work centered in three areas: 1) electrical modeling of the human cochlea; 2) electrical and psychophysical measures in human subjects with implants; and 3) preparation of a program project grant that was submitted to and funded by the NIH.

The goal of work in the first area is to construct a software model of the cochlea that predicts the patterns of current flow due to the stimulation of arbitrarily placed intracochlear electrodes. Human temporal bone sections were digitized and resistivities assigned to the major cochlear elements (e.g., bone, nerve, perilymph, endolymph). Finite element techniques were used to convert the anatomical and resistivity data to a set of equations representing a three-dimensional mesh of 512 by 512 by 46 nodes. Current sources were defined at nodes representing positions of intracochlear electrodes in implanted subjects and the equations solved. Maps of nodal potentials located along the length of the spiraling scala tympani displayed a monotonic reduction of potential from the stimulating electrode toward the base and a potential plateau toward the apex. For basal stimulating electrodes, "bumps" in the potential plateaus between 15 and 20 mm from the base indicated current pathways between turns in addition to the pathway along the length of the cochlea.

Measurements of potentials at unstimulated electrodes made in five human subjects implanted with intracochlear electrodes demonstrated the asymmetric potential distributions predicted by the model. The "bumps" predicted by the model were also present in the potential distributions measured during the stimulation of the most basal of the implanted electrodes in all five subjects. Psychophysical measures are continuing to determine if these asymmetric potential distributions are also reflected in aspects of the subject's perception.

September marked the beginning of an NIH Program Project Grant that supports individuals from Harvard and M.I.T. who investigate a wide range of issues related to cochlear implants.

References

- ¹ J. Rosowski, L.H. Carney, and W.T. Peake, "The Radiation Impedance of the External Ear of Cat: Measurements and Applications," submitted to *J. Acoust. Soc. Am.*
- ² J. Rosowski, J.H. Carney, T.J. Lynch, III, and W.T. Peake, "The Effectiveness of External and Middle Ears in Coupling Acoustic Power into the Cochlea," *Peripheral Auditory Mechanisms*, edited by J.B. Allen, J.L. Hall, A. Hubbard, S.T. Neely, and A. Tubis, New York: Springer-Verlag, 1986, pp. 3-12.

- ³ X-D. Pang and W.T. Peake, "How do Contractions of the Stapedius Muscle Alter the Acoustic Properties of the Ear?," *Peripheral Auditory Mechanisms*, edited by J.B. Allen, J.L. Hall, A. Hubbard, S.T. Neely, and A. Tubis, New York: Springer-Verlag, 1986, pp. 36-43.
- ⁴ D.M. Freeman, "Hydrodynamic Study of Stereociliary Tuft Motion in Hair Cell Organs," Ph.D. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1986.
- ⁵ D.M. Freeman and T.F. Weiss, "On the Role of Fluid Inertia and Viscosity in Stereociliary Tuft Motion: Analysis of Isolated Bodies of Regular Geometry," *Peripheral Auditory Mechanisms*, edited by J.B. Allen, J.L. Hall, A. Hubbard, S.T. Neely, and A. Tubis, New York: Springer-Verlag, 1986, pp. 147-154.
- ⁶ D.M. Freeman and T.F. Weiss, "Hydrodynamic Study of Stereociliary Tuft Motion," The IUPS Satellite Symposium on Hearing, San Francisco, California, July, 1986,
- ⁷ R.C. Kidd, "A Theoretical Study of the Signal Processing Properties of Calcium Channel Kinetics in the Peripheral Auditory System," S.M. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1986.
- ⁸ B. Delgutte, "Intensity Discrimination: Relation of Auditory-Nerve Activity to Psychophysical Performance," *J. Acoust. Soc. Am.* *80*, S48 (1986).
- ⁹ B. Delgutte, "Peripheral Auditory Processing of Speech Information: Implications from a Physiological Study of Intensity Discrimination," *In The Psychophysics of Speech Perception*, edited by B. Schouten (in press).
- ¹⁰ M.C. Brown, "Morphology of Labeled Efferent Fibers in the Guinea Pig Cochlea," *J. Comp. Neurol.* (in press).
- ¹¹ M.C. Brown, "Morphology of Labeled Efferent Fibers in the Guinea Pig Cochlea," *J. Comp. Neurol.* (in press).
- ¹² M.C. Liberman and M.C. Brown, "Physiology and Anatomy of Single Olivocochlear Neurons in the Cat," *Hear. Res.* *24*, 17 (1986).
- ¹³ J.J. Guinan, Jr. and M.P. McCue, "Asymmetries in the Acoustic Reflexes of the Cat Stapedius Muscle," *Hear. Res.* (in press).
- ¹⁴ J.J. Guinan, Jr., "Effect of Efferent Neural Activity on Cochlear Mechanics." *Scand. Audiol. Suppl.* *25*, 53 (1986).
- ¹⁵ M.L. Gifford and J.J. Guinan, Jr., "Effects of Electrical Stimulation of Medial Olivocochlear Neurons on Ipsilateral and Contralateral Cochlear Responses," submitted to *Hear. Res.*
- ¹⁶ W.B. Warr, J.J. Guinan, Jr., and J.S. White, "Organization of the Efferent Fibers: The Lateral and Medial Olivocochlear Systems." *Neurobiology of Hearing: The Cochlea*, edited by R.A. Altschuler, D.W. Hoffman, and R.P. Bobbin. New York: Raven Press, 1986, pp. 333-348.

¹⁷ N.Y.S. Kiang, M.C. Liberman, W.F. Sewell, and J.J. Guinan, Jr., "Single-Unit Clues to Cochlear Mechanisms." *Hear. Res.* 22, 171 (1986).

18.2 Auditory Psychophysics and Aids for the Deaf

Academic and Research Staff

L.D. Braida, D.K. Bustamante, H.S. Colburn, L.A. Delhorne, N.I. Durlach, J.A. Frisbie, K.W. Grant, J.D. Koehnke, N.A. Macmillan, E.M. Luongo, K. Payton, W.M. Rabinowitz, C.M. Reed, B. Schneider, V.W. Zue, P.M. Zurek

Students

C.A. Bickley, C.R. Corbett, D.A. Frost, R.F. Goldberg, Y. Ito, G. Kline, D.F. Leotta, M.E. Machado, C. Passaro, M.H. Power, P.M. Peterson, L. Stephens, R.M. Uchanski

18.3 Role of Anchors in Perception

National Science Foundation (Grants BNS 83-19874 and BNS 83-19887)

Louis D. Braida, Nathaniel I. Durlach, Neil A. Macmillan, L. Stephens

This research seeks to provide a unified theory for identification and discrimination of stimuli which are perceptually one-dimensional (e.g., sounds differing only in intensity, Braida and Durlach, 1986). During this period, we examined the identification and (fixed-level) discrimination of synthetic consonant-vowel syllables differing in VOT or in place of articulation.

Stimuli were constructed using Klatt's (1979) software synthesizer. The VOT stimuli consisted of an 85-msec bilabial stop with a VOT ranging from 0 to 36 msec, followed by a 273-msec vowel, /a/. There were nine stimuli; step size was 6 msec, except for the stimuli with a VOT between 12 and 24 msec, where it was 3 msec. The place continuum was created by varying the linear transitions of the first three formants over the first 40 msec. All stimuli in this set were 400 msec in duration.

In the VOT experiments, sensitivity reached a peak in mid-range in all tasks. The peak in fixed discrimination implies a region of high basic sensitivity near 15 msec, since labeling, phonemic or otherwise, plays only a minor role in that task. Similar results were obtained in the place-of-articulation experiments: two peaks appear in the fixed data, one between each phonemic category, and sensitivity follows the same qualitative pattern in all tasks.

The relative size of context-coding noise can be estimated by comparing sensitivity in identification and fixed discrimination tasks. For VOT, context-coding variance is roughly 2/3 the size of sensory variance; for place of articulation the variances are roughly equal. Thus the ability to identify these consonants could not be much better, given the inevitable sensory limitations.

For VOT, sensitivity in identification tasks is best for the extreme /ba/ and /pa/. For place, sensitivity is best near the center of a phonemic category. The best remembered consonants on these continua appear to be the “best” exemplars; other syllables are identified with reference to those prototypes.

It is interesting to compare the present results with those from the vowel studies, and also with our laboratory’s experiments of the tone-intensity continuum. First, of the three continua, only consonants exhibit peaks in the basic (sensory) sensitivity function: peaks are observed on vowel and intensity continua only when memory limitations are significant. Second, the range (in jnds) of consonant continua is not necessarily less than that of vowel continua. Ades (1977) speculated that some processing differences between consonants and vowels might arise from differences in this range. But the total number of jnds between /ba/ and /pa/ is about the same as between /i/ and /l/. Third, context variance is greater for vowels than for consonants; that is, it is easier to label consonants than vowels. Notice that this is true even for two-category continua, where the jnd range is the same. An intensity range of 6 JND’s has a context variance roughly 1.5 times the sensory variance, while a range of 15 JND’s has a context variance roughly 4 times the sensory variance. For comparable ranges, vowels are more similar to intensity than are consonants with respect to the growth of context variance. Fourth, the stimuli with regard to which identification judgments are made, are different for all three continua. For consonants, they are prototypes; for vowels, stimuli near category boundaries are remembered best; and for intensity, stimuli at the edges that serve as anchors.

The experimental work has required two types of analysis: decision-theoretic (to separate measures of sensitivity from measures of response bias) and statistical (to estimate the reliability of differences in measures of sensitivity and bias. We (Macmillan and Kaplan, 1986) have been developing a technique, based on log-linear models, which integrates these two analyses. In simple detection paradigms log-linear parameters measure the sensitivity and bias parameters of Choice Theory. In experiments with multiple conditions or observers, comparison of different log-linear models permits statistical testing of hypotheses.

References

- ¹ L.D. Braida and N.I. Durlach, “Peripheral and Central Factors in Intensity Perception,” in *Functions of the Auditory System*, edited by G.M. Edelman, W.E. Gall and W.M. Cohen. New York: Wiley, 1986.
- ² D.H. Klatt, “Software for a Cascade-Parallel Formant Synthesizer,” *J. Acoust. Soc. Am.* 77, Supp. 1, S7 (1985).
- ³ A.E. Ades, “Vowels, Consonants, Speech and Nonspeech,” *Psych. Rev.* 84, 524 (1977).
- ⁴ N.A. Macmillan and H.L. Kaplan, *Log-Linear Models Provide Answers to Signal-Detection Questions*. New Orleans, Louisiana: Psychonomic Society, 1986.

Publications

- Goldberg, R.F., "Perpetual Anchors in Vowel and Consonant Continua," M.S. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1986.
- Goldberg, R.F., N.A. Macmillan and L.D. Braida, "A Perpetual-Anchor Interpretation of Categorical Phenomena on a Vowel Continuum (abstract)," *J. Acoust. Soc. Am.* 77, Supp. 1, S7 (1985).
- Macmillan, N.A., L.D. Braida and R.F. Goldberg, In *Psychophysics of Speech Perception*, edited by M.E.H. Schouten (Martinus Nijhoff, in press).
- Rosner, B.S., "Perception of Voice-Onset Time: a Signal-Detection Analysis," *J. Acoust. Soc. Am.* (1984).
- Sachs, R.M. and K.W. Grant, "Stimulus Correlates in the Perception of Voice-Onset Time (VOT): II. Discrimination of Speech with High and Low Stimulus Uncertainty (abstract)," *J. Acoust. Soc. Am.* 60, S91 (1976).
- Watson, C.S., D. Kewley-Port and D.C. Foyle, "Temporal Acuity for Speech and NonSpeech Sounds: the Role of Stimulus Uncertainty (abstract)," *J. Acoust. Soc. Am.* 77, Supp. 1, S27 (1985).

18.4 Hearing Aid Research

National Institutes of Health (Grant 6 R01 NS 12846)

Louis D. Braida, Diane K. Bustamante, Lorraine A. Delhorne, Nathaniel I. Durlach, Matthew H. Power, Charlotte M. Reed, 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 1) effects of noise on intelligibility; 2) amplitude compression; 3) frequency lowering; and 4) clear speech.

Effects of Noise on Intelligibility

We have completed a study (Zurek and Delhorne, 1986) to determine the extent to which the difficulty experienced by impaired listeners in understanding noisy speech can be explained merely on the basis of elevated detection thresholds. Twenty impaired ears of fourteen subjects (age 50 years or younger), spanning a variety of audiometric configurations with average hearing losses to 75 dB, were tested for reception of consonants in a speech-spectrum noise. Speech level, noise level, and frequency-gain characteristic were varied to generate a range of listening conditions. Results with impaired listeners were compared to those of normal-hearing listeners tested under the same conditions but with masking noise added to simulate approximately the impaired listeners' elevated thresholds. Each group of impaired listeners with similar audiograms was simulated with a group of normal-hearing listeners. In addition to this direct com-

parison, results were also compared by computing articulation indices for the various listening conditions, treating hearing loss as resulting from an internal additive noise. The articulation index takes into account the minor differences between actual and simulated thresholds.

The results indicate that masking noise is effective in producing roughly the same level of consonant intelligibility for normal listeners as hearing loss does for impaired listeners. Although there were individual differences, there were no impaired listeners whose results were clearly and consistently outside the range of masked-normal results. When equated for articulation index, 6% of the data points from impaired listeners fell below the range of data from normal listeners, whereas 7% of the impaired points fell above the normal range. When performance at a constant value of articulation index is plotted as a function of the average hearing loss of the listener, there is no clear dependence. Finally, the consonant confusions exhibited by the impaired were well-simulated with noise masking, a result that has been seen in other studies (Wang, Reed, and Bilger 1978; Fabry and Van Tasell, 1986). Thus, our conclusion based on this sample of non-elderly subjects with mild-to-moderate hearing losses is that the primary limitation in understanding noisy speech, aside from the noise itself, is the limited audibility of the speech signal due to elevated thresholds.

Amplitude Compression

Our work on amplitude compression can be subdivided into four projects: 1) Multi-band Amplitude Compression; 2) Principal-Component Amplitude Compression; 3) Adaptive Automatic Volume Control; and 4) Peak-Clipping Amplitude Compression.

Multiband Amplitude Compression

To estimate the optimal range of speech levels to present to an impaired listener, intelligibility was measured with three compression systems that differed parametrically in the amount of speech energy presented above the listener's elevated detection thresholds (DeGennaro, 1982). The three systems provided amplification to place 25, 50, or 90 percent of the short-term amplitude distributions in each of 16 frequency bands within the listener's residual auditory area. With band gains set to achieve these varying degrees of audibility, listeners then selected compression ratios to achieve maximum intelligibility and long-term listening comfort while listening to octave bands of compressed running speech. The reference linear amplification system was constructed by having subjects set octave bands of speech to their most comfortable listening level. The four systems were compared in terms of CVC intelligibility.

The principal results of this study, from three ears (of two subjects) with bilateral flat losses and dynamic ranges of 18-33 dB, were as follows: 1) the subjects generally responded to increases in the range of audible speech, from 25 to 50 to 90 percent, with higher effective compression ratios. Measurements of amplitude level distributions suggest that the subjects selected compression ratios that resulted in 1% peak speech levels being below discomfort thresholds; 2) Speech intelligibility scores with the three reference systems were, at best, equivalent to those of the comparison linear system. Further, the best performance with compression was achieved with moderate compression ratios, even though a higher percentage of speech energy was placed above threshold with high compression ratios; 3) Consonant error patterns indicated that while compression improved distinctions between stops and fricatives relative to linear

amplification, distinctions among stops and among fricatives, which are often considered to be dependent on spectral cues, were degraded. It was concluded that it may be necessary not only to improve audibility but also to preserve and/or enhance spectral distinctions.

Principal-Component Amplitude Compression

Experiments were conducted (Bustamante, 1986) with four sensorineural hearing-impaired listeners to test the effectiveness of principal-component amplitude compression (see also Sec . D-1-a). All subjects had bilateral flat losses, roughly normal discomfort levels, and reduced dynamic ranges of 18-31 dB. Two implementations, involving compression of the first principal component (PC1) or compression of both the first and second principal components (PC1 and PC2), were compared to linear amplification (LA), independent compression of multiple bands (MBC), and wideband compression (WC). The MBC system placed 50 percent of the speech energy in each of 16 frequency bands above the subject's detection threshold. The WC system provided high-frequency emphasis of the speech signal prior to the wideband compression amplifier, similar to the most promising compression system studied by Henrickson (1982). The linear amplification system was constructed by having subjects set octave bands of speech to their most comfortable listening level. Frequency shaping was applied in a similar manner to the post-compression output of the two compression systems and the WC system. Speech tests were conducted at two levels: each listener's most comfortable listening level (MCL) and at 10-15 dB below MCL. The five systems were compared on the basis of intelligibility measures using both CVC nonsense syllables (600 trials per system-level combination) and Harvard sentences (I.E.E.E., 1969; 200 keywords per system-level combination).

The major results of this study were that: 1) compression of short-term overall level with compression of PC1 and WC maintains speech intelligibility equivalent to that with LA at MCL and over a wider range of levels than is possible with LA; 2) compression of spectral tilt (by compressing both PC1 and PC2) degrades speech intelligibility; and 3) independent compression of multiple bands with MBC can provide intelligibility comparable to or better than that with compression of PC1 and WC at MCL, but generally does not maintain the same level of performance over a 10-15 dB range of input levels. Vowel scores were high with all processing schemes while consonant scores were significantly lower, particularly for the two more severely impaired listeners. Both compression of PC1 and WC improved consonant scores for these listeners relative to LA but the gains were not large, indicating that either these compression schemes are not capable of making all speech sounds audible or that making speech sounds audible is not sufficient to compensate for all hearing impairments.

Preliminary Study of Adaptive Automatic Volume Control

Automatic Volume Control (AVC) systems are amplitude compressors designed to maintain a relatively constant output level without affecting short-term level variations, as might benefit listeners with peaked articulation functions or limited tolerance ranges for speech (e.g., Poliakoff, 1950). Since AVC systems tend to amplify background noise during pauses in speech, they are likely to prove distracting in many listening situations. To address this problem we have begun to study adaptive AVC systems which can adjust their characteristics automatically on the basis of estimates of speech and noise levels. Bristol (1983) developed an AVC system that combined a compressor having

fixed characteristics with an expander (intended to reduce output levels during pauses in speech, e.g., McAulay and Malpass, 1980) having a variable expansion threshold. Two expansion ratios (ERs) and two compression processors were studied: a wideband system, and a multiband system with independent compression in each of 16 critical-bandwidth channels.

The systems were evaluated by a normal-hearing listener using conversational sentences presented in a background of steady flat-spectrum Gaussian noise ($S/N = +10$ dB). Measurements indicated that the expansion processing increased the difference between output levels when speech was present and when speech was absent from 10 dB to 17 dB ($ER = 2.5$) and 31 dB ($ER = 4.5$) for both compression systems. However the processing introduced some perceptual distortions and reduced keyword intelligibility scores (without processing: 94% in quiet, 84% in noise; compression alone: 74% multiband, 69% wideband; expansion alone: 79% for $ER = 2.5$, 70% for $ER = 4.5$). However, when expansion was combined with multiband compression, there was little additional reduction in scores over that with compression alone (74% for $ER = 2.5$ and 70% for $ER = 4.5$). These measurements, which were meant to measure the perceptual distortions introduced by the systems rather than to estimate their potential benefit under conditions of varying speech and noise levels, suggest that the expansion processing does not improve intelligibility by attenuating noise during pauses in speech, but may nevertheless be beneficial if used in conjunction with improved multiband AVC compression.

Peak-Clipping Amplitude Compression

In this project we (Siletto, 1984) have begun assessing infinite clipping as a means of amplitude compression by determining: 1) the compression of the range of speech levels when clipping is followed by filtering (as is presumed to occur when clipped speech is analyzed by the ear); and 2) the spectral distortions due to clipping and how these can be minimized by pre-filtering. We have also done preliminary experiments with a three-band clipping system (Hildebrant, 1982). Clipping followed by filtering compresses the range of third-octave speech-level distributions (between 10% and 90% cumulative levels) to 10-15 dB, compared to input ranges of 30-40 dB. This compression is relatively independent of the characteristics of the filter following the clipper. Despite the radical distortion of the input wave, spectral distortions are not severe. An empirical rule of thumb is that if the input spectral envelope has a global maximum at some frequency f , then there will be a maximum near f in the spectral envelope of the output. While there may be additional local maxima added or changed by clipping, generally the largest peak is preserved.

This rule led to the design of a system that separates the first three speech formant regions prior to infinite clipping (Hildebrant, 1982). The output of each of the three clippers is then filtered over the same range as its respective input filter to remove out-of-band distortion. The three channels are added after this post-filtering to form the signal that is presented to the listener. The system was tested under two conditions with recordings of PB-50 word lists and two normal-hearing listeners. In the first condition multi-talker babble was added to the speech prior to processing. The effect of this processing can be summarized as an effective decrease in signal-to-noise ratio of about 6 dB, relative to unprocessed speech. Our results with this system, which has some similarity to a system described by Thomas and Ravindran (1974), do not corroborate their results indicating intelligibility enhancement of noisy speech by clipping. The

second condition was an attempt to simulate a reduced dynamic range. A white noise was added to processed speech; this noise elevated tone-detection thresholds to be approximately constant across frequency. Discomfort threshold was simulated by a visual peak indicator that was lit when the signal exceeded a threshold set to be 25 dB above detection threshold. The listeners adjusted the level of the processed or unprocessed speech (prior to addition with the threshold noise) so that the indicator lit very infrequently when running speech was presented. The unprocessed speech was passed through a whitening filter to produce a flat long-term speech spectrum. With a 25-dB dynamic range, both subjects scored 4% correct words with whitened speech, while with 3-band clipped speech they scored 33 and 40% correct words. Additional tests using Harvard sentences produced scores of 16 and 18% correct key words with whitened speech and 98% correct for both subjects with 3-band clipped speech. These results demonstrate that it is possible for an amplitude-compression system to yield superior intelligibility, at least for normal listeners, relative to a linear system with properly-chosen frequency-gain characteristic. It is also clear that the benefits of an amplitude compressor will be seen only when the listener's dynamic range is markedly smaller than the normal range of speech amplitudes.

Frequency Lowering

We have conducted work on a vocoder-based system for lowering the frequency content of natural speech (Posen, 1984) as well as on the development of low-frequency artificial codes for speech sounds (Foss, 1983; Power, 1985).

Vocoder-Based Frequency Lowering

The vocoder-based method was modeled after a system described by Lippmann (1980) in which speech levels in high-frequency bands are used to modulate low-frequency bands of noise, which are then added to lowpass-filtered speech. In the real-time processing scheme implemented for our studies, the four high-frequency bands were each 2/3-octave wide with center frequencies from 1 to 5 kHz, while the four low-frequency noise bands were each 1/3-octave wide with center frequencies from 400 to 800 Hz. The level of each low-frequency band of noise was controlled by the level of the corresponding high-frequency band of speech (using the obvious monotonic relationship). The modulated noise bands were added to the unprocessed speech signal such that the 10% cumulative level of each noise band was 12 dB below the 10% cumulative level of speech in the same 1/3-octave band. The combined signal was then lowpass filtered at 800 Hz to simulate a sharply sloping, high-frequency hearing loss.

Performance using this lowering system was evaluated relative to performance on lowpass-filtered speech with a cutoff frequency of 800 Hz. Two normal-hearing listeners were trained and tested on the identification of 24 consonants in CV syllables spoken by two male and two female speakers. Results indicated equivalent overall performance (roughly 57% correct) for the two conditions. If only the stops, fricatives, and affricates are considered, however, an advantage of 4-7 percentage points was obtained with frequency lowering, a result consistent with that of Lippmann (1980) whose testing included only this set of consonants. For the remaining consonants (i.e., the semivowels and nasals), an advantage of 10-13 percentage points was obtained for lowpass filtering. Based on these results, which are similar to those observed in our previous studies of warped, lowered speech (Reed et al., 1983, 1985), the frequency-

lowering system was modified to suppress the processing of sounds dominated by low-frequency energy. The energy in the spectrum above 1400 Hz was compared to that in the spectrum below 1400 Hz, and processing occurred only if the ratio of high to low frequency energy exceeded a threshold value of about 4 dB. Post-training tests with the modified system indicated an increase in overall performance of 7 percentage points over lowpass filtering (and over the original lowering system). Performance on the semivowels and nasals was similar for the modified system and for lowpass filtering, and was 10 percentage points higher than for the original processing scheme. A slight advantage was also observed for the reception of fricatives and stops through the modified system compared to the original system, suggesting that the processing of vowels may have had a masking effect on the weaker noise signals resulting from the lowered consonants. The major difference in confusion patterns between the modified lowering system and lowpass filtering lies in improved identification of the features affrication and duration (both of which are related to the perception of high-frequency consonants) through the lowering system. Transmission of information regarding place of articulation was equally poor for lowering and filtering.

Artificial Low-Frequency Codes

Artificial frequency-lowering codes were developed for 24 consonants (C) and 15 vowels (V) for two values of lowpass cutoff frequency F (300 and 500 Hz). Individual phonemes were coded by a unique, non-varying acoustic signal confined to frequencies less than or equal to F . The consonant codes were created by varying frequency content (both center frequency and bandwidth), amplitude, duration, and intervals of silence to specify various abstract properties (i.e., voicing, manner, and place) that serve to classify the consonants. The vowels were generated by varying the spectral shape of a ten-tone harmonic complex with 50-Hz spacing for $F=500$ Hz and 30-Hz spacing for $F=300$ Hz, where the relative amplitudes of the components were derived from measurements of natural speech sounds. The vowels were further coded as “long” (220 msec) or “short” (150 msec). Diphthongs were generated by varying the frequency and amplitude of the individual components over the 300-msec stimulus duration.

Performance on the coded sounds was compared to that obtained on single-token sets of natural speech utterances lowpass filtered at F . Identification of coded Cs and Vs was examined in normal-hearing listeners for $F=500$ and 300 Hz (Foss, 1983; Power, 1985). Subjects trained to asymptotic performance with each stimulus set (24 Cs or 15 Vs) and type of processing (artificial coding or lowpass filtering). For a set of 24 consonants in C-/a/ context, performance on coded sounds averaged 90% correct for $F=500$ Hz and 65% for $F=300$ Hz, compared to 75% and 40% for lowpass filtered speech. For a set of 15 vowels in /b/-V-/t/ context, performance on coded sounds averaged 85% correct for $F=500$ Hz and 65% for $F=300$ Hz, compared to 85% and 50% for the lowpass filtered sounds. The overall results thus indicate that it is possible to construct low-frequency codes that provide better performance than lowpass filtered natural sounds with the same cutoff frequency (at least for the case in which each phoneme is represented by a single token). Furthermore, for the specific codes considered, the coding advantage is greater for consonants than vowels.

In conjunction with this work, we have begun to investigate the effect of token variations on the intelligibility of speech. The studies of Foss (1983) and Power (1985) included tests of both one-token and three-token sets of filtered stimuli. For consonant and vowel identification at both values of F , performance on the single-token sets was

substantially better than on the multiple-token sets, with the size of the effect averaging about 20 percentage points. An additional study was conducted to examine further the effect of number of tokens on identification of consonants in lowpass-filtered CV syllables (DuBois, 1984). The number of tokens per consonant (all produced by the same female talker) ranged from one (one utterance of each syllable in C-/a/ context) to three (three utterances of each syllable in C-/a/ context) to nine (three utterances per syllable in each of three contexts--C-/a/, C-/i/, and C-/u/). As the number of tokens increased from one to three to nine, identification scores decreased from 79 to 64 to 57 percent correct. Thus, the size of the effect appears to be substantial for small numbers of tokens.

We have also begun to investigate the ability of subjects to comprehend streams of coded consonants and vowels (Reed et al., 1986). Results have been obtained with $F=500$ Hz for a CV identification experiment in which C was chosen at random from a set of 24 Cs and V was chosen at random from a set of 15 Vs. Performance averaged across two subjects indicated an average C identification score of 67% correct and an average V identification score of 71%, scores that are somewhat lower than those observed in the fixed-context experiments described above. The percentage of syllables in which both C and V were identified correctly averaged 50%. It should be noted, however, that the two subjects tested in these experiments showed very different performance: whereas the average recognition score for one subject was 40%, for the other subject it was 90%. Also, the difference in performance between the tests with fixed and variable context was much larger for the subject with the lower scores.

Speaking Clearly

In addition to the publication of our first report on the analysis of acoustic differences between clear and conversational speech (Picheny et al., 1986), work completed on our clear-speech project includes detailed duration measurements, expansion of a phonetically-labeled speech database, development of a software system for time-scale modifications of speech, and perceptual evaluation of time-modified speech.

Uchanski et al. (1985) describe detailed measurements of segment durations in both conversational and clear speech. The data, which were obtained from three speakers, show that a distinctly non-uniform lengthening of duration occurs as a speaker changes from a conversational to a clear speaking style. Semivowels and unvoiced fricatives are lengthened the most (75% or greater increase) while voiced stops and short vowels are lengthened the least (less than 30% increase). Results from ANOVA on the duration data indicate that stress is not a statistically significant factor and that prepausal lengthening is statistically significant for many groups of sound segments. Segment position at a phrase boundary has a significant effect on duration in both speaking modes. The prepausal effect accounts for roughly 30% of the total variance in duration measures. The significance of the prepausal effect implies that grammatical structure is important in the production of both conversational and clear speech.

Measurement of temporal and spectral characteristics of conversational and clear speech requires an adequate database. During this grant period an additional 660 nonsense sentences (Picheny et al., 1985) were digitized from analog tape and phonetically labeled. These 660 sentences comprise recordings from three speakers and both conversational and clear speaking styles. Including the 300 previously labeled and

digitized sentences (50 sentences spoken conversationally and clearly by three speakers), our database now contains a total of 960 sentences.

A software system was developed to produce non-uniformly time-scaled sentences. The first part of this system matches the phonetic sequences of the same sentences spoken conversationally and spoken clearly. From the label-matching procedure, a time alignment of the conversational and clear sentences is generated. The time alignment information is used to interpolate linearly across short-time magnitude spectra. Finally, a new time-scaled speech signal is created from the modified magnitude spectra.

A perceptual experiment was performed to isolate the effect of segment duration on intelligibility of conversational and clear speech. The time-scaling system just described was used to modify conversational sentences to have segment durations typical of clear sentences and also to modify clear sentences to have segment durations typical of conversational sentences. In all, four hearing-impaired subjects listened to nonsense sentences produced or modified in five ways: 1) unprocessed conversational speech; 2) unprocessed clear speech; 3) unprocessed conversational speech with repetition of the utterance; 4) non-uniformly sped-up clear speech; and 5) non-uniformly slowed-down conversational speech. The intelligibility scores averaged over the four subjects (roughly 550 items per condition per listener) were:

Condition	1	2	3	4	5
Score (%)	77	91	82	71	74

These results confirm the difference in intelligibility between naturally-produced clear and conversational speech. Repetition of a conversational sentence improves intelligibility, but not nearly as much as speaking clearly. (If the repetition had involved distinct utterances, this improvement would undoubtedly have been greater, but we do not know how much greater). Unfortunately, both modified productions, slowed-down conversational and sped-up clear speech, are perceived less well than natural conversational speech. However, the degradation in performance with these modified productions is less than the degradation seen with uniformly time-scaled speech (Picheny et al., 1986). One interpretation of this difference is that time scaling at the segmental level more closely approximates the natural changes from conversational to clear speech.

18.5 Additional Work

Simulation of Hearing Impairment

In a preliminary study (Gilbert, 1984), a biased rectifier was used as a detector in a model of a sensorineural hearing loss. The rectifier parameters were varied to reflect different levels of hearing loss. The output of the rectifier was calculated numerically to generate predictions of recruitment, temporal integration at absolute threshold, and tone masking. With appropriate choices of parameters, the model was able to predict

recruitment and temporal integration, but it failed to predict empirically-observed patterns of tone masking in hearing-impaired listeners.

Use of Articulatory Signals in Automatic Speech Recognition

Automatic speech recognition systems generally attempt to determine the spoken message from analysis of the acoustic speech waveform. In this research we evaluated the performance of the IBM Speech Recognition System (Jelinek, 1985) when the input included measurements of selected articulatory actions occurring during speech production. The system achieved significant recognition rates (for isolated words in sentences) when the acoustic signal was disabled and the input was restricted to articulatory signals similar to those sensed by users of the Tadoma method of tactile-speech recognition (Reed et al., 1985b). In other tests the availability of articulatory inputs improved recognition performance when the acoustic signal was sufficiently degraded by additive white noise. Improvements were observed independent of whether the recognition system made use of the likelihood that words would appear in the vicinity of the other words in a given sentence. (Portions of this work were conducted at, and supported by, the I.B.M. T.J. Watson Research Center.)

References

- ¹ P.M. Zurek and L.A. Delhorne, "Consonant Reception in Noise by Listeners with Mild and Moderate Hearing Impairment," (submitted to J. Acoust. Soc. Am. 1986).
- ² S. DeGennaro, "Multi-Channel Syllabic Compression for Severely Impaired Listeners," Ph.D. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1982.
- ³ D.K. Bustamante and L.D. Braida, "Compression Limiting for the Hearing Impaired," (submitted to J. Rehab. Res. and Dev. 1986).
- ⁴ K.A. Siletto, "Evaluation of Infinite Peak Clippings as a Means of Amplitude Compression," M.S. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1984.
- ⁵ E.M. Hildebrant, "An Electronic Device to Reduce the Dynamic Range of Speech," S.B. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1982.
- ⁶ M.P. Posen, "Intelligibility of Frequency Lowered Speech Produced by a Channel Vocoder," S.M. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1984.
- ⁷ K.K. Foss, "Identification Experimentation on Low-Frequency Artificial Codes as Representation of Speech," S.B. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1983.
- ⁸ M.H. Power, "Representation of Speech by Low Frequency Artificial Codes," S.B. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1985.

- ⁹ C.M. Reed, B.L. Hicks, L.D. Braida, and N.I. Durlach, "Discrimination of Speech Processed by Low-Pass Filtering and Pitch-Invariant Frequency Lowering," *J. Acoust. Soc. Am.* **74**, 409 (1983).
- ¹⁰ C.M. Reed, L.D. Braida, N.I. Durlach and K.I. Schultz, "Discrimination and Identification of Frequency-Lowered Speech in Listeners with High-Frequency Hearing Impairment," *J. Acoust. Soc. Am.* **78**, 2139 (1985).
- ¹¹ C.M. Reed, M.H. Power and N.I. Durlach, "Development and Testing of Artificial Low-Frequency Speech Codes," (to be submitted to *J. Speech Hear. Res.* 1986).
- ¹² M.A. Picheny, N.I. Durlach and L.D. Braida, "Speaking Clearly for the Hard of Hearing. III: an Attempt to Determine the Contribution of Speaking Rate on Differences in Intelligibility between Clear and Conversational Speech," (to be submitted to *J. Speech Hear. Res.* 1986).
- ¹³ R.M. Uchanski, C.M. Reed, N.I. Durlach and L.D. Braida, "Analysis of Phoneme and Pause Durations in Conversational and Clear Speech," *J. Acoust. Soc. Am.* **77**, Supp. 1 (1985).
- ¹⁴ M.A. Picheny, N.I. Durlach and L.D. Braida, "Speaking Clearly for the Hard of Hearing. I: Intelligibility Differences between Clear and Conversational Speech," *J. Acoust. Soc. Am.* **78**, 96 (1985).
- ¹⁵ E.I. Gilbert, "Investigation of Sensorineural Hearing Impairment Using a Center-Clipping Model," S.B. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1984.

Publications

- Braida, L.D., "Articulation Testing Methods for Evaluating Speech Reception by Impaired Listeners," in *Speech Recognition by the Hearing Impaired*, edited by E. Elkins, ASHA Reports **14**, 26 (1984).
- Braida, L.D., "Speech Processing for the Hearing Impaired," Proceedings of the Second International Conference on Rehabilitation Engineering, Ottawa, Canada, 1984.
- Braida, L.D., M.A. Picheny, J.R. Cohen, W.M. Rabinowitz and J.S. Perkell, "Use of Articulatory Signals in Speech Recognition," (to be submitted to *J. Acoust. Soc. Am.*).
- Bustamante, D.K. and L.D. Braida, "Principal Component Amplitude Compression of Speech for the Sensorineural Hearing Impaired," Proceeding of the 7th Annual Conference of the I.E.E.E. Engineering in Medicine and Biology Society, Chicago, Illinois, 1985, p. 1096.
- Bustamante, D.K. and L.D. Braida, "Principal-Component Amplitude Compression for the Hearing Impaired," (submitted to *J. Acoust. Soc. Am.* 1986).

- Bustamante, D.K., E. Hildebrant, K. Silletto, P.M. Zurek, L.D. Braida and N.I. Durlach, "New Amplitude Compression Techniques," Proceedings of the International Conference on Rehabilitation Engineering, Ottawa, Canada, 1984, p. 293.
- DeGennaro, S., L.D. Braida and N.I. Durlach, "Multichannel Syllabic Compression for Severely Impaired Listeners," *J. Rehab. Res. and Dev.* 23, 17 (1986).
- Milner, P., L.D. Braida, N.I. Durlach and H. Levitt, "Perception of Filtered Speech by Hearing-Impaired Listeners," in *Speech Recognition by the Hearing Impaired*, edited by E. Elkins, ASHA Reports 14, 30 (1984).
- Picheny, M.A., N.I. Durlach and L.D. Braida, "Speaking Clearly for the Hard of Hearing. II: Acoustic Characteristics of Clear and Conversational Speech," *J. Speech Hear. Res.* (1986).

Theses

- Bristol, N., "Automatic Volume Control and Noise Suppression for Hearing Aids," S.M. Thesis, Dept. Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1983.
- Bustamante, D.K., "Multichannel Syllabic Compression for Severely Impaired Listeners," Ph.D. Diss., Dept. Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1986.
- Laves, A.L., "Consonant Intelligibility and Acoustic Analysis in Clear and Conversational Speech," S.B. Thesis, Dept. Electr., Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1982.

18.6 Multimicrophone Monaural Aids for the Hearing-Impaired

National Institutes of Health (Grant 1 RO1 NS 21322)

H. Steven Colburn, Cathleen R. Corbett, Nathaniel I. Durlach, Patrick M. Peterson, William M. Rabinowitz, Patrick M. Zurek

The goal of this work is the development of systems that sample the acoustic environment at multiple points in space to form a single-channel output providing enhanced speech intelligibility for the hearing-impaired. Work in the past year has included studies of: 1) an adaptive beam-steering algorithm for interference reduction; and 2) coding of multiple messages for monaural presentation.

1) Adaptive processing schemes vary the complex weightings applied to the microphone signals so that the beam pattern tracks changes in the acoustic environment. Our research in this area (Peterson, 1987; Peterson et al., 1987) has focused on the constrained adaptive beamforming method of Griffiths and Jim (1982). Constrained adaptive beamformers are based on the assumptions that the target and interference are uncorrelated and that the target direction is known, and operate to minimize total output power under the constraint that signals from the target direction are preserved. Minimization of total output power then implies minimization of interference and

maximization of target-to-interference ratio. Although the Griffiths-Jim method adapts more slowly than some other methods (e.g., Cioffi and Kailath, 1984), it asymptotically bounds the performance that can be achieved by any constrained adaptive beamformer and is relatively simple to realize.

A two-microphone version of a Griffiths-Jim beamformer has been studied in three environments with different degrees of reverberation. Intelligibility tests were performed on normal-hearing subjects using sentences as the target signal and speech babble as the interference. The microphone signals were generated by passing the target and the interference through simulated room transfer functions (Peterson, 1986). It was assumed that the microphones were separated by 20 cm, that the microphones and sound sources were omnidirectional, and that the interference arose from a single source 45 degrees from the straight-ahead target. The simulated rooms were chosen to represent anechoic space (no reverberation), a living room (moderate reverberation) and a conference room (strong reverberation). Intelligibility of processed sentences was compared to that of unprocessed sentences presented binaurally. All system parameters were held fixed except the length of the adaptive filter; preliminary examination suggested that two lengths be tested, 10 ms and 40 ms. Longer filters can potentially remove more interference, but at the cost of more computation and longer adaptation times. In all cases, the intelligibility tests estimated the asymptotic (adapted) performance of the system. Empirical adaptation times, with the given system parameters, were of the order of one second.

The results of these tests indicate that under the conditions of zero-to-moderate reverberation, the interference reduction achieved by the array exceeds that achieved by the binaural auditory system. In the anechoic condition the interference was reduced by 30 dB, while in the living room environment the interference was reduced 9 dB by the 10-ms filter and 14 dB by the 40-ms filter. Furthermore, under the conference room condition, where reverberation was most severe, intelligibility performance was not degraded by processing. Research designed to determine the generalizability of these results and their implications for a practical hearing aid is now underway.

2) If we assume that directional channels can be built (i.e., that simultaneous spatially-diverse inputs can be resolved into separate signals), then we must determine how to code the multiple simultaneous signals into a monaural signal that facilitates attention to each of the individual messages. We have completed a study of filtering as a means of coding two or four sentence-length messages (Corbett, 1986). In particular, we evaluated the extent to which the presentation of individual messages in separate spectral regions enhances the joint intelligibility of all messages over that obtained when the wideband messages are simply summed. Different transfer functions were assigned to the different channels in such a way that intelligibility scores were roughly independent of which channel was chosen to carry the target message (and which channels carried the interfering messages). Overall performance was measured by the joint intelligibility, i.e., the intelligibility score averaged over talker/channel combinations used for the target. A given filter configuration was judged to be helpful if the joint intelligibility for this configuration was greater than the joint intelligibility for the reference case in which there was no filtering and all voices were summed in a wideband channel.

Although our results are preliminary, they suggest that, at least for the case of four different talkers speaking simultaneously, joint intelligibility can be improved by the use

of differential filtering. For the best filter configuration tested, the joint intelligibility was 39% compared to the wideband reference score of 22%. Some of our results, however, were less positive. For example, tests with two talkers, easy speech materials, and noise masking (to prevent saturation of intelligibility scores) showed no improvement for filtering. Nevertheless, based on all results to date, we believe that judiciously chosen filtering will provide an overall gain in joint intelligibility for many cases likely to be encountered in real environments.

References

- ¹ P.M. Peterson, "Using Linearly-Constrained Adaptive Beamforming to Reduce Interference in Hearing Aids from Competing Talkers in Reverberant Rooms," Proceedings of the International Conference on Acoustic Signal Processing, 1987.
- ² P.M. Peterson, N.I. Durlach, W.M. Rabonowitz and P.M. Zurek, "Multimicrophone Adaptive Beamforming for Interference Reduction in Hearing Aids," J. Rehab. Res. Dev. (1987).
- ³ L.J. Griffiths and C.W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," I.E.E.E. Trans. Antennas Propag. *AP-30*, 27 (1982).
- ⁴ J.M. Cioffi and T. Kailath, "Fast, Recursive Least Squares, Transversal Filters for Adaptive Filtering," I.E.E.E. Trans. Acoust. Speech Signal Process. *32*, 304 (1984).
- ⁵ P.M. Peterson, "Simulating the Response of Multiple Microphones to a Single Acoustic Source in a Reverberant Room," J. Acoust. Soc. Am. *80*, 1527 (1986).
- ⁵ C.R. Corbett, "Filtering Competing Messages to Enhance Mutual Intelligibility," M.S. Thesis, Dept. of Electr. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., 1986.