## Hearing Aid Research

### Sponsor

National Institutes of Health/National Institute on Deafness and Other Communication Disorders
Grant R01 DC00117

### Project Staff

Professor Louis D. Braida, Dr. Paul Duchnowski, Dr. Julie Greenberg, Dr. Kenneth W. Grant, Dr. Karen L. Payton, Dr. Christine M. Rankovic,

Ann K. Dix, Merry Brantley,

Paninah S. Fine, Joseph A. Frisbie, Raymond Goldsworthy Jean C. Krause

### Goals and Significance

Our long-term goal is to develop improved hearing aids for people suffering from sensorineural hearing impairments. Our efforts are focused on problems resulting from inadequate knowledge of the effects of various transformations of speech signals on speech reception by impaired listeners, specifically on the fundamental liimitations on the improvements in speech reception that can be achieved by processing speech. Our aims are:

To evaluate the effects of style of speech articulation and variability in speech production on speech reception by hearing impaired listeners.

To develop and evaluate analytical models that can predict the effects of a variety of alterations of the speech signal on intelligibility.

To develop and evaluate signal processing techniques that hold promise for increasing the effectiveness of hearing aids.

To assess the relative contributions of various functional characteristics of hearing impairments to reduced speech-reception capacity.

### Current Studies

### Characteristics of the Speech Signal

Our previous work has shown that sentences spoken ``clearly'' are more intelligible (roughly 17 percentage points) than those spoken ``conversationally'' for hearing-impaired listeners in a quiet background (Picheny et al., 1985) as well as for both normal hearing and hearing-impaired listeners in noise (Uchanski et al., 1996) and reverberation backgrounds (Payton et al., 1994). While producing clear speech, however, talkers often significantly reduce their speaking rate. A more recent study (Krause and Braida, 1995) has shown that talkers can be trained to produce a form of clear speech at normal rates (clear/normal speech). This finding suggests that acoustical factors other than reduced speaking rate are responsible for the high intelligibility of clear speech. Our recent work has been aimed at determining which acoustical properties of clear/normal speech contribute most to its high intelligibility through acoustic analysis of clear/normal speech and related signal processing transformations.

Initial acoustical measurements of clear speech at slow and normal rates (Krause, 2001) has shown that differences in acoustic characteristics of clear/slow speech relative to conv/normal

speech were consistent with previously reported results. Many of these differences, however, were not apparent when comparing clear/normal speech to conv/normal speech. Moreover, some of the acoustic characteristics (e.g. segment duration, pitch, and voice- onset time) retained in clear/normal speech differed dramatically between talkers, suggesting that different talker strategies exist for producing clear speech at normal rates.

Based on the results of this acoustical analysis, signal transformations were developed to alter three properties of conv/normal speech that were found to be altered by one or more talkers in clear/normal speech. In processed(A) speech, vowel formant energy was increased by raising formant amplitudes and widening formant bandwidths; in processed(B) speech, low- frequency modulations ($<$3--4Hz) were enhanced; and in processed(C) speech, F0 (pitch) average and range was increased, since this acoustical property was exhibited by the talker whose clear/normal speech was most robust to other degradations. These intelligibility enhancement schemes were evaluated singly and in combination by both normal hearing and hearing-impaired listeners participated in intelligibility experiments to evaluate whether listeners could derive intelligibility benefits from artificial manipulation of these acoustic properties. Although the speech- based STI predicted that a majority of these processing combinations would improve intelligibility over conv/normal speech presented in wideband noise to normal hearing listeners, actual experiments with normal hearing listeners revealed an advantage only for clear/normal speech and processed(A) speech. Moreover, hearing-impaired listeners did not obtain similar intelligibility benefits from clear/normal or processed(A) speech as reliably as normal hearing listeners in noise, although these conditions did provide a statistically significant benefit for some individual hearing-impaired listeners and talkers (Krause, 2001).

One possible explanation of these results is that the benefits of clear/normal speech may be related to age, since the hearing- impaired participants in this study were older (40 to 65 years) than the normal hearing participants (19 to 43 years). Some studies report an age-related decline in speech reception for elderly listeners (Arlinger and Gustafsson, 1991), particularly those with hearing impairments (Hargus and Gordon-Salant, 1995). Another possibility is that the intelligibility benefits of these conditions do not extend to hearing-impaired listeners and that the additive noise model for simulating impairment in normal hearing listeners is inadequate. Although this simulation is appropriate for many mild to moderate impairments, it may not represent the effects of more severe impairments accurately. To investigate these two possibilities further, additional intelligibility tests will be conducted to evaluate the intelligibility of clear/normal, clear/slow, conv/normal, and conv/slow speech for young hearing-impaired, elderly hearing- impaired, and elderly normal-hearing listeners. These tests will differentiate the effect of age and impairment factors and clearly identify which groups can receive benefit from clear speech at normal speaking rates.

Because the intelligibility advantage provided by formant processing was not as large as the advantage provided by clear/normal speech, additional acoustic properties of clear/normal speech that contribute to its high intelligibility must exist. Analysis of additional talkers may lend insight into these properties. Three additional talkers who produced clear/normal speech in Krause's intelligibility study (Krause, 1995) are available for such an analysis, since only two of the five talkers from that study had been analyzed initially (Krause, 2001). Analysis of these three talkers is currently underway and parallels measurements made previously both for the first two talkers (Krause, 2001) and for the acoustics of clear/slow speech (Picheny et al., 1986). In these studies, measurements were taken at three levels of detail: global, phonological, and phonetic.

Three global measurements (long-term spectra, pitch, and pause structure) have been completed to date, and no additional properties associated with clear/normal speech have been identified.

If the current analysis does not identify one or more of acoustic properties related to high intelligibility, a factor that should be considered is the complexity of the speech database used for acoustic analysis. While a sentence database was appropriate for the intelligibility experiments, the primary problem with using a sentence database for the purposes of acoustic analysis is the

presence of acoustic variability due to word positioning within sentences or phonetic context within words. For some acoustic properties, this variability could be large enough to mask the variability between conv/normal and clear/normal tokens. Therefore, after the current measurements are completed, a new database of conv/normal and clear/normal speech should be created that consists of sentences with a fixed number of phonetic contexts. This type of database would best satisfy the conflicting demands of acoustic analyses and intelligibility experiments. To capture various talker strategies in the new database, a large number of talkers should be recorded. An acoustic analysis of a database of this type is likely not only to identify additional acoustic properties associated with clear/normal speech but also to provide a comprehensive description of a variety of talker strategies. This information will be essential to the development of processing schemes that can provide robust intelligibility improvement for a variety of talkers and environments.

**Computational Model of Speech Intelligibility**

Based on the results of the pilot study we conducted Summer 2000, we modified the parameters of our hearing aid and hearing loss simulations. First, we reduced the attack time of the amplitude compression hearing aid to 5 ms. This was done to better match typical attack times of actual compression hearing aids. Second, on the hearing loss simulation, we modified the attenuation slope below normal-hearing thresholds to match the slope just above thresholds. This was done to reduce audible distortion artifacts. Third, we adjusted input levels to both the hearing aid and the hearing loss simulation to make sure we were correctly simulating the way a hearing aid might be used in the field (we set input levels to the hearing aid to be 70 dB SPL and input levels to the hearing loss simulation to be 89 dB SPL).

The STI analysis on the new data has not yet been completed. When analyzing speech processed through the hearing loss simulation, we found that the envelope spectra typically are enhanced relative to the original spectra. This is not consistent with what the intelligibility results led us to expect. When we used the output of the hearing aid as the input to our envelope spectra algorithm and included the impairment losses as internal noise at the elevated threshold levels, we obtained envelope spectra more consistent with the intelligibility results. Note that, for actual hearing-impaired listeners, we modeled the loss as internal noise and the predictions fit intelligibility data quite well (see Payton, et al., 1994).

Thus far we have evaluated our new MTF method with conditions for which we have theoretical MTF values against which we can compare our results. The conditions we have examined are speech in speech-shaped noise, speech in reverberation, speech in speech-shaped noise plus reverberation and speech in restaurant babble. In each case, the ratio of the magnitude of Sxy to Sxx follows the theoretical MTF very closely. An unexpected finding is that there is a peak in the MTF of speech in restaurant babble at very low modulation frequencies (.3-.4 Hz). While the analysis is not complete, we have observed that this frequency range is approximately the sentence rate of the speech materials. We will continue to investigate this. Figures 1-4 show speech Modulation Transfer Functions for the four conditions mentioned above (Houtgast's method, the ratio of the magnitude of Sxy to Sxx – referred to as Revised Ludvigsen, and Drullman's method are all plotted). We will be presenting this work at the Spring 2002 meeting of the Acoustical Society of America in Pittsburgh.

**Cochlear-Implant Research**

Once component of our work in the area of models of speech intelligibility concerns predicting the intelligibility of cochlear-implant processed speech. The goal of this effort is to develop a subject-independent metric that can be used to predict the maximum possible intelligibility performance for a particular cochlear-implant speech processing strategy. (Subject-dependent factors may

lead to lower performance for particular subjects.) Once such a metric is developed, it will be used to evaluate promising noise-reduction strategies for cochlear implant preprocessing. In this effort, our previous experience with hearing-aid users in two main areas (models of speech intelligibilty and signal processing algorithms for noise reduction) is being applied to benefit cochlear-implant users, a population for whom background noise affects speech intelligibility even more adversely than hearing-aid users.

The model of speech intelligibility under consideration is based on the speech transmission index (STI). STI was originally developed as a way of assessing room acoustics, and the original STI calculations are based on a system's response to specific test signals. Although these test signals are appropriate for assessing room acoustics, they are not suitable for many kinds of signal processing used in hearing aids and cochlear implants. As a result, several research groups, including our own, have attempted to develop methods for calculating STI based on the speech signal itself, rather than specific test signals.

We have completed an analysis of the various methods for speech-based STI calculation described in the literature, and we have established explicit relationships between the various speech-based STI calculations. This analysis revealed a number of issues that may hamper the performance of existing speech- based STI calculations for both noise reduction and cochlear-implant speech processing. We have developed improved methods for speech-based STI calculation that address these issues. These various methods for speech-based STI calculation will be compared to each other as well as actual intelligibility data for cochlear-implant users and normal-hearing subjects listening to a simulation of cochlear-implant speech processing. We have developed test procedures for these intelligibility experiments, and the appropriate software is currently being implemented.

**Signal Processing for Hearing Aids**

Work on noise reduction for hearing aids has progressed in two areas: A) Design, implementation and assessment of automatic gain control (AGC) algorithms for single-microphone noise reduction; B) Analysis of previously- obtained experimental results evaluating multi-microphone adaptive-array hearing aids.

Previous work in our lab has considered the use of automatic gain control algorithms specifically designed to reduce background noise, including modifications to the dual front-end AGC (Moore and Glasberg, Br. J. of Audiology 22, pp. 93-104, 1988). The main idea of the dual front-end AGC is to have two AGC components applied simultaneously; a slow-acting wideband automatic volume control, which determines the gain for most acoustic conditions, plus a fast- acting AGC with a higher threshold to provide transient suppression. Recent work in this area suggests a number of modifications that affect the design, implementation and parameter choices of the dual front-end AGC algorithm (Stone et al., JASA 106, pp. 3603-3619, 1999). We implemented several algorithms based on this newer version of the dual front-end AGC, and evaluating the algorithm with hearing-impaired subjects.

The dual front-end AGC system that we tested includes two optional features, a hold-timer, as proposed by Stone et al., and a signal-to-noise ratio (SNR) detector, as proposed in previous work performed in our lab. The purpose of the hold timer is to reduce pumping without extremely long recovery times. It effectively prevents gain fluctuations during speech and during brief pauses in speech. The purpose of the SNR detector is to have modify the release time of the slow acting AGC component so that it releases more quickly when strong speech (from the hearing-aid wearer's voice) is followed by weaker speech (from anther talker).

We evaluated five algorithms, four dual front-end AGC algorithms (all combinations of with and without the hold timer and with and without SNR detection) plus a linear reference condition with compression limiting. A set of twenty acoustic test conditions representative of everyday listening situations were selected. These conditions include speech in quiet at various levels, speech plus

multitalker babble at various signal-to-noise ratios, speech plus continuous environmental noises (for example, vacuum cleaner, hair dryer, running water) and speech plus transient environmental noises (for example, glass breaking, hammering, door slamming). Multiple stimuli corresponding to these conditions were processed by all five algorithms. Hearing-impaired subjects listened to the processed segments and rate each segment for subjective intelligibility and quality on 0-10 point scales.

Five hearing-impaired subjects participated in these experiments. Preliminary analysis of results for three subjects found no clear differences among the four dual frond-end AGC algorithms. Major differences did exist between the linear reference condition and the dual front-end systems. The direction of these differences are both subject- and condition-dependent. Future work will complete the analysis of data collected from the remaining subjects, and futher explore the implications of the different interference conditions.

With the completion of NIH Contract N01 DC-5-2107 for Hearing Aid Device Development, analysis and interpretation of experimental results obtained under that contract are being performed under the Hearing Aid Research grant.

Several array-processing algorithms were implemented and evaluated with experienced hearing-aid users. The arrays consisted of four directional microphones mounted broadside on a headband worn on the top of the listener's head. The algorithms included two adaptive array-processing algorithms, one fixed array-processing algorithm, and a reference condition consisting of binaural directional microphones. These algorithms were evaluated in noise conditions consisting of either one or three directional interferers. Two performance metrics were used: quantitative speech reception thresholds, measured by dynamically adjusting the signal-to-noise ratio of the test materials; and qualitative subject preference ratings for ease-of-listening, measured using a paired-comparison procedure.

Analysis of the experimental results revealed that on average, the fixed algorithm improved speech reception thresholds by 2 dB over the reference condition. The adaptive array processing algorithms provided 7-9 dB improvement over the reference condition. Subjects judging ease-of-listening generally preferred all array-processing algorithms to the reference condition. The results suggest that these adaptive algorithms should be evaluated further in more realistic acoustic environments.

**Characteristics of Sensorineural Hearing Impairment**

Previous studies that have examined the relationship between speech quality and intelligibility for amplitude compressed speech have been mixed. Some investigators have observed better speech intelligibility and higher speech quality ratings with compression (e.g., Humes et al., 1999). Other studies have found that ratings of speech quality decreased as compression ratio was increased, but found no effect of compression ratio on speech intelligibility (e.g., Boike and Souza, 2000). While it is difficult to reconcile the disparate results of these studies, differences in compression systems and fitting procedures may be a factor. In addition, differences in listeners suprathreshold processing abilities, including temporal resolution, frequency selectivity, and auditory dynamic range, may account for at least some of the variability in results across studies.

An alternative strategy to study the relationship between speech quality and intelligibility of amplitude compressed speech is to allow listeners with normal-hearing to experience selected perceptual effects of cochlear hearing loss. These effects are produced by simulations of cochlear hearing impairments that transform the sound that reaches the ear of listeners with normal-hearing to achieve specific changes in auditory abilities. These simulations create listeners with new types of hearing impairments that do not exist clinically. Simulation parameters are derived from absolute thresholds.

In this study, the perception of amplitude compressed speech processed to incorporate the effects of elevated thresholds and loudness recruitment were examined in normal-hearing listeners. Two hearing loss configurations were simulated: (1) a flat moderate loss and (2) a sloping mild to moderate hearing loss. The speech stimuli consisted of nonsense sentences that are syntactically correct but semantically meaningless (Picheny et al., 1985). Speech was presented in either a background of quiet or noise. The interfering background noises consisted of either speech-spectrum noise (i.e., random noise filtered to have the same long-term average rms as the speech stimuli) or multi-talker restaurant babble. Digital signal processing techniques (Simulink, 3.0.1) were used to simulate a linear-gain hearing aid and an amplitude compression hearing aid.

For the linear-gain hearing aid, frequency dependent gain for each hearing loss configuration was specified according to NAL-R. For the compression hearing aid simulation, signals were compressed independently in four non- overlapping frequency bands with center frequencies of 0.5, 1, 2, and 4 kHz. Amplitude compression in each band was implemented using a level detector that provided a control signal to the amplifier in each band. The compression threshold in each band was set at 20 dB below the speech input rms in that band to insure that the entire speech signal was compressed. Attack time and release time constants for each band were specified as 5 ms and 200 ms, respectively. Compression ratios for the two hearing loss configurations simulated were specified as follows: (1) Flat hearing loss- Compression ratio equal to 2:1, in all four channels. (2) Flat hearing loss- Compression ratio equal to 3:1, in all four channels. (3) Sloping hearing loss- Compression ratio equal to 1.5:1 in channels one and two, and 2.5:1 in channels three and four. Subsequent to the compression stage, all stimuli were processed with a high-frequency emphasis filter. The shape of the filter was based on the NAL-R target curve in order to increase high- frequency audibility. Gain for each compression hearing aid was adjusted so that output rms levels were equal to linear-gain hearing aid output levels (2.5 dB), for a 70 dB rms speech input level.

Level matching was verified for the broadband signals, and band-pass filtered signals using the 14-band filter bank used in the hearing loss simulation (see below) and for the broadband signals. This approach was taken to insure that the effect of the hearing loss simulation would be the same for all amplification conditions. The effects of elevated thresholds and loudness recruitment were simulated in normal- hearing listeners using a multiband expansion algorithm. With this approach, first investigated by Villchur (1974;1977), level-dependent attenuations are applied to sounds in different frequency bands to map tone levels at the hearing-impaired listener's detection thresholds to normal threshold levels. It is correctly referred to as an expansion simulation because output level changes are greater than input level changes, in dB. Because expansion simulations work by attenuating the input signal, they can be used to simulate hearing impairments more severe than can be simulated comfortably by masking noise.

In this study, digital signal processing techniques (Simulink, 3.0.1) were used to implement a 14-band system previously reported on by Duchnowski and Zurek (1995) to accurately model the perception of speech in noise by listeners with mild to moderate cochlear hearing loss. The simulator analyzed the input signal into 14 non-overlapping bands with filter bandwidths comparable to critical bandwidths . The lowest cut-off frequency and the highest cut- off frequency were 50 and 4500 Hz, respectively. Percent correct scores and categorical ratings of pleasantness and intelligibility were measured in all simulated loss-listeners. Results indicate that subjects were able to accurately rate the intelligibility of speech; categorical ratings of intelligibility were similar to objective measures. Results also showed that while compression (cr=2) significantly improved speech intelligibility (relative to linear amplification) for the simulated flat-loss listeners, both in quiet and multi-talker babble, the compressed speech was always rated as significantly less pleasant. This finding suggests, that categorical ratings of pleasantness alone should not be used for selecting compression ratio, if the goal is to maximize speech intelligibility.

## References

Aguayo, I. "Evaluation of different forms of compression in digital hearing aids," M.Eng. Thesis, Massachusetts Institute of Technology, Cambridge, MA, June 2001.

Arlinger, S., and Gustafsson, H. A. (1991). "Masking of speech by amplitude-modulated noise," Journal of Sound and Vibration. Vol. 15, no.3, pp. 441-445.

Boike, K. T., and Souza, P.E. (2000). "Effect of compression ratio on speech recognition and speech-quality ratings with wide dynamic range compression amplification," J. Speech Hear. Res. 43, 456-468.

Drullman, R. Festen, J. M., Plomp, R. (1984) "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. 95, 2670-2680.

Duchnowski, P., and Zurek, P. M. (1995). "Villchur revisited: Another look at automatic gain control simulation of hearing loss," J. Acoust. Soc. Am. 98, 3170-3181.

Goldsworthy, R. and Greenberg, J. "Using STI as a performance metric for cochlear implant users," presented at the Conference on Implantable Auditory Prostheses, Asilomar Conference Grounds, Pacific Grove, CA, August 2001.

Greenberg, J.E. and Zurek, P.M. "Microphone-array hearing aids," in Microphone Arrays: Techniques and Applications, Brandstein and Ward, eds., Springer, 2001.

Greenberg, J.E., Desloge, J.D., and Zurek, P.M. "Evaluation of array-processing algorithms for hearing aids," submitted to J. Acoustical Society of America.

Hargus, S. E., and Gordon-Salant S. (1995). "Accuracy of Speech Intelligibility Index Predictions for Noise-Masked Young Listeners with Normal Hearing and for Elderly Listeners with Hearing

Houtgast, T. and Steeneken, H. (1985) "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. 77, 1069-1077.

Humes, L.E., Christensen, L., Thomas, T., Bess, F.H., Hedley-Williams, A., and Bentler, R. (1999). "A comparison of the aided performance and benefit provided by a linear and a two-channel wide dynamic range compression hearing aid," J. Speech Hear. Res. 42, 65-79.

Krause, J.C. (1995). "The Effects of Speaking Rate and Speaking Mode on Intelligibility," S.M. Thesis, Massachusetts Institute of Technology,

Krause, J.C. (2001). "Properties of Naturally Produced Clear Speech at Normal Rates and Implications for Intelligibility Enhancement." PhD dissertaion, Massachusetts Institute of Technology,

Krause, J.C. and Braida, L.D. (1996). "Properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. 100 (4), S2828.

Payton, K. and Braida, L. (1999) "A method to determine the speech transmission index from speech waveforms," J. Acoust. Soc. Am. 106, 3637-3648.

Payton, K. L., Uchanski, R. M. and Braida, L. D. (1994): "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am., Vol. 95, 1581-1592.

Picheny, M. A., Durlach, N.I., and Braida, L.D. (1985). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. 28, 96-103.

Uchanski, R.M., Choi, S., Braida, L.D., Reed, C.M., vand Durlach, N.I. (1996). “Speaking Clearly for the Hard of Hearing IV: Further Studies of the Role of Speaking Rate," J. Speech and Hearing Res. 39 494-509.

Villchur, E. (1974). "Simulation of the effect of recruitment on loudness relationships in speech," J. Acoust. Soc. Am. 56, 1601-1611.

Villchur, E. (1977). "Electronic models to simulate the effect of sensory distortions on speech perception by the deaf," J. Acoust. Soc. Am. 62, 665-674.

## Auditory Perception and Cognition

### Sponsor

National Institutes of Health
Grant R01 DC 3909

### Project Staff

Dr. Andrew J. Oxenham (P.I.), Stephan Ewert[a], Zachary Smith, Joshua Bernstein, Hwa Jung Son, Michael Qin, Hector Penagos

### Introduction

Our aim is to further our basic understanding of normal and impaired hearing using behavioral, or psychophysical, techniques. This work is approached at two different levels. The first is concerned with finding behavioral measures of how the peripheral auditory system breaks down incoming sounds into their constituent frequency components. This work focuses on the perceptual effects of cochlear mechanics, in particular nonlinear filtering. The second level addresses how sounds, which are separated according to frequency content in the cochlea, are recombined at higher levels of auditory processing to form ‘auditory objects’ or ‘auditory streams’. The findings are used to place constraints on models of higher-level auditory processing and may act as guides to studies in search of the neurophysiological underpinnings of auditory perception. Long-term benefits of such research include improved auditory prostheses (hearing aids or cochlear implants), improved perceptually based speech recognition systems and computational auditory scene analysis.

### 1. Behavioral measures of phase response and peripheral compression in the human auditory system (S. Ewert and A. Oxenham)

This work represents an extension of the work described in last year’s report. Harmonic tone complexes with constant phase curvature were used to define the phase response of the auditory filters. In our current work, the assumption that auditory filtering can be approximated as having a constant phase curvature was tested by using maskers in different frequency ranges. It was found that, consistent with physiological data from other mammals, the curvature is negative for

[a] Department of Medical Physics, Universität Oldenburg, 26111 Oldenburg, Germany.

frequencies around characteristic frequency (CF) but tends to zero for frequencies well below CF. The new results provide information which can be incorporated into models of human peripheral auditory processing. The work on human frequency selectivity, described below, shows clearly that the oft-made assumption that animal models provide good approximations of human auditory processing is not always valid. The current technique offers a way to define peripheral auditory

processes in a noninvasive way, thereby providing the first data on the temporal response of human auditory filtering.

## 2. The perception and functional imaging of complex pitch (J. Bernstein, H. Penagos, A. Oxenham)

Pitch is a fundamental attribute of auditory sensation and has been the subject of research for well over a century. However, the mechanisms underlying the phenomenon remain a matter of vigorous debate. Most sounds that elicit a pitch sensation comprise several frequency components. One pressing question is how the auditory system computes a single pitch when confronted with a combination of tones of many frequencies. A number of theories have been proposed to account for such results. The purpose of our behavioral experiments is to distinguish between them. Our results to date show that, to contribute to a pitch percept, individual frequency components not only have to be perceptually separable (or resolved) from other tones, but that tone which are not normally resolved do not contribute to pitch perception, even if they are presented in a way that enables the ear to resolve them. These results can be interpreted as supporting “template” theories of pitch.

Earlier work has shown regions of the brain, particularly around Heschl’s Gyrus, which respond strongly to sounds with regular temporal waveforms, eliciting pitch (Griffiths *et al.*, 1998; Griffiths *et al.*, 2001). The question we asked was whether the increased activity, measured using functional magnetic resonance imaging (fMRI) was a reflection of stimulus temporal regularity, as was claimed, or whether it represented a response to a higher-level pitch sensation. We addressed this question in collaboration with Dr. Jennifer Melcher of MEEI by using stimuli that had the same temporal regularity in the acoustic waveform, but elicited very different pitch strengths. Pilot data suggest that, while the response in brainstem regions is determined by stimulus regularity, the response in cortical regions is determined more by perceived pitch strength than regularity. These findings, if confirmed, suggest that significant pitch processing takes place at or before primary auditory cortex.

## 3. Auditory chimeras (Z. Smith, A. Oxenham)

This work was performed in our laboratory in collaboration with Dr. Bertrand Delgutte of RLE and MEEI. We investigated the relative perceptual roles of stimulus envelope and fine structure by producing novel stimuli coined “auditory chimeras” by Dr. Delgutte. These stimuli are created by splitting the sound into frequency subbands, extracting the envelope and fine structure from each subband by means of a Hilbert transform, and then combining the envelopes of one sound with the fine structures of the second sound. We found that speech perception is dominated by information in the envelope, while melody perception and localization is determined primarily by the fine structure. This has the intriguing consequence that “what” sound is heard is determined by the envelope, but “where” it is heard is determined by the fine structure. We have suggested that this represents a possible acoustic basis for the proposed “what” and “where” pathways in the auditory cortex (e.g., Tian *et al.*, 2001). Also, the importance of the fine structure in pitch and localization suggests that such information may be of benefit to cochlear-implant patients. Currently, the fine-structure information is discarded in cochlear-implant processing.

### 4. Cochlear-implant simulations of speech reception in complex backgrounds (M. Qin, A. Oxenham)

The work described above on auditory chimeras and previous studies (Shannon *et al.*, 1995) have shown that good speech reception can be achieved with no fine-structure information and with severely degraded spectral resolution. Many of these studies have been carried out with normal-hearing listeners, using acoustic stimuli that are designed to simulate the processing of cochlear implants. Although speech in quiet and speech in a background of steady white noise seems to be robust to cochlear-implant processing, the same may not be true of speech in more complex backgrounds, such as amplitude-modulated noise or a competing talker. In such cases, audibility of the target speech may not be the primary factor. Instead, successful speech reception may depend on the ability to perceptually segregate the target from the masker. Our experiments using such simulations support this conjecture. Performance, even with very high spectral resolution, is severely degraded in the presence of complex maskers when fine-structure information is eliminated. The results provide important information for future cochlear-implant algorithms, as well as supplying basic information on the cues used by the auditory system in perceptually segregating competing sources.

### 5. Informational masking and auditory feature detection (H. Jung Son, A. Oxenham)

Most masking phenomena can be explained at a fairly peripheral level in terms of an interaction between the patterns of excitation on the basilar membrane between the masker and the target. In a striking departure from this general rule, it has been shown that very large masking effects can be produced by presenting maskers, whose frequencies vary randomly from trial to trial, simultaneously with a fixed-frequency target, even if the frequencies of the maskers are far

removed from that of the target (Neff and Green, 1987). This and related effects have been termed informational masking. Amplitude modulating the target has been shown to reduce informational masking somewhat (Neff, 1995). It was proposed that the modulation provided a cue with which the target could be perceptually segregated from the masker components. We tested this idea further by amplitude modulating the masker components but not the target. If perceptual segregation could account for the results, then a similar release from masking might have been expected. In fact, no release was observed. This, together with similar results from our other experimental conditions, leads us to conclude that the release from masking observed when the target was modulated was due to a form of feature detection: any modulation present was a reliable cue to detecting the target. This does not imply that the target was in some way perceptually segregated. When the maskers were modulated, the presence of the unmodulated target had very little effect on the overall amount of modulation. Thus, our results indicate that the interpretation in terms of perceptual segregation due to amplitude modulation was is probably false, and that the phenomenon is better regarded as a form of auditory feature detection. This work contributes towards a better understanding of how high-level acoustic features are processed in the auditory system.

## Publications

**Journal Articles, Published**

Oxenham, A. J. "Forward masking: Adaptation or integration?," *J. Acoust. Soc. Am.* 109: 732-741 (2001).

Oxenham, A. J., and Dau, T. "Modulation detection interference: Effects of concurrent and sequential streaming," *J. Acoust. Soc. Am.* 110: 402-408 (2001).

Oxenham, A. J., and Dau, T. "Reconciling frequency selectivity and phase effects in masking," *J. Acoust. Soc. Am.* 110: 1525-1538 (2001).

Oxenham, A. J., and Dau, T. "Towards a measure of auditory-filter phase response," *J. Acoust. Soc. Am.* 110: 3169-3178 (2001).

**Journal Articles Accepted for Publication**

Plack, C. J., Oxenham, A. J., and Drga, V. "Linear and nonlinear processes in temporal masking," *Acustica/Acta-Acustica.* Forthcoming.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. "Chimaeric sounds reveal dichotomies in auditory perception," *Nature.* Forthcoming.

Shera, C. A., Guinan, J. J., and Oxenham, A. J. "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," *Proc. Nat. Acad. Sci.* Forthcoming.

**Meeting Papers, Published**

Dau, T., and Oxenham, A. J. "Phase effects in masking in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* 109: 2466.

Delgutte, B., and Oxenham, A. J. "Auditory chimeras," *Midwinter Meeting of the Association for Research in Otolaryngology*, St. Petersburg Beach, Florida, February 4-8, 2001.

Oxenham, A. J., and Dau, T. "Behavioral estimates of the phase response of the auditory system," *Midwinter Meeting of the Association for Research in Otolaryngology*, St. Petersburg Beach, Florida, February 4-8, 2001.

**References Cited**

Griffiths, T. D., Buchel, C., Frackowiak, R. S., and Patterson, R. D. (1998). "Analysis of temporal structure in sound by the human brain," Nat. Neurosci. 1, 422-427.

Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., and Patterson, R. D. (2001). "Encoding of the temporal regularity of sound in the human brainstem," Nat. Neurosci. 4, 633-637.

Neff, D. L. (1995). "Signal properties that reduce masking by simultaneous, random-frequency maskers," J. Acoust. Soc. Am. 98, 1909-1920.

Neff, D. L., and Green, D. M. (1987). "Masking produced by spectral uncertainty with multi-component maskers," Perception and Psychophysics 41, 409-415.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," Science 270, 303-304.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). "Functional specialization in rhesus monkey auditory cortex," Science 292, 290-293.

# Tactile Communication of Speech

**Sponsor**

National Institutes of Health/National Institute on Deafness and Other Communication Disorders
Grant 2 R01 DC00126

**Project Staff**

Andrew R. Brughera, Lorraine A. Delhorne, Nathaniel I. Durlach, Seth M. Hall, Eleanora Luongo, Charlotte M. Reed, Mandayam A. Srinivasan, Han-Feng Yuan

**Goals and Significance**

The long-term goal of this research is to develop tactual aids for persons who are profoundly deaf or deaf-blind to serve as a substitute for hearing in the reception of speech and environmental sounds. This research can contribute to improved speech reception and production, language competence, and environmental-sound recognition in such individuals. This research is also relevant to the development of improved tactual and haptic displays for a broad class of applications (including virtual-environment and teleoperator systems in addition to sensory aids).

Research is being conducted in two major areas. Work in *Area 1* (Basic Studies of Human Touch) is designed to increase our knowledge concerning the transmission of information through the sense of touch. This research includes theoretical and experimental studies

concerned with dynamic information transfer as well as experimental work designed to increase our understanding of the psychophysical properties of the sense of touch. Work in *Area 2* (Tactual Displays of Speech and Environmental Sounds) is concerned with the application of tactual displays to sensory aids for persons who are profoundly deaf or deaf-blind. This research includes studies related to the processing and display of speech and environmental sounds through the tactual sense as well as studies concerned with evaluations of performance achieved through these displays.

**Current Studies**

*Basic Studies of Human Touch*
Experiments are being conducted to measure subjects' ability to identify sequences of multidimensional tactual signals. These experiments employ a set of 28 stimuli delivered through a multi-finger tactual stimulator (see Tan et al., 1999). The 28 stimuli are composed of 7 different waveforms presented to four different locations (thumb, middle finger, index finger, or all three digits stimulated simultaneously). The seven waveforms consist of 3 single-frequency stimuli, three double-frequency stimuli (resulting from all possible combinations of the three single-frequency waveforms), and one triple-frequency stimulus (resulting from the combination of the three single-frequency waveforms). Stimulus sets were created at two different durations: 250 msec (with single-frequency values of 2, 30, and 300 Hz) and 125 msec (with frequencies of 4, 30, and 300 Hz).

Subjects initially receive training on the identification of the seven waveforms presented at one location, followed by training on the identification of all 28 stimuli (7 waveforms and 4 locations). Training is conducted using a one-interval forced-choice procedure with trial-by-trial correct-answer feedback. Following the training, subjects are tested in an AXB paradigm and in an $X_1 \ldots X_n$ paradigm where n is gradually increased from 2 to 4. In the AXB paradigm, the subject's task is to identify the target stimulus X which is preceded and followed by maskers A and B (all selected at random from the same stimulus set). In the $X_1 \ldots X_n$ paradigm, the subject's task is to identify n stimuli (randomly selected from a given stimulus set) presented sequentially. The

duration of the stimuli was fixed at either 250 or 125 msec and seven values of inter-stimulus interval (ranging from 0 to 640 msec) were studied in both paradigms.

Data collection is in progress on a group of six subjects. For the 250-msec signals, four subjects have completed the AXB experiment and have begun testing in the $X_1X_2$ paradigm. Identification of the middle signal in the AXB sequence decreased with increasing presentation rate, falling from roughly 92% correct for an inter-stimulus interval (ISI) of 640 msec to 80% for an ISI of 0 msec. Performance on the identification of two sequential stimuli was somewhat worse than that observed in the AXB paradigm. For long ISI, identification of two stimuli was roughly 5 percentage points worse than identification of one stimulus; however, this difference increased to roughly 15 percentage points at short ISI. For the 125-msec signals, four subjects have completed the AXB experiment and are beginning testing in the $X_1X_2$ paradigm. Overall scores

with the 125-msec signals are less than those obtained at 250 msec, although the same general trends are observed at both durations.

*Tactual Displays of Speech and Environmental Sounds*

Current work in the area of tactual displays of speech is concerned with the development of improved displays of consonantal voicing as a supplement to speechreading. This research is focused on the use of envelope signals as the basis for the extraction and display of cues to consonantal voicing. Research during the past year includes acoustic measurements of CVC syllables to determine voicing-related characteristics for display through the tactual sense. These acoustic measurements were conducted on the envelopes of a 500-Hz lowpass band and a 3000-Hz highpass band of speech. These bands were selected to capture the different patterns of spectral energy that are observed for voiced versus voiceless consonants (regardless of place or manner of production). Specifically, voiceless consonants tend to have significant energy above 3 kHz accompanied by less energy below 500 HZ, while voiced consonants tend to exhibit significant energy below 500 Hz (at least during part of their production) in addition to significant energy above 3 kHz. The timing of the onset of high-frequency relative to low-frequency energy differs for voiced relative to voiceless consonants: typically, this relative onset is longer for voiceless compared to voiced consonants. For these reasons, measurements were made of the onset asynchronization of the two envelopes. Envelope-onset asynchronization (EOA) was defined as the onset time of the high-frequency envelope minus the onset time of the low-frequency envelope.

Envelope-onset aysnchronization (EOA) was measured for 16 initial consonants in CVC syllables, constructed with 16 different medial vowels and with the final consonant selected at random from a set of 19 consonants. Measurements were made on 112 tokens of each initial consonant (based on CVC recordings from two different female talkers). The distribution of EOA values was examined for 8 pairs of voiced-voiceless cognates. Two observations were made from the resulting distributions: (1) there appears to be little overlap in the EOA values for the voiced versus the voiceless member of each pair and (2) there appears to be greater variability associated with the EOA values of voiceless compared to voiced consonants. The distribution of EOA was also examined for two major categories of “voiced” and “voiceless” by grouping the eight voiced consonants and the eight unvoiced consonants. These data indicate that EOA patterns are stable across different manners and places of consonant production and in a variety of vowel contexts.

Calculations were made to estimate the ability of an “ideal observer” to distinguish voiced from voiceless consonants on the basis of EOA values. The cumulative distribution of EOA for each consonant was derived from the probability-density functions and was fit by a Gaussian function. The Gaussian fitting determined the values of mean and standard deviation to minimize the sum of squares of the error. The estimated means and standard deviations were then used to derive

the performance metric d-prime assuming a two-interval, two-alternative forced-choice procedure. Across the 8 pairs of voiced-voiceless consonants, estimates of d-prime ranged from roughly 3 to 8, indicating high discriminability by the ideal observer. Thus, the EOA appears to provide a robust cue for voicing and is promising as the basis of a tactual display for supplementing speechreading.

Our current approach towards a tactual display of envelope-onset asynchronization involves the use of a multi-finger tactual stimulating device. The output of the high-frequency envelope band will modulate a sinewave presented to one digit and the output of the low-frequency envelope band will modulate a different sinewave presented to a second digit. The stimulus-onset asynchrony between the two channels will serve as a perceptual cue to voicing. When the high-frequency channel precedes the low-frequency channel (by a mean difference of roughly 152.5 msec averaged across all voiceless sounds), the stimulus is likely to be voiceless. When the two channels are roughly simultaneous (mean EOA of roughly 2.5 msec averaged across all voiced sounds), then the stimulus is likely to be voiced. Evaluations of the ability of human subjects to distinguish voiced from voiceless consonants under conditions of speechreading alone, the tactual display alone, and the combined display of speechreading plus tactual cue will be carried out in the coming year.

Work in the area of tactual displays of environmental sounds is concerned with the development of a survey to assess the interest of members of the deaf community in such devices. Previous efforts to develop tactual communication aids for the deaf have focused primarily on the reception of speech. Field studies of the adult users of tactual aids, however, have led to the observation that these aids are also highly useful in the reception of non-speech sounds in the environment (Reed and Delhorne, 1995). Thus, it may be possible that a broader class of deaf individuals, with little or no interest in speech reception, may nonetheless be interested in tactual displays for gaining information about acoustic events such as warning signals and sounds associated with nature.

A pilot survey has been developed for administration to a group of 20 deaf and hard-of-hearing students and staff of the National Technical Institute for the Deaf. The survey consists of a total of 70 questions in four different areas. The first section of the survey is concerned with obtaining information concerning the respondent's history of hearing loss, preferred modes of communication (i.e., sign language versus oral communication), and use of general support services. The second section is concerned with obtaining information about the respondent's current use of communication devices, including hearing aids, cochlear implants, tactile aids, and assistive-listening devices and/or alerting systems. The third section probes the respondent's interest in a variety of different types of environmental sounds, including sounds occurring in the home, outdoors, at work or school, as well as sounds related to music. The final section of the survey probes the respondent's attitudes towards the importance of a variety of different characteristics in the development of an "ideal" device for receiving information about acoustic environmental stimuli. The resulting pilot data will be summarized to determine relationships between characteristics of a respondent's personal history of deafness and communication and attitudes towards the reception of non-speech acoustic stimuli in the environment. The results of the pilot study, together with information gained from in-depth interviews with ten of the respondents, will guide us in producing a final version of the survey for distribution to a pool of several hundred deaf and hard-of-hearing subjects.

## Publications

Bratakos, M.S., Reed, C.M., Delhorne, L.A., and Denesvich, G., "A Single-Band Envelope Cue as a Supplement to Speechreading of Segmentals: A Comparison of Auditory versus Tactual Presentation." *Ear and Hearing*, 22*,* 225-235 (2001).

Reed, C.M., and Braida, L.D, "Frequency Compression," in *MIT Encyclopedia of Communication Science and Disorders (MITECSD)*, ed. Ray D. Kent (Cambridge: MIT Press), forthcoming.

### References

H.Z. Tan, N.I. Durlach, C.M. Reed, and W.M. Rabinowitz, "Information Transmission with a Multifinger Tactual Display," *Perception and Psychophysics,* 61: 993-1008 (1999).

Reed, C.M., and Delhorne, L.A., "Current Results of a Field Study of Adult Users of Tactile Aids," *Seminars in Hearing,* 16*,* 305-314 (1995).