Hearing Aid Research

Sponsor
National Institutes of Health Grant R01 DC00117

Academic and Research Staff
Professor Louis D. Braida, Dr. Julie Greenberg, Dr. Karen L. Payton

Visiting Scientists and Research Affiliates
Dr. Paul Duchnowski, Dr. Kenneth W. Grant, Dr. Jean C. Krause, Dr. Christine M. Rankovic

Graduate Students
Alan Asbeck, Raymond Cheng, Raymond Goldsworthy, Peninah S. Rosengard

Technical And Support Staff
Joseph A. Frisbie,

Our long-term goal is to develop improved hearing aids for people suffering from sensorineural hearing impairments and cochlear implants for the deaf. 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 limitations 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.

Studies and Results

Simulations of Sensorineural Hearing Loss

We have developed software hearing loss and hearing aid simulators that permit wide flexibility in control of parameters such as hearing thresholds, compression range, and attack and release times and allow the use of normal-hearing listeners in tests of hearing aid processing algorithms. The signals generated by these simulators have also been used to generate speech-based STI predictions of performance (see Sec. 3-c).

The Hearing Loss Simulation, which has 14-bands and includes recruitment, is a software version of the analog simulation of Duchnowski and Zurek (1995) and of the digital simulation of Lum and Braida (1997). This simulation allows us great flexibility to modify parameters, should we choose to simulate different impairment characteristics. Our current experiments simulated a flat, 50 dB, hearing loss and a sloping hearing loss. Both simulated losses included recruitment. In each of 14 bands, the desired impairment threshold was attenuated to normal-hearing threshold. Expansion occurred independently in each band until the sound level in a band reached the threshold of recruitment, above which the gain was unity.

The Hearing Aid Simulation allows compression, independently, in up to four bands with variable static compression ratios, attack times and release times. The hearing aid also optionally includes the NAL-R frequency gain characteristics. We can easily modify the parameters of this simulation to mimic a wide range of hearing aids. Together, the hearing aid and hearing loss simulations allow us to predict the effects of different aids on a wide variety of hearing losses and then to test the results on normal-hearing listeners.

Wide Dynamic Range Compression. Monaural simulations of a linear-gain hearing aid and a four-channel wide dynamic range compression hearing aid were tested in conjunction with expansion simulations of sensorineural hearing impairment (Rosengard et al., 2004a). The purpose of this study was two-fold: 1) to determine the extent to which four-channel, slow-acting wide dynamic range amplitude compression can counteract the perceptual effects of reduced auditory dynamic range, and 2) to examine the relationship between objective measures of speech intelligibility and categorical ratings of speech quality for speech processed with amplitude compression. Frequency dependent gain for the linear hearing aid simulation was specified according to NAL-R. For the compression hearing aid simulation, input signals were compressed independently in four non-overlapping bands. Attack time and release time constants for each band were specified as 5 ms and 200 ms, respectively. Hearing aid processed signals were then further processed by the flat and sloping hearing loss simulations described previously and presented to nine normal-hearing listeners. Parameters of the impairment simulations and the aid simulations were adjusted to correspond to the specified hearing losses. Speech intelligibility scores and subjective ratings of speech quality were determined when the listeners with simulated losses were aided by both the compression system and the linear system.

Results showed that moderate compression can provide a small but significant improvement in speech intelligibility, relative to linear amplification, for simulated-loss listeners with a flat moderate hearing loss. This benefit was found for speech at conversational levels, both in quiet and in a background of babble. Simulated-loss listeners with sloping hearing loss did not show any improvement. However, in all listeners, ratings of pleasantness decreased as compression ratio increased. These findings suggest that subjective measures of speech quality should be used in conjunction with objective measures of speech intelligibility in order to optimize both comfort and intelligibility.

Estimates of Cochlear Compression. On average, the range of loudness recruitment is a fairly orderly function of the degree of hearing loss. As such, most simulations of recruitment derive the slope of the expansion characteristic from the specified hearing threshold alone (e.g., Duchnowski and Zurek, 1995). Alternatively, the simulation recruitment curve could be derived from psychoacoustic functions that characterize the response of the peripheral auditory system. This approach would allow a hearing-loss simulation to be customized to the impairment of an individual listener.

Recent physiological and psychoacoustical evidence suggests that loudness recruitment may be due to a loss or reduction of compression on the basilar membrane (e.g. Yates et al., 1990; Moore et al., 1997). Thus, simulator input-output functions could be derived from behavioral measures of basilar-membrane compression measured in individual listeners. Toward this end, we compared different estimates of compression in listeners with normal and impaired hearing. Two behavioral forward-masking paradigms that have been used to estimate the magnitude of cochlear compression are growth of masking (GOM) and temporal masking (TM). Despite substantial procedural differences, both of these measures purport to examine the same phenomenon, namely, basilar-membrane compression. However, no study has yet made a direct comparison of the two methods in the same subjects. Such a comparison is of interest for at least two reasons. First, it tests the assumption that the two techniques do indeed provide estimates of the same underlying phenomenon; and second, it allows a direct test of which measure is more efficient, and hence more accurate for a given effort. The purpose of this study was to determine the following: 1) whether these two measures produce within-subject results that are consistent across a range of signal frequencies and 2) whether one of the two measures leads to a more efficient estimate of compression.

GOM functions and TM curves were measured in a group of five normal-hearing listeners and five hearing-impaired listeners at signal frequencies of 1000, 2000, and 4000 Hz (Rosengard et al., 2004b). Compression values were derived from the masking data and confidence intervals were constructed around these estimates. The results suggest that GOM is a more efficient estimator of cochlear compression in listeners with both normal and impaired hearing, although it may not be appropriate for signal frequencies below 2000 Hz. In addition, results for the TM curves suggested that listeners with more severe hearing loss may have impaired temporal resolution that cannot be accounted for by a loss of on-frequency cochlear compression.

Speech Recognition Thresholds in Fluctuating Backgrounds and Cochlear Compression. Previous studies of speech perception suggest that the ability to achieve a release from masking in a fluctuating background sound is dependent on the integrity of active mechanical processes (e.g., Peters et al., 1998). Other aspects of hearing impairment, such as loudness recruitment, reduced frequency selectivity, reduced rate of decay of forward masking, and larger-than-normal gap detection thresholds in narrow-band noise, have also been attributed to a loss or reduction of compression on the basilar-membrane (e.g., Yates, 1990; Oxenham and Bacon, 2003). To further explore the relationship between psychoacoustic measures related to cochlear compression and suprathreshold measures of speech perception thought to be dependent on cochlear compression, we measured speech recognition thresholds in a group of five normal hearing (without simulations) and five hearing-impaired listeners with flat mild to severe hearing-loss. There were six background noise conditions: quiet, steady-state speech-shaped noise, single interfering talker, temporally modulated speech-shaped noise, notch-filtered speech-shaped noise, and temporally modulated notch-filtered speech-shaped noise. We also measured frequency selectivity using a notched-noise paradigm and derived estimates of the magnitude of cochlear compression from growth of masking functions at signal frequencies of 1000, 2000, and 4000 Hz. (see section above). In agreement with previous research (Moore et al., 1999b), we found a strong correlation between the width of auditory filters and basilar-membrane compression values (rho=0.86; n = 10, p = 0.01). The correlations between speech recognition thresholds in each background and the mean absolute thresholds, auditory filter bandwidth, and cochlear compression value at signal frequencies of 1000, 2000, and 4000 Hz were calculated. The correlations between the speech recognition thresholds in each background and the minimum absolute threshold, auditory filter bandwidth and cochlear compression value were also calculated. With the exception of the correlation between speech recognition thresholds in quiet and minimum absolute thresholds (ρηο=0.91; n = 5, p = 0.03), none were significant. In addition to calculating the correlations between absolute performance on the speech recognition task and the psychoacoustic measures, the correlation between these measures and the maximum speech recognition threshold *difference* (i.e., speech recognition threshold in steady-state speech-shaped noise minus speech recognition threshold in temporally modulated notch-filtered noise) was also calculated. The correlation between maximum threshold difference and minimum auditory filter bandwidth (ρηο=-0.91; n = 5, p = 0.03) and maximum threshold difference and minimum cochlear compression value (ρηο=-0.95; n = 5, p = 0.01) were both significant. The strong correlations between these measures suggest that listeners' ability to make use of local maxima and minima in a background's amplitude and frequency spectra is dependent on the integrity of the cochlear active mechanism. Furthermore, the finding that the actual speech recognition threshold values were not correlated with either of the two measures of peripheral compression suggests that, rather than the absolute performance in a given background noise condition, it is the ability to take *additional* advantage of spectral and temporal dips that is dependent on the active mechanism. That the correlation was significant for *minimum* auditory filter bandwidths and *minimum* cochlear compression values (as opposed to values averaged over the three signal frequencies or values at one particular frequency) suggests that an intact active mechanism in any region important for speech perception (i.e., 500 to 4000 Hz) may be sufficient for dip listening. While a more recent study which examined the relationship between speech perception in modulated backgrounds and auditory filter bandwidths in a group of hearing impaired listeners failed to find a significant correlation, the above results are indicative of a possible link between these measures.

Customization of Loudness Recruitment Simulations. As discussed above, multiband expansion can be used to simulate the effects of elevated thresholds and loudness recruitment in normal-hearing listeners. (Moore and Glasberg, 1993; Nejime and Moore, 1997) described an algorithm that implements expansion by raising the envelope of a signal to a power *N*. The value of *N*, in a given frequency band, is derived from the absolute thresholds at the band center frequency. In the current study, a variation of the expansion algorithm used by Nejime and Moore (1997) was customized to simulate the speech reception results of two hearing impaired listeners described above. Rather than simply calculating *N* from absolute hearing thresholds alone, the expansion characteristic was derived from cochlear compression values of each listener estimated from growth of masking data (Rosengard et al., 2004b). An extension of the loudness model described by and Moore and Glasberg (1997) was used to convert these compression values to simulator expansion values (*N*).

The speech reception results of the 2 hearing-impaired listeners were simulated in a group of 3 normal-hearing subjects using the original algorithm (Nejime and Moore, 1997) and the customized algorithm. Speech recognition thresholds were measured in quiet and a background of steady-state speech-shaped noise and temporally modulated notch-filtered noise. For the simulations of both hearing-impaired subjects, the maximum speech recognition threshold *difference* (speech recognition threshold in stead-state noise minus threshold in modulated noise) measured in the normal-hearing listeners using the original algorithm was less than 2 dB. In contrast, the average difference between the two background noise conditions was substantially larger with the customized algorithm (approximately 8 dB) and more closely matched the actual hearing-impaired data (approximately 7 dB). This finding is consistent with the expectation that the ability to exploit momentary fluctuations in a background sound is dependent on the magnitude of compression.

Analytical Modelling. Asbeck (2003) modeled the perception of vowels that had been distorted by various hearing loss simulations and compared the results with data on vowel identification by hearing impaired listeners. The hearing loss simulations included elevated absolute thresholds, loudness recruitment (Moore and Glasberg, 1993), reduced frequency resolution (ter Keurs et al., 1992; Nejime et al., 1997), and reduced temporal resolution (Hou and Pavlovic, 1994). The identification model assumed that listeners discriminate between sounds based on spectral differences.

Five listeners with sensorineural hearing loss identified six vowels (in CVC context). The effect of hearing impairment was simulated using signal processing on the actual vowel waveforms. Confusion matrices predicted by the identification model were compared to data from the impaired listeners. Whereas temporal smearing had very little effect on predicted error rates, predicted error rates increased as the degree of frequency smearing increased, and increased further when frequency smearing was combined with recruitment. However the perception model appears to use spectral cues in a different way than human listeners, most likely because the discrimination model is appropriate for noise stimuli rather than vowel sounds. Even so, the frequency smearing models caused the total error rate to increase with increasing smearing bandwidth, and the results were generally consistent with the expected behavior.

Characteristics of the Speech Signal

Numerous studies (e.g. Gagne et al, 1998; Uchanski et al., 1996; Schum, 1996; Payton et al., 1994; Picheny et al., 1985) have shown that the intelligibility advantage provided by speaking clearly is a large (11 to 34 percentage points) and robust effect that is not only of independent of talker, presentation level, frequency-gain characteristic, and degradation (noise or reverberation), but also independent of age and of whether listeners are normal-hearing or hearing-impaired. While talkers typically reduce their speaking rate when producing clear speech, we have shown (Krause and Braida, 2002) 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 intelligibility advantage of clear speech. Our recent work has been aimed at determining which acoustical properties of clear/normal speech contribute most to its higher intelligibility through acoustic analysis of clear/normal speech and related signal processing transformations.

Initial acoustical measurements of clear/normal speech (Krause and Braida, 2004) have shown that only a subset of the acoustic properties found in clear speech at slower rates (clear/slow speech) are apparent in clear/normal speech. Moreover, some of the acoustic characteristics (e.g., segment duration, pitch, and voice-onset time) retained in clear/normal speech have differed dramatically between talkers, suggesting that different talker strategies exist for producing clear speech at normal rates. Additional intelligibility tests, conducted on conversational and clear speech of two talkers (at both slow and normal rates) in four additional listening environments (low-pass filtering, high-pass filtering, reverberation, and listeners whose first language is not English), also suggest that talkers use different strategies to achieve clear/normal speech (Krause and Braida, 2003). In these tests, the intelligibility advantage of clear speech at normal rates varied with talker as well as listening environment. Based on these results and the acoustical analysis, it appears that while most talkers use the same or similar strategies when producing clear/slow speech, many of them may use different strategies when producing clear/normal speech. Nonetheless, three global acoustic properties associated with clear/normal speech were identified as properties that were most likely linked to robust intelligibility improvements: increased energy near the second and third formants, higher average and greater range of F0, and increased modulation depth of low frequency modulations of the intensity envelope.

To estimate the relative contributions of the various properties to the intelligibility advantage provided by clear/normal speech, three signal-processing transformations of conversational speech were developed based on each of the three properties. In intelligibility experiments with both normal hearing and hearing-impaired listeners, these transformations (applied singly and in combination to conversational/normal speech), as well as naturally produced clear/normal and conversational/normal (conv/normal) speech, were evaluated. Replicating our earlier results with normal-hearing listeners, all five normal hearing listeners exhibited an intelligibility advantage for clear/norm over conv/normal speech. Of three older hearing-impaired listeners tested, the intelligibility advantage of clear/normal speech varied, with the two listeners with flat hearing loss benefiting and the listener with sloping loss showing no benefit. Although the speech-based STI (Payton and Braida, 1999) predicted that most of the combinations would improve intelligibility over conv/normal speech, when presented in wideband noise to normal hearing listeners, our experiments with normal hearing listeners revealed an advantage only for formant-processed speech. For the three hearing-impaired listeners tested to date, this condition also provided a statistically significant benefit over conv/normal speech for some combinations of individual hearing-impaired listeners and talkers. However, neither this condition's benefit nor the benefit of clear/normal speech was statistically significant for the hearing-impaired group (3 listeners) as a whole, suggesting that either 1) more hearing-impaired listeners must be tested in order to establish the significance of the benefit, or 2) the advantage of clear/normal speech may not be as reliable across hearing-impaired listeners as it is across normal hearing listeners in noise (Krause, 2001).

If testing of additional hearing-impaired listeners suggests that the intelligibility benefits of these conditions do not extend reliably to hearing-impaired listeners, two possible explanations should be explored. One question would be whether the benefits of clear/normal speech are related to age, since the hearing-impaired participants tested to date were older (40 to 65 years) than the normal hearing participants (19 to 43 years). An age-related decline in speech reception has been reported for elderly listeners (Arlinger and Gustafsson, 1991), particularly those with hearing impairments (Hargus and Gordon-Salant, 1995). Some studies cite temporal processing deficits in the aging auditory system as a likely cause of such speech reception difficulties, but the relationship has not yet been firmly established (Strouse et al., 1998). A second question would be whether the additive noise model for simulating impairment in normal hearing listeners (Zurek and Delhorne, 1987) is inadequate. Although this simulation is appropriate for many mild to moderate impairments, it may not represent the effects of more severe impairments accurately.

Another focus of our recent work relates to the role of key words audibility in the intelligibility advantage of clear speech. Although sentences were normalized for level in all previous studies, individual words within a sentence are likely to have varied in level. The relationship between these level differences and the corresponding intelligibility of key words had not been investigated. One study (Uchanski et al., 1996) examined the intelligibility of key words in isolation, excised from clear and conversational speech, and found that the intelligibility advantage of clear speech was preserved. However, the individual excised words were not normalized beyond the normalization performed at the sentence level. Motivated by some analyses of our database by Schutte (2003), we recently conducted a similar intelligibility experiment that did investigate the effect of normalization at the word level on the intelligibility of excised words in clear/slow, clear/normal, and conv/normal speech. In addition, the intelligibility of excised words with normalization at the sentence level for conv/normal and clear/slow speech were included in the experiment in order to facilitate comparisons with the earlier Uchanski et al. (1996) study. Key words excised from 50 sentences (roughly 175 words) were presented in each condition. The speech of four talkers was evaluated in these conditions by eight normal hearing listeners in the presence of wideband noise at a signal-to-noise ratio of 0 dB.

Averaged across listeners, results for excised words with no additional normalization beyond the sentence level were in agreement with the Uchanski et al. study (1996): the advantage of clear/slow speech over conv/normal speech was the same or greater for excised words presented in isolation than for the same words in sentence context. Any increases in the intelligibility advantage indicate that the excision process was less harmful to the intelligibility of clear/slow speech than to that of conv/normal speech. A similar result was reported by Uchanski et al. (1996). For conditions with normalization at the word level, the advantage of clear/slow speech over conv/normal speech remained the same or greater for three of four talkers and decreased by 11% (from a 36 point advantage to a 32 point advantage) for the remaining talker, while the advantage of clear/normal speech over conv/normal speech remained greater for one talker and decreased by 23% on average for the remaining three talkers (from an 18 point advantage on average to a 14 point advantage on average). Thus, the maximum benefit that could potentially be achieved by signal transformations based on acoustic properties of clear speech at normal rates is still quite substantial, even when audibility is controlled at the word level, Because signal-processing transformations to date have not achieved this potential advantage provided naturally by clear/normal speech, additional acoustic properties of clear/normal speech that contribute to its high intelligibility must exist. Our future analyses of clear and conversational speech (Sec. 4-c) will combine several approaches in order to identify these properties.

Models of Speech Intelligibility

Our work on models of speech intelligibility is based on the speech transmission index (STI). The STI can be computed theoretically from signal-to-noise ratios and/or room impulse responses, or based on the measurements on 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. Several research groups, including our own, have attempted to develop methods for calculating the STI based on the speech signal itself, the speech based STI or SSTI. Our work can be divided into two main thrusts: theoretical studies to improve SSTI methods; and application of SSTI methods to simulated hearing impairments, and hearing-aid, and cochlear-implant processed speech.

Theoretical studies to analyze and improve speech-based STI methods

This work focused on the relationship between our SSTI method (Payton and Braida, 1999) and alternative methods (Ludvigsen, 1987; Drullman et al., 1994). The original SSTI methods computed speech envelopes by passing speech through octave-band filters, squaring each filter output, and lowpass filtering the results to generate envelope signals for each band (Payton and Braida,1999; Drullman et al., 1994; Houtgast and Steeneken, 1985). In contrast, Ludvigsen estimated envelopes by computing the magnitude of the Fast Fourier Transform (FFT) for frames of windowed speech. The squared magnitudes of the FFT within an octave band were summed into octave-band intensity values.

All of the SSTI methods considered are computed from two envelopes, that of the probe or input signal and that of the response or degraded signal. Most SSTI methods use these envelopes to compute speech-based modulation transfer functions (SMTFs), Ludvigsen's method does not. We assessed SSTI methods that do use SMTFs by comparing the SMTFs to theoretical MTFs. Houtgast and Steeneken's (1985) original SMTF calculation is based on the ratio of the magnitude of the response envelope spectrum to the magnitude of the probe envelope spectrum, but this method suffers from artifacts at high modulation frequencies, particularly in low octave bands (Payton and Braida, 1999). Our original SSTI method (Payton and Braida, 1999) attempted to deal with this problem by truncating the SMTF based on coherence. Drullman's original method (1994) uses the ratio of the real part of the cross power spectrum to the auto power spectrum of the probe envelope. We proposed and evaluated a new method for computing the SMTF based on the ratio of the magnitude of the cross power spectrum to the auto power spectrum of the probe envelope (Payton et al., 2002). We selected conditions (e.g., speech in speech-shaped noise, speech in reverberation, speech in speech-shaped noise plus reverberation, and speech in restaurant babble) for which theoretical MTF values can be computed to provide a ''gold standard' against which to compare the results. For all these conditions, both the new method (referred to below as Payton's method) and Drullman's method, followed the theoretical MTF very closely while Houtgast and Steeneken's (1985) method suffered from the previously mentioned artifacts.

In addition, we performed a theoretical analysis of three SSTI methods (Payton, Drullman, and Ludvigsen) that established explicit mathematical relationships among the various methods. This analysis revealed several issues that make all of these existing methods unsuitable for particular classes of non-linear operations (e.g., simulation of hearing loss by expansion described in Sec. 3-a, conventional noise reduction algorithms, and coding strategies used in cochlear-implant speech processors). To address these issues, we developed four candidate SSTI methods: three of which are simple variations on the Payton, Drullman, and Ludvigsen methods that incorporate modified normalization procedures, and a novel fourth that is based on the normalized cross-correlation between the probe and response envelope signals. We compared these four methods for conditions of additive noise and reverberation as well as two classes of non-linear operations: spectral subtraction and an envelope selection procedure (N-of-M processing) that is used in some cochlear-implant speech processors. Results indicate that all four of the candidate SSTI methods produce qualitatively reasonable results and warrant further study (Goldsworthy and Greenberg, 2004).

Application of speech-based STI methods

Simulated hearing impairments and hearing-aids. We evaluated several SSTI methods for their ability to predict the effects of simulated hearing impairments and hearing aids on intelligibility by comparing the methods to the intelligibility data from normal-hearing listeners listening to speech processed by hearing-loss and hearing-aid simulations (Sec. 3-a, Rosengard et al., 2004a). Two different hearing losses were simulated, a 50-dB flat loss and a moderate sloping loss. For the flat-loss, we examined a linear gain condition and two amplitude compression conditions, with a compression ratio (CR) of 2 and 3. For the sloping loss, we examined a linear gain condition and one compression condition, with CRs of 1.5 in the low frequency bands (<1.5 kHz) and 2.5 in the high frequency bands (>1.5 kHz).

We first computed SMTFs for the hearing-aid and hearing-loss simulation conditions (Payton et al., 2002). For both the flat and sloping simulated losses with linear amplification, we were able to compute theoretical MTF values for comparison. Both Drullman and Payton SMTFs matched the theoretical MTFs across all modulation frequencies and for all noise conditions, while the original SMTFs demonstrated the deviations described above. For the flat-loss compression conditions, the Drullman and Payton methods produced SMTFs that were comparable and relatively flat as a function of modulation frequency. For the conditions including the sloping-loss with the frequency-dependent compression, SMTFs computed for the Payton method were the only ones with valid values. Drullman's SMTF produced negative values at modulation frequencies of 10 Hz because the response envelope was exactly out of phase with the input envelope. At other modulation frequencies, the Drullman and Payton method SMTFs were similar.

To assess the validity of the various SSTI methods, we compared them to intelligibility data collected with normal-hearing subjects listening to speech processed by the hearing-loss and hearing-aid simulations. (When the Drullman method produced negative SMTF values, they were set to zero.) Qualitatively similar results were observed for all three methods. None of the SSTI methods tested predicted STI values that were consistent with intelligibility scores across all processing conditions. All methods predicted that linear processing conditions would be more intelligible than compression conditions, but the actual intelligibility scores were similar across processing conditions for a fixed noise type. Within a class of processing (linear or compression), we observed reasonable consistency between STI predictions and subjective intelligibility scores. The STI computed using the original method demonstrated the most spread within a class. Despite the invalid SMTFs as described above, the STI computed using Drullman's method showed the least spread with the intelligibility data, while the STI computed using Payton's method was in the middle (Payton et al. 2002).

Cochlear-implant processed speech. A new component of our work uses the STI to predict the intelligibility of cochlear-implant (CI) processed speech. We have collected intelligibility data under a variety of conditions, specifically, acoustic degradations (reverberation and additive noise), N of M processing (used in some CI speech processors), spectral subtraction, and binaural noise reduction. Subjects included cochlear-implant users and normal-hearing subjects listening to a noise-vocoder simulation of speech processing.

The data from normal-hearing subjects with the CI simulation indicate that within a particular class of conditions (acoustic degradations, N-of-M processing, spectral subtraction), several of the candidate SSTI methods are good predictors of the intelligibility data, correctly capturing trends due to SNR, reverberation, different types of noise, number of active channels in N-of-M processing, and the control parameter in spectral subtraction

(Goldsworthy and Greenberg, 2004). We have recently completed collection of the intelligibility data from CI users with both Nucleus and Clarion devices. Analysis of those data is currently underway.

Integration Across Frequency Bands. In addition to work on models of speech reception based on the STI, we also pursued the development of more basic models of the identification of speech elements. In initial work (Ronan et al., 2004), we addressed the problem of predicting identification performance for consonants when two or more non-overlapping bands of speech are combined. We compared four models for predicting performance, Fletcher's independent errors model, Massaro's Fuzzy Logical Model of Perception (FLMP) , and Braida's Pre-Labelling and Post-Labelling Models of Integration (1991). New data on the identification of consonants in nonsense syllables that have been filtered into one or more frequency bands were obtained to test the models. The models were compared in terms of their ability to predict combined-band scores for consonants in initial and final position on the basis of single-band scores and were evaluated in terms of bias and chi-squared computed across listeners. For each of the 38 combined band conditions, at least two models were capable of predicting performance satisfactorily. The ability of the models to make satisfactory predictions of performance in a combined-band condition was more a matter of whether the condition included the lowest frequency band than on the frequency separation of the bands.

Signal Processing for Hearing Aids

Evaluation of Hearing Aids With Variations on Dual-Front-End Compression. This work considers the use of automatic gain control algorithms designed to reduce background noise, specifically the dual front-end AGC (Moore and Glasberg, 1988). The dual front-end AGC applies two AGC systems simultaneously; a slow-acting wideband automatic volume control, that 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 for the dual front-end AGC (Stone et al., 1999). We implemented several algorithms based on this newer version of the dual front-end AGC, and evaluated the algorithm with hearing-impaired subjects. The dual front-end AGC system that we tested includes two optional features: a hold-timer which reduces pumping without extremely long recovery times and a signal-to-noise ratio (SNR) detector (Martin, 1992) which modifies 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 another talker).

We evaluated five algorithms, four dual front-end AGC algorithms (all combinations of with/without the hold timer and with /without SNR detection) plus a linear reference condition with compression limiting under twenty acoustic test conditions representative of everyday listening situations. These conditions include speech in quiet at various presentation levels, speech plus multitalker babble at various signal-to-noise ratios, speech plus continuous environmental noises (for example, vacuum cleaner) and speech plus transient environmental noises (for example, door slamming). Five hearing-impaired subjects listened to the processed segments and rated each segment for subjective intelligibility and quality on an 11-point scale.

The results showed no clear differences among the four dual front-end AGC algorithms. Major differences did exist between the linear reference condition and the dual front-end systems; however, the direction of these differences varied with subject and condition. These mixed results (namely that two subjects preferred the linear reference, one subject preferred the AGC systems, and two subjects exhibited preferences that varied with condition) illustrate the difficulty in assessing AGC algorithms under laboratory conditions and in determining which types of hearing loss might be best helped by compression hearing aids.

Using Intermicrophone Correlation To Estimate the Range of SNR. Many hearing aid noise reduction algorithms estimate parameters of the noise during the brief pauses that occur naturally in speech. To perform such estimations, the algorithm must first identify the intervals when the SNR ratio is low. We previously proposed a two-microphone approach for the case where desired signal and noise arrive from different spatial locations which assumes that the desired signal and the noise arrive from distinct ranges of angles about array broadside. The microphone signals are bandpass filtered, then the intermicrophone correlation is computed and compared to a threshold to determine the range of the short-time SNR. System parameters include the center frequency and bandwidth of the filters and the threshold.

Previous work demonstrated the effectiveness of this approach, although the parameters were selected on an *ad hoc* basis. A theoretical analysis of this system in anechoic and reverberant environments (Koul, 2003; Koul and Greenberg, 2004) quantified the effect of the parameters on system performance and indicated how to select the parameters in order to optimize performance in a range of acoustic environments. Computer simulations with broadband noise sources confirmed the theoretical results, and additional computer simulations with speech demonstrated the practicality of this approach.

Feedback Reduction Algorithms. Work performed under this grant considered the design of feedback reduction algorithms for hearing aids and identified several promising algorithms warranting further evaluation. Those algorithms were implemented in a laboratory-based digital hearing aid system and evaluated with dynamic feedback paths and hearing-impaired subjects. The analysis and interpretation of the experimental data were performed under the current grant.

The feedback reduction study (Greenberg et al., 2000) included three experimental algorithms and one reference condition. One algorithm (CNN: Continuous adaptation with No Noise) estimates the acoustic feedback path by continuously adapting on the ambient signals, while the other algorithms (ONO: Open circuit, Noise when Oscillating) and (ONQ: Open circuit, Noise when Quiet) inject a probe noise sequence during particular intervals and estimate the feedback path during those intervals. The evaluation included measurements of maximum stable gain and subjective quality ratings. Seven experienced hearing-aid users with bilaterally-symmetric sensorineural hearing loss participated in this study as subjects.

Overall, CNN provided the largest increases in maximum stable gain relative to the reference algorithm. For subjects with moderately-severe losses below 1 kHz, CNN provided almost 13 dB of added gain, while ONO and ONQ provided 7-8 dB of added gain. Subjects with milder losses below 1 kHz received an average gain increase of 3 dB from CNN, and no significant benefit from ONO and ONQ.

Analysis of variance indicated that relative to the reference algorithm, CNN and ONO had no significant effect on subjective intelligibility and pleasantness ratings. Subjects found all algorithms equally pleasant and equally intelligible both at their preferred gain setting and at various levels relative to the maximum stable gain for that algorithm. In the case of the intelligibility ratings, the lack of an algorithm effect is presumably because the subjects did not need additional gain for the 60 dB SPL speech (in quiet) presented. However, it is likely that the additional gain would improve intelligibility in other listening situations. Based on these results, the CNN algorithm is the logical choice for feedback reduction in hearing aids: it provides the highest maximum stable gain values, and informal comments about the objectionable nature of probe noise support its choice over the ONO and ONQ algorithms.

Two-Microphone Adaptive-Array Hearing Aid With Monaural And Binaural Outputs. Traditional two-microphone array processors applied to hearing aids require binaural inputs and produce a monaural output with improved signal-to-noise ratio (SNR), but no binaural cues. Previous work with normal-hearing listeners has demonstrated the potential to obtain a tradeoff between localization and intelligibility with a lowpass/highpass (LP/HP) system that splits the signal into two frequency bands, preserving binaural cues at low frequencies while improving SNR at high frequencies. The LP/HP system and a conventional broadband array processor were evaluated with eight hearing-impaired subjects. Results show that the broadband processor improved speech reception thresholds (SRTs) by 10 dB relative to a binaural reference condition, while the LP/HP method improved an SRTs by 2 dB (Greenberg and Zurek, 2001). This was unexpected in light of both physical measurements and the performance of normal-hearing listeners. Note that in the previous study, the normal-hearing listeners did not use hearing-loss simulators. These results suggest that hearing-impaired subjects are more reliant on low frequency information than are normal-hearing listeners, and that effective array processing hearing aids must operate over the entire frequency range. Consequently, the LP/HP approach is not advisable for hearing aids, and alternative methods for improving intelligibility of speech in noise while preserving a sense of auditory space should be pursued.

Evaluation Of Microphone-Array Headband Hearing Aids. Several array-processing algorithms had been previously 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: SRTs, 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 conducted under this grant (Greenberg et al., 2003) revealed that whereas the fixed algorithm improved SRTs by 2 dB over the reference condition, the adaptive algorithms provided 7-9 dB improvement. Subjects judging ease-of-listening generally preferred the array-processing algorithms to the reference condition. The results suggest that these algorithms should be evaluated further in more realistic acoustic environments.

References

Arlinger, S. and Gustafsson, H.A. (1991). "Masking of speech by amplitude-modulated noise," J. Sound Vib., 15(3), 441-445.

Asbeck, A. (2003). “Analysis of Frequency-Smearing Models Simulating Hearing Loss,” M.Eng. thesis, Massachusetts Institute of Technology.

Braida, L. D. (1991). "Crossmodal integration in the identification of consonant segments," Q. J. Expt. Psych., 43A(3), 647-677.

Drullman, R., Festen, J., and Plomp, R. (1994). "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 AGC simulation of recruiting hearing loss," J. Acoust. Soc. Am., 98, 3170-3181.

Gagne JP, Rochette AJ, Charest M (2002). "Auditory, visual and audiovisual clear speech," Speech Communication, 37(3-4), 213-230.

Goldsworthy, R. and Greenberg, J. E. (2004). "Analysis of speech-based Speech Transmission Index methods with implications for non-linear signal processing," submitted to J. Acoust. Soc. Am.

Greenberg, J.E. and Zurek, P.M. (2001). "Microphone-array hearing aids," in Microphone Arrays: Techniques and Applications, edited by M. S. Brandstein and D.B. Ward. (Springer, New York).

Greenberg, J.E., Desloge, J.D., and Zurek, P.M. (2003). "Evaluation of array-processing algorithms for a headband hearing aid," J. Acoust. Soc. Am., 113, 1646-1657.

Greenberg, J.E., Zurek, P.M., and Brantley, M. (2000). "Evaluation of feedback-reduction algorithms for hearing aids," J. Acoust. Soc. Am., 108, 2366-2376.

Hargus, S.E.; Gordon-Salant, S. (1995). "Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment," J. Speech and Hearing Res, 38(1), 234-243.

Hou, Z., and Pavlovic, C. V. (1994). "Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility," J. Acoust. Soc. Am., 96,1325-1340.

Houtgast, T. and Steeneken, H.J.M. (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.

Koul, A. "Use of Intermicrophone Correlation in Estimating Signal to Noise Ratio," M.Eng. Thesis, Massachusetts Institute of Technology, Cambridge, MA, February 2003.

Koul, A. and Greenberg, J.E. (2004). "Using intermicrophone correlation to detect speech in spatially-separated noise," submitted to EURASIP Journal on Applied Signal Processing, special issue on DSP in Hearing Aids and Cochlear Implants.

Krause J.C., and Braida L.D. (2003). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am., 112(5), 2165-2172.

Krause, J.C., and Braida, L.D. (2004). "Acoustic Properties of Naturally Produced Clear Speech at Normal Speaking Rates," J. Acoust. Soc. Am. 115, (1), pp. 362-378

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

Krause, J.C., and Braida, L.D. (2003). "Effects of listening environment on intelligibility of clear speech at normal speaking rates," Ir. Audiology 2, (1), pp. 39-47**.**

Krause, J.C., and Braida, L.D. (2004a). "Acoustic Properties of Naturally Produced Clear Speech at Normal Speaking Rates," J. Acoust. Soc. Am. 115, (1), pp. 362-378

Ludvigsen, C. (1987). "Prediction of speech intelligibility for normal-hearing and cochlearly hearing-impaired listeners." J. Acoust. Soc. Am., 82(4), 1162-71.

Lum, D.S., and Braida, L.D. (1997). "DSP implementation of a real-time hearing loss simulator based on dynamic expansion," in W. Jesteadt ed. Modelling Sensorineural Hearing Loss, L. Earlbaum Assoc., Mahwah N.J. pp 113-130.

Lum, D.S., and Braida, L.D. (2000). "Perception of speech and non-speech sounds by listeners with real and simulated sensorineural hearing loss", J. Phonetics, 28, 343-366.

Martin, G.W. (1992). "Real-time multiband adaptive hearing aids," S.M. Thesis MIT.

Moore, B.C.J. and Glasberg, B.R. (1988). "A comparison of four methods of implementing automatic gain control in hearing aids," Brit. J. Audiol., 22, 93-104.

Moore, B.C.J. and Glasberg, B.R. (1993). "Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech." J. Acoust. Soc. Am., 94(4), 2050-2062.

Moore, B. C. J., and Glasberg, B. R. (1997). "A model of loudness perception applied to cochlear hearing loss," Aud. Neurosci. 3, 289-311.

Moore, B.C.J., Vickers, D.A., Glasberg, B.R. and Baer, T. (1997). "Comparison of real and simulated hearing impairment in subjects with unilateral and bilateral cochlear hearing loss." Brit. J. Audiol., 31, 227-245.

Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (1999b). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am., 106, 2761-2778.

Nejime, Y. and Moore, B.C.J. (1997). "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise." J. Acoust. Soc. Am., 102(1), 603-615.

Oxenham, A. J., and Bacon, S. P. (2003). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing," Ear Hear., in press.

Payton, K. L. and L. D. Braida (1999). "A method to determine the speech transmission index from speech waveforms," J. Acoust. Soc. Am., 106(6), 3637-48.

Payton, K.L., Braida, L.D., Chen, S., Rosengard, P., and Goldsworthy, R. (2002). "Computing the STI using speech as a probe stimulus," in T. Houtgast and H. Steeneken (eds) Past, present, and future of the Speech Transmission Index, TNO Human Factors, Soesterberg, The Netherlands, 125-138.

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., 95,1581-1592.

Peters, R. W., Moore, B.C.J., and Baer, T. (1998). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people,"J. Acoust. Soc. Am., 103, 577-587.

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

Ronan, D, Dix, A, Shah, P. and Braida, L.D. (2004). "Integration of acoustic cues for consonant Identification across frequency bands", J. Acoust. Soc. Am. to be published in October 2004.

Rosengard, P.S., Payton, K.L., and Braida, L.D. (2004a). "Effects of slow-acting wide dynamic range compression on measures of intelligibility and ratings of speech quality in simulated-loss listeners," accepted pending revision for J. Speech Hear. Res.

Rosengard, P.S., Oxenham, A.J., and Braida, L.D., (2004b). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," submitted to J. Acoust. Soc. Am.

Schum, D. (1996). "Clear and conversational speech of young and elderly talkers," J. Am. Acad. of Audiology, 7, 212-218.

Schutte, K. T. "An investigation of clearly spoken speech and possibilities of intelligibility enhancement by redistribution of energy," S. M. Thesis, Dept. of Electrical Engineering and Computer Science, M.I.T., May 2003.

Strouse, A., Ashmead, D.H.,Ohde, R.N., and Grantham, D.W. (1998). "Temporal processing in the aging auditory system," J.Acoust. Soc. Am., 104 (4), 2385-2399).

ter Keurs, M., Festen, J., and Plomp, R. (1993). "Limited resolution of spectral contrast and hearing Loss for speech," J. Acoust. Soc. Am., 94, 1307-1314.

Uchanski, R.M., Choi, S., Braida, L.D., Reed, C.M., and 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., 3, 494-509.

Yates, G. K., Winter, I. M., and Robertson, D. (1990). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," Hearing Research, 45, 203-220.

Zurek, P.M. and Delhorne, L.A. (1987). "Consonant reception in noise by listeners with mild and moderate hearing impairment," J. Acoust. Soc. Am., 8, 1548-1559.

Zurek, P.M. and Greenberg, J.E. (2000) "Two-microphone adaptive array hearing aids with monaural and binaural outputs," in Proceedings of the Ninth IEEE DSP Workshop, Hunt, Texas, October, 2000.

PUBLICATIONS

Journal Articles

Krause, J.C., and Braida, L.D. (2003). "Effects of listening environment on intelligibility of clear speech at normal speaking rates," Ir. Audiology 2, (1), pp. 39-47**.**

Accepted for Publication

Krause, J.C., and Braida, L.D. (2004a). "Acoustic Properties of Naturally Produced Clear Speech at Normal Speaking Rates," J. Acoust. Soc. Am. 115, (1), pp. 362-378

Submitted for Publication

Ronan, D, Dix, A, Shah, P. and Braida, L.D. (2004). "Integration of acoustic cues for consonant Identification across frequency bands", J. Acoust. Soc. Am. to be published in October 2004.

Rosengard, P.S., Payton, K.L., and Braida, L.D. (2004a). "Effects of slow-acting wide dynamic range compression on measures of intelligibility and ratings of speech quality in simulated-loss listeners," accepted pending revision for J. Speech Hear. Res.

Rosengard, P.S., Oxenham, A.J., and Braida, L.D., (2004b). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," under revision for J. Acoust. Soc. Am.

Goldsworthy, R. and Greenberg, J. E. (2004). "Analysis of speech-based Speech Transmission Index methods with implications for non-linear signal processing," accepted pending revision by J. Acoust. Soc. Am.

Braida, L.D. and Krause, J.C. (2004). "Potential Clinical Applications of Automatically Produced Cued Speech", les Cahiers de l'Audition, in press.

Koul, A. and Greenberg, J.E. (2004). "Using intermicrophone correlation to detect speech in spatially-separated noise," submitted to EURASIP Journal on Applied Signal Processing, special issue on DSP in Hearing Aids and Cochlear Implants.

Sroka, J. J. and Braida, L. D. (2004). "Consonant Recognition by Humans and Machines," under review by Speech Communication.

Chapters in Books

Payton, K. L., Braida, L. D., Chen, S., Rosengard, P. and Goldsworthy, R., (2002). "Computing the STI using speech as a probe stimulus". in Past, Present and Future of the Speech Transmission Index, Soesterberg, Netherlands, TNO Human Factors, 125-138.

Reed, C.M., and Braida, L.D. (2003). "Frequency compression," in The M.I.T. Encyclopedia of Communications Disorders, edited by R.D. Kent, Cambridge:MIT Press.,471-475.
Published

Meeting Papers Presented

Theses

Asbeck, A. (2003). “Analysis of Frequency-Smearing Models Simulating Hearing Loss,” M.Eng. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Cheng, R. (2003). “Software for Psychoacoustic Experiment Design,” M.Eng. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Koul, A. (2003). "Use of Intermicrophone Correlation in Estimating Signal to Noise Ratio," M.Eng. Thesis, Massachusetts Institute of Technology, Cambridge, MA.