Hearing Aid Research

Sponsor:
National Institutes of Health Grants R01 DC00117, R01 DC007152,

Academic and Research Staff:
Professor Louis D. Braida, Dr. Joseph Desloge, Dr. Raymond Goldsworthy, Dr. Julie Greenberg, Dr. Karen L. Payton, Dr. Charlotte M. Reed

Visiting Scientists and Research Affiliates
Dr. Paul Duchnowski, Dr. Oded Ghitza, Dr. Kenneth W. Grant, Professor Ying-Yee Kong, Professor Jean C. Krause, Dr. Peninah S. Rosengard

Graduate Students
Chuping Liu, David Messing, Michael Naber, Mona Shretha,

Undergraduate Students
Jing Wang,

Technical and Support Staff
Loraine Delhorne, Denise Rosetti

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

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

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

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

Studies and Results

A. Role of Reduced Audibility

This research is concerned with analyzing the factors responsible for poor speech reception by listeners with hearing impairments, and with developing techniques for overcoming these degradations. To the extent the research is successful, it will help determine design goals for improved wearable hearing aids, establish new criteria and techniques for aural rehabilitation, and contribute to improved understand of both residual auditory function and speech perception.

We have made substantial progress over the past year toward our research goal of developing a greater understanding the role of reduced audibility in the speech-reception performance of listeners with moderate-to-profound degrees of hearing loss.

A-1. Simulations of Sensorineural Hearing Loss

We have implemented signal-processing algorithms to produce functional simulation of hearing loss in listeners with normal hearing. These algorithms, which have been developed in MATLAB, are based on the combination of two different approaches towards hearing-loss simulation: (a) the addition of a spectrally shaped masking noise to elevate thresholds of a normal-hearing listeners to match those of a given hearing-impaired listener; and (b) the use of multi-band expansion in which level-dependent attenuations are applied to sounds in different frequency bands to map tone levels at the hearing-impaired listener's threshold to the threshold of a normal-hearing listener and to mimic the rapid growth of loudness observed in hearing loss. In our current implementation, the use of additive-noise masking is limited to the first 40 dB of hearing loss at a given frequency and the remaining loss is simulated using multi-band expansion. With this combined approach, we are able to simulate hearing loss in the range of mild to profound loss using noise levels and signal presentation levels that do not exceed comfortable listening levels in normal-hearing listeners. Using a 3-alternative, adaptive, forced-choice procedure to measure pure-tone thresholds (implemented in a MATLAB-based platform referred to as AFC), we have been successful in matching the threshold elevations of listeners with severe-to-profound high-frequency and flat hearing losses in listeners with normal hearing.

A-2. Speech Testing in Background Noise

We have developed software for obtaining adaptive measurements of Speech Reception Threshold (SRT) for HINT sentences (Nilsson et al., 1994) as a function of the type of background noise (wideband noise which is either steady-state or square-wave-interrupted at a rate of 10 Hz) and noise level (60 or 80 dB SPL). This procedure measures the Speech-to-Babble (S/B) ratio required for 50%-correct reception of sentences using equivalent lists of phonemically balanced speech materials. Baseline data were obtained on a group of four listeners (three with normal hearing and one with a high-frequency loss above 2000 Hz) to examine the effects of noise type and noise level, repeatability of the S/B measurements, and finally the effects of repeated presentation of the same sentence lists across five test replications. The results indicated (a) averaged over subjects, a 10-dB release of masking for interrupted compared to steady-state noise, (b) less release of masking for the hearing-impaired compared to the normal-hearing subjects, (c) an average standard deviation of roughly 2.0 dB across the five lists presented within a given day on a given condition to a given subject, and (d) minimal effects of learning across the five replications of the experiment.

A-3. Signal Processing for Hearing-Aid Simulation

Signal-processing algorithms have been developed for simulation of two types of hearing-aid processing that will be included in the SRT testing. These processing modes include (a) linear amplification using the revised NAL guidelines for frequency-gain characteristic (Byrne and Dillon, 1986) and (b) multi-band amplitude compression amplitude based on a wide-dynamic range, instantaneous compressive system (Goldstein et al., 2003).

A-4. Review of Past Research on the Role of Audibility in Predicting Effects of Hearing Impairment

Although supra-threshold effects of hearing impairment are widely believed to be related to the decreased resolution on psychoacoustic tasks and the poorer speech-reception abilities of hearing-impaired listeners, the role of reduced audibility itself in explaining the consequences of hearing loss is as yet not completely understood. We have begun work on a comprehensive review of the literature on the effects of hearing loss on performance in psychoacoustic and speech-reception tasks to examine the evidence for supra-threshold deficits. Our review is organized into five major categories of studies that include (a) temporal resolution, (b) intensity resolution, (c) spectral resolution, (d) speech reception, and (e) correlational studies of speech and psychoacoustic abilities. In evaluating studies in which the performance of normal and hearing-impaired is compared, we will pay particular attention to three variables (audibility, presentation level, and subject age) whose control must be demonstrated before the reduced performance of hearing-impaired listeners can be attributed to supra-threshold factors. Our progress thus far includes the compilation of a working bibliography of over 300 papers in the areas described above and a draft of the first article in this series reviewing research in the area of temporal resolution.

B. Models of Speech Intelligibility

B.1 Further Development of the STI and NCM Models.

Substantial effort was focused on the development of the STI and NCM models for predicting speech reception by cochlear implant users. In particular, a component of the model was developed for comparing speech reception scores across multiple cochlear implant users as well as normal hearing subjects listening to a noise vocoder simulation of cochlear implant sound processing. The developed component is based on an efficiency factor that scales the metric values (Fletcher and Galt, 1950; Fletcher, 1953; Dugal et al., 1980). This efficiency factor was successfully used in the model framework to allow for across subject comparisons and will be investigated in the future for a wider range of subjects.

The speech transmission index (STI) was developed to predict the intelligibility of speech degraded by additive noise and reverberation. Its effectiveness for quantifying the reception of acoustically degraded speech is well established for listeners with normal and impaired hearing. Derivative metrics have been proposed to account for the effects of non-linear speech processing operations. In this work, the ability of STI-derived metrics to predict speech reception for cochlear implant users is investigated for acoustic degradations and for spectral subtraction processing. Speech reception scores were measured for implant users and for subjects with normal hearing listening to a simulation of sound processing for cochlear implants. Acoustic degradation conditions covered a variety of reverberation levels, noise levels, and noise types with differing degrees of modulation. Spectral subtraction conditions explored the effect of a control parameter within the algorithm. The results show the relative effects of acoustic degradations and spectral subtraction processing upon speech reception for implant users. The STI and the novel normalized correlation metric (NCM) were evaluated for their ability to predict measured speech reception. Both metrics perform fairly well, accurately predicting speech reception trends due to noise level, reverberation, and the benefit provided by spectral subtraction.

We submitted an article to the Journal of the Acoustical Society of America titled “Predicting the effect of acoustic degradation and spectral subtraction processing on speech reception by cochlear implant users.”

B2. Role of Discrimination on Speech Reception Performance

Preliminary psychoacoustic and speech reception data has been collected for one cochlear implant subject. The data focused on pure tone frequency discrimination, amplitude discrimination, and modulation frequency discrimination. The data will facilitate the decision process towards selecting the important psychoacoustic and speech reception testing battery that will be used in the first MSI experiments involving cochlear implant subjects. Substantial effort was extended on the software control and analysis functions for the collection of this data.

Psychoacoustic procedures The testing paradigm used in the preliminary data collection was a two alternative forced choice paradigm with a two down, one up decision rule. Such a paradigm converges to a 70.7% correct response criteria. Selected preliminary data for pure tone frequency discrimination tested at 440, 880, 1760 and 3520 Hz is given in Figure 1.

The stimuli consisted of two tones, a reference and a target. The reference frequency was randomized within a 4 semitone interval. So for the 440 Hz tone, the reference could be between 392 and 494 Hz. The target was always the higher tone. Both the target and reference were 500 ms long with a 200 ms gap between. Note that normal hearing performance on this task for non-musicians is approximately a quarter of a semitone. The pilot cochlear implant subject of Figure 1 showed relatively poor performance at the higher frequencies.

Selected preliminary data for modulation frequency discrimination tested at 110, 139, 175, and 220 Hz is given in Figure 2. The modulation envelope at the tested frequency was used to modulated an 880 Hz sinusoidal carrier.

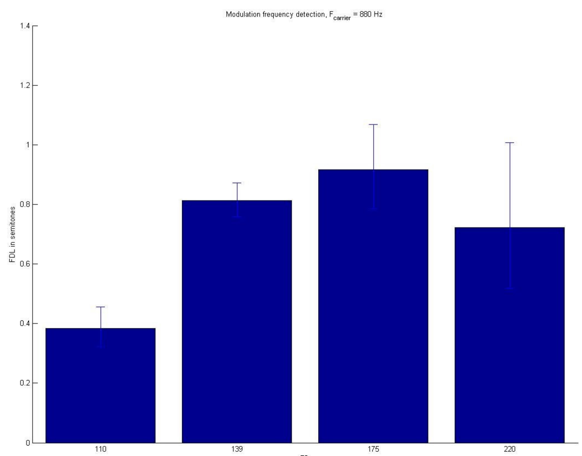


Figure 1: Pure tone frequency discrimination

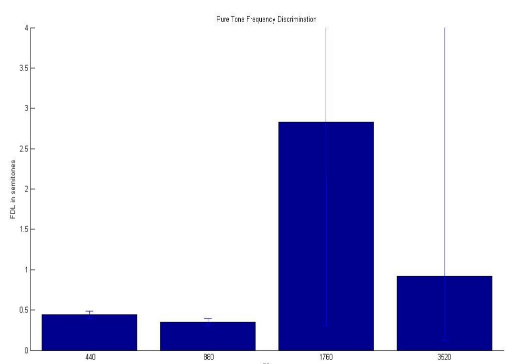


Figure 2: Modulation frequency discrimination

A two interval AFC paradigm, 2 down 1 up decision rule was again used. Reference is sinusoidally modulated at 100% depth using a modulation frequency within a 4 semitone range. That is, at 110 Hz the modulator could be between 98 and 124 Hz. The target stimuli had a higher modulator, and adaptively converged using standard techniques. No randomization in presentation level was used.

These results clearly indicate a high degree of F0 detection. The subject reports perceptually performing these task using pitch as a percept and not some secondary cue (such as level). That is, decisions are based on a percept of higher and lower pitch as the subject is familiar with the ranking of such a percept.

Speech reception procedures Consonant confusion: A set of consonants recorded at House Ear Institute are used. Figure 3 illustrates the consonant confusion pattern generated using 200 tokens, which amounts to 8 tokens for each of 25 consonants. Both male and female voices used.

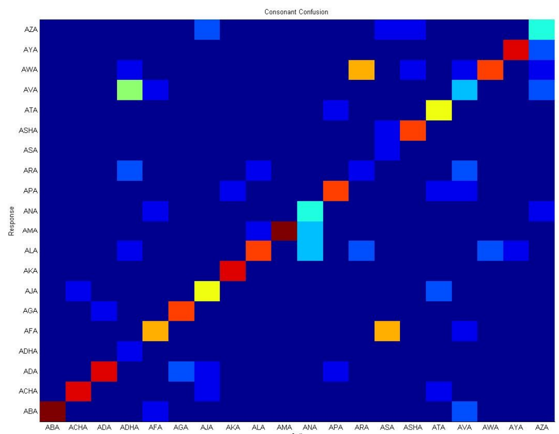


Figure 3: Consonant confusion.

Average percent correct for above was 65%. Certain confusions, such as 'ARA' versus 'ALA' can be easily interpreted as associated with an inability of the subject to discern differences in the second formant region of the speech signal, and perhaps directly related to the subjects relatively poor ability to detect frequency changes near 1760 Hz (Figure 1). But further testing and anlysis is required to confirm this hypothesis.

Vowel confusion: A set of vowels known as the Hillenbrand corpus recorded at House Ear Institute were used. Figure 4 illustrates the vowel confusion pattern generated using 216 tokens, which amounts to 24 tokens for the 9 tested vowels. Average percent correct was 75%.

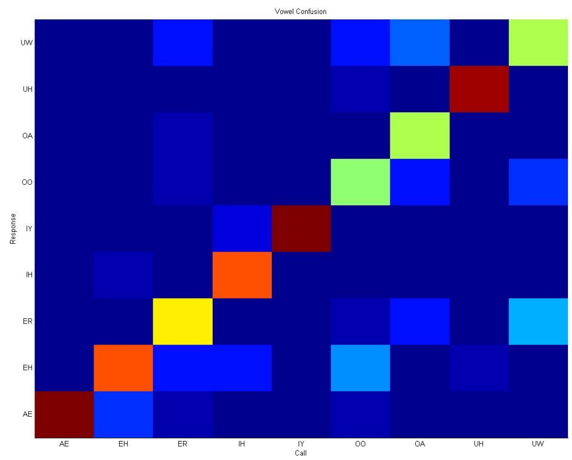


Figure 4: Vowel confusion.

Table 1. Phonetic mapping:

% AE='had' EH='head' ER='herd'
% IH='hid' IY='heed' OO='hood'
% OA='hoed' UH='hud' UW='whod'

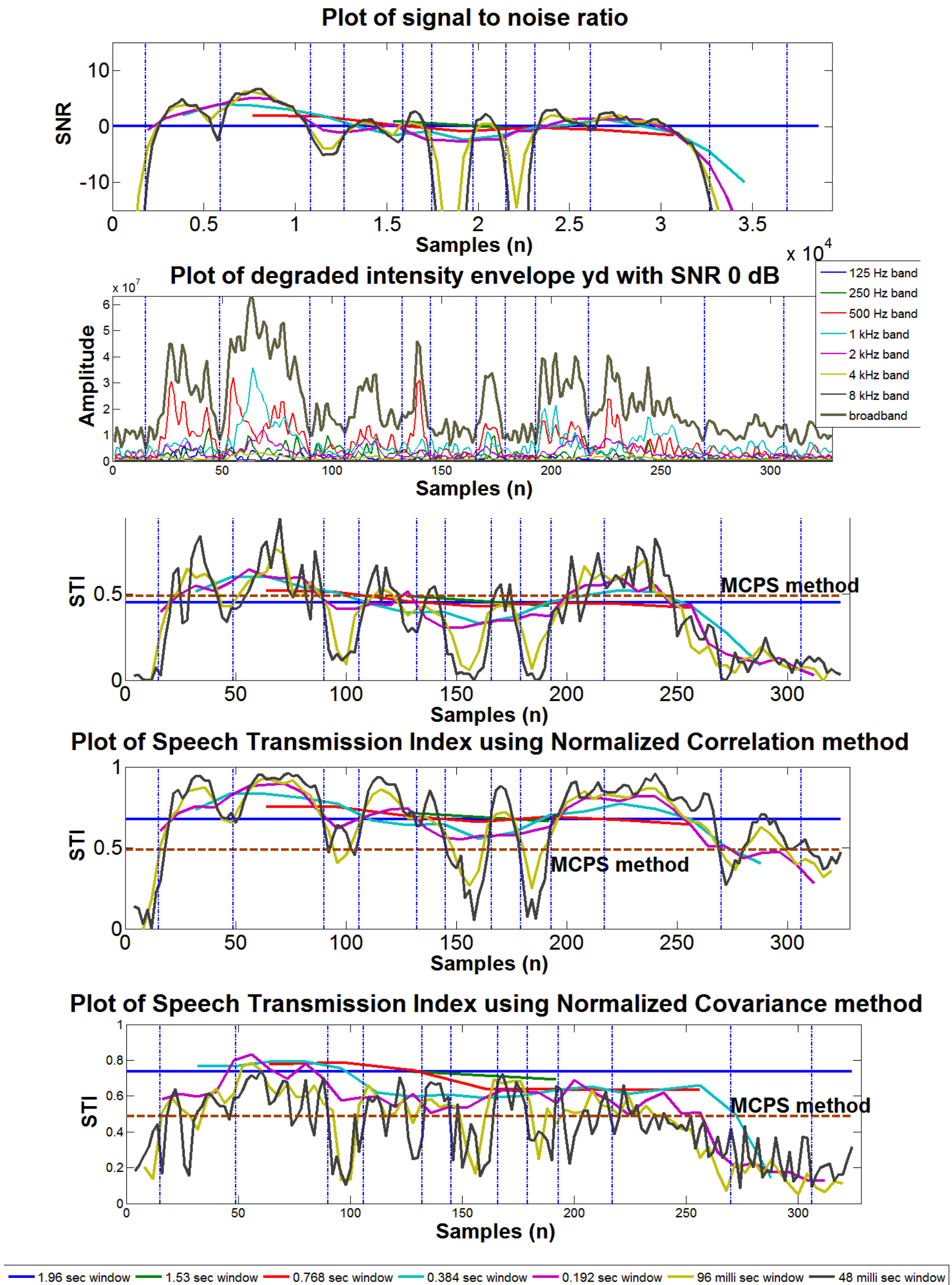
B.3. General comments

These psychoacoustic and speech reception tests lay a groundwork for the fundamental problem of the MSI grant which is to correlate observed psychoacoustic and speech reception measures within the STI and NCM models that are in development. Determination of the psychoacoustic measures that are most correlated to speech reception will naturally require substantial testing of numerous cochlear implant users in the following year.

Further, we have made active progress in acquiring necessary hardware to investigate cochlear implant psychoacoustics more directly by generating and streaming specific electrode patterns to implant recipients. The acquisition of this hardware is scheduled for June with several months estimated for developing the required code for controlling the experiments. While the code is being developed for this hardware interface, continued psychoacoustic tests and model development will be actively pursued.

B.4. Short- and Long-Time Computation of the Speech-Based STI

We have begun an investigation into the issue of short-time computation of the speech-based STI and comparison of results with long-term computation results and, for the simple case of additive noise, comparison with short-time SNR. To date, 0 dB SNR and 0 dB SNR plus reverberation acoustic conditions have been considered. The speech-based STI techniques analyzed include an envelope regression method [Ludvigsen, 1990], [Goldsworthy, 2004], a normalized correlation method [Goldsworthy, 2004] and a normalized covariance method [Holube, 1996], [Koch, 1992]. When speech plus noise is considered, all three techniques qualitatively track short-time fluctuations in SNR as seen below in Figs 1 and 2



Plot of signal to noise ratio
SNR
10
0
-10
0
0.5
1
1.5
2
2.5
3
3.5
Samples (n)
x 10^4
Plot of degraded intensity envelope yd with SNR 0 dB
x 10^7
Amplitude
Samples (n)
125 Hz band
250 Hz band
500 Hz band
1 kHz band
2 kHz band
4 kHz band
8 kHz band
broadband
STI
MCPS method
Samples (n)
Plot of Speech Transmission Index using Normalized Correlation method
STI
MCPS method
Samples (n)
Plot of Speech Transmission Index using Normalized Covariance method
STI
MCPS method
Samples (n)
1.96 sec window
1.53 sec window
0.768 sec window
0.384 sec window
0.192 sec window
96 milli sec window
48 milli sec window

Fig. 2

As noted in Goldsworthy [2004], the envelope regression method most closely fits the Magnitude Cross Power Spectral (MCPS) method of Payton et al. [2002] for the long term average STI. It also seems to track the short-term changes in RMS most closely.

When noise plus reverberation are used to degrade the speech, the results are more diverse, as shown in Fig. 3 below.

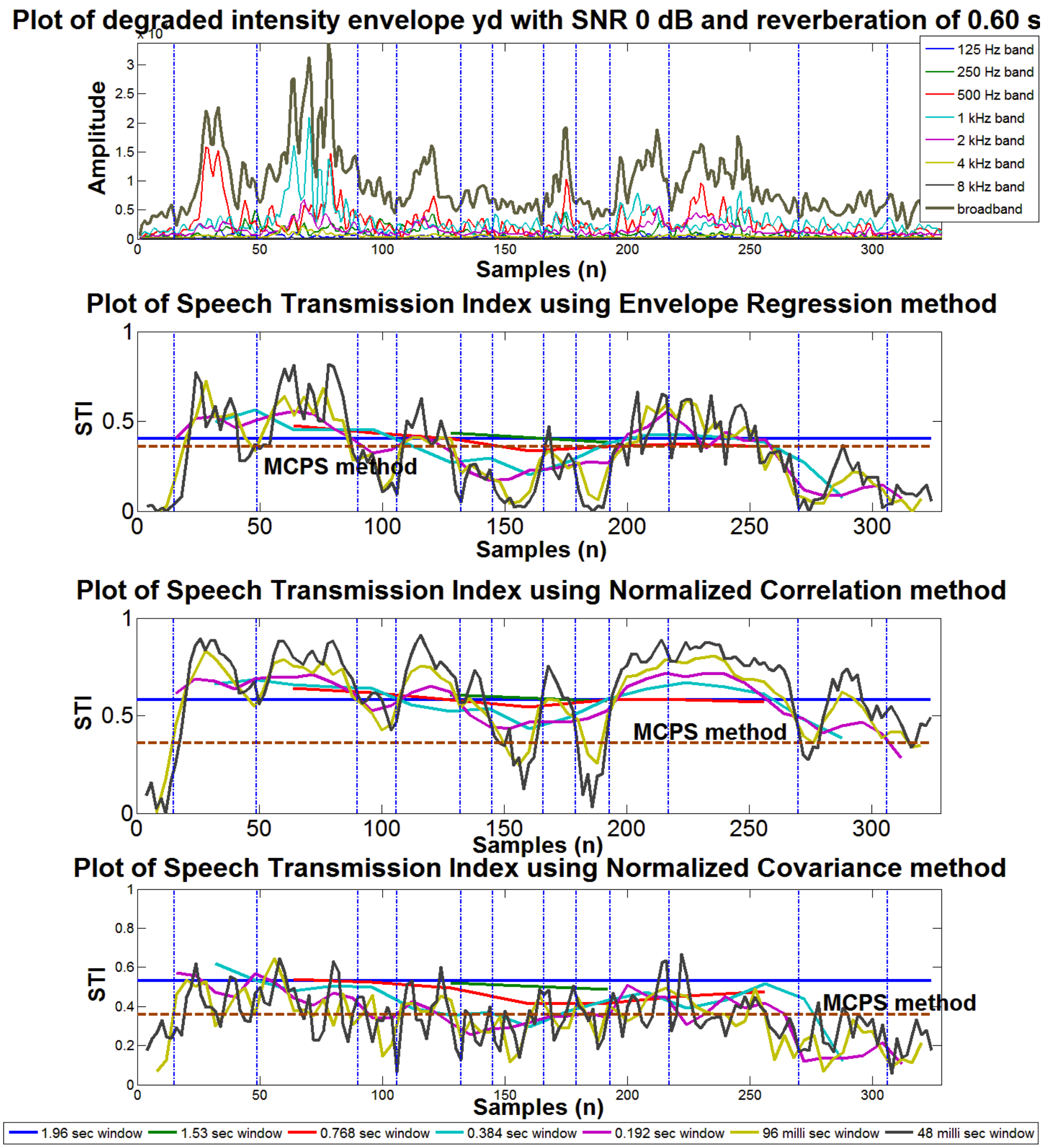


Fig. 3

In this case, both the Envelope Regression and the Normalized Correlation methods reflect the changes in the intensity envelope of the degraded signal whereas the Normalized Covariance fluctuates in an apparently random manner relative to the degraded stimulus (or the original stimulus for that matter).

B.5. Colloaborative Efforts

We have formalized our collaboration with Dr. Houben of amc.uva in the Netherlands who has collected a large amount of data from hearing-impaired listeners using amplitude compression and linear hearing aids with a wide range of parameter values (e.g. compression ratio, attack and release times, etc). He has sent us samples of his processed and unprocessed speech, along with sufficient information about the presentation levels, listener characteristics, and parameters used for the various conditions. We have begun to run his speech signals through our algorithms.

C. Application of Cortical Processing Theory to Acoustical Analysis

The goal of this research is to develop a machine which will use state-of-the-art non-linear peripheral auditory models (PAM) connected to a perceptually inspired model of template matching to (1) predict phonetic confusions made by normally-hearing listeners, and (2) predict intelligibility of distorted speech generated by passing naturally spoken speech through realistic communication systems.

Success in this project will contribute to and have significance for the following:

Revising models of auditory periphery by including the role of the descending pathway in making the cochlear response to speech sounds robust to degradation in acoustic conditions.

Establishing models of template-matching in the context of human perception of degraded speech. These models will provide guidance to physiological studies of cortical processing.
Enabling diagnostic assessment of speech intelligibility by using closed-loop models of the auditory periphery integrated with perception-based template matching.
Improving the performance of automatic speech recognition systems in acoustically adverse conditions.

C.1 Peripheral Models of Human Performance

A closed-loop model of the auditory periphery with efferent-inspired feedback was developed as a front-end in a system designed to predict human consonant confusions, under a variety of acoustic distortions. A template-matching operation was developed to serve as the back-end component, identifying an input pattern generated by the peripheral model by matching it against internal templates. Our approach was to tune the peripheral auditory model first, by means of a psychophysical task designed to reduce errors due to the back-end to a minimum (phone discrimination). To do so, we collected data from 9 human subjects. Each subject was presented a 2AFC DRT test based on real speech with noise and then a DRT test based on synthetic speech with noise. The results of these tests allowed us to optimize the parameters of the front-end system and tune the system to mimic human performance. Mimic results were evaluated using Chi-squared contingency table analysis per acoustic dimension to test for significant differences between human and machine performance. Data is currently being collected for a phone identification task with human subjects. The results of these tests will allow us to optimize the back-end template-matching operation.

References

Goldsworthy, R. and J. Greenberg (2004). “Analysis of speech-based speech transmission index methods with implications for nonlinear operations.” J. Acoust. Soc. Am. 116(6): 3679-3689.

Holube, I. and B. Kollmeier (1996). “Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model.” J. Acoust. Soc. Am. 100(3): 1703-1716.

Koch, R. (1992). Gehörgerechte Schallanalyse zur Vorhersage und Verbesserung der Sprachverständlichkeit (Auditory sound analysis for the prediction and improvement of speech intelligibility), Universität Göttingen.

Ludvigsen, C., C. Elberling, G. Keidser, T. Poulsen, (1990). “Prediction of intelligibility of non-linearly processed speech.” Acta Otolaryngol Suppl 469: 190-195.

Payton, K. L., L. D. Braida, S. Chen, P. Rosengard, R. Goldsworthy, (2002). Chapter 11. Computing the STI using speech as the probe stimulus. in Past Present and Future of the Speech Transmission Index. S. J. van Wijngaarden Ed. TNO Human Factors, The Netherlands: 125-138.

Publications

Micheyl, C. and Messing, D. Likelihood Ratio, Optimal Decision Rules, and Correct-Response Probabilities in a Signal Detection Theoretic, Equal-variance Gaussian Model of the Observer in the 4IAX paradigm”,. Perception and Psychophysics, 68 (5) 725-735, 2006.

Tramo, MJ, Cariani, PA, Koh, C.K., Makris, N., and Braida, L.D. Neurophysiology and neuroanatomy of pitch perception: Auditory cortex, Neurosciences and Music II: From Perception to Performance, Annals of the New York Academy of Sciences, 1060, 148-174, 2006.