Predicting OMNI/DIR Microphone Preferences

Ken W. Grant[1], Mounya Elhilali[2], Shihab A. Shamma[2], Brian E. Walden[1], Rauna K. Surr[1], Mary T. Cord[1], Andrew Dittberner[3]
[1]Walter Reed Army Medical Center, Army Audiology and Speech Center, Washington, DC 20307-5001, [grant@tidalwave.net]
[2]University of Maryland, Institute for Systems Research, College Park, MD 20742
[3]GN ReSound Coorporation, Chicago, IL

Abstract

Recent studies have shown that listener preferences for omnidirectional (OMNI) or directional (DIR) microphones depend largely on the characteristics of the listening environment, including the relative locations of the listener, signal sources and noise sources, and whether reverberation is present. Consequently, listeners must evaluate their environments and manually switch between microphone settings, a task that is hard for some patients to do. This study describes a strategy for automatically switching between microphone modes. Speech signals were presented at a variety of speech-to-noise ratios (SNRs) and in different spatial orientations. Monaural recordings made in both OMNI and DIR microphone processing modes were analyzed using a model of auditory processing that highlights important spectral and temporal dynamics in speech. Differences in the outputs of the auditory model for OMNI and DIR processing were expressed in terms of a spectro-temporal modulation index (STMI) of intelligibility. A modified version of the STMI (*m*STMI) was developed specifically for hearing-aid applications. Results show that the *m*STMI has the potential to predict speech intelligibility obtained through OMNI and DIR microphones as a function of SNR, and is sensitive to the spatial orientation of the listener, signal and noise.

Background

- To recognize speech at equivalent levels of performance, persons with sensorineural hearing loss require a more favorable speech-to-noise ratio (SNR) than do individuals with normal hearing.
- Hearing aids with directional microphones offer the potential to improve speech understanding in noise by improving the SNR.
- Nearly 30% of persons wearing manually switchable omnidirectional/directional hearing aids do not switch between the two microphone settings (Cord et al., 2002).

Automatic Switching Algorithms

- According to Walden et al. (2004), patients tend to prefer DIR mode when:
  - Noise is present.
  - Signal is in front.
  - Signal is near.
- One approach to automatic switching requires accurate scene analysis.
  - Is noise present?
  - Where are the noise and speech signals coming from?
  - Is there reverberation? How much?
  - Even if accurate scene analysis were possible, microphone preferences are not definitive (Walden et al., 2004).

Alternative Approach: Direct Microphone Comparisons

- Acoustic analysis of DIR and OMNI processed signals.
  - Compute index related to signal integrity or fidelity (e.g., *m*STMI).
    - Rarely have access to clean signal.
    - Must base comparisons on generic representation of clean speech.

Design and Methodology

Goals

- To develop an objective measure sensitive to differences in microphone mode under different speech-in-noise conditions and different spatial orientations of listener, speech, and noise.
- To compare objective measures to behavioral measures (Walden et al., in review).

Methods

- Acoustic (KEMAR) recordings of hearing-aid output measured in the right ear at 11 different speech-in-noise conditions in both OMNI and DIR microphone modes (see Figure 1).
  - *Hearing aid microphones were compensated for level.*
  - *Gain was adjusted to the average audiogram for patients tested by Walden et al. (in review).*
  - *Speech materials: IEEE sentences.*
  - *Noise: Three independent and uncorrelated speech-shaped noise sources.*
  - *Spatial Orientation: Four positions, 90-degree increments (see Figure 1).*
- Computational modeling of spectro-temporal modulations (see Figure 2).
- Compare STMI results with results from Walden et al. (in review).
  - *31 hearing-impaired listeners with sloping hearing loss.*
  - *11 different speech-in-noise conditions in both OMNI and DIR mode.*
  - *Speech from front, noise from sides and rear (see Figure 1, Front orientation).*
  - *Data included speech recognition scores for each microphone mode at each SNR.*

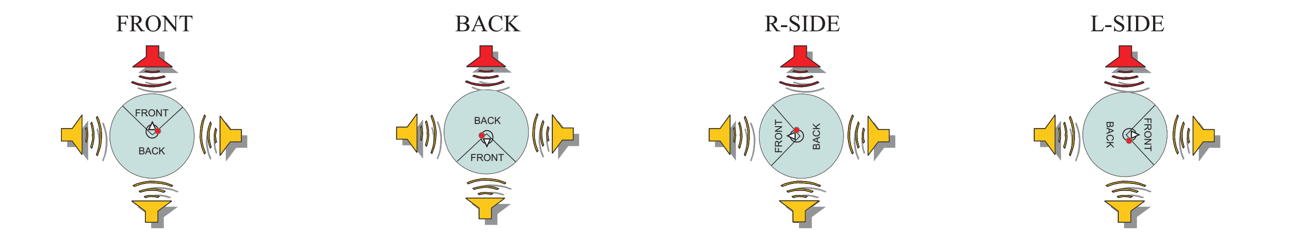

Figure 1. Spatial orientation of recording sessions. Speech is presented through the loudspeaker colored red at 65 dB SPL. Independent and uncorrelated speech-shaped noise is presented through the three loudspeakers colored yellow to produce 11 different speech-in-noise conditions. The red dot on the side of the listener's head indicates the location of the hearing aid (right ear).

MULTIRESOLUTION SPECTRO-TEMPORAL ANALYSIS OF SPEECH

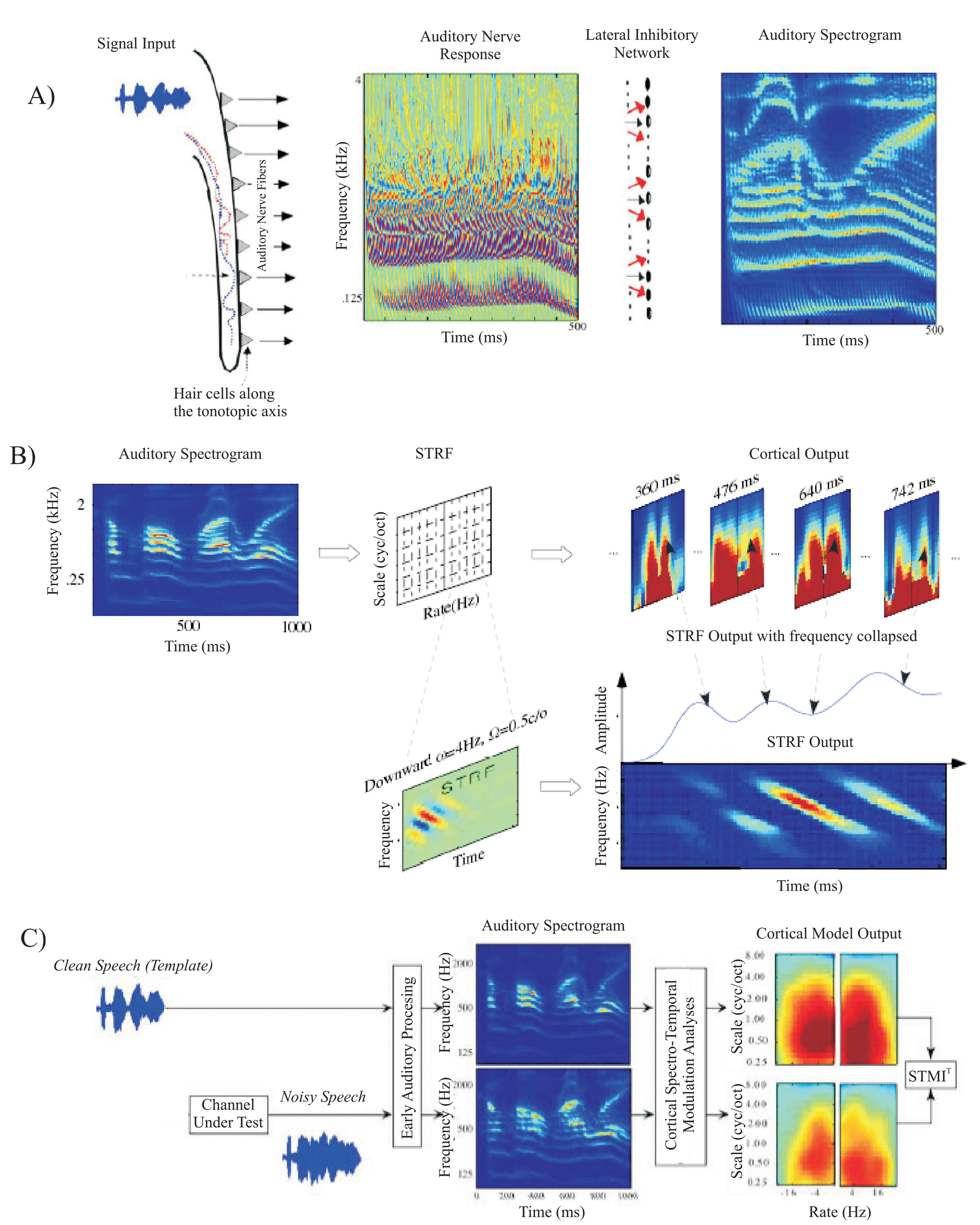

Figure 2. (A). Schematic of early auditory processing stages. The acoustic signal is analyzed by a bank of 128 constant-Q cochlear-like filters. The output of each filter is processed by a hair cell model followed by a lateral inhibitory network, and is finally rectified and integrated to produce the auditory spectrogram. (B). The auditory spectrogram is then processed through a bank of modulation-selective filters, each tuned to a range of temporal modulations (rate) and spectral modulations (scale). The response of one cortical spectro-temporal modulation filter is shown in the bottom panel along with the result of convolving it with the auditory spectrogram. (C). Schematic showing steps in computing STMI. A template is formed by processing clean speech through the model. In our case, 200 seconds of different male and female adult speakers extracted from the TIMIT data base (Garofolo, 1988) were used. The bottom channel displays the computation of the STMI after OMNI or DIR processing. The noisy signals result in a rate-scale plot (collapsed over tonotopic frequency and time for illustration purposes) that is distorted relative to the clean speech template. The STMI is a normalized metric reflecting the difference between template and test signal. (Figures adapted from Chi et al., 1999 and Elhilali et al., 2003).

Modifications to STMI (mSTMI)

- Collapse temporal and spectral modulations across tonotopic frequency.
  - *Avoids problem of treating frequency-gain shaping as distortion relative to clean, unamplified speech.*
- Weight difference between signal and template by normalized template.
  - *Serves to highlight differences in spectral-temporal regions where the template has significant energy and de-emphasizes other regions.*

Let X and T be the cortical representations of the test signal and template, respectively. We compute a new normalized mapping for the test and template representations, and weight them by the template T as follows:

$$\hat{X} = X - E[X] \bullet T \qquad \hat{T} = T - E[T] \bullet T \qquad \text{where } E[A] \text{ is the mean of matrix A}$$

$$STMI^{T} = 1 - \left\| \frac{\hat{X} - \hat{T}}{\hat{T}} \right\| \qquad \text{where } A = [a_i] \text{ and } \|A\| = \sqrt{\sum a_i^{2}}$$

Results: mSTMI Calculations

mSTMI Computations for OMNI and DIR Modes: Effects of SNR and Spatial Orientation

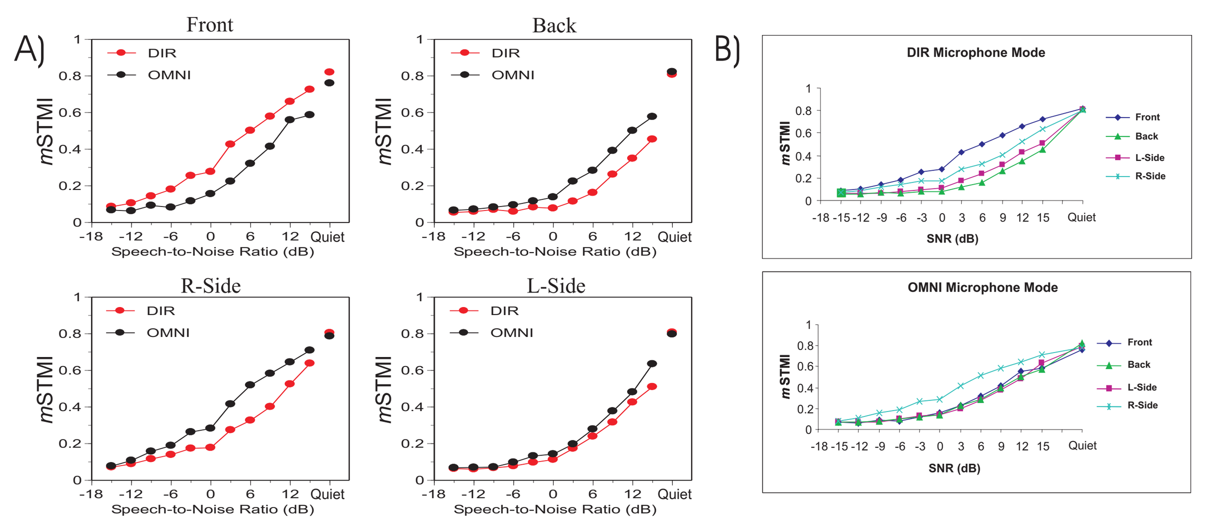

Figure 3. (A). *m*STMI as a function of speech-to-noise ratio. Each panel shows the results for IEEE sentences (1969) recorded through a KEMAR manikin wearing a modified Canta 770D hearing aid in the right ear at four different spatial orientations (see Figure 1). Each computation is based on the average of five 2-sec samples of speech. (B). Same data as (A) but with processing mode as the parameter. Top panel shows results for DIR microphone; bottom panel shows results for OMNI microphone.

For Hearing Aid on the Right Ear:

- DIR processing mode is better than OMNI only when signal is in front and noise is to the sides and back (Front).
- OMNI processing mode is better than DIR when signal is in back or when aided ear is nearest to signal (Speech on R-Side).
- Little or no difference is observed when aided ear is opposite signal and affected by head shadow (Speech on L-Side).
- *m*STMI values obtained with DIR processing show graded sensitivity for all spatial orientations.
- In OMNI mode, *m*STMI values are nearly identical for all orientations except when the speech is closest to the hearing aid (see Figure 3B, bottom panel, R-Side).
- These data show the utility of the *m*STMI as a possible tool for scene analysis. By monitoring the fidelity of the OMNI mode across ears, it is possible to determine where the predominant speech signal is in space. Similar $m\mathrm{STMI}_{\mathrm{OMNI}}$ values across ears suggest that the signal is in front, rear, above or below. Different $m\mathrm{STMI}_{\mathrm{OMNI}}$ values across ears suggest that the speech signal is coming from the side with the greater OMNI fidelity.

Results: Comparison to Behavioral Data

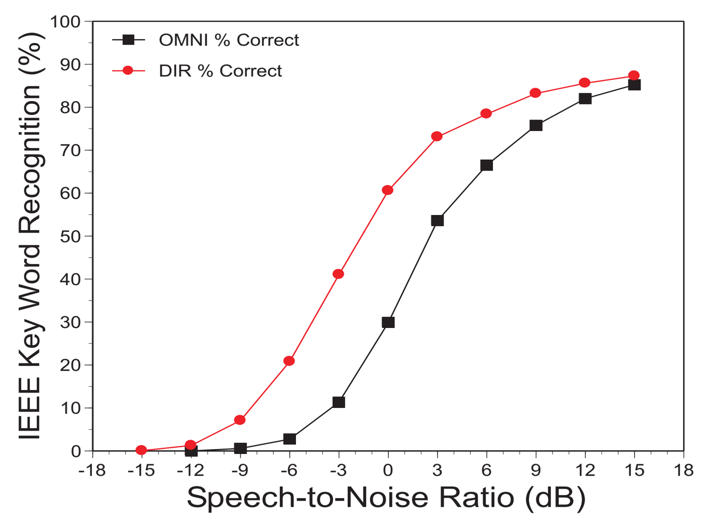

Figure 4. Percent correct word recognition as a function of speech-to-noise ratio for OMNI (black) and DIR (red) processing modes (from Walden et al., in review). Data are from 31 hearing-impaired subjects with the speech presented in front and uncorrelated speech-shaped noise from sides and back (see Figure 1 - Front).

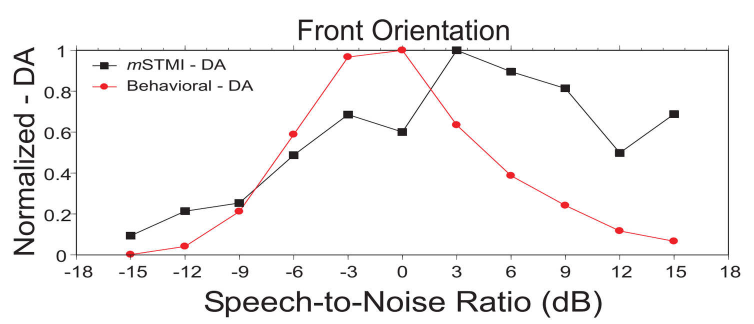

Figure 5. (Red). Average directional advantage (DA) for data shown in Figure 4. Data were obtained by subtracting OMNI percent-correct recognition scores from DIR percent-correct recognition scores and then normalized so that the maximum DA equals one. (Black). Normalized DA for the *m*STMI. Data were obtained by subtracting OMNI from DIR values shown in Figure 3A (front orientation). Note that the peak in the behavioral DA occurs at -3 dB SNR whereas the peak in the *m*STMI DA occurs at +3 dB SNR.

Why do these functions differ?

Discussion: mSTMI vs. Behavioral Measures

- *m*STMI based on normal-hearing auditory processing. Behavioral data from hearing-impaired subjects.
- *m*STMI measures fidelity, not percent correct word recognition.
- Percent correct scores are subject to ceiling and floor effects. Therefore, not all increases in SNR result in increases in intelligibility. Data indicate a plateau between 85-90% correct, even in quiet.
- Transformation between *m*STMI and percent correct is non-linear.
- To directly compare behavioral DA to *m*STMI differences, *m*STMI calculations need to be transformed to percent correct.
- This transform can be derived from behavioral scores (Walden et al., in review). Once transformed to percent correct scores, *m*STMI differences between DIR and OMNI processing modes can be expressed in similar terms as traditional DA scores.

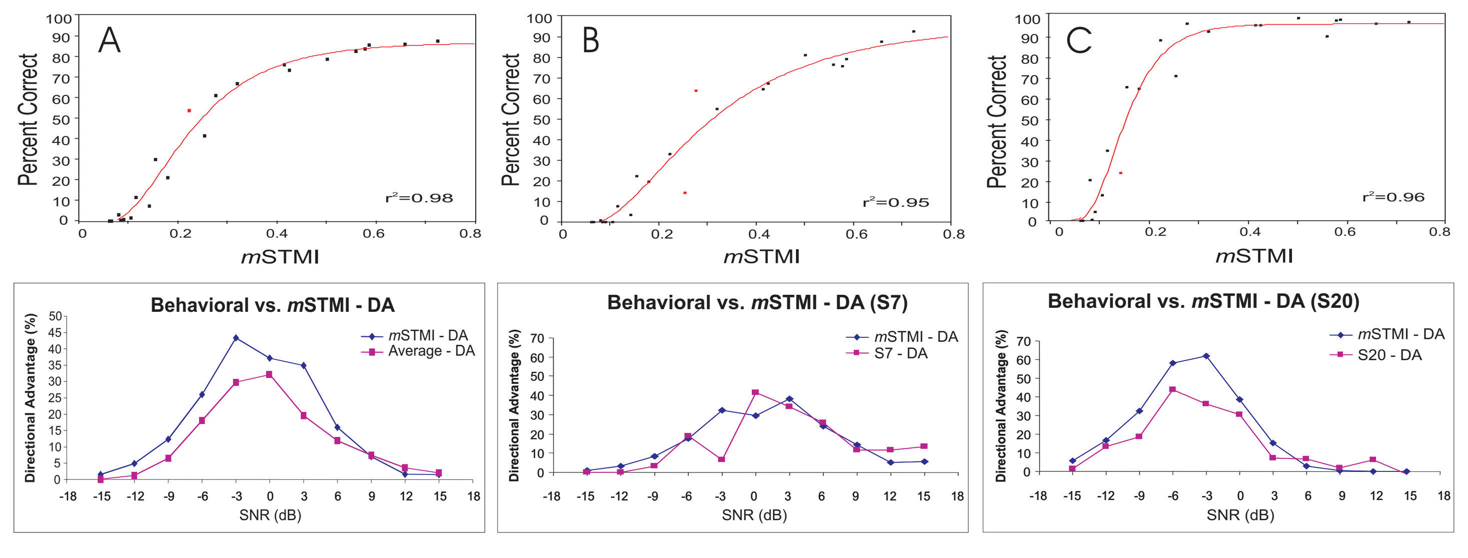

Figure 6. (A). *m*STMI values expressed as percent correct scores based on average data from Walden et al. (in review).The 22 SNR conditions (11 OMNI, 11 DIR) are fit with a logistic function. This function was then used to derive percent-correct equivalents for each *m*STMI value (see Figure 3A - Front). The predicted DA is shown in the bottom panel along with the average behavioral DA. (B, C). There were substantial individual differences in DA observed in the study by Walden et al. (in review).The same procedure in (A) can be used to map *m*STMI differences across OMNI and DIR processing modes onto percent-correct DA's. In panels (B) and (C), data for two hearing-impaired subjects are shown. In (B), subject S7's percent-score grows gradually as the SNR improves and the maximum behavioral DA is at 0 dB. In (C) subject S20's percent-correct score grows rapidly, reaching an asymptote at a *m*STMI of approximately 0.3. For this subject, further improvements in SNR have no effect on the behavioral DA (maximum at -6 dB), but are easily detected by *m*STMI analyses. In spite of these large individual differences, once the *m*STMI values are transformed to percent correct, the behavioral DA is more accurately predicted.

General Discussion and Conclusions

Summary

This study examined whether an acoustic analysis of OMNI and DIR processed speech in noise could predict behavioral measures of intelligibility across a range of SNRs and demonstrate sensitivity to spatial orientation of the speech and noise sources. For the tests conducted so far, the *m*STMI appears capable of distinguishing between microphone modes, is sensitive to spatial orientation, and is consistent with behavioral data.

Comparison with the Articulation Index and Speech Transmission Index

Previous studies (Dhar et al., 2003; Ricketts and Hornsby, 2003; Maj et al., 2004) have used various acoustically derived indices such as the Articulation Index (AI), the Speech Transmission Index (STI), and an intelligibility-weighted directivity index ($\mathrm{DI}_{\mathrm{AI}}$) to predict the behavioral advantage of directional microphones in noise and reverberation. Whereas these studies have demonstrated some success in accounting for directional advantages under controlled laboratory conditions, the methods employed were limited because the probe signals used to test the directional processing were restricted to speech-shaped noise modulated in highly prescribed ways (Ricketts and Hornsby, 2003; Maj et al., 2004), or to real ear measures using swept pure tones (Dhar et al., 2003). The problem with these particular choices of probe signals is that they do not provide guidance as to which microphone mode to choose when real speech, corrupted as it might be by noise and reverberation, arrives at the microphone ports.

A number of speech-based STI methods have been proposed to evaluate signal distortions such as dynamic amplitude compression, envelope expansion, envelope clipping, phase jitter, etc., as well as to analyze differences in the way noise and reverberation affect clear and conversational speech (Payton et al., 1994, 2002; Hohmann and Kollmeier, 1995; Drullman, 1995; Payton and Braida, 1999; Goldsworthy and Greenberg, 2004). In general, these methods have not fared very well at explaining the behavioral consequences of joint spectro-temporal distortions that are common to many non-linear operators (as in channel phase-distortion, amplitude clipping, or phase jitter). Further, many of these methods rely on a direct comparison between clean and noisy and/or reverberant speech tokens in order to determine the transmission loss caused by a particular environment or device (e.g., hearing aid). In many cases, the clean and degraded speech samples originated from the same tokens (same talker, same linguistic materials). However, in practical applications, it is not reasonable to assume that a clean version of the degraded speech signal will be available. Instead, corrupted speech samples must be evaluated and compared to a generalized clean-speech template that includes examples of male and female speakers and various signal levels. The purpose of this template is to obtain a generalized spectro-temporal representation of clean speech that can be compared to unknown speech samples distorted in a variety of unpredictable ways. Some of these distortions will negatively affect intelligibility whereas others may not. For the purpose of this study, we chose to use the biologically motivated STMI based on the auditory model of Chi et al. (1999). This model has been previously shown to be consistent with the STI model for speech corrupted by noise and reverberation as well as for a variety of non-linear distortions (Elhilali, et al., 2003). In this study, the STMI was modified to handle new challenges posed by speech processed through hearing aids while maintaining the previously published performance levels.

Future Directions

The conditions tested in this study involved continuous speech-shaped noise and a single speech source. Further evaluations of the *m*STMI are planned with more realistic noise sources as well as multiple speech targets. Initial results with these more realistic soundscapes appear promising.

References

Chi, T., Gao, M., Guyton, M.C., Ru, P., and Shamma, S.A. (1999). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Amer. 106, 27192732.

Cord MT, Surr RK, Walden BE, Olson L. (2002). "Performance of directional microphone hearing aids in everyday life," *Journal American Academy Audiology* 13, 295-307.

Dhar, S., Humes, L.E., Calandruccio, L., Barlow, N.N., and Hipskind, N. (2004). "Predictability of speech-in-noise performance from real ear measures of directional hearing aids," Ear Hear. 25, 147-158.

Drullman, R. (1995). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. 97, 585-592.

Elhilali, M., Chi, T., and Shamma, S.A. (2003). "A spectro-temporal modulation index (STMI) assessment of speech intelligibility," Speech Commun. 41, 331-348.

Garofolo, J. (1988). *Getting Started with the DARPA TIMIT CD-ROM: An Acoustic Phonetic Continuous Speech Database*, National Institute of Standards and Technology NIST, Gaithersburgh, M.D.

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

Hohmann, V., and Kollmeier, B. (1995), "The effect of multichannel dynamic compression on speech intelligibility," J. Acoust. Soc. Am. 97, 1191-1195.

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.

Institute of Electrical and Electronic Engineers (1969). *IEEE recommended practice for speech quality measures*. New York: IEEE.

Maj, J-B., Wouters, J., and Moonen, M. (2004). "Noise reduction results of an adaptive filtering technique for dual-microphone behind-the-ear hearing aids," Ear Hear. 25, 215-229.

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.

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

Payton, K.L., Braida, L.D., Chen, S, Rosengard, P., and Goldsworthy, R. (2002), "Computing the STI using speech as a probe stimulus," *Past, Present and future of the speech transmission index*. (TNO Human Factors, Soesterberg, The Netherlands), pp. 125-138.

Ricketts T.A., Hornsby W.Y. (2003). "Distance and reverberation effects on directional benefit," *Ear Hear* 24, 472-484.

Steeneken, H.J.M., and Houtgast T. (1980), "A physical method for measuring speech transmission quality," J. Acoust. Soc. Am. 67, 318-326.

Walden, B.E., Grant, K.W., Surr, R.K., Summers, W.V., Cord, M.T., and Dyrlund, O. (In review). "Effect of signal-to-noise ratio on directional microphone benefit, preference, and loudness," Journal American Academy Audiology.

Walden BE, Surr, RK, Cord MT, Dyrlund O. (2004). "Predicting hearing aid microphone preference in everyday listening," Journal American Academy Audiology 15, 365-396.

This work was partially supported by a grant from AFOSR to the University of Maryland. We would also like to thank GN ReSound for their support of the earlier study by Walden et al. (in review) and for supplying the modified CANTA 770D hearing aid used to make the OMNI and DIR recordings. The opinions and assertions presented are the private views of the authors and are not to be construed as official or as necessarily reflecting the views of the Department of the Army or the Department of Defense.