An Objective Measure for Selecting Microphone Modes in OMNI/DIR Hearing Aid Circuits

Ken W. Grant,[1] Mounya Elhilali,[2] Shihab A. Shamma,[2] Brian E. Walden,[1] Rauna K. Surr,[1] Mary T. Cord,[1] and Van Summers[1]

Objectives: **Studies have shown that listener preferences for omnidirectional (OMNI) or directional (DIR) processing in hearing aids 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. Many modern hearing aids incorporate algorithms to switch automatically between microphone modes based on an analysis of the acoustic environment. Little work has been done, however, to evaluate these devices with respect to user preferences, or to compare the outputs of different signal processing algorithms directly to make informed choices between the different microphone modes. This study describes a strategy for automatically switching between DIR and OMNI microphone modes based on a direct comparison between acoustic speech signals processed by DIR and OMNI algorithms in the same listening environment. In addition, data are shown regarding how a decision to choose one microphone mode over another might change as a function of speech to noise ratio (SNR) and spatial orientation of the listener.**

Design: **Speech and noise signals were presented at a variety of SNR's and in different spatial orientations relative to a listener's head. Monaural recordings, made in both OMNI and DIR microphone processing modes, were analyzed using a model of auditory processing that highlights the spectral and temporal dynamics of speech. Differences between OMNI and DIR processing were expressed in terms of a modified spectrotemporal modulation index (mSTMI) developed specifically for this hearing aid application. Differences in mSTMI values were compared with intelligibility measures and user preference judgments made under the same listening conditions.**

Results: **A comparison between the results of the mSTMI analyses and behavioral data (intelligibility and preference judgments) showed excellent agreement, especially in stationary noise backgrounds. In addition, the mSTMI was found to be sensitive to changes in SNR as well as spatial orientation of the listener relative to signal and noise sources. Subsequent mSTMI analyses on hearing aid recordings obtained from real-life environments with more than one talker and modulated noise backgrounds also showed promise for predicting the preferred microphone setting in varied and complex listening environments.**

(Ear & Hearing 2008;29;199–213)

Introduction

Persons with mild to moderate hearing loss or worse often require a more favorable signal to noise ratio (SNR) to understand speech than do individuals with normal hearing (Bronkhorst & Plomp, 1990; Dubno, et al., 1984). As a result, many hearing aids offer listeners the option to use directional (DIR) microphones instead of omnidirectional (OMNI) microphones, which have been demonstrated to improve the SNR under certain listening environments (Blamey, et al., 2006; Chung, 2004; Ricketts & Henry, 2002; Valente, et al., 1995). Recent laboratory and field studies (Cord, et al., 2002; Surr, et al., 2002; Walden, et al., 2004) have shown that the perceived benefit of DIR microphones, when compared with OMNI microphones, depends primarily on the characteristics of the listening environment. Specifically, OMNI microphones tend to be preferred in quiet listening situations or in the presence of background noise when the signal source is not located in front of the listener or is moving. On the other hand, DIR microphones tend to be preferred when background noise is present, and the signal is located close to and in front of the listener.

For a variety of reasons (e.g., poor manual dexterity, uncertainty as to when to switch between microphone modes, forgetfulness), roughly one-third of hearing-impaired patients fit with manually selected OMNI/DIR hearing aids fail to use the DIR mode even when the acoustic environment suggests that a clear microphone preference and a benefit in speech understanding would likely result (i.e., when noise is present, the signal is in front of the listener, and the signal is near). That is, some patients tend to leave their hearing aids permanently set in the default OMNI mode, regardless of the listening environment (Cord, et al., 2002).

To address this problem, most modern hearing aids that offer DIR processing use algorithms that

[1]Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC; and [2]Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland.

0196/0202/08/2902-0199/0 • Ear & Hearing • Copyright © 2008 by Lippincott Williams & Wilkins • Printed in the U.S.A.

automatically select the "preferred" microphone mode by continuously monitoring the listening environment and adapting the hearing aid parameters so that they are optimized for the particular acoustic conditions (Blamey, 2006; Fabry, 2006; Palmer, et al., 2006). Unfortunately, there have been very few studies that have explicitly evaluated patient satisfaction with automatic directionality circuits. Does the hearing aid switch at the right time and situation? Does it improve the SNR of the desired target signal, or is the wrong signal sometimes enhanced? How easy is it for the patient to override the automatic decision and return to manual control? The difficulty in answering these questions stems in large measure from not knowing precisely when the hearing aids switch between OMNI and DIR modes, how often and when listeners were in each particular acoustic environment, and what the listener's preference was at the time the hearing aid chose one mode over the other. Despite these methodological challenges, some recent studies suggest that whereas hearing aid algorithms used for automatic directionality continue to improve, they are still likely to make errors and select DIR processing mode in situations where OMNI would be preferred, and vice versa. For example, Palmer et al. evaluated the Siemens Triano3 ™ in a field trial of 49 individuals with moderate to severe sloping hearing loss. The primary questions addressed in this study were (1) whether adaptive directionality was preferred over fixed directionality or OMNI processing; (2) whether listeners prefer one of the three microphone modes (OMNI, fixed directionality, or adaptive directionality) for most everyday listening situations; and (3) whether the preferred microphone mode could be predicted from the particular acoustics of the listening situation. Results showed that roughly one-third of the patients could not distinguish between automatic directionality (either fixed or adaptive) and OMNI processing modes. For the remaining patients, when given the opportunity to choose between OMNI mode and automatic directionality, approximately half chose to be in OMNI mode more often than automatic directionality, even though automatic directionality includes the option to set the microphone mode to OMNI processing. To interpret these data one has to consider the two possible states that could have occurred with the automatic directionality algorithm used in this study. For a particular acoustic environment, if the automatic directionality algorithm resulted in placing the aid in OMNI mode (which probably happened quite often) listeners would have no basis for choosing reliably between OMNI and automatic DIR, and we would expect a roughly equal split between the amount of time spent in one mode or the other. On the other hand, if the algorithm selected the DIR mode and the listening situation favored directionality, listeners should have selected the automatic DIR mode more often than the OMNI mode. Therefore, the fact that the OMNI mode was used as often as the DIR mode suggests that the algorithm must have been wrong some proportion of the time when it chose to implement DIR processing instead of OMNI processing (e.g., a noisy environment where the talker was not in front of the listener) and vice versa. These errors can occur for at least two reasons: (1) the scene analysis classifier could have made an error, or (2) the rules linking a particular scene with a particular microphone mode may not be correct (e.g., not all noisy environments should result in DIR processing). Additionally, there could be a general bias on the part of some listeners for preferring OMNI over DIR (perhaps relating to localization and other qualitative differences between the two processing modes).

A somewhat similar finding was reported by Fabry (2006) in an evaluation of a variety of different wearable digital aids equipped with automatic directionality. Sixty-three hearing-impaired subjects rated several aspects pertaining to the switching behavior of the hearing aids including how often the instrument switched between microphone modes. Approximately 25% of the judgments were perceived as either too seldom or too often, suggesting some dissatisfaction with the algorithm decision.

Blamey et al. (2006) reported results for an automatic adaptive directional aid in conjunction with the adaptive dynamic range optimization amplification scheme. Data from eight patients included speech intelligibility, speech and spatial quality, and situational preferences. As with the Palmer et al. (2004) and Fabry (2006) studies, subjects could not distinguish between the adaptive directionality and OMNI mode of processing approximately one-third of the time (not too surprising considering that the adaptive directionality will often select the OMNI processing mode). When subjects were asked to rate their preferences for the two processing modes in different listening environments, 17% chose the OMNI mode of processing over the automatic adaptive directionality. As discussed earlier, the fact that a proportion of subjects still preferred OMNI processing over automatic directionality indicates that some of the decisions made by the algorithm must have been wrong.

Finally, Walden et al. (2004) estimated the time period listeners spent in different acoustic environments characterized by, whether noise and/or reverberation was present, the location of the primary talker and the distance of the primary talker. Listeners reported that roughly 35% of the time background noise was present and the talker was near and in front. Although not guaranteed, these are the most likely conditions where DIR processing would be favored over OMNI processing. Recent estimates from our own clinic, based on data logging capabilities found in some hearing aids, suggest that automatic directionality is activated between 5 and 17% of the time, substantially <35% anticipated by Walden et al. (2004).

Taken together, the studies referenced above suggest that (1) differences between OMNI and DIR processing in many real-world environments are subtle and difficult for subjects to hear, and (2) the decision as to when to switch between OMNI and DIR processing modes, depending on the attributes of the acoustic environment, continues to be a difficult decision with room for improvement.

Many automatic DIR hearing aids operate based on a "scene analysis" approach. This requires that hearing aids classify accurately a number of acoustic environmental attributes. For example, the algorithm may have to identify an acoustic signal of interest in the environment (e.g., speech, music, sounds of nature), whether noise was present or not, the location of the signal(s) of interest and the noise(s), an estimate of the distance of signal(s) and noise(s) from the listener, and an estimate of the degree of reverberation in the environment. Each of these estimates must then be consulted before automatically making a decision as to the preferred microphone setting at a given time.

Recent attempts to accurately identify and label different auditory scenes by a categorization of acoustic features extracted from signals recorded from the environment have shown considerable promise (e.g., quiet versus noise, speech versus music, Buchler, 2002; speech versus nonspeech, Mesgarani, et al., 2006; noisy speech versus clean speech, Mesgarani & Shamma, 2005). However, even with modern advances in computational auditory scene analysis, and greater and greater accuracy in classifying specific acoustic scenes, the chance of making a mistake and selecting the wrong microphone mode for a given listening environment may still be unacceptably high. This is because the acoustic correlates differentiating one auditory scene from another with regard to key dimensions of interest, for example, the number of sources, the source type (noise or speech), the location of the various sources, and the distance of each source from the listener, are not well understood or easily measured in real world environments. This is particularly true in reverberant environments where the signal source locations can be obscured by room reflections. Further, the rules for determining the relationship between the acoustic scene and optimal microphone selection are not well established. DIR microphones are not preferred in all noisy environments, and substantial differences in microphone preferences across hearing-impaired listeners may exist.

Other methods for determining microphone switching rules besides a detailed acoustic scene analysis also exist. Some of these methods use relatively simple criteria such as a measure of the overall incoming signal level, an estimation of the SNR, the spectral shape of the incoming sound, and a determination of whether wind noise is present or absent (e.g., see Chung, 2004 for a review; Blamey, 2006). More infrequently, acoustic analyses can be made in parallel for the different microphone modes under consideration. In these cases, a switching decision can be made based on an estimation of the SNR of the signal produced by OMNI and DIR circuits or on a variety of other variables (e.g., whether wind noise is present, whether moderate to high noise levels are detected).

The current manuscript focuses specifically on this latter approach of automatic switching between OMNI and DIR microphone modes. In this "direct comparison" approach, acoustic signals are sampled through both microphone settings of the hearing aid simultaneously or in near succession, and whichever signal (DIR or OMNI) is the "cleaner," less noisy signal is chosen to be delivered to the listener. When the target signal is speech, the primary decision may be which processed signal (OMNI or DIR) has characteristics that come closest to clean speech, uncorrupted by noise and/or reverberation. To make this decision, an analysis of the processed signals must be performed for each microphone mode, and the results compared with a model of clean speech. For practical applications (e.g., a hearing aid worn in everyday listening environments), a model, or "template" of clean speech, must be developed because access to an uncorrupted version of the current speech signal is unavailable.

Several different analysis metrics that might serve well for predicting the benefits of different microphone modes have been suggested. For example, Dhar et al. (2004), Ricketts and Hornsby (2003), and Maj et al. (2004) have used acoustically derived indices such as the Articulation Index (AI), the Speech Transmission Index (STI), and an intelligibility weighted Directivity Index (DI_{AI}) in an attempt to predict the behavioral advantage of DIR microphones in noise and reverberation. Although these studies have demonstrated some success in accounting for DIR advantages (DAs) under controlled laboratory conditions, the methods used were limited because the probe signals and analysis methods used to test the signal quality of DIR versus OMNI processed hearing aid outputs were restricted to speech-shaped noise modulated in highly prescribed ways (Maj, et al., 2004; Ricketts & Hornsby, 2003), or to real ear measures using swept pure tones (Dhar, et al., 2004). Unfortunately, these particular choices of probe signals and analysis methods do not provide guidance as to which microphone mode to choose when real speech, corrupted 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, and to analyze differences in the way noise and reverberation affect clear and conversational speech (Drullman, 1995; Goldsworthy & Greenberg, 2004; Hohmann & Kollmeier, 1995; Payton & Braida, 1999; Payton, et al., 1994, 2002). In general, these methods have not fared very well at explaining the behavioral consequences of joint spectrotemporal distortions that are common to many nonlinear operators (as in channel phase-distortion, amplitude clipping, or phase jitter). More importantly, however, these methods tend to rely on a comparison between clean and noisy and/or reverberant speech tokens to determine the transmission loss caused by a particular environment. In many cases, the clean and degraded speech samples originated from the same tokens (same talker, same linguistic materials). As previously mentioned, for the purposes of developing a strategy for automatically switching between DIR and OMNI microphone modes in a variety of everyday listening environments, it is neither reasonable nor practical to assume that a clean version of the degraded speech signal will be available. In the approach described in this article, corrupted speech samples, processed through both microphone modes, are compared with a generalized clean-speech template that includes examples of male and female speech and various signal levels. This template is a generalized representation of clean speech that can be compared with unknown speech samples distorted in a variety of unpredictable ways.

The goal of the present investigation was to compare acoustic signals captured at the output of the hearing aid microphone (after OMNI or DIR processing) to a clean-speech template, make a decision based on this comparison as to whether the signal processed by the OMNI microphone or DIR microphone would be the more intelligible and most likely preferred signal for the listener, and evaluate the accuracy of this decision against behavioral data that included intelligibility measures and microphone preferences through both microphone modes (Walden, et al., 2005). To achieve these goals, we chose to use the spectrotemporal modulation index (STMI) based on the auditory model of Chi et al. (1999, 2005). Specifically, we examined whether an STMI analyses of OMNI and DIR speech recorded in laboratory and real-world environments with both noise and reverberation could be related to behavioral measures of intelligibility and preference across a range of SNR, speech/noise source locations, and environmental configurations. To the extent that acoustic measures of DIR microphone processing can be closely related to behavioral measures of DIR microphone performance (e.g., intelligibility and preference), more effective automatic microphone switching algorithms might be developed. In exploring this possibility, the STMI was modified to handle speech processed through hearing aids while maintaining its efficacy as a metric for analyzing a variety of linear and nonlinear acoustic distortions, previously reported by Elhilali et al. (2003). Modifications to the STMI, acoustic recording procedures, and behavioral measures are described below. The results of the modified STMI analyses were compared with the behavioral results of Walden et al. to determine if behavioral preferences of OMNI/DIR microphone modes can be predicted from the acoustic output of the hearing aid alone.

MATERIALS AND METHODS

Experiment 1: Relation Between Objective and Subjective Laboratory Measures of Speech Intelligibility in Noise

Acoustic Recordings • A modified GN ReSound Canta770D hearing aid that provided direct access to the processed signals just before the hearing aid receiver was used to make all acoustic recordings. The modified hearing aid was programmed with the audiogram + fitting algorithm of the Aventa 1.2 software using the binaurally averaged hearing threshold data from 31 hearing-impaired subjects tested by Walden et al. (2005). Because DIR processing typically results in a loss of gain (relative to OMNI mode) in the low frequencies, the low-frequency gain was adjusted using the "max boost" feature to equalize the outputs between the two microphone settings. For recording purposes, the hearing aid was placed behind the right ear of a Knowles Electronics Mannequin for Acoustic Research (KEMAR). The mannequin was positioned in the center of four loudspeakers located at 90° intervals around the head. Figure 1 shows the basic geometry of the recording setup. The loudspeakers were mounted on the walls of a double-walled sound-treated audiometric suite with dimensions 8′ wide × 10′ long × 7′ high. The floor of the booth was carpeted and the walls and ceiling were sound treated to reduce reverberation. The loudspeakers

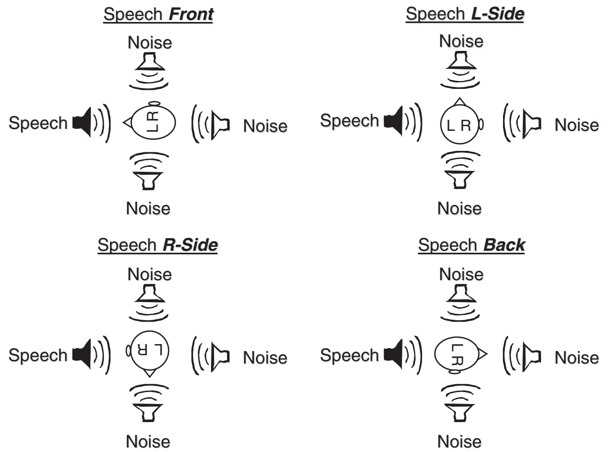

Fig. 1. Recording arrangement. The speech was presented from the loudspeaker labeled "speech". Uncorrelated speech-shaped noise was presented from the remaining three loudspeakers. The hearing aid was worn on the mannequin's right ear. The panel labels (e.g., "Speech R-Side") indicate the location of the speech signal relative to the mannequin's facing orientation.

were positioned in the center of each wall at a height equal to ear level for a seated subject (approximately 4′). This same exact audiometric suite was used in the Walden et al. study, where subjects recognized speech and rated OMNI and DIR preferences in a variety of SNR conditions.

In separate recordings, the mannequin was rotated so that it faced either the speech signal (0°) or one of three uncorrelated speech-shaped noise signals (90, 180, and 270°). Note that for three of these spatial orientations, the hearing aid was closer to one of the noise sources than to the speech. In other cases, head shadow effects played a role in shaping the recorded output, for example, when the hearing aid was closest to the rear speaker (i.e., the mannequin rotated to the right 90°).

The input speech signals consisted of IEEE/Harvard sentences (IEEE, 1969) spoken by a female talker of American English. The sentences were presented at a fixed level of 65 dBA (measured at the position of KEMAR's head). Eleven different SNRs, ranging from −15 to +15 dB in 3-dB steps, were used. The noise consisted of a white noise spectrally filtered to match the long-term spectrum of the entire set of 720 IEEE sentences. The signal coming from each of the three separate noise loudspeakers was equal in level but uncorrelated in phase. The SNR was determined as the difference between the speech level and the sum of the three noise signals measured at the position of the mannequin's head. The four loudspeakers (three noises, one speech) were controlled by separate programmable attenuators.

The output of the hearing aid (just before the hearing aid receiver) was routed to the microphone input of a laptop computer and digitized at 44.1 kHz (16 bits). Approximately 30 sec of speech were recorded at each SNR in each microphone settings (OMNI and DIR) and in all four spatial orientations. The recorded signals were further processed through a simulation of the Canta 770D receiver. No attempt was made to simulate the effects of occluded ear canal resonance that would have further shaped the recorded signals. This decision was made because it was felt that any automatic algorithm selection performed by the hearing aid would likely be based on differences among electronic signals (e.g., OMNI versus DIR) acquired from inside the hearing aid circuitry rather than from signals recorded in the patient's ear canal (Walden, et al., 2007).

Modified STMI Analysis • The acoustic recordings were analyzed using a modified version of the spectrotemporal modulation index (STMI) described by Elhilali et al. (2003). Conceptually, the STMI is a measure of the joint spectral and temporal modulations in speech as reflected by a model of auditory processing (Chi, et al., 1999, 2005). The model is inspired from known physiological findings of the peripheral and central mammalian auditory system, but is greatly simplified to allow for relatively fast computations. We chose to use the auditory processing model developed by Chi et al. primarily because it represents complex acoustic signals (such as speech) as a combination of spectral and temporal modulations. These slow rate modulations in amplitude and frequency have been demonstrated to be critically important for speech (Drullman, et al., 1994; Houtgast & Steeneken, 1985). The model extracts and highlights this particular attribute of dynamic acoustic signals through the use of a cortical model that uses tuned modulation filters. Thus, speech signals (and other complex signals) are represented in terms of a set of relevant cortical features that are related to both intelligibility and quality.

The auditory processing model starts with a peripheral stage (including cochlear filtering, hair-cell transduction, auditory-nerve, and cochlear-nucleus spectrotemporal sharpening) to produce a time-frequency pattern of activation called an auditory spectrogram (Fig. 2A). Next, the model proceeds to a finer analysis of the spectrogram via a bank of modulation-selective filters tuned to a range of temporal (rate) and spectral (scale) modulations. This analysis is implemented via a two-dimensional wavelet transform, whose parameters are derived from physiological data in animals and psychoacoustic results in humans (Chi, et al., 1999, 2005). The model yields a multi-dimensional representation of

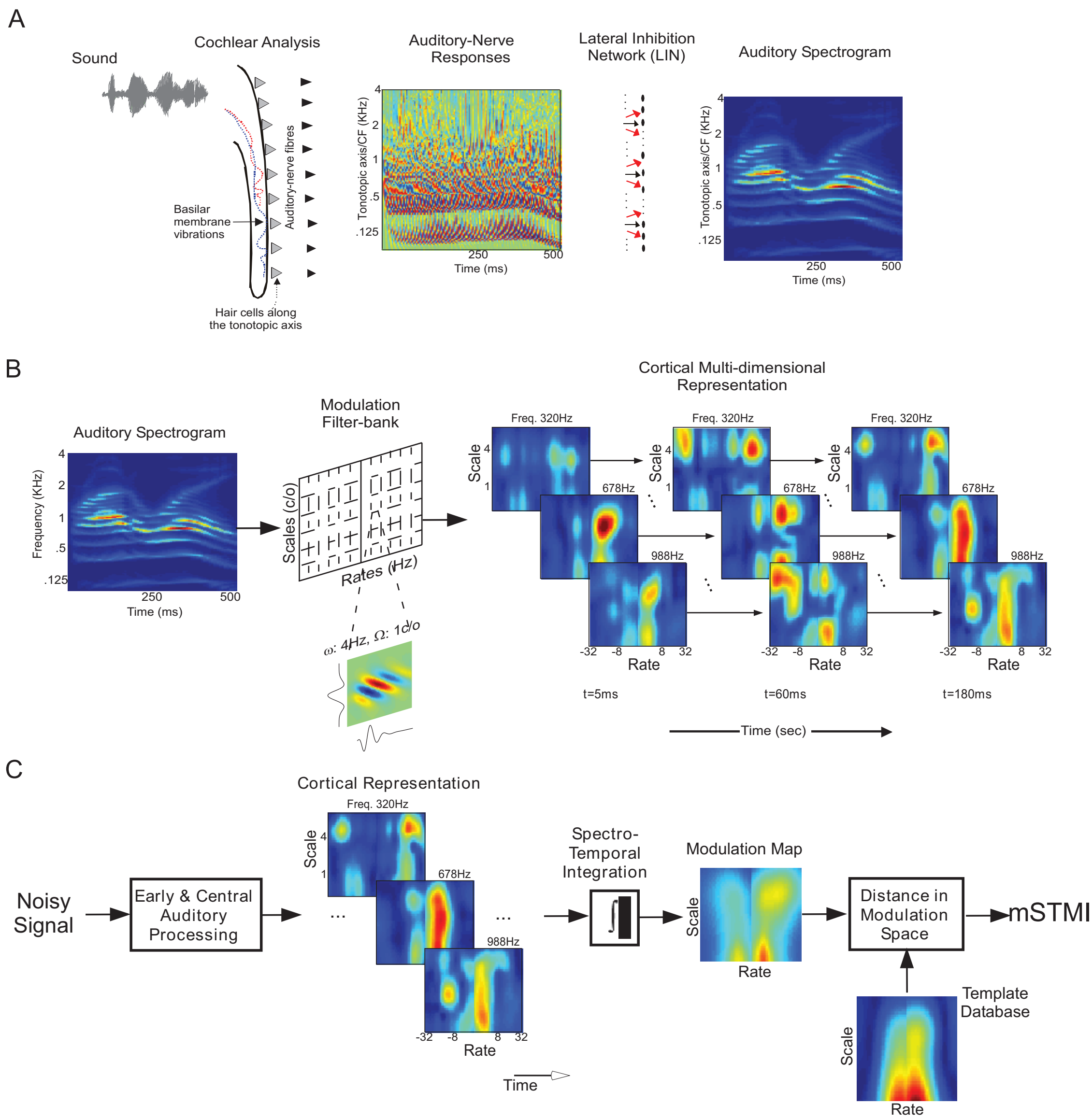

Fig. 2. Schematic of processing stages for STMI calculations. A, Early auditory processing stages. The acoustic signal is analyzed by a bank of 128 constant-Q filters (cochlear filtering stage). The output of each filter is processed by a hair cell model followed by a lateral inhibitory network, and 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 spectrotemporal modulation filter is shown along with the result of convolving it with the auditory spectrogram. C, Schematic showing steps in computing the STMI. A template is formed by processing clean speech through the model. In this study, 200 sec of speech extracted randomly from different male and female adult speakers from the TIMIT database (Garofolo, 1988) were used to create the clean-speech template. To evaluate a particular sample of speech (e.g., after OMNI or DIR processing), the noisy signal is processed in the same manner as the template. The result is a rate-scale plot (shown here collapsed over tonotopic frequency and time) that is distorted relative to the clean-speech template. The STMI is the normalized difference between template and test signal, weighted by the clean-speech template. (Figure adapted from Chi, et al., 1999, 2005 and Elhilali, et al., 2003).

speech signals that reflect the temporal and spectral modulation content of sound and their distribution in time and frequency (Fig. 2B). This auditory model is used to build a template or generalized representation of the average spectrotemporal modulation patterns found in natural speech. This template is derived from a long sample of conversational speech produced by many male and female speakers. The template is an average across various speakers and speaking styles of the typical spectral and temporal modulation content in conversational speech. Using this generalized template, the STMI measures the fidelity of the modulation content of any target speech signal relative to this generic template by computing a difference between the template and modulation map of the test signal. In order to focus the analysis specifically on the dynamic temporal and spectral modulations in the speech signal, the cortical output of the model for both template and test stimulus are modified by subtracting the model output due to the signal's base spectrum. The base spectrum is a stationary noise with a spectrum identical to that of the long-term spectrum of the signal being processed (usually about 20–30 sec of speech). This base signal is processed through the auditory model to yield its own multirate representation, which is then subtracted from the original signal's representation. The same baseline adjustment is also performed when constructing the clean-speech templates.

The STMI is based on the premise that any distortion of the modulation map of the generalized natural speech template reflects a loss of fidelity and, as a consequence, a loss in intelligibility. On the other hand, any manipulation or noise that does not disrupt significantly the integrity of this map would be relatively harmless to intelligibility. The STMI bears a great deal of similarity to another, more common intelligibility metric known as the STI. However, the STMI offers the advantage of using a generalized template that does not have to correspond directly to the specific speech samples being tested. In other words, newly acquired speech samples can be evaluated against the generic template to determine the relative loss in signal clarity as a result of environmental factors such as noise and reverberation. In addition, The STMI differs fundamentally from the STI in its sensitivity to joint spectrotemporal modulations, and hence in its ability to detect distortions that are inseparable along the temporal and spectral dimensions.

The basic steps in computing the STMI are depicted schematically in Figure 2. The top panel shows the early stage of processing where speech is analyzed by a bank of cochlear filters, a hair cell model, and lateral inhibitory network to produce a neural spectrogram. The middle panel shows the neural spectrogram processed by a bank of modulation selective filters to produce the multirate cortical representation. The multidimensional cortical representation is typically reduced to a 3D representation (spectral modulation, temporal modulation, and frequency) by integrating over time to produce either a generic template or the analyses of a specific sample of speech that is under investigation. The third panel in Figure 2 shows the actual STMI analyses as used for the current study. Here, the target noisy signal is processed as described earlier and the cortical representation is integrated over both frequency and time to produce a 2D representation of spectral and temporal modulations. The target representation is compared with a similarly reduced template and the normalized distance between target and template is interpreted as a measure of the loss in signal fidelity.

Three important modifications to the original STMI described by Elhilali et al. (2003) were used for this work. First, the clean-speech database derived from randomly selected TIMIT sentences (Garofolo, 1988) and used to build the generic speech template was extended by roughly a factor of 10 (from 20 sec to roughly 200 sec). This increase allows for a more realistic view of the variability in natural speech and makes the method for computing the STMI from generic templates more flexible and robust. In particular, by averaging individual differences among talkers (e.g., gender, age, physical characteristics of the vocal tract) and speaking styles, the impact of any one individual speaker is reduced as longer and more varied speech is used to develop the generic template. Throughout the current study, we used the same extended template for all experiments. This template was derived from the original TIMIT database without being processed through a hearing aid.

A second major modification to the STMI pertained to the number of dimensions used to characterize the spectrotemporal modulations in speech. In the original formulation (Elhilali, 2003), the template and target representations were described in three dimensions (tonotopic frequency × temporal modulation × spectral modulation) collapsed over stimulus duration. In the modified STMI, the three-dimensional mapping was reduced to a two-dimensional mapping by integrating the template across the tonotopic axis. This modification was essential to account for the high-frequency emphasis introduced by the hearing aid, but still enabled the model to capture the total amount of spectrotemporal modulation present in the sound. This frequency compensation could have equally been achieved by a change of the cochlear filtering stage in the model. However, we opted for using a spectral integration in the mSTMI to maintain the generality of the STMI technique. In fact, the mSTMI can use the same generic template for analyzing both regular and hearing aid processed speech, hence circumventing the need to change the analysis method depending on the circumstances and the nature of the signal.

A third modification to the original STMI computation was designed to emphasize the importance of particular modulation bands in clean speech. That is, by weighting any differences observed between the current speech sample and the clean-speech template by the template itself, modulation regions where the template has significant amounts of energy are highlighted, and regions where the template has low energy are deemphasized. This modification helped to further reduce the unwanted influence of spectral tilt introduced by the frequency-gain shaping of the hearing aid. In the original template-based STMI described by Elhilali et al. (2003), an Euclidean distance metric was used to describe differences between the clean-speech template and a given noisy speech template. In other words, $STMI^T = 1 - \frac{\|T - X\|^2}{\|T\|^2}$, where T and X are the 3D model outputs (temporal modulation × spectral modulation × tonotopic frequency) for the clean-speech template and noisy speech sample, respectively. In the modified STMI used in this study, a weighted distance metric was used: $STMI^T = 1 - \frac{\|T \bullet (X - T)\|}{\|T^2\|}$, where T and X are the 2-D model outputs. The shortest (Euclidian) distance is defined as $\|A\| = \sqrt{\Sigma_i A_i^2}$, where A_i is the individual element of the vector (or vectorized matrix) $\boldsymbol{A}$. Thus, in computing the modified STMI, the distance between the clean-speech template and the recorded speech samples processed through the hearing aid was weighted by a factor proportional to the degree of modulation within each band in the clean-speech template. The denominator represents the Euclidian norm (length) of the vectorized squared template profile T^2. These distances were obtained for each successive 2-sec interval.

The final mSTMI was computed as a moving average of five successive 2-sec intervals. Thus, in its current formulation, a microphone decision based on mSTMI differences for OMNI and DIR modes could be made at most once every 2 sec, after an initial 10 sec of speech. Further, for a change in the acoustic environment to be fully represented by the mSTMI analysis (with no overlap with the previous analysis), the test samples would have to be at least 10 sec apart.

Behavioral Measures of DIR Advantage • Data from Walden et al. (2005) was used to evaluate the mSTMI for predicting speech recognition scores and microphone preferences. These data consisted of measurements of sentence recognition in noise and microphone preference ratings by 31 hearing-impaired adults. IEEE/Harvard sentences were presented in a sound attenuated audiometric booth at SNR values between −15 and 15 dB in 3 dB steps. The speech signal was presented at a constant level of 65 dBA from a loudspeaker positioned approximately 4 ft. in front of the subject. Uncorrelated noise from three loudspeakers positioned to the sides and back of the subject (90, 270, and 180°) was added at varying levels to produce the different SNR conditions (see Fig. 1, "Speech Front"). For all conditions, the level of the noise from three loudspeakers was equated at the position of the listener's head. The intelligibility of each SNR condition was assessed under both ONMI and DIR microphone modes using 30 unique IEEE sentences per condition. Subjects responded verbally to each sentence and the number of correct key words was tallied.

In addition to speech recognition scores, Walden et al. (2005) also reported microphone preferences for each of the 11 SNR conditions. Approximately 40 sec of concatenated IEEE sentences were presented to subjects who manually switched several times between microphone modes and reported a preference for either OMNI or DIR processing or reported "no preference." Walden et al. showed that the preference data and the difference between speech recognition scores obtained for the OMNI and DIR modes (i.e., the DA) were highly correlated ($r > 0.9$), suggesting that, for the conditions tested, microphone preferences were determined largely by the relative intelligibility of speech through each microphone mode. For further details, see Walden et al. (2005).

RESULTS

Acoustic Analyses

The purpose of this study was to determine the feasibility of automatically choosing the best microphone mode for a given listening environment based on a direct comparison of acoustic information derived from the OMNI and DIR conditions within a single stable environment. Different test environments were created by changing the SNR of the background noise for a fixed-intensity, fixed-position speech signal. Recordings of hearing aid output were obtained by rotating, in 90° increments, a KEMAR dummy wearing a hearing aid on the right ear. As noted earlier, the mSTMI is a summary description of the differences in spectral and temporal dynamics between the test signal and a generic clean-speech template. The results of this analysis are shown in Figure 3.

The data in Figure 3A indicate that higher mSTMI scores occurred for the DIR processing mode (when compared with the OMNI mode) only when the speech signal originated from the front. When the signal was in back, or when the aided ear was nearest to the signal (e.g., speech on R-Side), the OMNI processing mode resulted in the greater mSTMI values. Little or no difference between microphone modes was observed when the aided ear was opposite to the signal and affected by head shadow (e.g., speech on L-Side).

When the data are plotted separately for each processing mode (Fig. 3B), the mSTMI values obtained with DIR processing showed a graded sensitivity for each of the four spatial orientations indicating a gradual shift in the SNR. In OMNI mode,

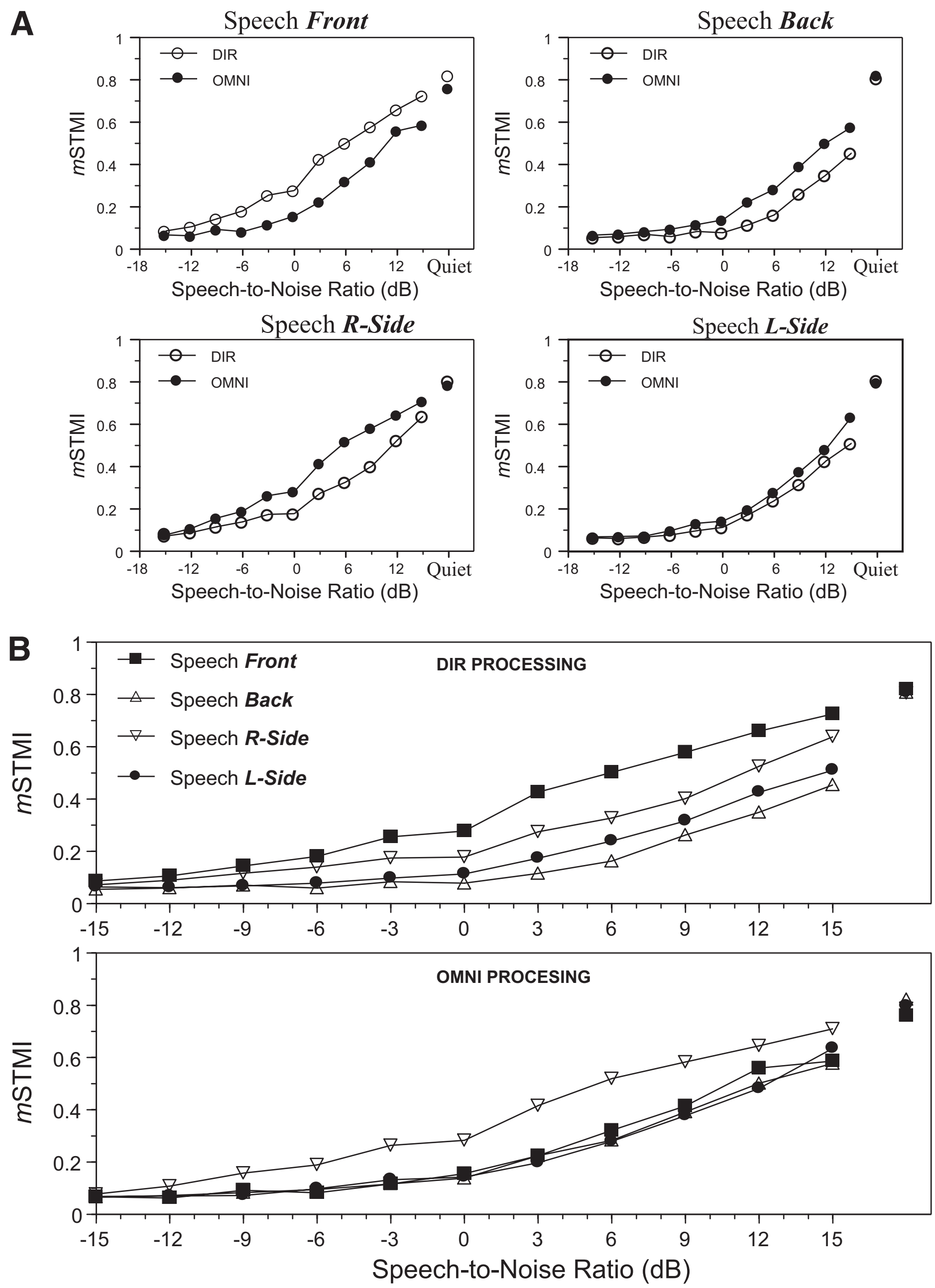

Fig. 3. A, mSTMI as a function of speech to noise ratio and spatial location of speech and noise sound sources. Each panel shows the results for IEEE sentences (1969) recorded through a KEMAR mannequin wearing a modified Canta 770D hearing aid in the right ear at a different spatial orientations (see Fig. 1). Each computation is based on the moving average of five 2-sec nonoverlapping 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.

however, mSTMI values were nearly identical for all orientations except when the speech was closest to the hearing aid (see Fig. 3B, bottom panel, Speech R-Side). These data show one potential use of the mSTMI as a tool for scene analysis. By computing the mSTMI in the OMNI mode across ears, it may be possible to determine where the predominant speech signal is in space. For example, similar $\mathrm{mSTMI}_{\mathrm{OMNI}}$ values across ears would suggest that the signal is in front, rear, above, or below the listener. Different $\mathrm{mSTMI}_{\mathrm{OMNI}}$ values across ears would suggest that the speech signal is coming from the side with the greater OMNI fidelity.

Comparison of Acoustic Analyses and Behavioral Data

Walden et al. (2005) measured speech intelligibility in noise and microphone preferences for a range of SNR conditions when speech was presented from the front loudspeaker and noise from the sides and rear. Their data showed a consistent advantage for the DIR mode over the OMNI mode for all SNR conditions tested. As can be seen in Figure 3A, mSTMI analyses obtained when the speech was presented from the front showed that the DIR microphone mode resulted in higher mSTMI values than the OMNI mode for all SNR's tested. Further,

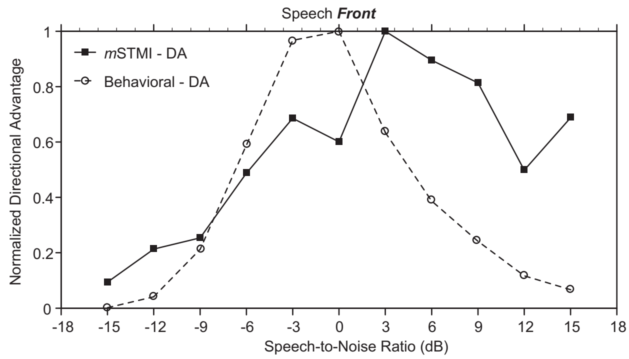

Fig. 4. Average directional advantage (DA) from Walden et al. (2005) and from the mSTMI analyses presented in Figure 3. Behavioral DA scores were obtained by subtracting OMNI percent-correct recognition scores from DIR percent-correct recognition scores. These differences were then normalized so that the maximum DA equals one. Normalized DA for mSTMI values were based on OMNI and DIR values shown in Figure 3A (Speech Front).

because Walden et al. showed that for these laboratory test conditions subject preference ratings were strongly associated with the DA (the difference in intelligibility between the DIR and OMNI modes), the mSTMI seems to be generally consistent with both the speech recognition data and preference ratings showing an advantage for DIR processing.

Figure 4 plots the DA, or the difference in speech intelligibility between DIR and OMNI modes, obtained by Walden et al. (2005). Also shown are the mSTMI differences formed by subtracting the $\text{mSTMI}_{\text{OMNI}}$ from the $\text{mSTMI}_{\text{DIR}}$ at each SNR value. To facilitate comparisons between objective and behavioral data, the difference scores (i.e., DIR-OMNI) from both the present study and the Walden et al. study were normalized so that the maximum difference (in either percent correct or mSTMI) was set to one.

One obvious similarity between objective (mSTMI) and behavioral data is that all DIR-OMNI difference scores are positive across the range of SNR's tested. In other words, when speech was presented from the front, speech intelligibility was always better, and subjects always preferred the DIR processing mode, even in relative quiet (Walden et al., 2005). Likewise, acoustic analyses showed that the DIR processed signals had the higher mSTMI for all SNR's. However, the peak in the average behavioral DA occurred at −3 dB SNR, whereas the peak in the mSTMI DA occurred at +3 dB SNR. In interpreting these differences, it is important to remember that the relationship between the mSTMI (or any other intelligibility index) and speech recognition scores depends to a large degree on the speech materials. As noted in the ANSI (1969) standard for calculating the AI (ANSI, Fig. 15), an AI of 0.4 results in recognition scores ranging from 100% correct for a small, closed set of phonetically balanced words to approximately 52% correct for a large set of nonsense syllables. These different mappings between intelligibility index and speech recognition score are due primarily to nonauditory factors in speech processing, such as lexical and contextual effects on word recognition, as well as ceiling and floor effects. In the case of the Walden et al. data, scores for the hearing-impaired subjects were limited by their speech-in-quiet recognition, which was roughly 85% correct. The point at which these subjects' average speech recognition scores approached asymptote was at an SNR of about 3 to 6 dB for the DIR condition and about 9 dB for the OMNI condition. At SNR's more favorable than these, the hearing-impaired subjects seemed not to be able to benefit from any further improvements in SNR. However, the mSTMI continues to register improvement in signal quality as the SNR is systematically increased from +3 to +15 dB SNR and beyond. Thus, the strength of the relation between the mSTMI advantage and the behavioral DA will depend on whether the OMNI or DIR speech recognition data have reached asymptotic performance.

To look at this issue more closely, the behavioral DA in percent correct and the objective DA in the mSTMI units, must be converted to the same unit of measure. This was done by converting the mSTMI scores shown in Figure 3 to percent correct and then computing the DIR-OMNI difference scores. To accomplish this, the 22 average percent correct scores (from Walden et al., 2005; 11 SNR conditions × 2 microphone processing modes) were plotted as a function of mSTMI and fitted with a logistic regression. The fitted curve was then used to convert the mSTMI scores to predicted percent correct. The results of this conversion are shown in Figure 5.

The top panel of Figure 5 shows the different percent correct scores obtained for the 22 test conditions as a function of mSTMI. The fitted logistic equation ($r^2 = 0.98$) was used to compute a percent correct value for a given mSTMI. Differences between OMNI and DIR percent correct estimates were then plotted as a function of SNR to construct a predicted DA analogous to the traditional DA obtained in behavioral studies. As can be seen in the bottom panel of Figure 5, the objective and behavioral DA now look quite similar showing peak differences between the two microphone modes at SNRs between −3 and 0 dB.

Walden et al. (2005) found that across individual subjects there was considerable variability in the SNR that produced the maximum DA. For example, one subject demonstrated a DA of 36% at an SNR of −3 dB and a DA of 6% at an SNR of +3 dB. In contrast, a different subject demonstrated a DA of 5% at an SNR of −3 dB and a DA of 35% at an SNR of +3 dB. In other words, at a particular SNR, some

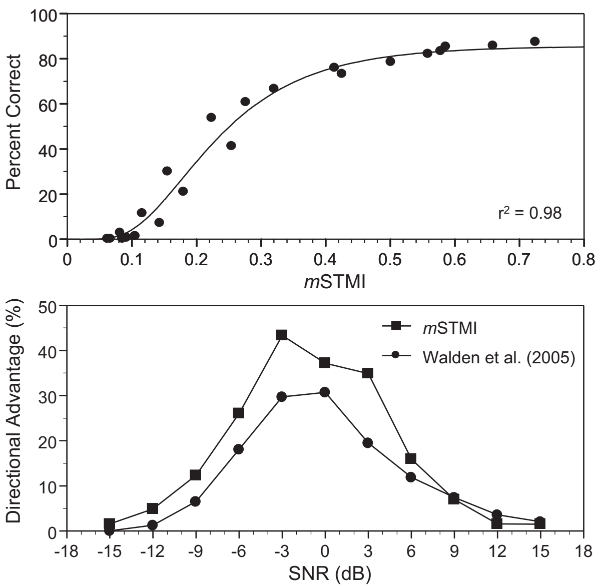

Fig. 5. (Top) mSTMI values expressed as percent correct speech recognition scores based on average data from Walden et al. (2005).The 22 SNR conditions (11 OMNI, 11 DIR) were fit with a logistic function. (Bottom) The predicted mSTMI directional advantage (DA) is shown along with the average behavioral DA from Walden et al. (2005).

subjects were able to benefit from DIR processing whereas other subjects were not. These differences in individual performance are not hard to understand once we consider the individual functions relating percent correct speech recognition to SNR. Subjects who demonstrate a significant DA for negative SNR's (and nearly no DA for more positive SNR's) tend to show a very steep increase in percent correct for SNR values between −12 and −3 dB, followed by a much more gradual increase in intelligibility for all subsequent improvements in SNR. In contrast, subjects who demonstrate a significant DA at more positive SNR's tend to show a more gradual increase in percent correct as a function of SNR extending out to much more favorable listening conditions before reaching an asymptote at their maximum speech recognition performance in quiet. Thus, it is possible to demonstrate excellent agreement between the behavioral and objective DA for individual subjects just as was done for the average across hearing-impaired subjects shown in Figure 5. To fit individual data, all that is required is to repeat the steps taken in Figure 5 using individual subject functions relating percent correct and mSTMI rather than the average function.

Experiment 2: Relation Between Objective and Subjective Microphone Preferences in Real-World Acoustic Environments: A Pilot Study

The data discussed thus far are based on laboratory measures of DA using fixed spatial positions for speech and noise sources and noises that are steady and unmodulated. However, the kinds of noise interference encountered in everyday listening environments are often much more complex, temporally dynamic, and quite often composed of competing speech (Cherry, 1953). The data from Walden et al. (2005) are further limited because there were no conditions where OMNI processing was better than DIR processing. The question posed in this second, brief pilot experiment is whether the direct comparison approach to automatic microphone selection based on a metric such as the mSTMI is as robust and informative in more complex sound environments as in relatively pristine laboratory conditions.

Acoustic Recordings • To further test the potential value of the mSTMI measure as a means for automatically predicting microphone preferences, additional recordings using the modified Canta 7 hearing aid were obtained in more realistic listening environments (Grant, et al., 2006; Walden, et al., 2007). These included male and female speakers positioned in front and to the sides and back of the listener having conversations in settings such as a hospital cafeteria, hospital lobby, moderately reverberant lunch room, next to an outdoor fountain, and seated in a small conference room. These sound environments were much more complex and varied than the previous laboratory samples that were composed of male and female speakers in motion with respect to each other. In addition, many of the recordings from this set were made in environments where events were somewhat unpredictable. Street noise could increase and decrease as could noise from wind and leaves. Therefore, the spectrum of the noise and the short-term SNR varied over the duration of the recording. Overall, the selection of the different sound environments was chosen to accomplish certain specific goals, namely a unanimous consensus preference for one microphone mode over the other, or ambiguity between microphone modes ("no preference"). Further, there had to be some stability over the course of 20 to 40 sec of recording to perform the planned acoustic analyses. If there were loud, abrupt changes in the environmental background sound, these recordings were eliminated.

The hearing aid was programmed to provide gain and compression characteristics in accordance with the audiogram + fitting algorithm of the Aventa 1.2 software with full compensation for the normal low-frequency roll-off in the DIR mode. The binaurally averaged hearing threshold data from 17 hearing-impaired subjects tested by Walden et al. (2004) were used to program the aid.

A total of 12, approximately 2-min recordings were made, alternating between OMNI and DIR modes every 10 sec. From these original 12 recordings, laboratory test materials were created, which were approximately 60 sec long, alternating four times between OMNI and DIR modes. For each of the original master recordings (representing approximately 2 min of a single listening environment), six different stimuli were created reflecting different time slices through the master recording (e.g., beginning, middle, and end). This was done to investigate the stability of a given listening situation with regard to eventual microphone preference decisions. A total of 72 test stimuli were constructed, with half of the test recordings starting with OMNI and half starting with DIR.

From these 72 stimuli, 12 were selected for further analyses using the mSTMI approach. These 12 stimuli were selected after being played monaurally to five normal-hearing subjects using an insert earphone (Eartone 3A) and being judged unanimously as representing four clear examples of where OMNI processing was preferable, four clear examples where DIR processing was preferable, and four clear examples where neither microphone mode was preferred. The primary question addressed by this experiment was whether the mSTMI analyses for these 12 recordings would correspond to the preference judgments given by the normal-hearing subjects.

mSTMI Analysis • Recorded speech samples alternated roughly every 10 sec between OMNI and DIR modes. These alternating segments were accumulated into two, approximately 20-sec sound samples (one for OMNI and one for DIR) before being subjected to the mSTMI analysis. The mSTMI analysis of the recorded materials was the same as described above for the laboratory measures. Results are shown in Figure 6.

As seen with the laboratory recordings, the mSTMI results for these more complex recordings were consistent with preference judgments made by normal-hearing subjects for those stimuli that were judged to have a preference. In comparing OMNI-preferred and DIR-preferred recordings, all of the mSTMI differences were in the expected direction (OMNI mSTMI values greater than DIR for OMNI preferred sites and DIR mSTMI values greater than OMNI for DIR preferred sites). The "no preference" recordings, however, showed more variability and some unexpected values, deviating sometimes substantially from a small (approximately zero) expected DIR-OMNI difference.

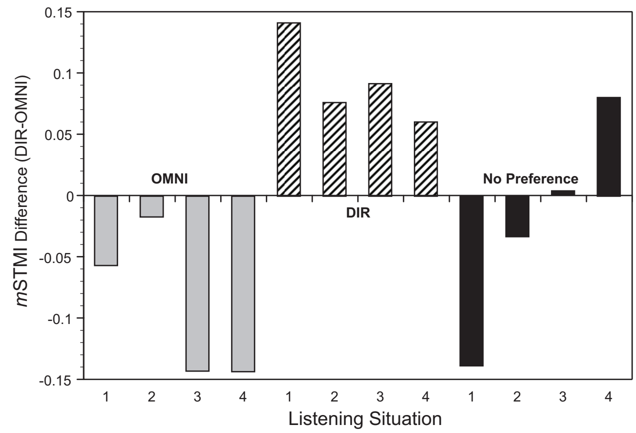

Fig. 6. mSTMI differences (DIR-OMNI) for 12 recordings made in complex, real-world listening environments. Four of the recordings were unanimously judged to be OMNI preferred (OMNI), 4 unanimously judged to be DIR preferred (DIR), and 4 were unanimously judged as no preference. Judges were five normally hearing adults.

DISCUSSION

The use of objective acoustic analyses on the processed output of hearing aids for the purpose of optimizing the choice between two or more potential processing algorithms for any given listening environment seems promising. In the current context, objective measures of how well the DIR or OMNI processed signal matched the joint spectral and temporal modulations inherent in clean speech were mostly consistent with subjects' preferences regarding the two microphone modes. This was true regardless of whether the recordings were made in the laboratory under relatively pristine conditions or in the real world under more naturalistic conditions with speech and noise dynamically changing over time. These measures can be made in near real-time in the background without requiring any input from the hearing aid user. If the analysis suggests that one signal is closer than the other to the spectrotemporal modulations found in clean, undistorted speech, that signal would be selected and passed on to the hearing aid receiver for the patient to hear. There seem to be several advantages to this direct comparison approach for switching between hearing aid processing options. First, the method is general and can be applied to any signal processing algorithm where the device must select among a number of processing options. Second, there is no need to estimate the details of the listening environment, such as how many sources there might be, which sources are speech and which are noise, the distance of these sources from the listener, or whether the environment is reverberant or not. And third, rules for interpreting changes in the acoustic environment so that decisions regarding which algorithm should be used need not be made. Instead, decisions are based completely on a direct comparison between the processed signal and a generic template of clean speech. Further, thresholds can be set such that differences in mSTMI values would have to exceed a particular level before any switching is implemented. Such thresholds might be desirable if it were to be determined that patients prefer to remain in OMNI mode when there is little or no preference between OMNI and DIR, as reported earlier by Walden et al. (2004), or if the frequency with which the device switches between two processing modes needs to be regulated if switching between modes is distracting to the wearer.

The mSTMI is an approach that can be used to monitor and analyze hearing aid processing for the purpose of determining the better algorithm to apply in any given listening environment at any given time. Other analysis tools, such as the STI and AI can also be used, provided that the analysis can proceed without requiring a sample of clean speech for making comparisons with the corrupted speech signal. In other words, some form of stored clean-speech template needs to be used. Thus far, procedures for estimating the STI or AI from canonical representations of clean speech have not been developed.

For the mSTMI method to be used effectively as the basis for deciding on the best speech processing strategy for a given listening environment, rules would have to be developed that take into account the magnitude of the mSTMI values and their differences across the various algorithms under consideration. For example, how large should the difference between one algorithm and another be before deciding to implement the preferential processing mode? Naturally, behavioral data sets showing preferences for various processing strategies would be consulted in establishing such rules, and given the likelihood of large individual differences regarding the preference for one or another algorithm, individual data might be required (Walden, et al., 2007). In experiment 1, the data provided by Walden et al. (2005) were consulted for the current project. However, in these laboratory test environments, structured to favor DIR processing (noise present, signal front and near), a DA was observed across a broad range of SNRs. Because Walden et al. did not test conditions that led to OMNI preferences or to "no preference" judgments, it was not possible to establish a definitive set of switching rules. All mSTMI values computed with speech emanating from the front loudspeaker suggested a DA. Likewise, all behavioral judgments collected by Walden et al. showed a preference for the DIR microphone mode. Further tests of the mSTMI are required under conditions more similar to real-world settings where the OMNI mode is often preferred. Preliminary results of this type (experiment 2) demonstrated fairly good agreement between mSTMI values and subjective preferences. Listening environments that resulted in a judgment of "no preference" were more difficult to capture using the current objective approach. Work is currently underway with a group of hearing-impaired subjects to determine whether the mSTMI is a good predictor of how subjects might respond to various hearing aid algorithms (such as OMNI/DIR processing, noise reduction). In this regard, it is important to remember that the current mSTMI procedure is based on normal auditory processing and not on impaired auditory processing. As such, all of the spectrotemporal modulations observed in the clean-speech template or in the specific signal under test are included in computing the mSTMI. It is quite possible that not all modulations in the clean-speech signal are informative or predictive of various outcome measures for hearing-impaired listeners because of a loss of audibility and a loss of spectral and temporal acuity that result from the hearing loss. Thus, it remains to be seen how robust the preferences are for different signal processing strategies across listeners with and without hearing loss (Walden, et al., 2007). If hearing-impaired individuals show distinctly different preferences than do normally hearing individuals, the auditory processing model used for the mSTMI analysis may have to be changed to reflect impaired processing. On the other hand, it may be the case that listeners, regardless of hearing status, rank processed speech signals similarly (e.g., signal A is better than signal B) even though the intelligibility of a given signal may be much worse for impaired listeners.

CONCLUSIONS

This study evaluated a means for selecting automatically between hearing aid signal processing strategies designed to improve the transmission of speech information. The decision as to which of a set of possible speech signals should be delivered to the listener's ear is based on which signal comes closest to matching a clean-speech template. This decision is made in a direct comparison mode rather than a scene analyses mode by comparing both the OMNI and DIR processed signals to a clean-speech template. The current process operates on approximately 10 sec of signal and is quite general in that it does not need to know anything about the target speech signal before it was corrupted by noise and/or reverberation.

In the first experiment, mSTMI calculations were performed on speech in a diffuse, stationary noise coming from a variety of different spatial locations and processed by either OMNI or DIR microphone modes. The OMNI microphone mode yielded the better fidelity speech signal in noise whenever the speech source was not in front of the listener. These objective predictions were well matched to the behavioral data reported by Walden et al. (2005) demonstrating the benefits of DIR processing for both intelligibility and preference when speech is presented from the front in a diffuse noise background. In a second experiment using much more varied and complex sound environments, the mSTMI also did very well in matching the preferences of a panel of five normal-hearing judges. These complex environmental sound recordings were analyzed in the exact same manner as were the more controlled laboratory recordings and compared with the same clean-speech template as before. The results were very promising in that the current version of the mSTMI was able to identify accurately the preferred processing scheme in all cases where there was a clear microphone preference. Work is underway to expand these tests with a much larger data set to determine the robustness of the direct comparison approach outlined in this study (Walden, et al., 2007).

ACKNOWLEDGMENTS

We thank GN ReSound for their support of the earlier study by Walden et al. (2005) and Oly Dyrlund and Andrew Dittberner from GN ReSound for their technical help in modifying the CANTA 770D hearing aids 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.

This work was partially supported by a grant from AFOSR to the University of Maryland and by a Cooperative Research and Development Agreement from GN ReSound.

Parts of the work reported here were presented at the annual convention of the American Academy of Audiology, April 2005, Washington, DC.

Address for correspondence: Ken W. Grant, PhD, Army Audiology and Speech Center, Walter Reed Army Medical Center, 6900 Georgia Avenue, N.W., Washington, DC 20307-5001. E-mail: grant@tidalwave.net.

Received August 14, 2006; accepted August 23, 2007.

REFERENCES

ANSI. (1969). *ANSI S3.5–1969. American national standard methods for calculation of the speech intelligibility index*. New York: American National Standards Institute.

Blamey, P. J., Fiket, H. J., & Steele, B. R. (2006). Improving speech intelligibility in background noise with an adaptive directional microphone. *J Am Acad Audiol, 17*, 519–530.

Bronkhorst, A. W., & Plomp, R. (1990). A clinical test for the assessment of binaural speech perception in noise. *Audiology, 5*, 275–285.

Buchler, M. C. (2002). *Algorithms for sound classification in hearing instruments*. PhD Thesis, Swiss Federal Institute of Technology, Zurich, Switzerland. ETH No. 14498: 1–138.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am, 25*, 975–979.

Chi, T., Gao, M., Guyton, M. C., et al. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am, 106*, 2719–2732.

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am, 118*, 887–906.

Chung, K. (2004). Challenges and recent developments in hearing aids: Part I. speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends Amplif, 8*, 83–124.

Cord, M. T., Surr, R. K., Walden, B. E., et al. (2002). Performance of directional microphone hearing aids in everyday life. *J Am Acad Audiol, 13*, 295–307.

Dhar, S., Humes, L. E., Calandruccio, L., et al. (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.

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am, 95*, 2670–2680.

Dubno, J. F., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am, 76*, 87–96.

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

Fabry, D. A. (2006). In one ear and synchronized with the other: automatic hearing instruments under scrutiny. *Hear Res,* 13, 36–38.

Garofolo, J. (1988). *Getting started with the darpa timit cd-rom: an acoustic phonetic continuous speech database*. Gaithersburgh, MD: National Institute of Standards and Technology (NIST).

Goldsworthy, R. L., & Greenberg, J. E. (2004). Related articles, analysis of speech-based speech transmission index methods with implications for nonlinear operations. *J Acoust Soc Am, 116*, 3679–3689.

Grant, K. W., Summers, V., Elhilali, M., et al. (2006). Predicting hearing aid microphone preferences in normal and hearing-impaired adults. *Am Audit Soc,* 31, 26.

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

Houtgast, T., & Steeneken, H. J. M. (1985). A review of the MTF-concept in room acoustics. *J Acoust Soc Am, 77*, 1069–1077.

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

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

Mesgarani, N., & Shamma, S. (2005). Speech enhancement based on filtering the spectrotemporal modulations. *Proc Int Conf Acoustic Speech Signal Processing, 1*, 1105–1108.

Mesgarani, N., Slaney, M., & Shamma, S. (2006). Discrimination of speech from non-speech based on multiscale spectro-temporal modulations. *IEEE Trans Audio Speech Lang Processing, 14*, 920–930.

Palmer, C., Bentler, R., & Mueller, H. G. (2006). Evaluation of a second-order directional microphone hearing aid: II. self-report outcomes. *J Am Acad Audiol, 27*, 190–201.

Payton, K. L., & 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., et al. (2002). Computing the STI using speech as a probe stimulus. In *Past, present and future of the speech transmission index* (pp. 125–138). Soesterberg, The Netherlands: TNO Human Factors.
Payton, K. L., Uchanski, R. M., & 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.
Ricketts, T., & Henry, P. (2002). Evaluation of an adaptive, directional microphone hearing aid. *Int J Audiol, 41*, 100–112.
Ricketts, T. A., & Hornsby, W. Y. (2003). Distance and reverberation effects on directional benefit. *Ear Hear, 24*, 472–484.
Surr, R. K., Walden, B. E., Cord, M. T., et al. (2002). Influence of environmental factors on hearing aid microphone preference. *J Am Acad Audiol, 13*, 308–322.
Valente, M., Fabry, D. A., & Potts, L. G. (1995). Recognition of speech in noise with hearing aids using dual microphones. *J Am Acad Audiol, 6*, 440–449.
Walden, B. E., Surr, R. K., Cord, M. T., et al. (2004). Predicting hearing aid microphone preference in everyday listening. *J Am Acad Audiol, 15*, 365–396.
Walden, B. E., Surr, R. K., Cord, M. T., et al. (2007). The robustness of hearing aid microphone preferences in everyday listening environments. *J Am Acad Audiol, 18,* 358–379.
Walden, B. E., Surr, R. K., Grant, K. W., et al. (2005). Effect of signal-to-noise ratio on directional microphone benefit and preference. *J Am Acad Audiol, 16*, 662–676.