J Am Acad Audiol 19:708–720 (2008)

Evaluation of a "Direct-Comparison" Approach to Automatic Switching in Omnidirectional/Directional Hearing Aids

DOI: 10.3766/jaaa.19.9.6

Van Summers*
Ken W. Grant*
Brian E. Walden*
Mary T. Cord*
Rauna K. Surr*
Mounya Elhilali†

Abstract

Background: Hearing aids today often provide both directional (DIR) and omnidirectional (OMNI) processing options with the currently active mode selected automatically by the device. The most common approach to automatic switching involves "acoustic scene analysis" where estimates of various acoustic properties of the listening environment (e.g., signal-to-noise ratio [SNR], overall sound level) are used as a basis for switching decisions.

Purpose: The current study was carried out to evaluate an alternative, "direct-comparison" approach to automatic switching that does not involve assumptions about how the listening environment may relate to microphone preferences. Predictions of microphone preference were based on whether DIR- or OMNI-processing of a given listening environment produced a closer match to a reference template representing the spectral and temporal modulations present in clean speech.

Research Design: A descriptive and correlational study. Predictions of OMNI/DIR preferences were determined based on degree of similarity between spectral and temporal modulations contained in a reference, clean-speech template, and in OMNI- and DIR-processed recordings of various listening environments. These predictions were compared with actual preference judgments (both real-world judgments and laboratory responses to the recordings).

Data Collection and Analysis: Predictions of microphone preference were based on whether DIR- or OMNI-processing of a given listening environment produced a closer match to a reference template representing clean speech. The template is the output of an auditory processing model that characterizes the spectral and temporal modulations associated with a given input signal (clean speech in this case). A modified version of the spectro-temporal modulation index (mSTMI) was used to compare the template to both DIR- and OMNI-processed versions of a given listening environment, as processed through the same auditory model. These analyses were carried out on recordings (originally collected by Walden et al, 2007) of OMNI- and DIR-processed speech produced in a range of everyday listening situations. Walden et al reported OMNI/DIR preference judgments made by raters at the same time the field recordings were made and judgments based on laboratory presentations of these recordings to hearing-impaired and normal-hearing listeners. Preference predictions based on the mSTMI analyses were compared with both sets of preference judgments.

Results: The mSTMI analyses showed better than 92% accuracy in predicting the field preferences and 82–85% accuracy in predicting the laboratory preference judgments. OMNI processing tended to be favored over DIR processing in cases where the analysis indicated fairly similar mSTMI scores across the two processing modes. This is consistent with the common clinical assignment of OMNI mode as the default setting, most likely to be preferred in cases where neither mode produces a

---

*Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC; †Institute for Systems Research, University of Maryland, College Park, MD

Van Summers, Ph.D., Army Audiology and Speech Center, Walter Reed Army Medical Center, 6900 Georgia Ave. NW, Washington, DC 20307-5001; Phone: 202-782-8610; Fax: 202-782-9228; E-mail: vsummers6@gmail.com

Portions of this article were presented at the meeting of the American Auditory Society, March 2006, Scottsdale, Arizona.

substantial improvement in SNR. Listeners experienced with switchable OMNI/DIR hearing aids were more likely than other listeners to favor the DIR mode in instances where mSTMI scores only slightly favored DIR processing.

Conclusions: A direct-comparison approach to OMNI/DIR mode selection was generally successful in predicting user preferences in a range of listening environments. Future modifications to the approach to further improve predictive accuracy are discussed.

Key Words: Automatic switching, directional processing, hearing aids, microphone preferences

Abbreviations: AllData = scoring method using OMNI, DIR, and No Preference judgments; DIR = directional; HI = hearing impaired; IEEE = Institute of Electrical and Electronic Engineers; mSTMI = modified spectro-temporal modulation index; NH = normal hearing; NoPref = No Preference; OmDir = scoring method using only OMNI and DIR preference judgments; OMNI = omnidirectional; SNR = signal-to-noise ratio; STMI = spectro-temporal modulation index; WD = weighted difference

Sumario

Antecedentes: Los auxiliares auditivos hoy día aportan opciones de procesamiento tanto direccionales (DIR) como omnidireccionales (OMNI), con el modo activo actual seleccionado automáticamente por el dispositivo. El enfoque más común para este cambio automático involucra un "análisis acústico de la escena" donde se utilizan estimados de varias propiedades acústicas del ambiente de escucha (p.e., tasa de señal-ruido [SNR], nivel de sonido global), como la base para las decisiones de cambio.

Propósito: El estudio actual fue llevado a cabo para evaluar un enfoque alternativo de "comparación directa" al cambio automático, que no exige presunciones sobre cómo el ambiente de escucha puede relacionarse con las preferencias de micrófono. Las predicciones de preferencia de micrófono se basaron en cómo el procesamiento DIR- o OMNI- de un ambiente de escucha dado produjo una equivalencia mayor a una plantilla de referencia que representaba las modulaciones temporales y espectrales presentes en el lenguaje limpio.

Diseño de la Investigación: Estudio descriptivo y de correlación. Las predicciones de las preferencias OMNI/DIR se determinaron con base en la similitud entre las modulaciones espectrales y temporales contenidas en una plantilla de referencia de lenguaje limpio, y en grabaciones procesadas para OMNI- y DIR- en varios ambientes de escucha. Estas predicciones fueron comparadas con los juicios reales de preferencia (tanto juicios del mundo real como respuestas de laboratorio a los registros).

Recolección y Análisis de los Datos: Las predicciones de preferencia de micrófono se basaron en si el procesamiento DIR- u OMNI- de una ambiente de escucha dado producían un equivalente a una plantilla de referencia que representaba lenguaje limpio. La plantilla es la producción de un modelo de procesamiento auditivo que caracteriza las modulaciones espectrales y temporales asociadas con una señal dada de ingreso (lenguaje limpio en este caso). Una versión modificada del índice de modulación espectro-temporal (mSTMI) se usó para comparar la plantilla con las versiones procesadas tanto de DIR- como de OMNI- de un ambiente de escucha dado, conforme fue procesado a través del mismo modelo auditivo. Estos análisis se llevaron a cabo en grabaciones (originalmente recolectadas por Walden y col., 2007) de lenguaje procesado OMNI- o DIR-, producido a partir de un rango de situaciones cotidianas de escucha. Walden y col. reportaron juicios de preferencia OMNI/DIR realizados por calificadores, al mismo tiempo que se hicieron las grabaciones de campo y los juicios basados en las presentaciones de laboratorio de estas grabaciones para sujetos hipoacúsicos y normoyentes. Las predicciones de preferencia con base en los análisis mSTMI fueron comparadas con ambos grupos de juicios de preferencia.

Resultados: Los análisis mSTMI mostraron una exactitud mejor de 92% en predecir las preferencias de campo y una exactitud del 82–85% en predecir los juicios de preferencia del laboratorio. El procesamiento OMNI tendió a ser preferido sobre el procesamiento DIR en casos donde el análisis indicó puntajes MSTMI bastante similares en ambos modos de procesamiento. Esto es consistente con la asignación clínica común del modo OMNI como la configuración por defecto, posiblemente a ser preferida en casos donde ninguno de los modos logra una mejoría sustancial en la SNR. Los sujetos con experiencia con auxiliares auditivos OMNI/DIR intercambiables tuvieron mejor posibilidad de escoger el modo DIR que otros sujetos, en situaciones donde los puntajes mSTMI sólo favorecieron levemente el procesamiento DIR.

Conclusiones: El enfoque de comparación directa en el modo de selección OMNI/DIR tuvo éxito en general para predecir preferencias del usuario en un rango de ambientes de escucha. Se discuten modificaciones futuras al enfoque para mejorar la exactitud de predicción.

Palabras Clave: Cambio automático, procesamiento direccional, auxiliares auditivos, preferencias de micrófono

Abreviaturas: AllData = método de calificación que usa juicios OMNI, DIR o sin preferencia; DIR = direccional; HI = hipoacúsico; IEEE = Instituto de Ingenieros Eléctricos y Electrónicos; mSTMI = Índice modificado de modulación espectral-temporal; NH = audición normal; NoPref = Sin preferencia; OmDir = método de calificación utilizando sólo juicios de preferencia OMNI y DIR; OMNI = omnidireccional; SNR = tasa señal-ruido; STMI = Índice de modulación espectral-temporal; WD = diferencia ponderada

Hearing-impaired (HI) listeners frequently report difficulty understanding speech in noisy or reverberant listening situations (Dubno et al, 1984; Helfer and Wilber, 1990; Kochkin, 1993, 1994). Distortions in internal processing due to hearing loss, including reduced frequency selectivity and loss of peripheral compression, make these listeners particularly susceptible to the negative effects of noise and/or reverberation in the listening environment (Moore, 1995). As a result, HI listeners generally require a better signal-to-noise ratio (SNR) than listeners with normal hearing (NH) in order to show similar speech performance (Plomp, 1978; Suter, 1985). Hearing aids may provide little benefit in these difficult listening situations since they amplify competing background sounds along with the signal of interest, leaving the SNR of the amplified stimulus unchanged relative to the input. In particular, omnidirectional (OMNI) processing, which provides equal amplification to sounds coming from all directions, will not alter the original SNR.

Directional (DIR) processing, on the other hand, can improve the SNR of the amplified signal in some situations. DIR processing can amplify sounds coming from in front of the listener relatively more than sounds coming from other locations (Ricketts, 2001). This will provide benefit when the listener is facing the signal of interest. In situations where the listener is focused on a signal arriving from some other location, DIR processing may reduce the SNR relative to the original input (and relative to omnidirectional processing). In addition, directional processing will not improve the SNR when both signal and competing noise are located in front of the listener or in highly reverberant listening situations where substantial competing energy originating from the side or back arrives from the front via reflections. Finally, directional processing may not provide benefit in quiet environments where the listener is facing the signal of interest, since there is little energy arriving from locations other than the front.

Since, for most HI listeners, directional processing is beneficial in some listening situations and omnidirectional processing is preferable in others (Walden et al, 2004; Walden et al, 2007), current directional hearing aids generally provide both processing modes, and switching between microphone modes is either done manually by the user or automatically by the device based on an analysis of the incoming sound. Under manual switching, the user selects the currently preferred mode by flipping a toggle on the hearing aid or a remote control. A problem with this approach is that listeners may not be aware that a change in microphone mode could be beneficial in a given listening situation if they do not actively switch modes. In addition, listeners may find manual switching and active comparison of the two modes burdensome and inconvenient. As a result, they may leave their devices in the default OMNI mode permanently. Cord et al (2002) estimated that about one-third of listeners fit with manually switchable OMNI/DIR hearing aids tend to leave their instruments in the default OMNI mode regardless of the listening situation. Obviously these patients receive no benefit from the (unused) DIR mode.

The potential problems with manual switching make automatic switching an attractive alternative. Here, analyses of the incoming sound by the hearing aid leads to automatic selection of the current processing mode. Most commercially available hearing aids equipped with both OMNI and DIR processing now provide automatic switching and use a "scene analysis" approach to determine the appropriate microphone mode (Chung, 2004; Fabry and Tchorz, 2005; Blamey et al, 2006; Palmer et al, 2006). In scene analysis, the acoustic environment is sampled and examined for specific acoustic properties likely to favor either OMNI or DIR processing (Walden et al, 2004). The automatic switching decisions made by currently available hearing aids may be accurate enough to make automatic switching preferred over manual switching by many patients (Olson et al, 2004). Nevertheless, a great deal of improvement in these switching decisions may still be possible. Palmer et al (2006) collected field ratings of satisfaction with a current hearing aid that provides automatic mode selection based on scene analysis. The device provided three processing modes: fixed omnidirectional processing, automatic OMNI/DIR switching with a fixed (hypercardioid) polar pattern during directional processing, and automatic switching with an adaptive polar pattern. Forty-nine HI subjects compared the three processing modes in a variety of real-world listening situations. About one-third of the subjects were unable to hear any differences across the processing modes. Of the remaining subjects, approximately half preferred fixed omnidirectional processing over automatic switching most of the time, and half preferred automatic switching (with either fixed or adaptive directionality) more frequently. Presumably the subjects preferring the fixed OMNI mode did so because the switching algorithm selected the less preferred microphone mode in some listening situations.

"Direct-comparison" provides an alternative to scene analysis as an approach to automatic switching. This approach involves comparing DIR- and OMNI-processed versions of the input signal in terms of one or more acoustic properties or relative to some reference template. For example, parallel analyses of DIR- and OMNI-processed versions of a given input may be compared to determine which processing mode is currently producing output having the higher SNR. A recent report from Grant et al (2008) provided preliminary evidence that a direct comparison approach using a version of the spectro-temporal modulation index (STMI) (Elhilali et al, 2003) may support accurate automatic switching between OMNI and DIR modes. In Experiment 2 of Grant et al (2008), recordings of DIR- and OMNI-processed versions of speech produced in different real-world environments were presented to listeners with normal hearing. Experimental stimuli were prepared that alternated between 10 sec samples of each processing mode, presenting each microphone mode twice (OMNI-DIR-OMNI-DIR or DIR-OMNI-DIR-OMNI). Eight stimuli were identified where listeners unanimously preferred either the OMNI- or the DIR-processed portions of the stimulus (four OMNI, four DIR). A modified STMI (mSTMI) analysis of these stimuli accurately predicted subject preferences in all eight cases. That is, for stimuli where OMNI-processed portions were preferred, the mSTMI analyses indicated that these portions were less noisy (more closely approximated a clean speech template) than DIR-processed regions. For stimuli where DIR processing was preferred, the mSTMI analysis indicated that DIR processing produced a cleaner signal. While promising, the preliminary results reported by Grant et al (2008) involved only normal-hearing listeners and a limited number of recordings and real-world environments (the eight recordings represented only four different environments, two recordings per environment). The current study describes a more extensive test of the mSTMI as a predictor of listener preferences for OMNI versus DIR processing, using a larger set of real-world listening conditions and testing both HI and NH listeners.

In a recent report, Walden et al (2007) made recordings of OMNI- and DIR-processed speech produced in a range of everyday listening environments. Walden et al reported preference judgments (OMNI, DIR, or No Preference [NoPref]) made "in the field" at the time the recordings were made and laboratory judgments based on comparisons of the recorded, OMNI- and DIR-processed signals. The laboratory stimuli from Walden et al (2007) were based on monaural field recordings and were presented monaurally in the laboratory. The authors reported close agreement between original field preferences and laboratory preferences based on these recordings. That is, the recordings apparently captured much of the important information influencing the original field preferences. The current article will examine whether mSTMI analyses of these same recordings can be used to accurately predict the preference data reported by Walden et al.

METHOD

Participants

Walden et al (2007) reported preference data from 30 HI and 10 NH subjects. The individual preference data from one NH subject were not available for the present analysis. Therefore, preference data from the 30 HI and 9 NH subjects tested by Walden et al were examined here. Two of the HI listeners also served as field raters who identified real-world listening environments where either OMNI or DIR processing was preferred or where neither microphone mode was clearly preferable (see "Field Recordings" section below). All 39 subjects provided laboratory preference judgments of recorded speech samples collected in the real-world environments.

To be included in the study, HI listeners had to demonstrate a directional advantage of at least 15% (i.e., a 15% improvement in speech recognition performance for DIR processing re OMNI processing) under laboratory testing where speech was presented from the front and competing noise originated from 90°, 180°, and 270° azimuths (see Walden et al, 2007). The 30 HI listeners formed three groups of ten listeners. One group (hereafter, the COHORT group) included the two field raters and eight additional listeners experienced with switchable OMNI/DIR devices. These listeners had extensive experience in making microphone preference judgments in everyday listening situations as a result of participation in an earlier study (Walden et al, 2004). A second group of ten listeners (hereafter, the DIR group) had been fit with manually switchable OMNI-DIR hearing aids prior to enrollment in the study and reported making use of each processing mode in daily living. Although these subjects were experienced users of hearing aids with manually selected OMNI and DIR processing, they were not experienced subjects in making microphone preference judgments. The third group of listeners (hereafter, the OMNI group) were experienced users of

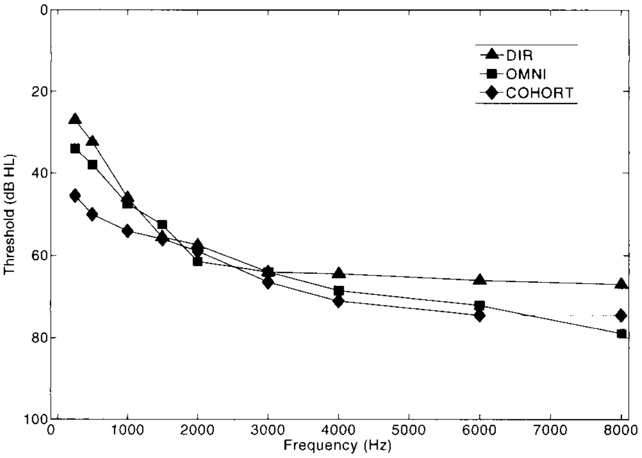

Figure 1. Mean test-ear audiometric thresholds between 250 and 8000 Hz for each HI group.

omindirectional hearing aids who had no prior experience with directional processing. Listeners in the DIR and OMNI groups had not previously participated in studies requiring direct comparisons of OMNI- and DIR-processed signals.

HI listeners had moderate to severe gradually sloping bilateral hearing impairments, all of which fell within the fitting range of the hearing aid used in the field portion of the study (see "Field Recordings" section below). Mean audiograms of test ears for listeners in the COHORT, DIR, and OMNI groups are plotted in Figure 1. In general, mean thresholds were slightly better in the DIR group than in the other two groups and slightly better in the OMNI group than in the COHORT group. Pure tone averages (based on thresholds at 500, 1000, and 2000 Hz) were 45.3 dB HL for the DIR, 49 dB HL for the OMNI, and 54.3 dB HL for the COHORT group. Audiometric thresholds for the three groups were compared in a two-way ANOVA using group and frequency as predictor variables. The results showed significant main effects of both group (F[2,243] = 12.99, $p < 0.0001$) and frequency (F[8,243] = 50.86, $p < 0.0001$). The group by frequency interaction was not significant (F[16,243] = 1.41, p = 0.14). Post hoc analyses using the Tukey HSD (honest significant difference) procedure indicated that all three group means significantly differed from one another with the DIR group having better thresholds than the other groups and with the COHORT group having the poorest thresholds. The nine NH listeners had air conduction thresholds of 15 dB HL or less at the audiometric frequencies between 250 and 6000 Hz.

Field Recordings

The methods used in making the recordings of real-world listening situations are described in detail in Walden et al (2007). In brief, each field rater was fit with a modified version of the manually switchable Canta 770D hearing aid (GN Resound Corporation). This modified aid allowed for direct recording of the DIR- and OMNI-processed versions of the hearing aid input just prior to processing by the hearing aid receiver. Each field rater identified two real-world listening situations where DIR processing was preferred over OMNI, two environments where OMNI processing was preferred and two where neither processing mode was clearly preferable. Thus a total of 12 real-world listening environments were recorded (3 preference categories × 2 listening situations per category × 2 field raters). Field recordings were two minutes long and alternated between OMNI and DIR processing approximately every 10 sec. General descriptions of each recording site along with talker and noise characteristics (talker and noise location, talker distance, etc.) appear in Table 1 (table 2 from Walden et al, 2007).

Preparation of Laboratory Stimuli

Prior to editing, each field recording was processed with software simulating the Canta 770D receiver. As described by Walden et al (2007), each two-minute recording was then edited to produce six overlapping 40 sec stimuli representing a given listening environment (12 listening situations × 6 stimuli per situation = 72 total stimuli). These 40 sec stimuli alternated between DIR and OMNI processing every 10 sec (D-O-D-O or O-D-O-D). The original field recordings also included the tones produced by the hearing aid to indicate which processing mode was currently active. Samples of these recorded tones were extracted and used to mark transitions between processing modes in the laboratory stimuli, with the first and third intervals always preceded by a single tone and the second and fourth intervals by two tones. That is, for each stimulus, the tone sequence was always 1-2-1-2. Given that half of the stimuli began with DIR processing and half began with OMNI, the number of tones did not provide a cue to processing mode.

Collection of Laboratory Preferences

Laboratory preference judgments were collected in a sound-treated booth. Laboratory stimuli were presented monaurally via an insert receiver at a comfortable listening level determined individually for each listener. Listening levels were determined using concatenated IEEE (Institute of Electrical and Electronic Engineers, 1969) sentences recorded through the test hearing aid while mounted on a Knowles Electronics Mannequin for Acoustic Research (KEMAR). Subjects

Table 1. Characteristics of the 12 Everyday Listening Environments Used for Recordings

	Talker						Background Noise			
	Location			Distance (feet)			Location			
Recording Site	Front	Side	Back	<3	3–10	>10	Front	Back	All Around	Loudness
				OMNI-Preferred Sites						
Outdoor Garden			x		x				x	Moderate
Car (passenger)			x	x					x	Soft
Office (10′ × 12′)		x		x				x		Moderate
Hospital Cafeteria		x		x					x	Loud
				DIR-Preferred Sites						
Conference Room (14′ × 18′)	x					x		x		Moderate
Hospital Cafeteria	x			x					x	Soft
Outside of hospital	x			x					x	Soft
Hospital Lobby	x			x					x	Loud
				No-Preference Sites						
Lunch Room (13′ × 15′)		x		x			x			Soft
Outdoor Garden		x		x					x	Moderate
Classroom (19′ × 25′)	x					x	x			Loud
Conference Room (14′ × 18′)	x				x		x			Soft

Note: The loudness judgments were made by the field raters.

indicated perceived loudness of the IEEE sentences using a loudness scale adapted from the contour test of the Independent Hearing Aid Fitting Forum protocol (Valente and Van Vliet, 1997). Loudness labels ranged from "very soft" (label 1) to "uncomfortably loud" (label 7). A bracketing procedure using 5 dB steps was used to identify the level associated with "comfortable" (label 4). The 72 stimuli were presented at the subject's comfortable listening level in random order. For each stimulus, subjects indicated whether they preferred samples 1 and 3 (preceded by one tone), samples 2 and 4 (preceded by two tones), or had no clear preference. For each stimulus, the processing mode associated with the preferred sample, or a No Preference response, was recorded for each subject. Subjects generally responded after the second or third 10 sec portion of a test item; they seldom needed the entire 40 sec sample to make a preference judgment. The software allowed a given stimulus to be repeated, but this was seldom necessary. The test equipment was calibrated prior to each test session.

mSTMI Analyses

DIR and OMNI portions of each laboratory stimulus were analyzed using a modified version (Grant et al, 2008) of the STMI analysis described by Chi et al (1999) and Elhilali et al (2003). STMI scores range between 0 and 1 and indicate degree of similarity between the spectro-temporal modulations present in a given acoustic input and in a reference standard (clean speech, in this case). The comparison is between outputs of an auditory processing model with several stages representing both peripheral (cochlear) and more central processing. Peripheral stages of the model represent cochlear filtering, hair-cell transduction, and auditory-nerve spectro-temporal sharpening. These stages 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 (density) modulations. The model yields a multidimensional representation of a given signal, reflecting the temporal and spectral modulation content of the signal and its distribution in time and frequency (Fig. 2B). The model was used to build a template of the average spectro-temporal modulation patterns found in natural speech. This "clean speech" template was based on analysis of a large sample of conversational speech produced in quiet by adult speakers. The sample included approximately 100 speakers and included an equal number of males and females (Texas Instruments/Massachusetts Institute of Technology [TIMIT]-database [Garofolo, 1988]). STMI analyses were used to measure the similarity of the modulation content of target speech signals to the clean-speech template by computing the normalized distance between the spectro-temporal modulations in the template and the modulation pattern produced by the test signals. The STMI bears a great deal of similarity to another, more common intelligibility metric known as the STI (Speech Transmission Index). The STMI differs fundamentally from the STI in its sensitivity to joint spectro-temporal 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

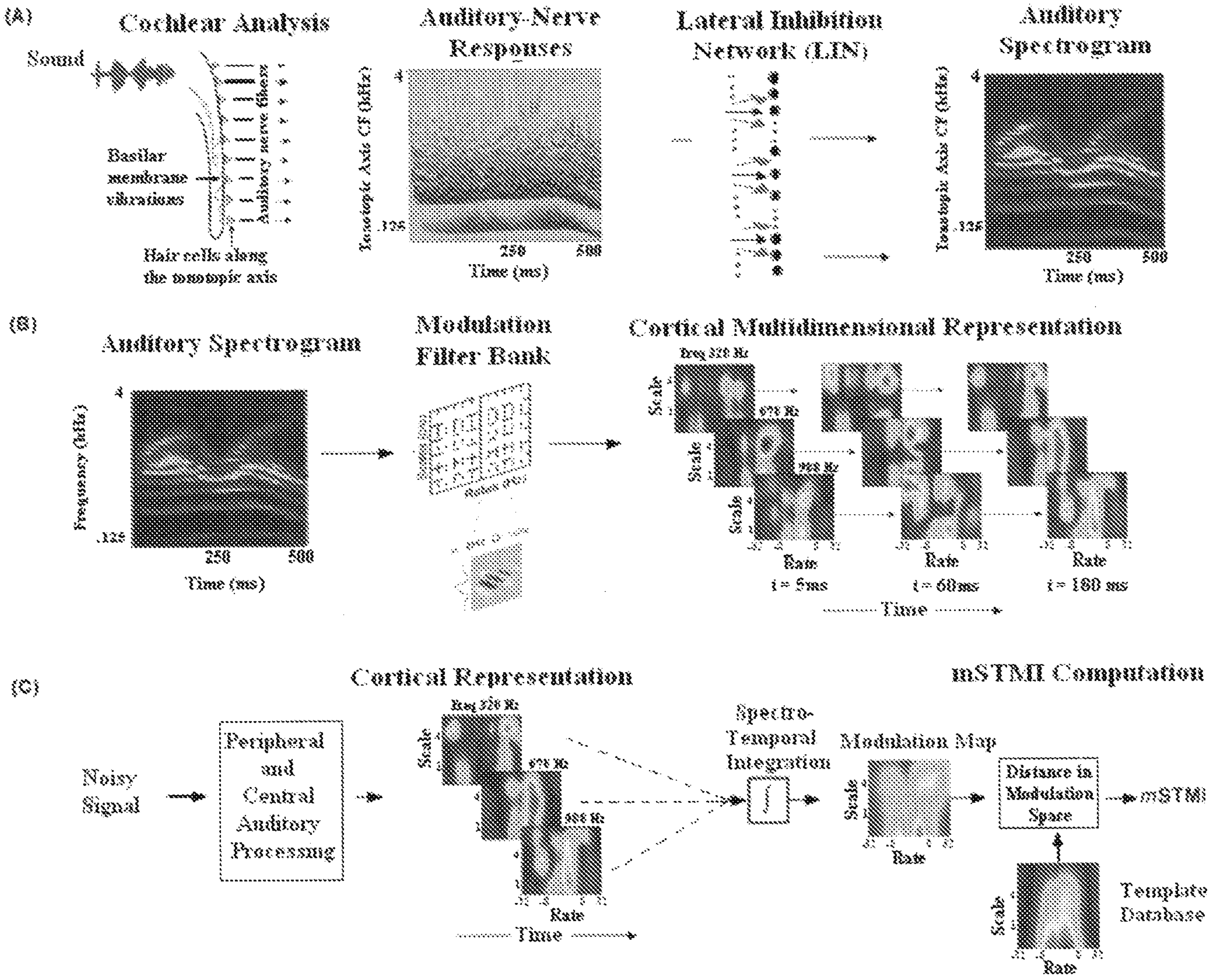

Figure 2. Processing stages for mSTMI calculations. *(A)* Early processing stages. The incoming acoustic signal is analyzed by a bank of 128 constant-Q filters (cochlear filtering stage). The output of each filter is processed by a hair cell model followed by a lateral inhibitory network, rectified, and integrated to produce the auditory spectrogram. *(B)* The auditory spectrogram is 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 along with the result of convolving it with the auditory spectrogram. *(C)* Schematic showing steps in computing the STMI. Reference template formed by processing clean speech through the model. To evaluate a particular speech sample (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 collapsed over tonotopic frequency and time) that is distorted relative to the clean speech template. The mSTMI reflects the normalized distance between template and test signal, weighted by the clean speech template. (Figure adapted from Grant et al (2008); see Chi et al (1999), Elhilali et al (2003), and Grant et al (2008) for further description of STMI model and mSTMI modification.)

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 noisy target signal is processed as described above 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 to a similarly reduced template, and the normalized distance between target and template is interpreted as a measure of similarity to clean speech.

Grant et al (2008) described three significant modifications to the original STMI calculations to

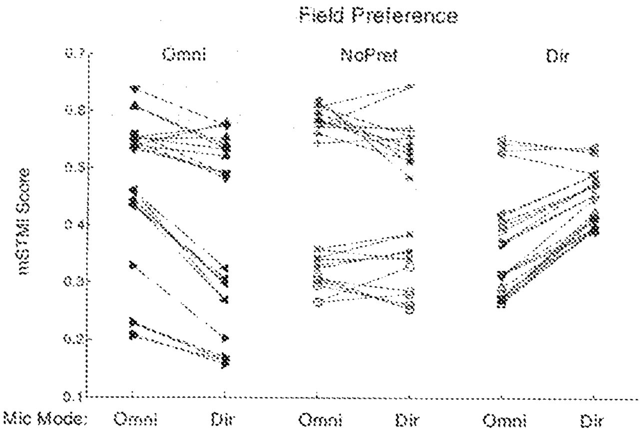

Figure 3. mSTMI scores for OMNI- and DIR-processed portions of the 72 stimuli. Scores for DIR and OMNI portions of a given stimulus are connected by dashed lines. Scores for stimuli from each field-preference category are separated along the abscissa, and scores for the same listening situation (i.e., scores for stimuli based on the same original field recording) are represented by the same symbol.

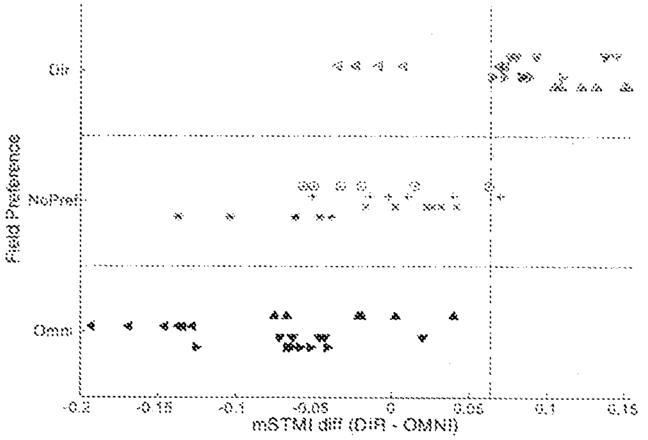

Figure 4. Field preferences as a function of mSTMI difference scores (DIR-OMNI). Scores for the same listening situation (i.e., scores for stimuli based on the same original field recording) are represented by the same symbol. The dashed line at 0.063 indicates the optimal cutoff score for preference prediction using either the OmDir or AllData scoring method.

produce the mSTMI measure used in their study and here. The modifications: (1) extended the clean speech template to more accurately represent natural variability in normal speech (including gender, speaking styles, content); (2) confined the comparison between the template and target signal to spectro-temporal regions that are most relevant in clean speech; and (3) reduced the extent to which spectral shaping introduced by the hearing aid was treated as distortion.

RESULTS

Relating mSTMI Scores to Field Preferences

Prior to examining how well the mSTMI scores predicted laboratory preference judgments, the scores were compared with the original field preferences (provided by the two field raters). The mSTMI scores for the DIR and OMNI portions of each of the 72 stimuli are plotted in Figure 3.

The connected points in the left-hand portion of Figure 3 represent listening situations where OMNI processing was preferred in the field. For these stimuli, nearly all of the lines show a negative slope, indicating higher mSTMI scores for OMNI-processed portions of the stimuli than DIR regions. Conversely, in the right-hand part of the figure, representing situations where DIR processing was preferred in the field, almost all of the lines have a positive slope (higher mSTMI scores for DIR than OMNI processing). For stimuli receiving NoPref field judgments (central portions of the figure), mSTMI scores were often fairly similar across processing modes, and the slopes of the lines are generally fairly flat. There was considerable overlap in mSTMI scores across the three field preference categories. As a result, absolute mSTMI scores for either the OMNI or DIR portions of the stimuli (or average mSTMI scores across modes) clearly could not be used to directly predict preferences.

However, DIR-OMNI difference scores, based on these absolute scores, may allow predictions of preferences. The original field preference for each stimulus is plotted as a function of mSTMI difference score in Figure 4. The ordinate of the figure is divided into three separate regions, representing the three preference categories. Assuming that the processing mode producing the higher mSTMI score will be preferred, difference scores were expected to be positive when DIR processing was preferred, negative when OMNI processing was preferable, and near zero for NoPref stimuli. This expected pattern was generally borne out: most of the OMNI-preference stimuli (lower third of the figure) produced difference scores less than −0.05 (M = −0.072), most of the DIR-preference stimuli (upper third) had scores greater than 0.05 (M = 0.084), and most of the NoPref stimuli had scores between −0.05 and 0.05 (M = −0.02). A Kruskal-Wallis one-way ANOVA by ranks on the difference scores allowed a high level of confidence that scores across the three groups did not represent the same underlying population (H[2] = 305.7, $p < .00001$).

Accuracy of the mSTMI difference scores in predicting field preferences was examined using two different scoring methods. Both involve setting a criterion for the DIR-OMNI cutoff score and predicting preferences based on comparing measured scores with the cutoff score. Since NoPref is not a selectable hearing aid mode, the most appropriate processing mode for stimuli judged NoPref is somewhat unclear. Therefore, the first scoring method was limited to only the OMNI and DIR preference data (OmDir method). Using this

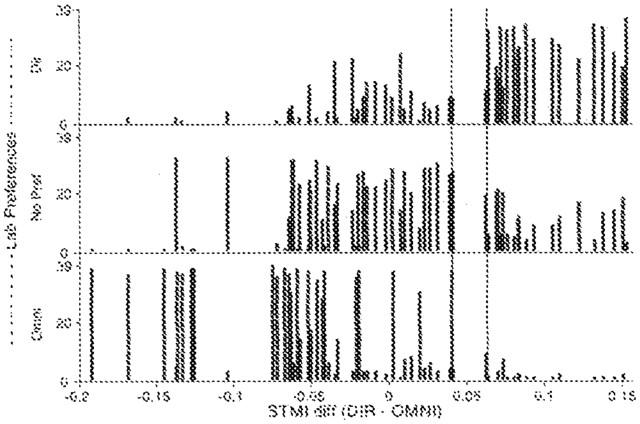

Figure 5. Number of OMNI, NoPref, and DIR preference judgments as a function of mSTMI difference scores (DIR-OMNI) for the 72 lab stimuli. Symbols plotted in the upper panel represent DIR preference judgments. Middle and lower panels represent NoPref and OMNI responses, respectively. For a given stimulus, the total number of preference ratings across panels equals 39. Solid and dashed vertical lines indicate optimal cutoff scores for OmDir and AllData scoring methods.

method, and setting the cutoff score at 0.0, 21 of 24 OMNI-preference stimuli and 21 of the 24 DIR preferences were correctly categorized based on mSTMI differences (87.5% correct overall). Accuracy was improved if the cutoff score was changed to be slightly positive. Using a cutoff of 0.063 (dashed line in Fig. 4), all 24 OMNI-preference stimuli were correctly categorized and 20 of 24 DIR-preference stimuli were assigned correctly (91.7% overall accuracy).

The second method of scoring included the NoPref stimuli in the analysis (AllData method). Under this method, NoPref stimuli were treated as correctly categorized when assigned an OMNI preference. The rationale used was that the OMNI mode is widely viewed as the default microphone mode most appropriate for the majority of listening situations and may therefore be the best choice in the absence of a clear DIR preference. The 0.063 cutoff score was again optimal with overall accuracy at 93% (67 of 72 correct). Thus, either scoring method produced better than 90% accuracy in predicting field preferences based on mSTMI difference scores. The next section examines how well the mSTMI results predicted the laboratory preference data.

Relating mSTMI Scores to Laboratory Preferences

Each of the 72 laboratory stimuli received preference ratings from all 39 subjects. The panels of Figure 5 indicate the number of DIR, NoPref, and OMNI judgments for each stimulus, plotted as a function of mSTMI difference score for that stimulus. The upper panel of the figure indicates DIR preferences; the middle and lower panels represent NoPref and OMNI preferences, respectively. Weighted difference (WD) scores were determined for each laboratory preference category by scaling the difference scores for items receiving a given preference response by the number of these responses. For example, consider the item with the largest positive difference score of 0.153 (see Fig. 5). This item received 36 of the 818 DIR preference judgments made across all 72 stimuli. Therefore, the 0.153 difference score was scaled by 36/818 and summed with the weighted difference scores for all other items to determine the WD score for the DIR preference category. WD scores were −0.062, 0.005, and 0.068 for items receiving OMNI, NoPref, and DIR laboratory preference responses. A Kruskal-Wallis one-way ANOVA by ranks provided clear support for differences in WD scores across preference categories ($H[2] = 699868.9$, $p < .00001$).

The OmDir and AllData scoring methods described earlier were used to examine how accurately mSTMI difference scores predicted laboratory preferences. As in the analysis of the field data, for each scoring method, the cutoff score maximizing correct predictions was a positive DIR-OMNI difference. For the OmDir method, a cutoff score of 0.04 (dotted line in Figure 5) led to 85% correct prediction. For the AllData scoring, a cutoff of 0.063 led to 82% accuracy.

These 82–85% accuracy levels, for laboratory ratings, are lower than the 92–93% accuracy seen in prediction of the field ratings. One factor contributing to this reduction is that unlike the field preference data, where each stimulus had a single preference rating, each laboratory stimulus received separate ratings from 39 different subjects. Most of the stimuli (54 out of 72) received some OMNI and some DIR ratings. Using the current approach, a given mSTMI difference score cannot simultaneously predict an OMNI preference for some subjects and a DIR preference for others. As a result, this approach cannot accurately predict preferences with 100% accuracy when subject preferences vary on a single stimulus. Instead, the highest accuracy attainable is 100% accuracy in matching the most frequently selected preference rating for each stimulus. When the laboratory data were rescored using this standard, accuracy levels increased to approximately 93% correct for both the OmDir and AllData scoring methods (optimal cutoff scores were 0.04 and 0.073, respectively).

Effects of Hearing Status on mSTMI Cutoff Scores and Predictive Accuracy

In the analyses described above, a single cutoff score was used as the criterion value to evaluate preferences for all 39 subjects. If, however, optimal cutoff scores varied across subjects, the use of the same criterion

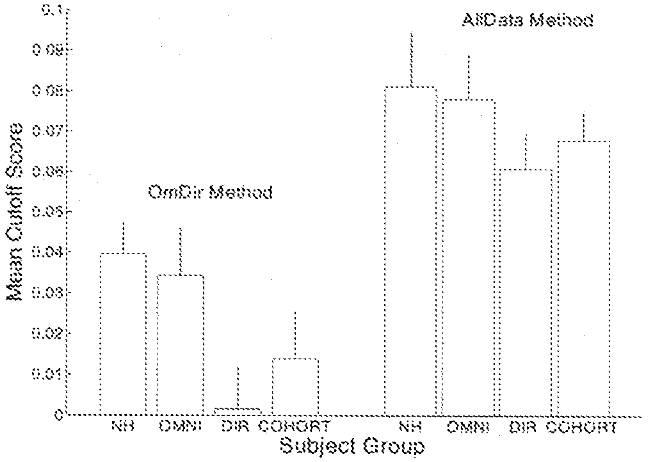

Figure 6. Mean optimal cutoff scores for each subject group under each scoring method. Error bars represent standard errors of means.

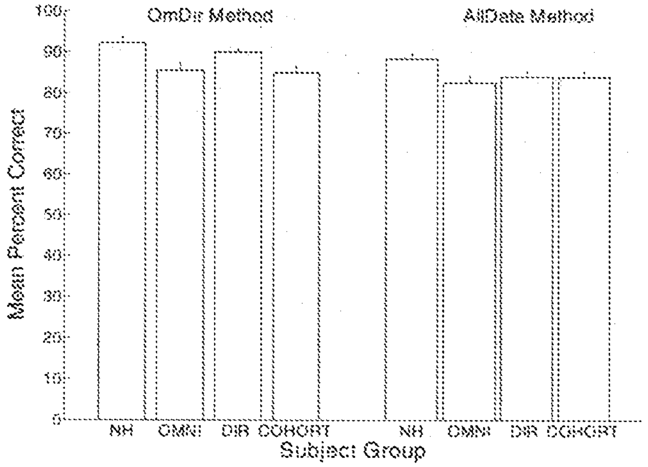

Figure 7. Mean percent-correct prediction scores for each subject group under each scoring method. Error bars represent standard errors of means.

score for all subjects would reduce the predictive accuracy of the analyses. To examine this possibility, optimal cutoff scores were determined for each subject individually, and predictive accuracy was determined based on these individual cutoff values. The use of individually optimized cutoffs had only a small positive effect on mean predictive accuracy with the 85 and 82% accuracies observed originally increasing to 88 and 85% (OmDir and AllData methods, respectively). Predictive accuracy was between 75 and 96% (OmDir method) and between 74 and 100% (AllData method) when optimal cutoff scores were used for each individual listener.

Although these data do not indicate large improvements in predictive accuracy using individually optimized cutoff scores, there was evidence of differences in optimal cutoffs across the four groups of listeners. Figure 6 shows mean optimal cutoff values for the four groups of listeners using each scoring method. Under each method, the NH and Omni groups had similar mean values that were higher than the mean cutoffs for the Cohort and Dir groups. A two-way analysis of variance was carried out on these data, treating group and scoring method as predictors of cutoff scores. The results indicated significant main effects of both group and scoring method ($F[3,35] = 2.91$, $p = 0.048$; and $F[1,35] = 55.39$, $p < 0.00001$, respectively). The group by scoring method interaction was not statistically significant ($F[3,35] = 0.76$, $p = 0.52$). Post hoc tests involving pairwise comparisons of mean cutoff scores for different groups did not indicate statistically significant differences using either the OmDir or AllData method. However, an analysis of the data comparing cutoff scores for subjects experienced with directional processing (the Cohort and Dir subjects) versus subjects with no such prior experience (NH and Omni subjects) indicated significantly higher cutoff scores for the subjects who had not experienced directional processing previously. That is, a two-way ANOVA similar to the original analysis but defining group membership based on experience with directional processing (i.e., NH and Omni subjects versus Cohort and Dir subjects) showed a significant main effect of group on optimal cutoffs ($F[1,37] = 8.07$, $p = 0.007$) and a nonsignificant group by scoring method interaction ($F[1,37] = 1.14$, $p = 0.32$). The higher cutoff scores observed for NH and Omni subjects indicate that these listeners required larger mSTMI differences favoring DIR processing than the other subjects before indicating a DIR preference. This pattern may indicate that experience with DIR-processed signals allowed "acclimatization" to directional processing that increased the benefit provided by this processing and made it more likely to be preferred in certain listening environments (Gatehouse, 1992; Munro and Lutman, 2003).

There were also group differences in how accurately mSTMI difference scores predicted preferences. For both scoring methods, mean percent correct scores were slightly higher for the NH group than for any of the groups of hearing-impaired listeners (see Fig. 7). Given that the auditory model underlying the mSTMI analyses assumed normal, unimpaired auditory processing, greater accuracy in predicting preferences for the NH listeners may not be surprising. A two-way group by scoring method ANOVA, in which group membership was based on hearing status (NH listeners versus hearing-impaired listeners) produced a significant main effect of group membership on predictive accuracy ($F[1,37] = 7.39$, $p = 0.01$). The group by scoring method interaction was not significant ($F[1,37] = 0.07$, $p > 0.75$).

Differences in audiometric thresholds across the three HI groups did not have a clear effect on optimal cutoff scores or the accuracy of preference predictions. That is, the largest group differences in quiet thresholds were between the Dir and Cohort group (see Fig. 1), yet mean cutoff scores and mean percent correct scores were similar for these two groups (Figs. 6 and 7).

DISCUSSION

For most HI listeners, neither OMNI nor DIR processing is likely to be preferred across all listening situations (Ricketts et al, 2003; Walden et al, 2004). Therefore, current hearing aids commonly include both processing modes and provide automatic mode selection, allow manual selection, or supply both of these options. For a large percentage of hearing aid users, automatic selection of the appropriate processing mode is an important advance over manual selection by removing the requirement that the listener continuously monitor the environment and assess the appropriate microphone mode. However, a recent report by Palmer et al (2006) suggests that current approaches to automatic switching might still be greatly improved.

Current implementations of automatic switching can be broadly characterized as using either scene analysis or direct-comparison approaches (Flynn, 2006). The majority of currently available hearing aids providing automatic switching use scene analysis as the basis for microphone mode selections (Chung, 2004). This approach involves sampling and acoustic analyses of the input signal to identify characteristics of the listening environment associated with OMNI or DIR preferences. Scene analysis involves two assumptions: that certain characteristics of the acoustic environment reliably predict preferences and that these characteristics can be accurately monitored via acoustic analyses. The decision rules for automatic switching in current devices primarily focus on the overall level and SNR of the input and whether significant wind noise is present (Chung, 2004). That is, these analyses do not attempt to identify either signal or noise location. Instead, the analyses generally assume that the signal is located in front of the listener. This may lead to an inappropriate microphone mode selection when the signal of interest is to the side or behind the listener (e.g., while driving an automobile and conversing with a passenger).

In contrast to scene analysis, the direct comparison approach does not involve assumptions concerning scene characteristics or their identification. Instead, this approach assumes that whichever processing mode is currently producing output more similar to a desired template (a "clean speech" template in the present case) will be preferred. The current data represent a generally successful initial effort to develop an approach to automatic switching based on direct comparison of DIR- and OMNI-processed signals. Future refinements of the current methods would be likely to improve on this performance.

Based on earlier analyses of the same preference data examined here, Walden et al (2007) concluded that microphone mode preferences for a given stimulus were fairly consistent across HI listeners, suggesting little need for customization of automatic switching algorithms to individual patients. Nevertheless, the differences in optimal cutoff scores across the different groups of HI listeners reported earlier (Fig. 5) suggest some benefit of individually optimizing the criteria for automatic switching. None of the current approaches to automatic switching effectively account for listener-specific differences in microphone preferences. Assuming that individual differences in microphone mode preferences are real and are not incorporated into current algorithms, manual override of the automatic selection should be available to the hearing aid user. This manual override may eventually contribute to the improvement of automatic switching. That is, future hearing aids could incorporate learning algorithms that make use of manually entered preferences to individualize automatic switching decisions (and many other device parameters, see Dillon et al, 2006). These algorithms would presumably assign high decision weights to acoustic/auditory features closely associated with individual preference judgments. The group difference in optimal cutoff scores (Fig. 5) provides a specific example where a manual override and an associated learning algorithm could improve directional switching. Listeners experienced with directional processing were more likely than listeners with no such experience to prefer DIR processing in cases where this microphone mode produced only a small difference in mSTMI scores. To accommodate these differences, a learning algorithm could adjust the decision criteria as experience with (and acceptance of) directional processing increased.

The 93% accuracy of the mSTMI difference scores in predicting the most frequent preference judgment for each laboratory stimulus was not as high as the predictive accuracy of the field preference data. That is, the original field preferences correctly predicted the most frequent laboratory preferences with nearly 100% accuracy using either scoring method (43 of 43 cases using OmDir scoring, 71 of 72 cases for AllData scoring). This indicates that the recorded stimuli contained acoustic cues that allowed very close agreement with field preferences but that some of these cues were not effectively incorporated into the mSTMI-based predictions. Identifying and appropriately weighing these cues may allow modifications to the current mSTMI analyses (or alternative analyses) that produce more accurate preference predictions.

This would not be the case if the critical cues were unavailable in the processed recordings (e.g., cross-modal auditory-visual cues such as timing relationships between visual and acoustic events or binaural cues).

It was originally assumed that whichever processing mode produced the higher mSTMI score would be preferred and that a criterion difference score of 0 would be optimal for predicting preferences. Instead, the criterion score producing the most accurate predictions was consistently positive across the two scoring methods and all 39 subjects. This suggests a general preference for OMNI processing unless the DIR mode produces an mSTMI difference score of approximately 0.06. This is consistent with assignment of OMNI processing as the default mode in manually switchable aids with DIR processing available for specific listening situations (Walden et al, 2004). It is also consistent with OMNI preferences in quiet listening situations (Kuk, 1996; Preves et al, 1999; Walden et al, 2004), since DIR processing should have little effect on mSTMI scores (or SNR) in quiet. The reasons for preferring Omni processing, when neither microphone mode produces a substantial mSTMI benefit, are not completely clear. One possibility is that OMNI processing may allow greater access to the full range of sounds in the surrounding environment, providing a greater feeling of "connectedness" to that environment. Processing noise associated with DIR processing may also contribute to OMNI preferences (Ricketts, 2001).

There are several ways that the clean speech template used in the current analysis might be altered to improve preference predictions. The current template was based on a large sample of clean speech from both male and female adults. It seems likely that this generic clean speech template might produce less accurate predictions than a template that more precisely represented the speech characteristics of the current target talker. For example, separate templates for male and female talkers, or much more specific templates associated with specific listeners (e.g., spouse or other frequent companion), might improve predictive performance.

The auditory modeling used to produce the clean speech template might also be modified in ways that improve predictive accuracy. Currently, the processing used to produce the template is intended to represent a normal, unimpaired auditory system. The modeling of an unimpaired processor may account for the more accurate prediction of preferences for NH than HI listeners (see Fig. 6). If so, auditory modeling that represented processing by either a generic or specific impaired listener (reduced signal audibility, reduced frequency resolution, etc.) might produce a more appropriate template and more accurate preference predictions.

This article examined a direct-comparison approach to automatic switching based on mSTMI difference scores. Alternatively, absolute mSTMI scores could be used to support switching decisions under a scene analysis approach. For several commercially available hearing aids using scene analysis, the estimated SNR in the listening environment is an important factor in microphone mode selection (Chung, 2004). The mSTMI scores for omni-processed stimuli may allow fairly accurate predictions of SNRs (see Grant et al, 2008, fig. 5b, lower panel). However, absolute mSTMI scores were a poor predictor of microphone preferences in the current data. As seen in Figure 1, there was extensive overlap in absolute mSTMI scores across the three preference categories. These data suggest that for listening situations with similar external SNRs, DIR processing is likely to be preferred in some cases, and OMNI preferences are likely in other instances (Ricketts et al, 2003; Walden et al, 2004; Walden et al, 2007). For this data set, and presumably for a range of real-world listening situations, external SNRs may be a poorer predictor of preferences than relative SNRs of DIR- and OMNI-processed versions of the input.

CONCLUSIONS

A "direct-comparison" approach to automatic selection of directional or omnidirectional hearing aid processing was examined, using mSTMI analyses (Grant et al, 2008) to predict the preferred microphone mode. The approach differs from more frequently used "acoustic scene analysis" approaches to automatic switching between microphone modes in that it does not require assumptions about relationships between acoustic scene characteristics and microphone mode preferences.

The mSTMI analysis of field recordings of DIR- and OMNI-processed signals, made in a variety of listening environments, allowed nearly 100% accuracy in predicting original "in-the-field" DIR/OMNI preferences from two raters. The analysis allowed 82–85% accuracy at predicting individual preferences from 39 subjects based on laboratory comparisons of the DIR- and OMNI-processed recordings of each environment.

Subjects tended to prefer OMNI processing in cases where the two microphone modes led to fairly similar mSTMI scores, consistent with the current view that OMNI mode should be treated as the default mode and suggesting that DIR processing may only be preferred in instances where it produces a substantial improvement in SNR over OMNI processing.

Hearing-impaired listeners experienced with directional processing were more likely to prefer DIR processing than other listeners in instances where the mSTMI analysis indicated a small DIR benefit. This finding suggests that acceptance and overall use of DIR mode may increase with experience.

REFERENCES

Blamey PJ, Fiket HJ, Steele BR. (2006) Improving speech intelligibility in background noise with an adaptive directional microphone. *J Am Acad Audiol* 17:519–530.

Chi T, Gao M, Guyton MC, Ru P, Shamma SA. (1999) Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am* 106:2719–2732.

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 MT, Surr RK, Walden BE, Olson L. (2002) Performance of directional microphone hearing aids in everyday life. *J Am Acad Audiol* 13:295–307.

Dillon H, Zakis JA, McDermott H, Keidser G, Dreschler W, Convery E. (2006) The trainable hearing aid: what will it do for clients and clinicians? *Hear J* 59:30–36.

Dubno JR, Dirks DD, Morgan DE. (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 SA. (2003) A spectro-temporal modulation index (STMI) assessment of speech intelligibility. *Speech Commun* 41:331–348.

Fabry D, Tchorz J. (2005) Results from a new hearing aid using acoustic scene analysis. *Hear J* 58:30–36.

Flynn MC. (2006) A comparison of decision-making processes in directional systems. *Hear Rev* 13:94–99.

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

Gatehouse S. (1992) The time course and magnitude of perceptual acclimatization to frequency responses: evidence from monaural fitting of hearing aids. *J Acoust Soc Am* 92:1258–1268.

Grant KW, Elhilali M, Shamma SA, Walden BE, Surr RK, Cord MT, Summers V. (2008) An objective measure for selecting microphone modes in OMNI/DIR hearing-aid circuits. *Ear Hear.*

Helfer KW, Wilber LA. (1990) Hearing loss, aging, and speech perception in reverberation and noise. *J Speech Hear Res* 33:149–155.

Institute of Electrical and Electronic Engineers. (1969) *IEEE Recommended Practice for Speech Quality Measures*. New York: Institute of Electrical and Electronic Engineers.

Kochkin S. (1993) Marke track III: why 20 million in U.S. don't use hearing aids for their hearing loss. *Hear J* 46:20–27.

Kochkin S. (1994) Marke track IV: Impact on purchase intent of cosmetics, stigma and style of hearing instrument. *Hear J* 47:29–36.

Kuk FK. (1996) Subjective preference for microphone types in daily listening environments. *Hear J* 49:29–35.

Moore BCJ. (1995) *Perceptual Consequences of Cochlear Damage*. Oxford, UK: Oxford University Press.

Munro KJ, Lutman ME. (2003) The effect of speech presentation level on measurement of auditory acclimatization to amplified speech. *J Acoust Soc Am* 114:484–495.

Olson L, Ioannou M, Trine TD. (2004) Appraising an automatically switching directional system in the real world. *Hear J* 57: 32–38.

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

Plomp R. (1978) Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J Acoust Soc Am* 63: 533–549.

Preves DA, Sammeth CA, Wynne MK. (1999) Field trial evaluations of a switched directional/omnidirectional in-the-ear hearing instrument. *J Am Acad Audiol* 10:173–284.

Ricketts T. (2001) Directional hearing aids. *Trends Amplif* 5:139–176.

Ricketts T, Henry P, Gnewikow D. (2003) Full time directional versus user selectable microphone modes in hearing aids. *Ear Hear* 24:424–439.

Suter AH. (1985) Speech recognition in noise by individuals with mild hearing impairment. *J Acoust Soc Am* 78:887–900.

Valente M, Van Vliet D. (1997) The Independent Hearing Aid Fitting Forum (IHAFF) protocol. *Trends Amplif* 2:6–35.

Walden BE, Surr RK, Cord MT, Dyrlund O. (2004) Predicting hearing aid microphone preference in everyday listening. *J Am Acad Audiol* 15:365–396.

Walden BE, Surr RK, Cord MT, Grant KW, Summers V. (2007) The robustness of hearing aid microphone preferences in everyday listening situations. *J Am Acad Audiol* 18:358–379.