Psychometric Comparison of the Hearing in Noise Test and the American English Matrix Test

Jumana Harianawala*
Jason Galster*
Benjamin Hornsby†

Abstract

Background: The hearing in noise test (HINT) is the most popular adaptive test used to evaluate speech in noise performance, especially in context of hearing aid features. However, the number of conditions that can be tested on the HINT is limited by a small speech corpus. The American English Matrix test (AEMT) is a new alternative adaptive speech in noise test with a larger speech corpus. The study examined the relationships between the performance of hearing aid wearers on the HINT and the AEMT.

Purpose: To examine whether there was a difference in performance of hearing aid wearers on the HINT and the AEMT. A secondary purpose, given the AEMT's steep performance-intensity function, was to determine whether the AEMT is more sensitive to changes in speech recognition resulting from directional (DIR) microphone processing in hearing aids.

Research Design: A repeated measures design was used in this study. Multiple measurements were made on each subject. Each measurement involved a different experimental condition.

Study Sample: Ten adults with hearing loss participated in this study.

Data Collection and Analysis: All participants completed the AEMT and HINT, using adaptive and fixed test formats while wearing hearing aids. Speech recognition was assessed in two hearing aid microphone settings—omnidirectional and fixed DIR. All testing was conducted via sound field presentation. Performance on HINT and AEMT were systematically compared across all test conditions using a linear model with repeated measures.

Results: The results of this study revealed that adult hearing aid users perform differently on the HINT and AEMT, with adaptive AEMT testing yielding significantly better (more negative) thresholds than the HINT. Slopes of performance intensity functions obtained by testing at multiple fixed signal-to-noise ratios, revealed a somewhat steeper slope for the HINT compared with the AEMT. Despite this steeper slope, the benefit provided by DIR microphones was not significantly different between the two speech tests.

Conclusions: The observation of similar DIR benefits of the HINT and AEMT suggests that the HINT and AEMT are equally sensitive to changes in speech recognition thresholds following intervention. Therefore, the decision to use the AEMT or the HINT will depend on the purpose of the study and/or the technology being investigated. Other test related factors such as available sentence corpus, learning effects and test time will also influence test selection.

Key Words: adaptive test, AEMT, HINT, performance-intensity functions, speech in noise test

Abbreviations: AEMT = American English Matrix test; DIR = directional; ES = effect size; HINT = hearing in noise test; OMNI = omnidirectional; PI = performance intensity; REARs = real ear–aided responses; RMANOVA = repeated measures analysis of variance; SNR = signal-to-noise ratio; SRT50 = speech reception threshold at 50% sentence recognition

*Starkey Hearing Technologies, Eden Prairie, MN; †Department of Hearing and Speech Sciences, Vanderbilt Bill Wilkerson Center, Vanderbilt University Medical Center, Nashville, TN

Corresponding author: Jumana Harianawala, Starkey Hearing Technologies, Eden Prairie, MN 55346; Email: jumana_harianawala@starkey.com

INTRODUCTION

Speech-in-noise tests provide a direct method to assess an individual's speech recognition abilities. In the hearing impaired population, these tests can be used to document changes in the individual's speech recognition that may result from amplification, assistive technologies and/or rehabilitation programs. Adaptive tests such as the hearing in noise test (HINT; Nilsson et al, 1994) have a long history of use in research and clinical settings (Pumford et al, 2000; Ricketts and Mueller, 2000; Valente et al, 2000; Mendel, 2007; Soli and Wong, 2008). The HINT's popularity may be because of its high test–retest reliability (Wilson et al, 2007; Vermiglio, 2008; Yund and Woods, 2010; Stuart and Butler, 2014), the simple up-down test procedure, and quick administration time.

The original HINT corpus uses Bench–Kowal–Bamford sentences recorded in American English. These are everyday sentences that have been equalized in terms of length, naturalness, intelligibility, and phonemic distribution, to maximize measurement reliability (Nilsson et al, 1994). The sentences, spoken by a male talker, are five to seven syllables in length and are rated at a first grade reading level (e.g., She took off her fur coat). The listener must repeat the entire sentence to receive a correct rating. Sentences are presented in a speech shaped, steady-state noise which is held constant in level while the speech signal is adaptively varied to find the signal-to-noise ratio (SNR) needed for speech reception threshold at 50% sentence recognition (SRT50).

The HINT consists of 25 lists, each containing 10 sentences. The test developers recommend using two lists (20 sentences) to improve reliability (Nilsson et al, 1994; Soli and Wong, 2008). However, using two lists per condition reduces the number of conditions that can be tested without replicating lists. This is an important limitation in clinical research settings, where testing across multiple conditions, or sessions, is often essential. The availability of only 12 pairs of HINT lists (24 lists) may not be sufficient for some experimental designs. Repeating the tests lists to accommodate more conditions, or improve reliability, has been contraindicated by studies that have demonstrated learning effects across test sessions (Wilson et al, 2003; Yund and Woods, 2010). These reasons suggest the need for an alternative test with a larger corpus that provides similar ease of administration and psychometric properties to the HINT.

The American English Matrix Test (AEMT; Kollmeier et al, 2015) is another adaptive speech-in-noise test consisting of 36, 20 sentence, lists. Like the HINT, the AEMT uses a speech-shaped, steady-state noise that is held constant as the sentence level is varied to converge to a 50% correct SRT. The AEMT uses sentences with a fixed syntactic structure (name-verb-numeral-adjective-object; e.g., Rachel wants four pretty chairs). Sentences have been pre-constructed by making random combinations of five words, one from each category of an inventory of 50 words (10 words per category). Grammatically correct but semantically unpredictable, the AEMT sentences are less redundant than the everyday HINT sentences. The low semantic predictability minimizes learning effects after sufficient practice (Rose, 2013; Kiolbasa, 2015). Compared with the HINT, the larger sentence corpus and small learning effects of the AEMT may make it a reasonable option for clinical investigators interested in assessing speech recognition under multiple test conditions.

On any speech-in-noise test, the perceptual benefit that a listener will receive from an improvement in speech level or SNR relies on the slope (gradient) of the tests performance intensity (PI) function. The PI function indicates how speech recognition changes with modifications in speech level/SNR and for many speech materials, it is sigmoidal in shape. The steeper the slope of the PI function, the more benefit a listener receives from a gains in speech level or SNR (MacPherson and Akeroyd, 2014). Therefore, a steep PI function can enhance our ability to detect meaningful differences between conditions where SNR or speech levels vary. The AEMT PI function has been reported to have a slope of 13.3%/dB (Kollmeier et al, 2015) which is steeper than that of the American English HINT (10.6%/dB; Soli and Wong, 2008). This means, compared with AEMT, a larger change in speech level or SNR will be required to see performance differences on the HINT between different experimental conditions (e.g., microphone modes in hearing aids).

As such, the primary goal of this study was to examine performance differences on the HINT and AEMT in a group of adult hearing aid wearers. A comparison between the HINT and AEMT is warranted because the AEMT's reportedly steep PI function and the ability to repeat the test with minimal learning effects may make the AEMT a good alternative test to other adaptive speech recognition tests. Hearing health professionals often use speech in noise tests to demonstrate the benefit of amplification and other hearing aid–related features. To determine the clinical utility of both tests, we decided to examine the effect of hearing aid microphone mode (omnidirectional [OMNI] and fixed directional [DIR]) on speech recognition as measured by both tests. We chose to vary microphone mode because switching between OMNI and DIR modes can change the SNR at the listener's ear (e.g., Ricketts, 2001). We hypothesized that, compared with the HINT, the AEMT may be more sensitive to potential benefits in speech recognition offered by DIR microphones due to its reportedly steeper PI function.

Finally, when comparing the performance of single hearing aid features, such as directionality, an adaptive task would be appropriate. However, the use of multiple

fixed input SNRs may be more appropriate for certain experiments that aim to compare two *different* nonlinear systems (e.g. compression). Using an adaptive task or a single fixed input SNR in such experiments could be misleading due to potential differences in the underlying hearing aid architectures (Naylor and Johannesson, 2009). To make across- and within-technology comparisons, testing with an adaptive and a fixed task at multiple SNRs with the same test material would be ideal. Hence, for prospective experimental designs, we obtained a representative data set to enable selection of SNRs for experiments that may want to adapt either the HINT or the AEMT for fixed administration. To do so, we measured the average aided PI functions using HINT and AEMT test materials.

For the present study, the research questions were as follows:

- What is the difference between adaptively measured HINT and AEMT SRT50 obtained from adults with hearing loss listening in OMNI and DIR microphone modes?
- Does DIR benefit, as measured by changes in adaptive SRT50, vary between the HINT and AEMT?
- When measured using multiple fixed SNRs, does the slope of the aided PI function differ between the HINT and AEMT?

METHODS

Participants

Participants were 10 adults (six males and four females) with symmetrical sensorineural hearing loss ranging in age from 31 to 79 years (M = 68.4 years; standard deviation = 13.7 years). A power analysis was performed to estimate sample size, based on data from an internal study using HINT (N = 30), comparing OMNI with DIR microphone mode. The effect size (ES; Cohen's d) in this study was 0.99, which is considered to be large using Cohen's (1988) criteria. With an alpha = 0.05 and power = 0.80, the projected sample size needed with this ES (GPower 3.1; Faul et al, 2007) was approximately N = 10. In the absence of similar AEMT data, our proposed sample size of 10 was considered adequate to at least observe differences between microphone modes.

The mean four-frequency pure-tone average (500–4000 Hz octave frequencies) was 45 dB HL for both right and left ears. The right and left average, minimum, and maximum audiometric thresholds are shown in Figure 1. All participants had normal middle ear function as verified with tympanometry and middle ear compliance measurements.

Participants were native English speakers and had not been tested on the HINT in the past year. Nine of the 10 participants were experienced hearing aid
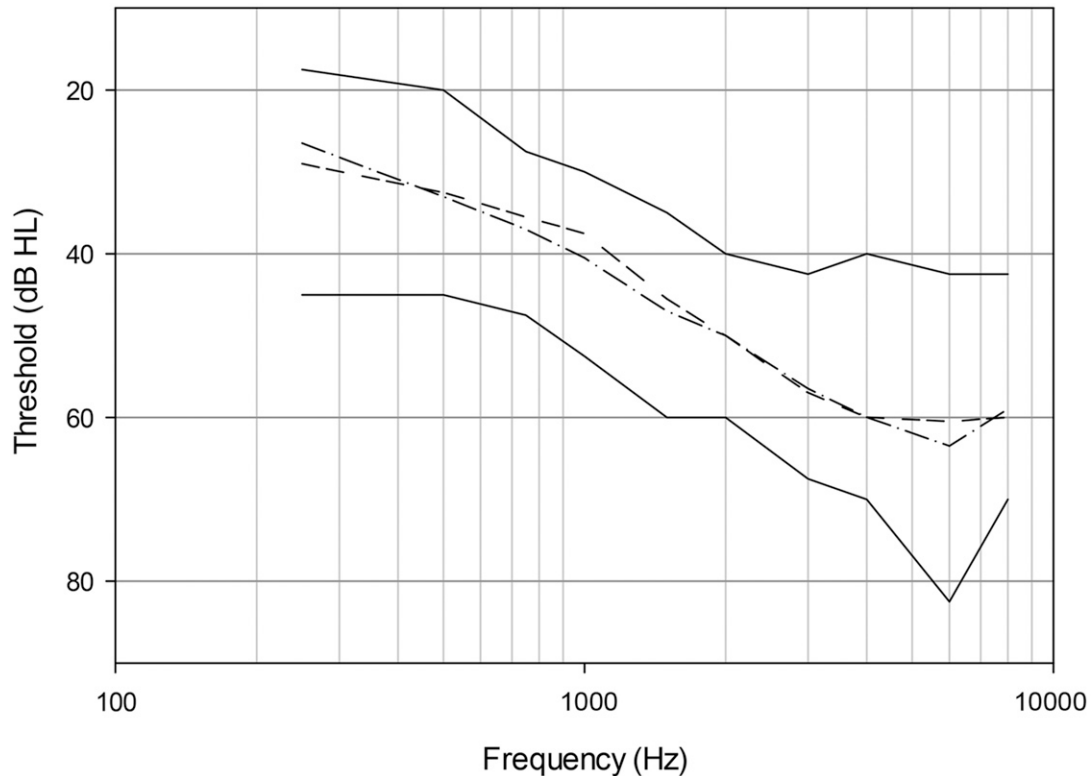

Figure 1. Mean audiometric thresholds for 10 participants. The center dotted lines show the mean right and left ear thresholds. The solid line at the top and bottom show the collective maximum and minimum thresholds, respectively.

users. Median hearing aid experience was eight years (range = 1–36 years). All participants signed an informed consent form before their participation in the study. All participants were compensated for their participation and travel to and from the research facility.

Test Materials

Commercially available sentence corpus for the HINT (Maico Diagnostics, Eden Prairie, MN, 2004) and AEMT (HörTech gGmbH, Oldenburg, Germany, 2015) and their respective speech-shaped noise were used in this study. For both tests, the speech was played from a speaker placed at 0° azimuth in front of the participant and uncorrelated samples of the respective speech-shaped noises were played continuously from seven other speakers placed around the participant (45°, 90°, 135°, 180°, 225°, 270°, and 315° azimuth). The noise was calibrated by, first, equating the levels from each loudspeaker, independently, at the position of the listener's head. Then the noise level was measured while the noise was presented simultaneously from all loudspeakers. The level of each loudspeaker was adjusted equally to create an overall noise level of 65 dBA. Spectral analyses of the digital speech materials from each test and their respective speech shaped noises are shown in Figure 2.

HINT

Test stimuli consisted of 25 lists of 10 sentences spoken by a male talker in general American dialect (Nilsson et al, 1994; Maico Diagnostics, 2004). Two lists of 10 sentences each were used for each administration. The participants were instructed to listen and repeat aloud the sentences heard or understood. No feedback was provided. The testing was conducted by presenting the spectrally matched speech-shaped noise at a fixed level of 65 dB (A). An adaptive up-down strategy was used to determine the sentence presentation levels. The first sentence was presented at 55 dB (A) and was increased in level in 4 dB steps, if needed, until it was repeated correctly. Subsequent sentences were presented only once, with presentation levels dependent on the accuracy of the participant's response. Presentation level of the sentences were attenuated by 2 dB following a correct response and increased by 2 dB following an incorrect response. For a response to be scored correct the entire sentence had to be repeated exactly with the exception of verb tense (is/was, are/were, and has/had) and article (a/the) variations. The test was concluded at the 21st response for each administration. Speech levels for the 5th to 21st responses were averaged and then subtracted from the noise level (65 dBA) to calculate the participant's adaptive 50% speech reception threshold (SRT50).

For the fixed administration, the background noise level was held constant at 65 dBA and the speech level was adjusted to achieve the desired SNR. The test was conducted at six pre-determined SNRs (+3, 0, −3, −6, −9, and −12). Two lists of 10 sentences were administered at each SNR. Similar to the adaptive test, complete sentences were scored and each sentence was allotted five percentile points for a correct response.
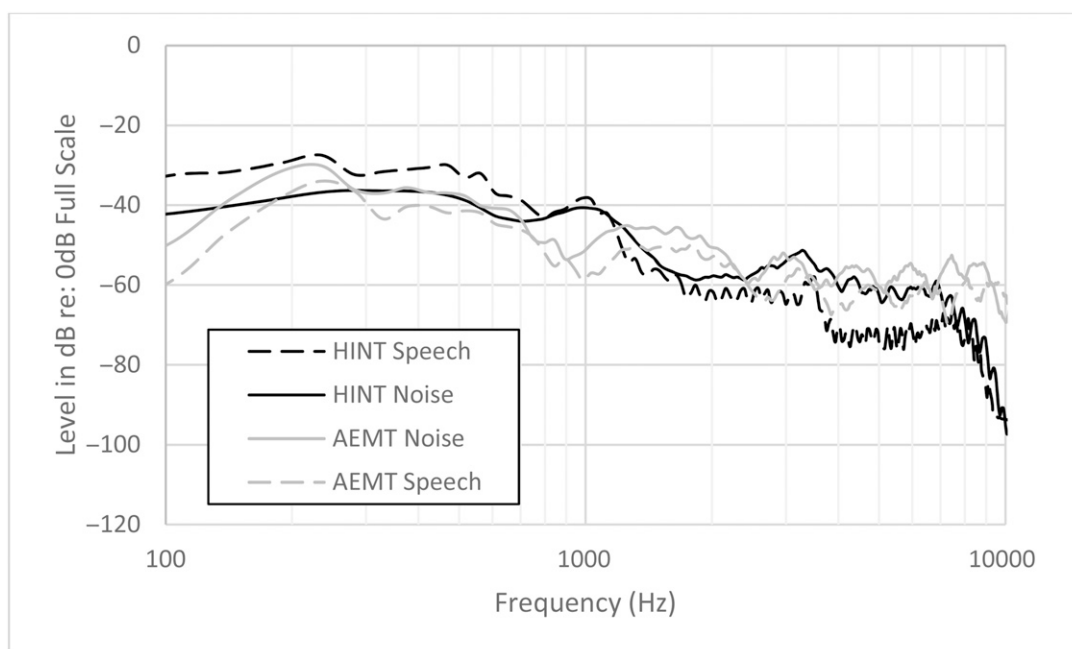

Figure 2. Spectrum of test material plotted in dB referenced to 0 dB full scale. HINT materials are shown in black and AEMT materials in gray. Solid line represent the noise spectrums and dotted lines represent the speech spectrum (list 1s of each test).

Incorrect responses scored zero points. The final percent correct score at each SNR was obtained by adding up the number of correctly repeated sentences in a list pair. A maximum score of 100% was possible at each SNR.

AEMT

The AEMT consisted of 36 lists (20 sentences/list) that have been optimized such that the speech elements of each sentence have high homogeneity in terms of intelligibility at test SNRs (Kollmeier et al, 2015; HörTech gGmbH, 2015). Unlike the HINT, the AEMT sentences are not recorded by a single talker but are a concatenation of words spoken by multiple female talkers. However, the coarticulation between words present in natural speech is preserved by segmenting the single words such that the coarticulation to each subsequent word is included and resynthesized (See Kollmeier et al (2015) for details regarding stimulus development). For test administration, digital copies of the AEMT sentences and corresponding speech-shaped noise were exported from the manufacturer's CD and saved to a computer hard drive. Digital copies were then presented via a commercial software program (Adobe Audition CS6, San Jose, CA). The test was conducted by presenting the first sentence at 0 dB SNR. The spectrally matched noise was fixed at a level of 65 dB SPL (flat weighting). A flat weighting was selected to stay with the manufacturer's recommendation and to have results that can be directly compared with other studies. Each sentence has five target words. The participant was instructed to repeat back all five words. The first five sentences of a twenty-sentence list are scored differently than the remaining 15 sentences. See Table 1 for the test's adaptation rules. After the final sentence was presented, the presentation level for what would have been the 21st sentence was calculated. The speech levels corresponding to responses 12–21 were averaged. The adaptive SRT50 was then calculated by subtracting the averaged speech level from the noise level of 65 dB.

Similar to the HINT, the AMET was administered at predetermined fixed SNRs (+3, 0, −3, −6, −9, and −12) by keeping the noise level constant at 65 dB SPL and adjusting the speech level to achieve the desired SNR. A list of 20 sentences was presented at each SNR. Each word of every sentence in the list was scored and allotted one percentage point. With a list of 20 sentences and five words in each sentence, a maximum of 100% correct could be obtained on a single list.

Hearing Aid Fitting

All participants were fitted with Starkey Hearing Technologies Muse i2400 behind-the-ear hearing aids. The Muse i2400 is a 24-channel wireless hearing aid

Table 1. Speech Level Changes for the Manual Adaptive Level Control of the Matrix Test

Correctly Understood Words in the Previous Sentence	Change of Speech Level	
	For Sentence 2–5	For Sentence 6–21
5	−3 dB	−2 dB
4	−2 dB	−1 dB
3	−1 dB	0 dB
2	+1 dB	0 dB
1	+2 dB	+1 dB
0	+3 dB	+2 dB

Note: Recreated from English (US) matrix test. Manual of audio CD for research and development (p.6) by HörTech gGmbH. Copyright (2015) by HörTech gGmbH Oldenburg.

with wide dynamic range compression. The same bilateral set of hearing aids was used for all participants. Hearing aids were coupled to the participants' ears using custom acrylic half shell occluded ear molds with standard size 13 tubing.

During the study, the OMNI and DIR microphone modes were programmed into the two memories of the hearing aid and were accessed by pushing a button on the faceplate. The memory was identified by the number spoken by a male voice (e.g., "one" or "two") after pushing the button. Memory one was programmed to OMNI and memory two to DIR. In the assessed DIR mode, the microphone had a fixed cardioid polar pattern. We confirmed the DIR microphone was functioning in the test box using the DIR test module of the Audioscan Verifit II (Etymotic Design, Inc, Dorchester, Ontario, Canada). This was carried out before beginning the study and on completion of the study. Settings in memory one and two were matched in all respects, except the microphone mode. All other automatic and adaptive features except feedback suppression were disabled throughout the study.

The hearing aids were programmed to e-STAT, Starkey's proprietary fitting formula using the fitting software default settings. The e-STAT prescription for hearing aid gain is derived from a combination of NAL-NL1 targets and empirical offsets that differ slightly as a function of device style and acoustic coupling to the ear (Scheller and Rosenthal, 2012). Real ear–aided responses (REARs; Figure 3) were measured via probe microphone measurements using the Audioscan Verifit II; the International Speech Test Signal was used as a test stimulus with presentation levels of 50, 65, and 80 dB SPL. In Figure 3, the solid lines show the e-STAT targets for 50, 65, and 80 dB SPL speech inputs. The dashed lines show the corresponding REARs. The targets and REARs have been averaged across participants and across right-left ears. As seen in the figure, the REARs were within 5 dB of targets from 100 to 5000 Hz and within 10 dB of targets from 6000 to 8000 Hz.
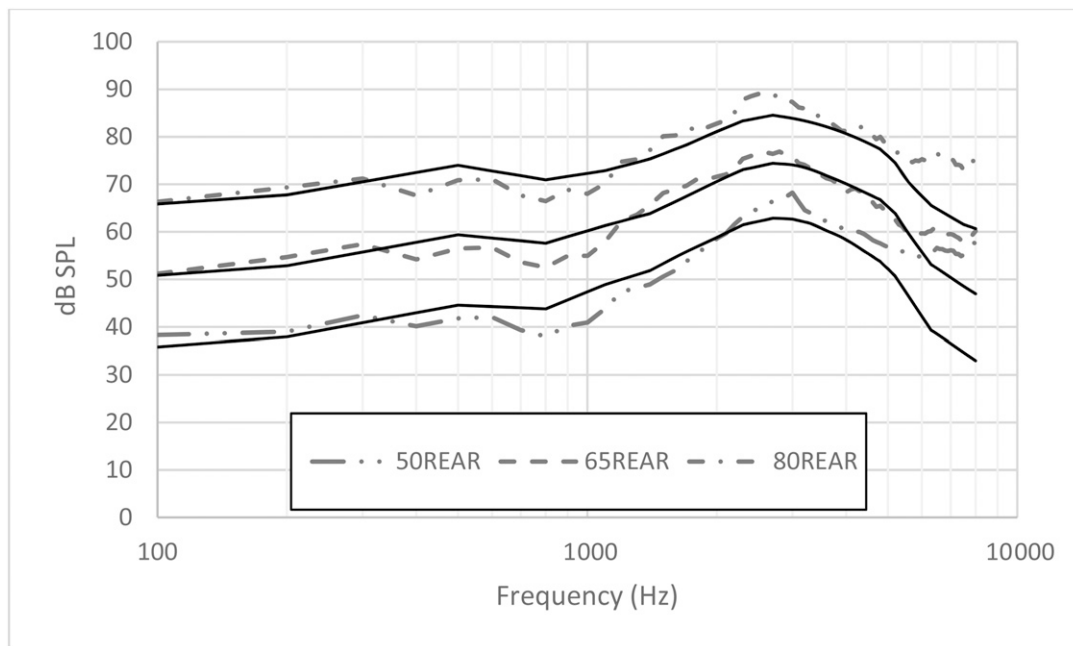
Figure 3. Average REAR for soft, medium and loud international speech test signal input. The long dash double dot line shows the 50 dB SPL REAR, the small dash line shows the 65 dB SPL REAR and the single dash dot line shows 85 dB SPL REAR. The solid lines show average eSTAT (Starkey proprietary) targets for 50, 65, and 80 dB inputs. The targets and REARs are averaged across participants and across right and left ears.

Test Set-Up and Procedures

Testing was conducted over two visits separated by a maximum of two weeks. Each session took no longer than two hours. Testing was completed using adaptive and fixed SNR procedures. Hearing aid fitting and adaptive testing was carried out in the first test session and fixed SNR testing was conducted in the second test session. Two of the microphone modes, OMNI and DIR, were tested on each test. Both test (HINT/AEMT) and microphone mode (OMNI/DIR) were counterbalanced across sessions. Two HINT lists (20 sentences) and one AEMT list (20 sentences) were administered in each condition. Before test administration, practice was provided on both tests following recommendations in the test manuals—two lists of HINT practice sentences (20 sentences total) and two lists of the AEMT sentences (40 sentences total). Because AEMT does not have separate practice lists, the list used for training were not included during test administration. At the end of session one, a total of four adaptive SRT50s (HINT-OMNI, HINT-DIR, AEMT-OMNI, and AEMT-DIR) were obtained for each participant. For adaptive administrations of both tests, we observed that it took approximately two minutes to complete the HINT (20 sentences) and six minutes to complete the AEMT (20 sentences).

In the second session, the HINT and AEMT were administered at predetermined fixed SNRs. The same six SNRs (+3, 0, −3, −6, −9, and −12) were used for both HINT and AEMT testing. These SNRs were selected, based on pilot data from three adults with hearing loss, to limit floor and ceiling effects. Two HINT lists (20 sentences) and one AEMT list (20 sentences) were administered at each SNR in both hearing aid conditions resulting in 24 scores (2 tests*6 SNRs*2 modes) for each participant.

The presentation order of the test (HINT and AEMT) and microphone mode (OMNI and DIR) were counterbalanced to control for order effects. All test lists used were randomized. Because of the limited number of HINT lists, one fixed SNR condition required repetition of one pair of HINT lists (20 sentences). We systematically varied the SNR chosen to receive the repeated list across participants. Two randomly selected 10-sentence lists were combined to create the extra pair of 20 sentences. Care was taken to ensure these lists had not been presented as a pair previously. Their position in the pair (1st half versus 2nd half) was also counterbalanced.

Stimulus presentation, randomization, and scoring for the adaptive HINT was managed through an automated, computer-controlled MATLAB-based testing software developed at Starkey Hearing Technologies. For all other test administrations (Fixed SNR HINT, Adaptive AEMT, and Fixed SNR AEMT), presentation of speech was managed through a digital audio workstation (Adobe Audition CS6). All stimuli were presented in the sound field via Genelec 8030B speakers (Lisalmi, Finland). Test presentation was controlled via a Dell

Precision T3500 (Round Rock, TX) desktop computer and a Lenovo T440 laptop computer. Two multichannel DSP sound cards (RME Multiface II, Haimhausen, Germany) were used to route the speech and noise signals. For both HINT and AEMT, care was taken to ensure the noise was not interrupted when changing the level of the speech signal. The speech stimulus was always played 20 seconds after the onset of noise to ensure all hearing aid processing had stabilized before testing. All testing was conducted in a double-walled sound booth.

RESULTS

Adaptive Speech Recognition Thresholds

The mean adaptive SRT50s, standard deviations and confidence intervals for each speech recognition test and microphone mode are shown in Table 2. Because the standard versions of the HINT and AEMT follow an adaptive protocol, the 50% points (SRT50) for each participant were obtained by averaging trials as specified by the respective test instructions. A two-way repeated measures analysis of variance (RMANOVA) was conducted to evaluate the difference between adaptive SRT50s across the two tests for each microphone mode. The results of the two-way RMANOVA showed

Table 2. Mean SNR, SD, and CI Obtained on Adaptive HINT and AEMT for Two Microphone Modes

Test	OMNI	DIR	Benefit
HINT			
Mean (SNR dB)	−0.9	−4.8	3.9
SD (dB)	1.3	1.3	1.8
95% CI of mean	±0.81	±0.81	±1.11
AEMT			
Mean (SNR dB)	−5.5	−8.4	2.9
SD (dB)	1.5	1.9	1.3
95% CI of mean	±0.93	±1.21	±0.79

Notes: DIR benefit (in dB) was computed by subtracting OMNI scores from the DIR scores. CI = confidence interval, SD = standard deviation.

significant main effects of test and microphone mode. For adult hearing aid wearers aided HINT and AEMT adaptive SRT50s, averaged across microphone modes, were significantly different [$F_{(1,9)} = 84.45$, $p < 0.001$, $\eta_p^2 = 0.904$]; the mean adaptive SRT50 obtained on the AEMT was more negative (lower) than that obtained with the HINT. Within each test, the DIR mode was better (i.e., lower SRT50s) than OMNI [$F_{(1,9)} = 74.86$, $p < 0.001$, $\eta_p^2 = 0.893$]. According to Cohen's (1988) convention for partial eta squared, an ES of 0.01 is considered small, 0.09 medium, and 0.25 large (Vacha-Haase and Thompson, 2004). Therefore, the ESs of $\eta_p^2 = 0.905$
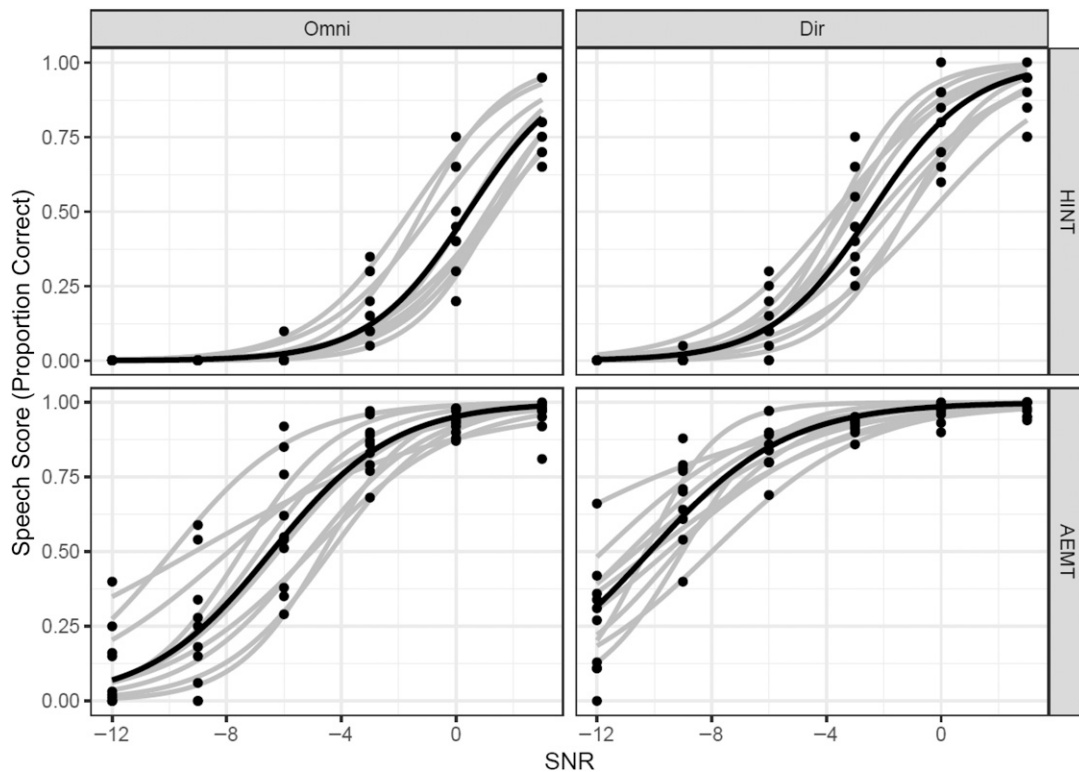

Figure 4. Individual and average logistic functions measured on the HINT and AEMT for Omni and Dir conditions. The gray lines show individual logistic fits for each participant and the superimposed black line shows the average PI function across all participants. The filled circles show individual data at each measured SNR in dB shown on the x-axis. The y-axis shows data as proportion correct.

and $\eta_p{}^2 = 0.893$, for differences in test and microphone modes, respectively, suggests high practical significance. By contrast to our original hypothesis, the interaction between the tests and microphone modes was nonsignificant [$F_{(1,9)} = 2.46$, $p = 0.151$, $\eta_p{}^2 = 0.215$], suggesting no differential effect of speech test on DIR benefit.

Fixed Speech Recognition Thresholds

For each test, percent correct scores obtained at the six fixed SNRs were used to generate PI functions by fitting a two-parameter logistic function to each participant's data and to the mean data. Figure 4 shows PI functions for the HINT and AEMT data (filled symbols) for both OMNI and DIR microphone modes. Performance (proportion correct) is shown on the y-axis in relation to stimulus level, in this case SNR on the x-axis. The gray lines show the individually fitted PI function for each participant and the black solid line shows the average PI function across all participants.

The average SRT50 calculated from the fits (henceforth referred to as the Fixed SRT50) are listed in Table 3. A two-way RMANOVA was conducted to evaluate the difference between fixed SRT50s across the two tests for each microphone mode. The results of the two-way RMANOVA once again showed significant main effects of test [$F_{(1,9)} = 180.73$, $p < 0.001$, $\eta_p{}^2 = 0.953$] and microphone mode [$F_{(1,9)} = 226.75$, $p < 0.001$, $\eta_p{}^2 = 0.962$]. The interaction between the tests and microphone modes was nonsignificant [$F_{(1,9)} = 1.81$, $p = 0.211$, $\eta_p{}^2 = 0.168$]. These results are consistent with results obtained with adaptive SRT50s and confirm the observed differences in test (AEMT more negative) and microphone mode (DIR more negative). Even in the fixed SRT50 analysis, there was no significant difference in DIR benefit on the two tests.

Table 3 also shows the slopes of the PI function around the 50% points calculated from the polynomial fits following logistic regression. The individual slopes around the 50% points from each participant for each condition (HINT-OMNI, HINT-DIR, AEMT-OMNI, and AEMT-DIR) were subjected to a two-way RMANOVA to evaluate the difference between the slopes of the PI functions across the two tests for each microphone mode. Results

showed a significant main effect of test [$F_{(1,27)} = 13.57$, $p < 0.001$, $\eta_p{}^2 = 0.244$] with slopes for the HINT functions being significantly steeper than slopes for the AEMT functions. The finding of steeper HINT functions is in contrast to our original hypothesis. The main effect of microphone mode and the interaction between speech test and microphone mode were not statistically significant. This suggests no significant difference in slopes between OMNI and DIR microphone mode.

DISCUSSION

The primary purpose of this study was to examine differences in the aided SRT50s of hearing aid wearers on the HINT and AEMT when listening in OMNI and DIR microphone modes. We were also interested in whether the purportedly steeper psychometric function of the AEMT resulted in greater measured DIR benefit for our group of adults with hearing loss. In this assessment, we considered administration constraints, psychometric properties, and test corpus. For this discussion, we will reflect on the psychometric properties of each test while drawing comparisons with past literature. Finally, we will discuss implications regarding the use of the HINT or AEMT, in the context of hearing aid clinical research.

Adaptive Speech Recognition Thresholds

In this study, mean adaptive HINT SRT50s for hearing aid wearers were −0.9 and −4.8 dB SNR for the OMNI and DIR microphone modes, respectively. These HINT thresholds were consistent with the ranges reported in literature for OMNI (−1 to 8 dB SNR) and DIR (−5.3 to 4 dB SNR) when tested in similar, multiple noise source, conditions (Ricketts and Dhar, 1999; Ricketts, 2001; Bentler, Palmer, et al, 2004; Bentler, Tubbs, et al, 2004; Kuk et al, 2005; Nordrum et al, 2006; Valente et al, 2006). The mean DIR benefit obtained using the adaptive HINT protocol was 3.9 dB which also falls within the 3–6 dB range of improvement in SNR provided by DIR microphones (Ricketts, 2001).

Given the AEMT was only recently developed, at this time peer-reviewed publications describing its performance are limited. Only three references

Table 3. Average Slopes and Fixed SRT50s Obtained from PI Functions of HINT and AEMT

Test	Avg SRT50 (SNR dB)	SD SRT50 (dB)	Avg Slope (%/dB)	SD of slope (%/dB)
HINT				
OMNI	0.5	1.2	14.5	2.0
DIR	−2.3	1.2	14.4	3.1
AEMT				
OMNI	−6.7	1.9	11.6	3.2
DIR	−10.5	2.0	10.4	4.4

Notes: SD associated with each metric is shown as well. SD = standard deviation.

specifically focused on the AEMT were found in literature, two of which were nonpeer-reviewed dissertation theses (see Rose, 2013; Kiolbasa, 2015). Rose (2013) tested normal-hearing individuals and Kiolbasa (2015) evaluated cochlear implant recipients. The most directly relevant work was a podium presentation at the 2016 Audiology Now! meeting (Zokoll et al, 2016) reporting results from a multicenter validation study. In this study, 60 adults with hearing loss completed the adaptive AEMT under headphones. Testing was conducted at four different test sites. Only one site, the University of Iowa, enrolled test participants with hearing losses similar to those of participants in the present study. Their results revealed average unaided SRT50s of ~ -5 dBSNR. The average adaptive AEMT SRT50 in the present study with hearing aids in OMNI mode was -5.5 dBSNR. Although this difference is small, it suggests the aided AEMT SRT50s observed in the present study are reasonable in comparison with limited information in this area (i.e., Zokoll et al, 2016). The reasons for the minimal difference between the aided AEMT SRT50 in the present study and the unaided AEMT SRT50 in the UI data set of Zokoll et al study are unclear. However, it may reflect the limited benefit provided by OMNI microphones in noisy settings (Walden et al, 2000) or simply reflect random individual differences in the test populations.

As noted previously, the OMNI and DIR AEMT SRT50s were significantly lower (more negative) than HINT SRT50s. A number of factors may be responsible for the observed differences in SRT50 on these two tests. For instance, the contribution of context to closed set tasks, as compared with open, affects the nature and the difficulty of the task with individuals typically performing better on closed sets, given the constrained nature of the task (Clopper et al, 2006). Potential differences in the clarity of the recorded speech (i.e., between talker differences on the AEMT and HINT) could clearly impact the SRT50 (e.g., Hood and Poole, 1980; Bess, 1983). In addition, although both the HINT and AEMT use speech-shaped noise as a masker; as seen in Figure 2, there is a slight difference in the spectrum of the speech and the corresponding noise for both test materials. This difference, coupled with potential talker differences, could create frequency-specific differences in speech and noise levels (SNRs) for the two tests. As a result, the amount of audible speech information (e.g., as calculated using the speech intelligibility index; ANSI, 1997) at any given wideband SNR could differ between tests, potentially impacting the SRT50.

In terms of DIR benefit, we hypothesized that benefit observed with the adaptive AEMT would be larger when compared with the adaptive HINT. By contrast to our hypothesis, DIR benefit on the AEMT ($M = 2.9$ dB) was not statistically different from the benefit

observed on the HINT ($M = 3.8$ dB). It is possible that the differences in gradient of the PI function, although statistically significant, may not be large enough to produce a meaningful perceptual difference.

Fixed Speech Recognition Thresholds

Analysis of the fixed SRT50s for the HINT and AEMT for both microphone modes (OMNI and DIR) demonstrated comparable outcomes as observed with adaptive SRT50s. The slopes of the PI function around the 50% point (calculated from the fixed SNR data), were significantly higher for the HINT ($\sim$14% per dB) than for AEMT ($\sim$11% per dB). This finding was unexpected, given the normative PI functions of 10.6% per dB for the HINT and 13.3% per dB for the AEMT. One possible reason for this finding could be the fact that normative PI functions reported in the literature are generally measured on younger normal-hearing listeners. Our participants were older with hearing loss. They may have relied on top-down processes to reinforce their speech understanding in noise, which would increase the slope of the HINT PI function (MacPherson and Akeroyd, 2014). The absence of sematic context in the AEMT limits the use of top-down information in effect flattening the AEMT PI function. This reasoning is further supported by Pichora-Fuller et al (1995) suggestion that older listeners can benefit more from context than younger listeners. Another reason for this finding could be the small sample size used in the present study, which could also have biased the slope estimates (Levitt, 1970). In addition, there was no significant difference in the slopes between OMNI and DIR modes within each test. This finding is expected given that the DIR microphone used in the present study only produced a linear SNR change.

Clinical Applications

Hearing aid algorithms such as noise reduction and DIR microphones have markedly different output characteristics at different input SNRs. Some algorithms are most effective at more positive SNRs whereas others function primarily in poorer SNRs (Ricketts, 2001; Smeds et al, 2015). Typical daily listening environments of hearing aid users most often have SNRs that range from +5 to +15 dB SNR (Smeds et al, 2015). To increase face validity, one would ideally administer tests at SNRs encountered in daily life. However, some existing sentence tests are poor predictors of real-world performance because scores for many individuals are near ceiling when tested at positive SNRs (Nilsson et al, 1994). For example, because participants are able to achieve near 100% correct performance on the AEMT at SNRs as low as -2 dB, as seen in Figure 4, the outcomes on the AEMT may be considered a less

ecologically valid test than the HINT. Researchers are investigating ways to modify speech tests with the goal of increasing real-world relevance by shifting the SRT to more positive SNRs (Bernstein, 2011; Best et al, 2013; Laugesen et al, 2013). Although, none of these have been validated for clinical use.

Often times, hearing health-care professionals have to also demonstrate a need for, and benefit from, amplification. It is very common for clinicians to look at differences between aided and unaided SRT50s to demonstrate this benefit. For adults with hearing loss, our results suggest that the HINT, with its steeper PI function, may be more effective at detecting small benefits resulting from amplification and variations in hearing aid processing.

In addition to comparing signal processing features and demonstrating benefit with hearing aid use, the SRT50 is often used by clinicians to monitor changes in speech recognition ability, candidacy for cochlear implant recipients and as justification for disability compensation. In some of these scenarios, a clinician may encounter an individual where a SRT50 on the HINT cannot be obtained because of the severity of their hearing loss (i.e., the impact of their loss is so severe they cannot understand 50% of the speech at any SNR). For these individuals the AEMT may offer advantages given mean SRT50s on the AEMT are substantially better (more negative) than HINT SRT50s for the same individual. This may allow the clinician to obtain a valid SRT50 using the AEMT even for individuals with very poor speech recognition abilities who are unable to complete HINT testing.

Although, we did not systematically study test administration time in this study, we observed that it took approximately six minutes to manually administer the AEMT compared with the HINT which took about two minutes. As indicated in the methods, we used manual scoring and administration of the AEMT in the present study compared with the commercially available automated software version. It is expected that software-driven administration would reduce test times. Even so, Zokoll et al (2016) reported an average administration time of four minutes to run the AEMT using the software version. Furthermore, both Rose (2013) and Kiolbasa (2015) reported significant training effects, which reinforces the need for a practice trial when administering the AEMT. Both HINT and AEMT developers recommend the use of two full lists as training before test administration, meaning that practice trials for the HINT may require two minutes and AEMT practice trials may require eight minutes.

Limitations

To the best of our knowledge, this is the first study to compare performance of hearing aid users on the HINT and AEMT. Thus, while providing important information, the study has some limitations. First, a post hoc power analysis revealed that the study was sufficiently powered to detect differences in SRT50 between tests (power = 1) and between microphone modes (power = 0.9) but was underpowered to observe differences in DIR benefit across the two tests (power = 0.4). A total of ~25 participants would be needed to obtain statistical power at the recommended 0.80 level (Cohen, 1988). Future studies looking to observe differences in DIR benefit on the HINT and AEMT should consider using a sample size larger than was used in the present study. Second, the participants selected in the present study for the most part had a mild to moderately severe sensorineural hearing loss. The outcomes observed in this study may vary for participants with more, or less, severe losses. All participants were fit with the same device to minimize variations due to technology differences and we did not allow for acclimatization between the hearing aid fitting and speech recognition testing. That said, we expect effects of acclimatization to be minimal and similar across participants. All participants were given the same practice conditions and test conditions were counterbalanced across participants. Furthermore, the differences in slopes observed here for the HINT and AEMT may be specific to the test-setup and participant sample (age and hearing loss) used in the present study. The results from the present study cannot be used to validate other experimental designs that might exclude any of the protocol procedures completed here. A future study with a larger sample size and a wider age range may allow us to better understand the factors that influence the slope of the PI function. Finally, we did not test participants unaided, so we are unclear on the effects of hearing aid use on the slopes of the PI function.

CONCLUSION

The purpose of this study was to examine the differences in SRT50 obtained on the two speech-in-noise tests (HINT and AEMT) when using OMNI and DIR microphone processing in a group of adults with hearing loss. Both tests have similar adaptive procedures, providing a SNR at which listeners understand 50% of the stimuli, rather than a percent correct score. Overall, the results of this study revealed that adults with hearing loss when tested with hearing aids perform differently on the HINT and AEMT, with the AEMT yielding significantly more negative (better) thresholds than the HINT. Despite differences in the tests slopes around the 50% point (HINT was steeper than AEMT), the benefit provided by DIR microphones was not statistically different between the tests. Because of this, we recommend that the investigators consider the SNR at which they expect to see improvements

with a given technology and the severity of the individual's speech recognition deficit when selecting between the HINT or AEMT. There remains a need for speech recognition in noise tests that can be administered at ecologically valid levels, although eliciting a range of performance that is not constrained by ceiling or floor effects.

REFERENCES

Bentler RA, Palmer C, Dittberner AB. (2004) Hearing in noise: comparison of listeners with normal and aided impaired hearing. *J Am Acad Audiol* 15:216–225.

Bentler RA, Tubbs JL, Egge JLM, Flamme GA, Dittberner AB. (2004) Evaluation of an adaptive directional system in a DSP hearing aid. *J Am Acad Audiol* 13:73–79.

Bernstein JGW. (2011) Controlling signal-to-noise ratio effects in the measurement of speech intelligibility in fluctuating maskers. *Proc Int Symp Auditory Audiological Res* 3:33–44.

Bess FH. (1983) *Clinical assessment of speech recognition*. In: Konkle D, Rintelmann W, eds. Principles of Speech Audiometry. Baltimore, MD: University Park Press, 127–201.

Best V, Keidser G, Buchhotz J, Freeston K. (2013) Psychometric effects of adding realism to speech in noise test. *Proc Mtgs Acoust* 19:1–6.

Clopper CG, Pisoni DB, Tierney AT. (2006) Effects of open set and closed set task demands on spoken word recognition. *J Am Acad Audiol* 17(5):331–349.

Cohen J. (1988) *Statistical power analysis for the behavioral sciences*. 2nd ed. New York, NY: Lawrence Erlbaum Associates, Publishers.

Faul F, Erdfelder E, Lang AG, Buchner A. (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedicalsciences. *Behav Res Methods* 39:175–191.

Hood JD, Poole JP. (1980) Influence of speaker and other factors affecting speech intelligibility. *Audiology* 19:434–455.

Kuk F, Keenan D, Sonne M, Ludvigsen C. (2005) Efficacy of an open fitting hearing aid. *Hear Rev* 12(2):26–32.

Kiolbasa AM. (2015) Evaluation of the American English matrix test with cochlear implant recipients (Doctoral Thesis). http://digitalcommons.wustl.edu/cgi/viewcontent.cgi?article=1709&context=pacs_capstones. Accessed April 19, 2017.

Kollmeier B, Warzybok A, Hochmuth S, Zokoll MA, Uslar V, Brand T, Wagener KC. (2015) The multilingual matrix test: principles, applications, and comparison across languages: a review. *Int J Audiol* 54(Suppl 2):3–16.

Laugesen S, Ronne FM, Jensen NS, Sorgenfrei MG. (2013) Validation of a spatial speech-in-speech test that takes signal-to-noise ratio (SNR) confounds into account. Paper presented at International Symposium on Auditory and Audiological Research,

Nyborg, Denmark. https://www.researchgate.net/publication/265077872. Accessed August 9, 2017.

Levitt H. (1970) Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49(2):467–477.

MacPherson A, Akeroyd MA. (2014) Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey. *Trends Hear* 18:1–26.

Mendel LL. (2007) Objective and subjective hearing aid assessment outcomes. *Am J Audiol* 16:118–129.

Naylor G, Johannesson RB. (2009) Long term signal to noise ratio at the input and output of amplitude-compression systems. *J Am Acad Audiol* 20:161–171.

Nilsson M, Soli SD, Sullivan JA. (1994) Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 95(2):1085–1099.

Nordrum S, Erler S, Garstecki D, Dhar S. (2006) Comparison of performance on the hearing in noise test using directional microphones and digital noise reduction algorithms. *J Am Acad Audiol* 15:81–91.

Pichora-Fuller MK, Schneider BA, Daneman M. (1995) How young and old adults listen to and remember speech in noise. *J Acoust Soc Am* 97(1):593–608.

Pumford JM, Seewald RC, Scollie SD, Jenstad LM. (2000) Speech recognition with in the ear and behind the ear dual microphone hearing instruments. *J Acad Audiol* 11:23–35.

Ricketts TA. (2001) Directional hearing aids. *Trends Amplif* 5(4):139–176.

Ricketts T, Dhar S. (1999) Comparison of performance across three directional hearing aids. *J Am Acad Audiol* 10:180–189.

Ricketts T, Mueller HG. (2000) Predicting directional hearing aid benefit for individual listeners. *J Am Acad Audiol* 11:561–569.

Rose HA. (2013) Reliability of the American English matrix test (Doctoral Thesis). https://mdsoar.org/bitstream/handle/11603/1949/TSP2012Rose.pdf?sequence=1&isAllowed=y. Accessed April 19, 2017.

Scheller T, Rosenthal J. (2012) Starkey hearing technologies' e-STAT fitting formula: the rationale behind the rationale. *Innovations* 2(2):41–45.

Smeds K, Wolters F, Rung M. (2015) Estimation of signal to noise ratios in realistic sound scenarios. *J Am Acad Audiol* 26:183–196.

Soli SD, Wong LN. (2008) Assessment of speech intelligibility in noise with the hearing in noise test. *Int J Audiol* 47(6):356–361.

Stuart A, Butler AK. (2014) No learning effects observed for reception thresholds for sentences in noise. *Am J Audiol* 23:227–231.

Vacha-Haase T, Thompson B. (2004) How to estimate and interpret various effect sizes. *J Couns Psychol* 51(4):473–481.

Valente M, Mispagel KM, Tchorz J, Fabry D. (2006) Effect of type of noise and loudspeaker array on the performance of omnidirectional and directional microphones. *J Am Acad Audiol* 17:398–412.

Valente M, Schuchman G, Potts LG, Beck LC. (2000) Performance of dual microphone: in the ear hearing aids. *J Am Acad Audiol* 11: 181–189.

Vermiglio AJ. (2008) The American English hearing in noise test. *Int J Audiol* 47(6):386–387.

Walden BE, Surr RK, Cord MT, Edward B, Olson L. (2000) Comparison of benefits provided by different hearing aid technologies. *J Am Acad Audiol* 11:540–560.

Wilson RH, Bell TS, Koslowski JA. (2003) Learning effects associated with repeated word recognition measures using sentence materials. *J Rehabil Res Dev* 40(4):329–336.

Wilson RH, McArdle RA, Smith SL. (2007) An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J Speech Lang Hear Res* 50:844–856.

Yund EW, Woods DL. (2010) Content and procedural learning in repeated sentence tests of speech perception. *Ear Hear* 31(6):769–778.

Zokoll M, Kollmeier B, Allen P, Folkeard P, Miller C, Ramkissoon I, Adams E, Bentler R. (2016) *Validation of the American English matrix sentence test in noise*. Paper Presented at the Annual Meeting of Audiology Now, Phoenix, AZ. https://www.eventscribe.com/2016/AudiologyNow/aaSearchByPresentation.asp?BCFO=P|SE|CE. Accessed January 19, 2017.