Sentence Recognition in Steady-State Speech-Shaped Noise versus Four-Talker Babble

Andrew J. Vermiglio*
Caroline C. Herring*
Paige Heeke*
Courtney E. Post*
Xiangming Fang†

Abstract

Background: Speech recognition in noise (SRN) evaluations reveal information about listening ability that is unavailable from pure-tone thresholds. Unfortunately, SRN evaluations are not commonly used in the clinic. A lack of standardization may be an explanation for the lack of widespread acceptance of SRN testing. Arguments have been made for the utilization of steady-state speech-shaped noise vs. multi-talker babble. Previous investigations into the effect of masker type have used a monaural presentation of the stimuli. However, results of monaural SRN tests cannot be generalized to binaural listening conditions.

Purpose: The purpose of this study was to investigate the effect of masker type on SRN thresholds under binaural listening conditions.

Research Design: The Hearing in Noise Test (HINT) protocol was selected in order to measure SRN thresholds in steady-state speech-shaped noise (HINT noise) and four-talker babble with and without the spatial separation of the target speech and masker stimuli.

Study Sample: Fifty native speakers of English with normal pure-tone thresholds ($\leq$ 25 dB HL, 250–4000 Hz) participated in the study. The mean age was 20.5 years (SD 1.01).

Data Collection and Analysis: All participants were tested using the standard protocol for the HINT in a simulated soundfield environment under TDH-50P headphones. Thresholds were measured for the Noise Front, Noise Left, and Noise Right listening conditions with HINT noise and four-talker babble. The HINT composite score was determined for each noise condition. The spatial advantage was calculated from the HINT thresholds. Pure-tone threshold data were collected using the modified Hughson-Westlake procedure. Statistical analyses include descriptive statistics, effect size, correlations, and repeated measures ANOVA followed by matched-pairs t-tests.

Results: Repeated measures ANOVA was conducted to investigate the effects of masker type and noise location on HINT thresholds. Both main effects and their interaction were statistically significant ($p < 0.01$). No significant differences were found between masker conditions for the Noise Front thresholds. However, for the Noise Side conditions the four-talker babble thresholds were significantly better than the HINT noise thresholds. Overall, greater spatial advantage was found for the four-talker babble as opposed to the HINT noise conditions ($p < 0.01$). Pearson correlation analysis revealed no significant relationships between four-talker babble and HINT noise speech recognition performances for the Noise Front, Noise Right conditions, and the spatial advantage measures. Significant relationships ($p < 0.05$) were found between masking noise performances for the Noise Left condition and the Noise Composite scores.

Conclusions: One cannot assume that a patient who performs within normal limits on a speech in four-talker babble test will also perform within normal limits on a speech in steady-state speech-shaped noise test, and vice-versa. Additionally, performances for the Noise Front condition cannot be used to predict performances for the Noise Side conditions. The utilization of both HINT noise and four-talker babble maskers, with and without the spatial separation of the stimuli, may be useful when determining the range of speech recognition in noise abilities found in everyday listening conditions.

*Department of Communication Sciences and Disorders, East Carolina University, Greenville, NC; †Department of Biostatistics, East Carolina University, Greenville, NC

Corresponding author: Andrew J. Vermiglio, Department of Communication Sciences and Disorders, East Carolina University, 3310P Health Science Building/Mail Stop 668, Greenville, NC 27834; Email: vermiglioa@ecu.edu

Key Words: audiology, pure-tone thresholds, spatial advantage, speech recognition in noise ability

Abbreviations: HINT = hearing in noise test; HRTF = head-related transfer function; KEMAR = Knowles Electronics Mannequin for Auditory Research; QuickSIN = Quick speech-in-noise; SD = standard deviation; SNR = signal-to-noise ratio; SRN = speech recognition in noise; WIN = Words-in-Noise

INTRODUCTION

In the field of audiology, pure-tone threshold testing has been considered the "gold-standard" for the evaluation of auditory function (Shargorodsky et al, 2010; Baiduc et al, 2013). However, according to the American Academy of Otolaryngology and the American Council of Otolaryngology, pure-tone audiometry was considered a tentative best-form of hearing assessment until the development of well-designed speech recognition in noise (SRN) tests became available (AAO-ACO, 1979). The authors stated that tests designed to evaluate auditory function in quiet would not provide an accurate measure of the ability to hear speech in everyday noisy environments. This has been demonstrated in multiple studies where individuals with normal pure-tone thresholds have reported speech perception difficulties in the presence of noise (King, 1954; Saunders and Haggard, 1989; Vermiglio, Soli, et al, 2017).

A number of protocols have been developed to assess the ability to recognize speech in noise. These protocols include the speech perception in noise test (Kalikow et al, 1977), the SRN test developed by Plomp and Mimpen (1979), the Hearing in Noise Test (HINT; Nilsson et al, 1994; Vermiglio, 2008), the Words-in-Noise (WIN) test (Wilson, 2003), the Quick Speech-in-Noise (QuickSIN) test (Killion et al, 2004), the Listening in Spatialized Noise-Sentences (LiSN-S) test (Cameron and Dillon, 2007), and the AzBio (Spahr et al, 2012). Even with the development of multiple protocols, SRN evaluations have not received widespread use in clinical settings. Mueller (2016) reported that although the HINT is very popular among researchers, the QuickSIN is the most commonly used SRN test in audiology clinics. However, according to Mueller in a survey of 107 dispensing audiologists, only 10% reported routinely using the QuickSIN. ASHA (2015) noted in a survey of 1,811 audiologists that only 30% routinely conduct a validation of outcomes using SRN testing.

One of the obstacles to the widespread implementation of SRN testing has been a lack of standardization. In a review of the variables for SRN protocols, Theunissen et al (2009) reported that two of the most commonly used maskers are steady-state speech-shaped noise and multitalker babble. Arguments have been made regarding the appropriateness of these maskers. Carhart and Tillman (1970) suggested that steady-state maskers, such as white noise or speech-spectrum noise, "may prove inadequate and unsatisfactory because they probably elicit less enhancement of masking than does competing speech." Wilson, Carnell, et al (2007) noted that multitalker

babble has more face validity than steady-state speech-shaped noise because listeners with hearing loss complain of difficulty understanding speech in noise, especially when the noise is composed of multiple talkers as found in a restaurant or other social settings. Killion et al (2004) stated that, "the use of continuous noise has the advantage of reducing the variability in noise level and the disadvantage that it is less representative of everyday speech-in-noise situations than babble noise." On the other hand, Soli and Wong (2008) noted that babble noise may introduce the confound of "informational masking."

Clinical SRN tests have been used to investigate the effect of masker type on SRN ability. Wilson, Carnell, et al (2007) used the WIN to measure the ability to recognize speech in the presence of speech-spectrum noise versus six-talker babble for participants with normal pure-tone thresholds. The results demonstrated that speech recognition performances were better in babble as opposed to speech-spectrum noise. Jin and Liu (2012) evaluated sentence recognition in long-term speech-shaped noise and 12-talker babble for participants with normal pure-tone thresholds. SRN ability was determined by using a nonstandardized version of the HINT. Consistent with the findings of Wilson et al, Jin and Liu reported better SRN performances in 12-talker babble than in long-term speech-shaped noise. In both of these studies, the target speech and masker stimuli were delivered monaurally. However, monaural SRN ability cannot be generalized to binaural listening conditions (Vermiglio, Griffin, et al, 2017).

Bronkhorst and Plomp (1990) recommended binaural SRN assessment in listening conditions where the target speech and masker are spatially separated. These authors argued that a practical approach to SRN testing would be to model the stimuli and listening conditions after circumstances in daily life. In this regard, they reasoned that sentences would be a better choice than word lists. They also selected 0° as the most natural azimuth for the presentation of the target speech. The authors recommended an evaluation of binaural cues on SRN ability by including test conditions with and without the spatial separation of the speech and masker stimuli. This protocol was designed to allow for the generalization of test results to listening experiences in daily life. Bronkhorst and Plomp (1988) evaluated SRN performances where the target speech was presented at 0°, and the masker was delivered from seven azimuths; 0 to 180° in 30° steps. The "spatial advantage" represents the improvement in SRN performance as the masker is spatially separated from the target speech.

The authors reported that the greatest spatial advantage was found when the masker was presented from 90°. Bronkhorst and Plomp also demonstrated that both interaural level differences and interaural time delays play a role for the spatial advantage.

Purpose

A review of the literature has shown that SRN evaluations reveal information about listening ability in daily listening environments unavailable from pure-tone thresholds (King, 1954; Saunders and Haggard, 1989; Vermiglio, Soli, et al, 2017). Unfortunately, SRN evaluations are not commonly used in the clinic (ASHA, 2015; Mueller, 2016). A lack of standardization may be an explanation for the lack of widespread acceptance of SRN testing. Arguments have been made for the utilization of steady-state speech-shaped noise vs. multitalker babble (Theunissen et al, 2009). Previous investigations into the effect of masker type have used a monaural presentation of the stimuli (Wilson, Carnell, et al, 2007; Jin and Liu, 2012). However, results of monaural SRN tests cannot be generalized to binaural listening conditions (Vermiglio, Griffin, et al, 2017).

The purpose of this study was to investigate the effect of masker type on SRN thresholds under binaural listening conditions with and without the spatial separation of the target speech and masker stimuli. The HINT protocol was selected to measure SRN thresholds in steady-state speech-shaped noise (HINT noise) and four-talker babble. The binaural listening conditions included presentations of the masker stimuli from 0°, 90° and −90°. Specifically, the following was hypothesized:

- Performances in four-talker babble would be significantly better than performances in HINT noise.
- The spatial advantage for the HINT noise would be less than the spatial advantage found for four-talker babble.
- Significant relationships would be found between four-talker babble and HINT noise thresholds.
- Significant relationships would be found between four-talker babble and HINT noise spatial advantages.

METHODS

Permission to conduct this research study was obtained from the East Carolina University Institutional Review Board. All participants had pure-tone thresholds (≤25 dB HL from 250 to 8000 Hz) with the exception of one participant who had pure-tone thresholds of 30 dB HL for 8000 Hz for both ears and for 6000 Hz for the left ear only. Fifty native speakers of American English (47 females and three males) participated in this study. The participants ranged in age from 19 to 25 yr with an average age of 20.5 yr (standard deviation [SD] 1.01). This convenience sample was made up of undergraduate students in the Department of Communication Sciences and Disorders at East Carolina University.

The HINT was used to measure the ability to recognize speech in HINT noise and four-talker babble. Short, simple sentences were presented in noise at a fixed level of 65 dBA. Each HINT threshold was measured using a single list of 20 sentences. Sentence lists were randomly selected from a set of 12 lists. Testing was conducted in a simulated soundfield environment using Knowles Electronic Mannequin for Auditory Research (KEMAR) head-related transfer functions (HRTFs). The target sentences were presented at 0° for each test condition. The maskers were presented at 0°, 90°, and −90° for the Noise Front, Noise Right, and Noise Left conditions, respectively. In addition, the HINT composite scores were determined for each masker type. The HINT composite score is an average of the three noise thresholds where the Noise Front threshold is weighted twice using the formula: (2 × Noise Front + Noise Right + Noise Left)/4. This score provides a single index of overall speech recognition in background noise (Vermiglio, 2007; Soli and Wong, 2008). All test conditions were randomized. Telephonic TDH-50P headphones were used to deliver the stimuli.

The noise conditions presented virtually under headphones are illustrated in Figure 1. There were two noise conditions. The standard steady-state speech-shaped noise for the HINT has the same spectrum as the long-term average spectrum of the HINT sentences (Nilsson et al, 1994). For the four-talker babble condition, the same Auditec source file licensed for use with the QuickSIN was used. The four-talker babble includes full sentences and phrases (Auditec: personal communication March 29, 2017). For the present study, the four-talker babble and HINT noise maskers were equated in root mean square level. Figure 2 shows the spectra for the two maskers. Although no attempt was made to spectrally match the four-talker babble to the HINT noise, the spectra of the two maskers were relatively close. The overall shape of the spectra is similar, although not identical. Even though the HINT noise is spectrally matched to the long-term spectra of the HINT target sentences, the target speech and HINT noise are not matched in spectra once the noise azimuth changes from 0° to ±90°. In addition, the HINT noise presented to the right and left ears are spectrally different for the Noise Side conditions. This is due to the presence of the head shadow effect by way of the KEMAR HRTFs.

The HINT uses an adaptive protocol where the level of the sentence presentations varies based on the response of the participant. The participant's task is to listen to and repeat the sentence heard in the presence of the noise. If the participant correctly repeats the sentence,
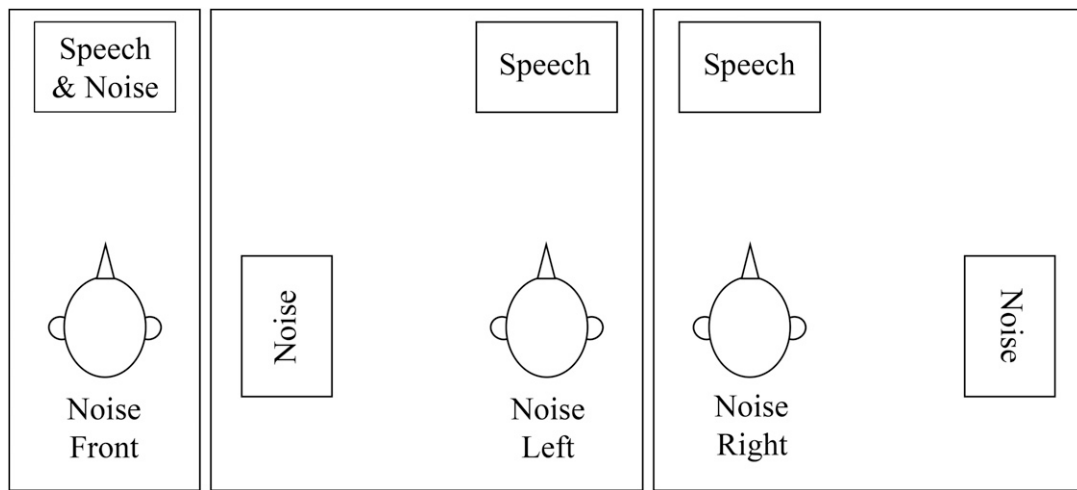
Figure 1. HINT soundfield conditions simulated under headphones using KEMAR HRTFs.

the level of the speech for the following sentence is decreased. If the participant incorrectly repeats the sentence, the level of the speech for the following sentence is increased. A 4-dB step size is used for the first four sentence presentations. A 2-dB step size is used for the remaining sentences. The HINT threshold is the signal-to-noise ratio (SNR) where a participant recognizes 50% of the sentences. The "variability" associated with the HINT threshold is also determined. This is the SD of the SNRs used for each test run. It is a measure of the stability or consistency of the subject's responses. The spatial advantage was determined by subtracting the Noise Side threshold (Noise Right or Noise Left) from the Noise Front threshold for each masker condition. This represents the improvement in SRN ability when the target-speech and masker stimuli are spatially separated. The HINT test was administered using commercially available software provided by the House Ear Institute in Los Angeles, CA.
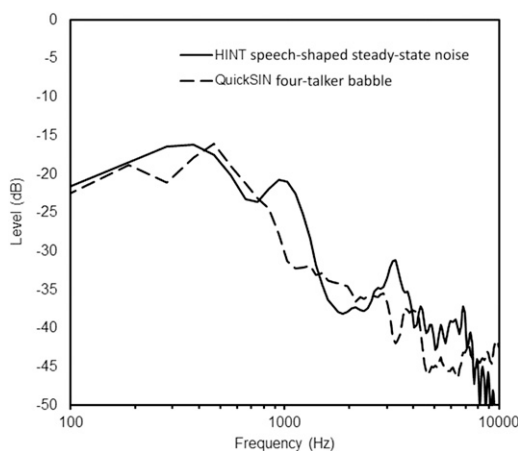
Figure 2. Spectra of the HINT steady-state speech-shaped noise and the four-talker babble from the QuickSIN test.

Statistical analyses were conducted using the JMP Pro (V.12) software. The matched pair *t*-test was conducted to determine if there were statistically significant differences between the thresholds for the masker conditions. The correlation coefficient was found between HINT noise and four-talker babble measures. Repeated measures analysis of variance was used to investigate the within-subject effects of masker type and masker location on threshold, followed by pairwise comparisons with Bonferroni correction.

RESULTS

The descriptive statistics for the variability of each threshold are presented in Table 1. Recall that the variability is the SD for each threshold run. The average mean variability across all HINT conditions was 2.05 dB. This indicates that the participant responses were consistent across the test conditions. The descriptive statistics for all HINT measures are presented in Table 2. The more negative the threshold in dB SNR, the better the SRN performance. For example, the average binaural Noise Front threshold for the four-talker babble condition was -1.89 dB SNR. This indicates that when the speech signal is 1.89 dB below the level of the masker, the participants on average recognize 50% of the sentences. For the HINT noise, the difference between the maximum and minimum thresholds for the Noise Front condition was 3.6 dB. Nilsson and Soli (1994) reported that a 1 dB change in threshold represents a 10% change in intelligibility. Therefore, the 3.6 dB difference in maximum and minimum thresholds corresponds to a 36% change in intelligibility. The average difference between the maximum and minimum thresholds for the Noise Side conditions was 6.05 dB. A 6.05 dB difference corresponds to a 60.5% change in intelligibility between the maximum and minimum

Table 1. Descriptive Statistics of the Variability for All HINT Thresholds

	Four-Talker Babble Front	HINT Noise Front	Four-Talker Babble Right	HINT Noise Right	Four-Talker Babble Left	HINT Noise Left
Mean	2.00	1.88	2.23	2.06	2.13	1.99
SD	0.50	0.28	0.51	0.55	0.47	0.40
n	50	50	50	50	50	50
Maximum	3.7	2.7	3.6	3.9	3.3	3.2
Minimum	1.2	1.4	1.3	1.4	1.3	1.4
Range	2.5	1.3	2.3	2.5	2	1.8

Note: Variability = SD for each HINT threshold run.

performances. Recall that all of the participants had normal pure-tone thresholds: even so, the range in SRN ability is notably large especially for the Noise Side conditions. The spatial advantage measures are also presented in Table 2. Greater spatial advantages were found for the four-talker babble condition when compared with the HINT noise conditions ($p < 0.01$).

A repeated measures analysis of variance was conducted to investigate the effects of masker type and masker location (Noise Front, Noise Left, and Noise Right) on HINT thresholds and revealed that both main effects and their interaction were statistically significant (F values are 39.466 for type, 1,270.077 for position, and 13.214 for interaction; all p-values <0.0001). This indicates there were significant differences in threshold scores between HINT noise and four-talker babble performances and among the three locations of noise. The interaction was significant, which means there were performance differences among combinations of noise type and location. Figure 3 demonstrates the interaction between masker type and the location of the noise for SRN ability.

Post hoc analysis was conducted by pairwise comparisons. The matched-pairs t-test results with Bonferroni corrections and effect size are presented in Table 2. No significant difference was found between the average

four-talker babble and HINT noise thresholds for the Noise Front condition. A scatter-plot of the Noise Front thresholds is presented in Figure 4. Data above the diagonal line represent performances that were better (more negative) for the HINT noise than the four-talker babble masker and vice versa. A HINT noise advantage was found for 50% (25) of the participants. In other words, half of the participants had better performances in HINT noise than in the four-talker babble. A four-talker babble advantage was found for 44% (22) of the participants. Three of the participants (6%) demonstrated no masker advantage.

For the Noise Side conditions, performances were significantly better ($p < 0.01$) for the four-talker babble than HINT noise masker (Table 2). The differences were −1.63 and −1.23 dB for the Noise Right and Noise Left conditions, respectively. Scatter-plots of the Noise Side thresholds are presented in Figure 5. Data above the diagonal lines represent performances that were better (more negative) for the HINT noise condition than the four-talker babble condition and vice versa. For the Noise Left thresholds, a HINT noise advantage was found for 24% (12) of the participants and a four-talker babble advantage was found for 76% (38) of the participants. For the Noise Right thresholds, a HINT noise advantage was found for 16% (8) of the participants and a four-talker

Table 2. Descriptive and *t*-Test Statistics for All HINT Measures with Bonferroni Correction

| Variable	Noise Type	Mean (dB SNR)	SD	Minimum (dB SNR)	Maximum (dB SNR)	Difference (dB)	Effect Size	*p*	Bonferroni Correction
Noise Front threshold	Four-talker babble	−1.89	1.18	−4.10	1.60	0.13	0.10	0.4926	1.0000
	HINT noise	−2.02	0.87	−3.70	−0.10				
Noise Right threshold	Four-talker babble	−10.76	2.07	−16.60	−5.60	−1.63	0.74	<0.0001	<0.0007
	HINT noise	−9.14	1.32	−12.30	−6.10				
Noise Left threshold	Four-talker babble	−9.88	1.67	−12.60	−5.20	−1.23	0.70	<0.0001	<0.0007
	HINT noise	−8.65	1.33	−11.00	−5.10				
Noise composite score	Four-talker babble	−6.11	1.03	−8.18	−3.93	−0.65	0.73	<0.0001	<0.0007
	HINT noise	−5.46	0.83	−6.93	−3.33				
Spatial advantage (Right)	Four-talker babble	8.87	2.30	2.60	13.70	1.75	0.65	<0.0001	<0.0007
	HINT noise	7.12	1.31	4.10	9.10				
Spatial advantage (Left)	Four-talker babble	7.98	1.87	1.50	11.80	1.36	0.58	0.0002	0.0014
	HINT noise	6.63	1.46	2.90	9.40				
Average spatial advantage	Four-talker babble	8.43	1.79	3.10	12.10	1.56	0.70	<0.0001	<0.0007
	HINT noise	6.87	1.25	3.50	9.50				
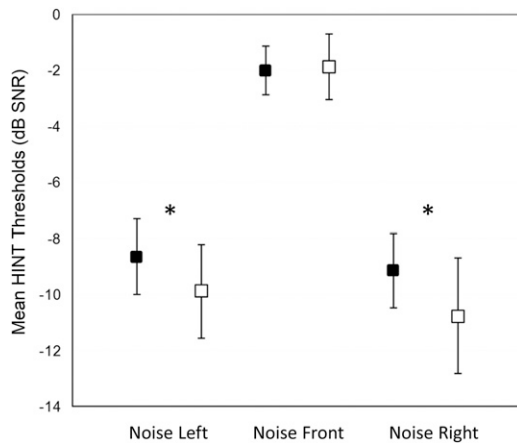
Figure 3. Mean thresholds for each HINT and noise condition. Filled and open squares represent thresholds with HINT noise and four-talker babble, respectively. SDs are represented by the vertical bars. The asterisks denote statistically significant differences between masker pairs for each HINT condition.

babble advantage was found for 82% (41) of the participants. One of the participants demonstrated no masker advantage for the Noise Right thresholds.

Noise Left versus Noise Right performances for the four-talker babble and HINT noise conditions are displayed in Figure 6. Data above the diagonal lines represent performances that were better (more negative) for the Noise Right condition than the Noise Left condition and vice versa. For the four-talker babble condition, a Noise Right advantage was found for 66% (33) of the participants and a Noise Left advantage was found for 32% (16) of the participants. One of the participants demonstrated no masker side advantage. For the HINT noise condition, a Noise Right advantage was found for 66% (33) of the participants and a Noise Left advantage
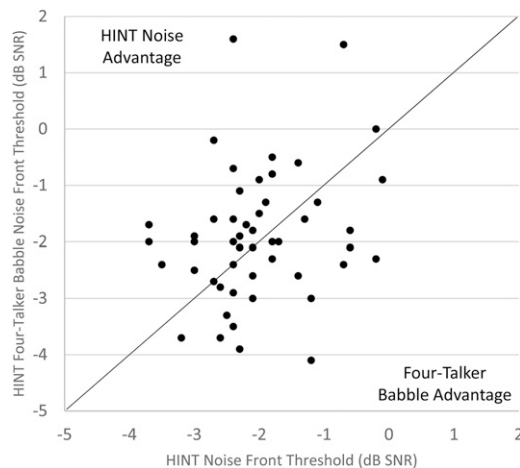

Figure 4. Scatter-plot of HINT Noise Front thresholds for HINT noise vs. four-talker babble listening conditions (n = 50). Data above the diagonal line represent performances that were better (more negative) for the HINT noise than the four-talker babble conditions and vice versa.

was found for 34% (17) of the participants. A stronger relationship was found between the Noise Side thresholds for the HINT noise ($r = 0.51$, $p < 0.05$) than the four-talker babble condition ($r = 0.34$, $p < 0.05$). Small but statistically significant improvements were found for the Noise Right over the Noise Left thresholds for the HINT noise (0.49 dB, $p < 0.05$) and the four-talker babble (0.89 dB, $p < 0.05$) conditions. For the Noise Right condition, the right ear is the unshadowed ear and the left ear is the shadowed ear (Figure 1). Because of the presence of the head-shadow effect, the left ear has a better SNR than the right ear, especially for the high frequencies. The opposite occurs for the Noise Left condition.

For the Noise composite scores, performances in four-talker babble were significantly better (0.65 dB, $p < 0.05$) than the performances in HINT noise (Table 2). Recall that the Noise composite score represents the overall HINT performance where the Noise Front and Noise Side conditions are equally weighted. The scatter-plot in Figure 7 displays the composite scores for the four-talker babble versus HINT noise listening conditions. A HINT noise advantage was found for 24% (12) of the participants, and a four-talker babble advantage was found for 71% (38) of the participants. The fifth percentile is used as the cut-point for normal HINT performance (HEI, 2007). Thresholds below (or more positive than) the fifth percentile are considered below normal limits. The fifth percentile cut-points for the four-talker babble (-4.41 dB SNR) and HINT noise (-4.10 dB SNR) are represented by the dashed lines in Figure 7. The dashed lines delineate four quadrants (I–IV). Data in quadrant I (88%) are within normal limits for both noise conditions. Data in quadrant II (8%) are within normal limits for the HINT noise condition and below normal limits for the four-talker babble condition. Quadrant III represents performances below normal limits for both masker conditions (0%). Data in quadrant IV (4%) represent performances below average for the HINT noise condition and within normal limits for the four-talker babble condition. The data in quadrants II and IV demonstrate that 12% of the performances were within normal limits for one masker condition and below normal limits for the other masker condition.

Table 3 shows the four-talker babble advantage (HINT noise threshold minus the four-talker babble threshold) for each HINT threshold and the composite score. A positive result indicates that mean SRN performances were better with the four-talker babble than HINT noise. Although a relatively small range for four-talker babble advantage was found for the composite score (3.95 dB), a relatively large range was found for the Noise Right (11.8 dB) and Noise Left (9.1 dB) conditions.

The correlation coefficients for the HINT noise versus four-talker babble performances are presented in Table 4. No significant correlations between masking conditions
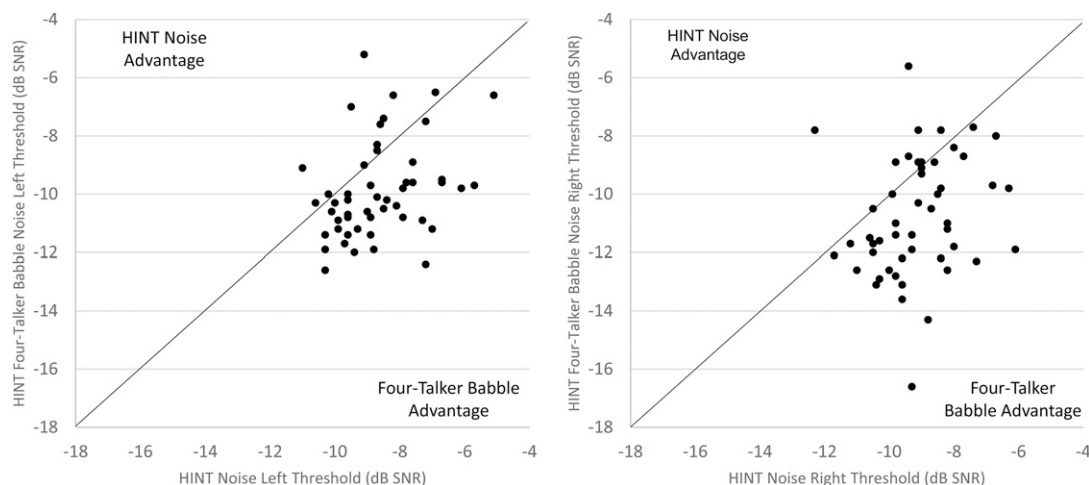
Figure 5. Scatter-plot of HINT Noise Side thresholds for HINT noise vs. four-talker babble listening conditions (n = 50). Data above the diagonal line represent performances that were better (more negative) for the HINT noise than the four-talker babble conditions and vice versa. Data for the Noise Left thresholds are on the left and data for the Noise Right Thresholds are on the right.

were found for the Noise Front or Noise Right conditions. Statistically significant correlations were found between masker conditions for the Noise Left thresholds ($r = 0.33$, $p < 0.05$) and Noise composite scores ($r = 0.56$, $p < 0.01$). No significant correlations between masker types were found for any of the spatial advantage measures.

DISCUSSION

Recall that the first hypothesis stated that performances in four-talker babble would be significantly better than performances in HINT noise. This was supported by results for the Noise Side conditions where the four-talker babble thresholds were, on average, 1.43 dB better than the HINT noise thresholds ($p < 0.01$). However, it was not supported by the results for the Noise

Front thresholds. The second hypothesis stated that the spatial advantage for the HINT noise would be less than the spatial advantage found for the four-talker babble. This was supported by the results. The spatial advantage for the four-talker babble was 1.56 dB greater than the spatial advantage for the HINT noise ($p < 0.01$). The third hypothesis stated that significant relationships would be found between four-talker babble and HINT noise thresholds. This was supported by the results for the Noise Left thresholds and the composite scores ($p < 0.05$). However, it was not supported by the results for the Noise Front and Noise Right thresholds. The fourth hypothesis stated that significant relationships would be found between the spatial advantages for the four-talker babble and HINT noise conditions. However, this was not supported by the results.
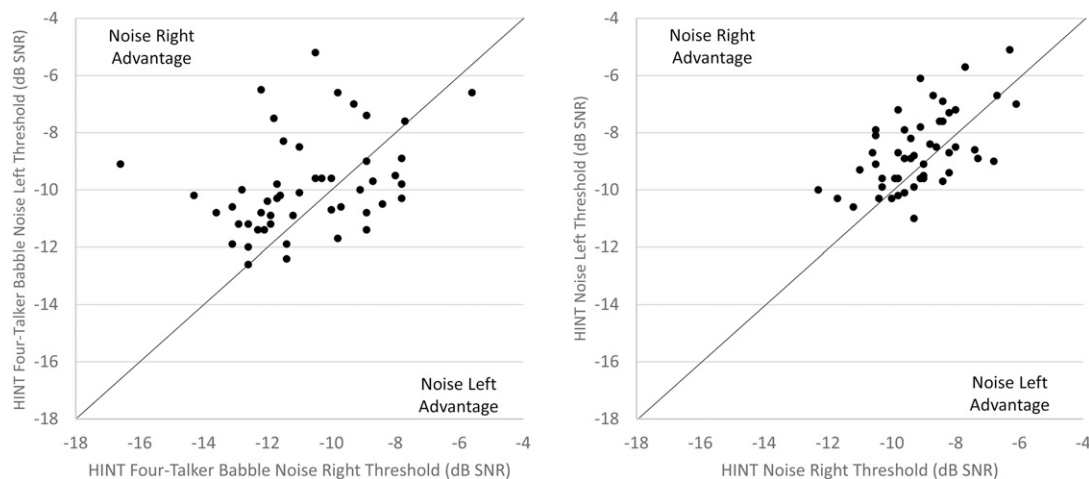

Figure 6. Scatter-plot of the Noise Left vs. Noise Right thresholds (n = 50) for both maskers. Data for the four-talker babble condition are on the left and data for the HINT noise condition are on the right. Data above the diagonal line represents performances that were better (more negative) for the Noise Right than the Noise Left thresholds and vice versa.
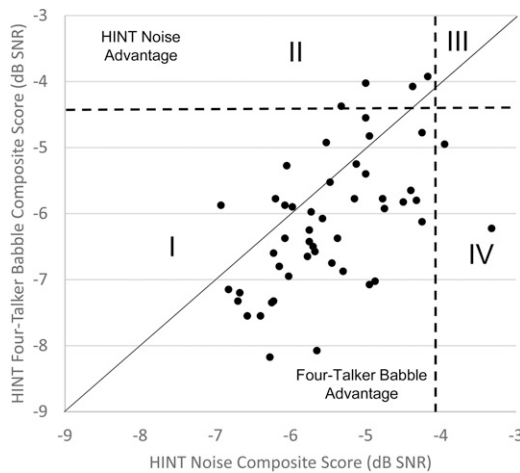
Figure 7. Scatter-plot of HINT composite scores for HINT noise vs. four-talker babble listening conditions (n = 50). Data above the diagonal line represent performances that were better (more negative) for the HINT noise than the four-talker babble conditions and vice versa. The dashed lines represent the fifth percentile cut-points for the HINT noise and four-talker babble thresholds.

A comparison of the HINT composite scores for the four-talker babble and HINT noise conditions revealed a small but statistically significant difference (-0.65 dB, $p \leq 0.01$). Most participants (76%) demonstrated better speech recognition ability in four-talker babble than in HINT noise (Figure 6). This is consistent with Wilson, Carnell, et al (2007) who reported a 2.3 dB improvement in speech recognition ability in six-talker babble as opposed to steady-state speech-shaped noise. Likewise, Jin and Liu (2012) reported better speech recognition performances in 12-talker babble than steady-state speech-shaped noise (with the same spectrum of the noise as the 12-talker babble). The average performance in 12-talker babble was 31% better than in steady-state speech-shaped noise. Improvements in multitalker babble relative to steady-state speech-shaped noise are thought to be due to the opportunities for listening in the gaps of the fluctuating amplitude of the babble masker.

Similar to babble noise, fluctuating noise also gives listeners an opportunity to listen in the gaps. Middelweerd et al (1990) reported better SRN performances

in fluctuating noise than in speech-shaped steady-state noise. Stuart and Butler (2014) investigated HINT sentence recognition in broadband noise with and without amplitude modulation. According to Stuart and Phillips (1996), the random gating of the interrupted noise, with durations from 5 to 95 msec, was designed to mimic the amplitude modulations found in speech. The stimuli were delivered monaurally (without KEMAR HRTFs). For the first of a series of five trials, every one of the participants performed better with the interrupted noise than the continuous (steady-state) noise. Stuart and Butler noted that the improved performance in interrupted noise was due to the participants' auditory temporal resolution. By contrast, no significant difference was found between masker performances for the Noise Front condition for the present study. In the Stuart and Butler study, the interrupted noise was analogous to a single-talker babble minus the semantic content. In the present study, the four-talker babble most likely gave fewer opportunities for participants to listen in the gaps than the interrupted noise in the study by Stuart and Butler. In the present study, the HINT Noise advantage found for a number of participants (Figure 7) may be interpreted as evidence that the four-talker babble provided semantic interference (Carhart et al, 1969) to the SRN performances. It may also be interpreted as evidence of the inability of the listener to take advantage of listening in the gaps due to poor temporal resolution (Middelweerd et al, 1990; Stuart and Butler, 2014).

Both Killion et al (2004) and Wilson, Carnell, et al (2007) have commented that babble is a closer representation of noise encountered in daily life than steady-state noise. Killion et al used four-talker babble with the QuickSIN, and Wilson et al used six-talker babble with the WIN. This begs the question regarding the number of talkers in crowd noise that patients may encounter in daily life. Of course, social gatherings may include greater numbers of talkers than those found in the babble used with the QuickSIN and WIN tests. A review of eight of the smallest restaurants in the United States indicated seating capacities at a single table that ranged from 10 to 16 (Ferst, 2016). On the extreme end, the maximum seating capacity for a restaurant is currently 6,014 at the Damascus Gate restaurant in Syria (Turnbull, 2008).

An argument could be made that to generalize SRN test performance to communication in daily life, the multitalker babble should include a larger number of talkers than found in the QuickSIN or WIN babble. However, as the number of talkers increases, the opportunities for listening in the gaps decreases. The decision to use multitalker babble or steady-state speech-shaped noise should depend on the primary research or clinical question. For example, Dickson et al (1946) selected a presumably steady-state aircraft noise for the SRN evaluation of Royal Air Force candidates. The Royal Air

Table 3. Descriptive Statistics of the Four-Talker Babble Advantage (HINT Noise Minus Four-Talker Babble Thresholds) for All Measures

	Noise Front (dB)	Noise Right (dB)	Noise Left (dB)	Composite Score (dB)
Mean	-0.13	1.63	1.23	0.65
SD	1.29	2.19	1.76	0.89
n	50	50	50	50
Maximum	2.9	7.3	5.2	2.9
Minimum	-4	-4.5	-3.9	-1.05
Range	6.9	11.8	9.1	3.95

Table 4. Pearson Correlation Coefficients and *p*-Values (in Parentheses) between HINT Noise and Four-Talker Babble Performances

	Noise Front	Noise Right	Noise Left	Noise Composite	Spatial Advantage (Right)	Spatial Advantage (Left)	Average Spatial Advantage
HINT noise vs. Four-talker babble	0.24 (0.0983)	0.23 (0.1191)	**0.33 (0.0185)**	**0.56 (< 0.0001)**	−0.06 (0.6713)	0.02 (0.9172)	−0.03 (0.8327)

Note: Statistically significant relationships are in bold font.

Force SRN protocol did not evaluate the candidate's ability to recognize speech in a babble masker, nor did it address the issue of listening in the gaps of fluctuating noise. However, the selected masker was consistent with the primary goal: to determine the ability of the candidates to recognize speech in their potential work environment.

The Optimization of a SRN Protocol

According to the AAO-ACO (1979), speech perception assessments have not been standardized because of the plethora of the test variables. As noted earlier, arguments have been made for steady-state speech-shaped noise versus multitalker babble. Another parameter under consideration is the spectral-matching of the maskers to a reference speech spectrum. However, there is no standardization for the spectral-matching of masker stimuli in the literature. Some studies have used the target speech for the reference spectrum (Middelweerd et al, 1990; Bronkhorst and Plomp, 1992; Nilsson et al, 1994). Other studies have identified a babble masker for the reference spectrum (Sperry et al, 1997; Wilson, Carnell, et al, 2007; Jin and Liu, 2012; Rosen et al, 2013). Moreover, some studies have used the maskers "as is" without any attempts to match their spectra to a reference speech stimulus (Simpson and Cooke, 2005; Lee et al, 2015).

The decision to spectrally match the maskers to a reference spectrum appears to be by convention. However, some investigators have provided a rationale for the decision to use a speech-shaped noise. Plomp and Mimpen (1979) developed a SRN test using noise with the same long-term average spectrum as the spectrum of 130 target sentences. They reasoned that because one or more talkers may be considered the main source of interference in everyday listening environments, noise with the same spectrum as speech should be adopted as the standard noise for their evaluations. They also noted that this spectrum is similar to the spectrum of traffic noise. Nilsson et al (1994) modeled the development of the HINT after the SRN protocol described by Plomp and Mimpen (1979). Nilsson et al reported that their decision to use steady-state speech-shaped noise was based on a study by Prosser et al (1990). Prosser et al evaluated SRN performances in "speech noise," cocktail party noise, traffic noise, and continuous discourse. The speech noise

data showed the steepest intelligibility functions, from 0 to 5 dB SNR, when compared with the other maskers. Nilsson et al surmised that this indicated that the speech-shaped noise was the most sensitive masker to changes in speech discrimination. However, the Prosser et al data also revealed that the speech-shaped noise data failed to show a delineation between SRN performances for young participants with normal pure-tone thresholds and older participants with hearing loss. By contrast, statistically significant differences in SRN performances were found between these two groups for the cocktail party noise, traffic noise, and continuous discourse masker. There was no indication that these maskers were spectrally matched to a reference speech stimulus.

Implications for Clinical Work and Research

Regardless of the reasoning behind the development of a test protocol and materials, the ultimate test of a SRN protocol should be its ability to detect the presence and absence of a target disorder. The disorder could be an SRN disorder (Middelweerd et al, 1990; Vermiglio, Soli, et al, 2017), elevated pure-tone thresholds (Wilson, McArdle, et al, 2007), a lesion of the central auditory nervous system (Sinha, 1959; Richburg et al, 2017), or perhaps auditory neuropathy (Berlin, 2012). Clinicians and researchers should know the diagnostic accuracy of the available SRN protocols. This will enable users to select the test(s) with the greatest validity.

A diagnostic accuracy study requires the procurement of a reasonable reference standard test for the independent verification of the target disorder (Bossuyt et al, 2003). For example, Vermiglio, Soli, et al (2017) have argued that self-report is a reasonable reference standard for a SRN disorder. Previous investigators have used self-report as a reference standard to delineate between the presence and absence of various target disorders such as tinnitus (Schaette and McAlpine, 2011), pain (Stilma et al, 2015), hearing loss (Beasley, 1940; Steinberg et al, 1940), and an SRN disorder (Saunders and Haggard, 1989; Middelweerd et al, 1990; Zhao and Stephens, 2006). Vermiglio et al used self-report as a reference standard for the determination of the diagnostic accuracy of both HINT thresholds and pure-tone average. Two groups of participants with normal pure-tone thresholds participated. The groups were matched for age and

had a range from 24 to 53 yr. The participants were placed in the control group (n = 22) or the disordered group (n = 25) based on their self-report of the ability to recognize speech in a noisy environment such as a crowded restaurant. The authors reported sensitivity of 88% and 28% for the HINT composite score and pure-tone average, respectively. They also reported that the specificity was 77% and 95% for the HINT composite score and pure-tone average, respectively. Overall, the HINT demonstrated greater diagnostic accuracy for the detection of a SRN disorder than the pure-tone threshold measure.

The results of the present study have implications for SRN evaluations for fitness-for-duty and for special classroom accommodations. The ability to recognize speech in noise on the job or in a classroom should be measured directly and not be inferred from normal pure-tone thresholds. SRN results should be compared with normative data to determine the presence of an SRN disorder. Intervention for an SRN deficit may include auditory training (Sweetow and Sabes, 2006), a mild gain hearing aid with a directional microphone (Kuk et al, 2008), or a frequency modulation system (Johnston et al, 2009). The benefit of intervention should be evaluated using a soundfield SRN protocol. Pre- and post-intervention thresholds should be measured to document the extent of the benefit.

Results from the present study may be used to counsel patients regarding the potential improvement in SRN ability when the target speech and masker sources are spatially separated. This may help guide the patient toward improving the favorability of their daily listening environments. Poorer performances in a babble masker than steady-state noise may imply a temporal resolution disorder and/or the presence of semantic interference. This may prompt further evaluations in these specific areas.

Limitations of the Present Study

The goal of the present study was to evaluate SRN ability in HINT noise and four-talker babble for young participants with normal pure-tone thresholds. The ability to recognize sentences was evaluated using the HINT adaptive protocol in a virtual soundfield environment. The HINT protocol was modeled in part after the work by Plomp and Mimpen (1979). The results of the present study may not be similar to results obtained using different protocols, speech materials, source locations for the stimuli, types of HRTFs, masker types, or alternative forms of spectral-shaping of the maskers. Moreover, the results may not be generalizable to older adult participants or participants with elevated pure-tone thresholds. The effect of these variables on speech recognition in HINT noise and four-talker babble should be addressed in future work.

CONCLUSION

According to the present results, a SRN deficit in four-talker babble may exist in the presence of normal SRN ability in HINT noise and vice versa. It may be appropriate to use both types of maskers when evaluating SRN ability. The results of the present study revealed poor relationships between the Noise Front and Noise Side thresholds. In other words, Noise Side performances may not be inferred from Noise Front thresholds. This supports the utilization of a SRN protocol with and without the spatial separation of the target speech and masker stimuli.

REFERENCES

AAO-ACO. (1979) Guide for the evaluation of hearing handicap. *JAMA* 241(19):2055–2059.

ASHA. (2015) 2014 audiology survey report: clinical focus patterns. www.asha.org. Accessed March 3, 2017.

Baiduc RR, Poling GL, Hong O, Dhar S. (2013) Clinical measures of auditory function: the cochlea and beyond. *Dis Mon* 59(4): 147–156.

Beasley WC. (1940) Characteristics and distribution of impaired hearing in the population of the United States. *J Acoust Soc Am* 12:114–121.

Berlin CI. (2012) Auditory neuropathy spectrum disorder, OAEs, ABR, and more. http://www.audiology.org/news/interview-charles-berlin-phd. Accessed August 25, 2016..

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HC, Lijmer JG; Standards for Reporting of Diagnostic Accuracy Group. (2003) The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Croat Med J* 44(5):639–650.

Bronkhorst AW, Plomp R. (1988) The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J Acoust Soc Am* 83(4):1508–1516.

Bronkhorst AW, Plomp R. (1990) A clinical test for the assessment of binaural speech perception in noise. *Audiology* 29(5):275–285.

Bronkhorst AW, Plomp R. (1992) Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *J Acoust Soc Am* 92(6):3132–3139.

Cameron S, Dillon H. (2007) Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). *Ear Hear* 28(2):196–211.

Carhart R, Tillman TW. (1970) Interaction of competing speech signals with hearing losses. *Arch Otolaryngol* 91(3):273–279.

Carhart R, Tillman TW, Greetis ES. (1969) Perceptual masking in multiple sound backgrounds. *J Acoust Soc Am* 45(3):694–703.

Dickson ED, Simpson JF, Fry DB, Swindell GE, Brown REC. (1946) A new method of testing the hearing efficiency of aviation candidates. *J Laryngol Otol* 61(3):139–203.

Ferst D. (2016) Eight of the smallest restaurants in the U.S. https://www.tastingtable.com/dine/national/smallest-restaurants-usa-talula-blanca-minibar. Accessed August 28, 2017.

HEI. (2007) *HINT Pro User's and Service Manual.* Illinois: Mundelein.

Jin SH, Liu C. (2012) English sentence recognition in speech-shaped noise and multi-talker babble for English-, Chinese-, and Korean-native listeners. *J Acoust Soc Am* 132(5):EL391–EL397.

Johnston KN, John AB, Kreisman NV, Hall JW 3rd, Crandell CC. (2009) Multiple benefits of personal FM system use by children with auditory processing disorder (APD). *Int J Audiol* 48(6):371–383.

Kalikow DN, Stevens KN, Elliott LL. (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am* 61(5):1337–1351.

Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S. (2004) Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 116(4 Pt 1):2395–2405. Erratum in: *J Acoust Soc Am* 2006;119(3):1888

King PF. (1954) Psychogenic deafness. *J Laryngol Otol* 68(9):623–635.

Kuk F, Jackson A, Keenan D, Lau CC. (2008) Personal amplification for school-age children with auditory processing disorders. *J Am Acad Audiol* 19(6):465–480.

Lee JY, Lee JT, Heo HJ, Choi CH, Choi SH, Lee K. (2015) Speech recognition in real-life background noise by young and middle-aged adults with normal hearing. *J Audiol Otol* 19(1):39–44.

Middelweerd MJ, Festen JM, Plomp R. (1990) Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology* 29(1):1–7.

Mueller HG. (2016) *Signia expert series: speech-in-noise testign for selection and fitting of hearing aids: worth the effort?* http://www.audiologyonline.com. Accessed March 3, 2017.

Nilsson M, Soli S. (1994) Norms for a headphone simulation of the hearing in noise test: comparison of physical and simulated spatial separation of sound sources. *J Acoust Soc Am* 95(5):2994.

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.

Plomp R, Mimpen AM. (1979) Improving the reliability of testing the speech reception threshold for sentences. *Audiology* 18(1):43–52.

Prosser S, Turrini M, Arslan E. (1990) Effects of different noises on speech discrimination by the elderly. *Acta Otolaryngol Suppl* 476:136–142.

Richburg CM, Atcherson SR, Findlen UM, Wakefield S, Benafield NJ. (2017) (Central) auditory processing disorder grand rounds: multiple cases, multiple causes, multiple outcomes. *Am J Audiol* 26(3):202–225.

Rosen S, Souza P, Ekelund C, Majeed AA. (2013) Listening to speech in a background of other talkers: effects of talker number and noise vocoding. *J Acoust Soc Am* 133(4):2431–2443.

Saunders GH, Haggard MP. (1989) The clinical assessment of obscure auditory dysfunction–1. Auditory and psychological factors. *Ear Hear* 10(3):200–208.

Schaette R, McAlpine D. (2011) Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. *J Neurosci* 31(38):13452–13457.

Shargorodsky J, Curhan SG, Curhan GC, Eavey R. (2010) Change in prevalence of hearing loss in US adolescents. *JAMA* 304(7):772–778.

Simpson SA, Cooke M. (2005) Consonant identification in N-talker babble is a nonmonotonic function of N. *J Acoust Soc Am* 118(5):2775–2778.

Sinha PS. (1959) The role of the temporal lobe in hearing. Master of Science, McGill University. Montreal Neurological Institute.

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

Spahr AJ, Dorman MF, Litvak LM, Van Wie S, Gifford RH, Loizou PC, Loiselle LM, Oakes T, Cook S. (2012) Development and validation of the AzBio sentence lists. *Ear Hear* 33(1):112–117.

Sperry JL, Wiley TL, Chial MR. (1997) Word recognition performance in various background competitors. *J Am Acad Audiol* 8(2):71–80.

Steinberg JC, Montgomery HC, Gardner MB. (1940) Results of the world's fair hearing tests. *J Acoust Soc Am* 12(2):533–562.

Stilma W, Rijkenberg S, Feijen HM, Maaskant JM, Endeman H. (2015) Validation of the Dutch version of the critical-care pain observation tool. *Nurs Crit Care*.

Stuart A, Butler AK. (2014) No learning effect observed for reception thresholds for sentences in noise. *Am J Audiol* 23(2):227–231.

Stuart A, Phillips DP. (1996) Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. *Ear Hear* 17(6):478–489.

Sweetow RW, Sabes JH. (2006) The need for and development of an adaptive Listening and Communication Enhancement (LACE) program. *J Am Acad Audiol* 17(8):538–558.

Theunissen M, Swanepoel W, Hanekom J. (2009) Sentence recognition in noise: variables in compilation and interpretation of tests. *Int J Audiol* 48(11):743–757.

Turnbull J. (2008) The (current) largest restaurant in the world. June 26, 2008. http://googlesightseeing.com/2008/06/the-current-largest-restaurant-in-the-world/. Accessed August 28, 2017.

Vermiglio AJ. (2007) *Speech Recognition in Noise. (AuD Doctoral project).* Mt. Pleasant, MI: Central Michigan University.

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

Vermiglio AJ, Griffin S, Courtney P, Fang X. (2017) An evaluation of the World Health Organization and American Medical

Association Ratings of Hearing Impairment and simulated single-sided deafness. *J Am Acad Audiol,* 1–14 (Published electronically prior to print on July 20, 2017).

Vermiglio AJ, Soli SD, Fang X. (2017) An argument for self-report as a reference standard in audiology. *J Am Acad Audiol* (Published electronically prior to print on May 11, 2017).

Wilson RH. (2003) Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *J Am Acad Audiol* 14(9):453–470.

Wilson RH, Carnell CS, Cleghorn AL. (2007) The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *J Am Acad Audiol* 18(6):522–529.

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(4):844–856.

Zhao F, Stephens D. (2006) Distortion product otoacoustic emissions in patients with King-Kopetzky syndrome. *Int J Audiol* 45(1):34–39.