Development of a Test Battery for Evaluating Speech Perception in Complex Listening Environments: Effects of Sensorineural Hearing Loss

Sandeep A. Phatak,[1] Benjamin M. Sheffield,[1,2] Douglas S. Brungart,[1] and Ken W. Grant[1]

Objective: To evaluate the speech-in-noise performance of listeners with different levels of hearing loss in a variety of complex listening environments.

Design: The quick speech-in-noise (QuickSIN)-based test battery was used to measure the speech recognition performance of listeners with different levels of hearing loss. Subjective estimates of speech reception thresholds (SRTs) corresponding to 100% and 0% speech intelligibility, respectively, were obtained using a method of adjustment before objective measurement of the actual SRT corresponding to 50% speech intelligibility in every listening condition.

Results: Of the seven alternative listening conditions, two conditions, one involving time-compressed, reverberant speech (TC+Rev), and the other (N_0S_π) having in-phase noise masker (N_0) and out-of-phase target (S_π), were found to be substantially more sensitive to the effect of hearing loss than the standard QuickSIN test. The performance in these two conditions also correlated with self-reported difficulties in attention/concentration during speech communication and in localizing the sound source, respectively. Hearing thresholds could account for about 50% or less variance in SRTs in any listening condition. Subjectively estimated SRTs (SRTs corresponding to 0% and 100% speech intelligibility) were highly correlated with the objective SRT measurements (SRT corresponding to 50% speech intelligibility).

Conclusions: A test battery that includes the TC+Rev and the N_0S_π conditions would be useful in identifying individuals with hearing loss with speech-in-noise deficits in everyday communication.

Key words: Hearing loss; Quick speech-in-noise; Speech recognition in noise.

(Ear & Hearing 2018;XX;00–00)

INTRODUCTION

Good hearing is crucial for our ability to socialize, work, communicate, and stay connected to the world around us (Garstecki 1987). A normal-hearing listener can communicate even in very complex listening scenarios using various auditory skills and processes, such as the ability to segregate the sound of interest from the interfering noise, attending to one out of multiple simultaneous audio signals, modulation masking release, spatial release from masking, etc. (Freyman et al. 2001; Dubno et al. 2002; Buss et al. 2009; Litovsky 2012). These auditory skills are known to degrade with increasing hearing loss.

Hearing difficulties due to hearing loss are often first apparent in difficult listening environments that involve noisy reverberant spaces, multitalker conversations, rapid speaking rates, and other complicating factors that are typically not present in the clinical spaces or in the standardized tests used to assess hearing loss (Pienkowski 2016). This may help explain why clinical assessments of hearing loss often do not correlate with the everyday speech-in-noise communication difficulties reported by patients (Hannula et al. 2011; Jerger 2011).

One common issue with many clinical speech tests is that they fail to assess binaural function, which is critical for speech perception in complex environments. The ability to use binaural cues is known to be affected by hearing loss and may very well relate to speech-in-noise difficulties reported by listeners with hearing loss (Carhart 1965, McArdle et al. 2012). A few clinical speech recognition tests like the listening in spatialized noise (Cameron et al. 2011) and the hearing in noise test (Nilsson et al. 1994) measure some aspects of binaural speech-in-noise performance. However, many of the most widely used standardized clinical tests, such as the quick speech-in-noise (QuickSIN) (Killion et al. 2004) and words in noise (Wilson 2003; Wilson & Burks 2005), rely exclusively on multitalker babble maskers presented diotically with the target speech. Another issue with clinical speech tests is the lack of visual speech cues, which are inherent to face-to-face communication. The ability to process and integrate visual cues, which may or may not be correlated with hearing loss, can greatly influence speech recognition performance. Very few standardized speech tests are currently available to clinically evaluate audiovisual speech perception in listeners with hearing loss.

To overcome these problems, Brungart et al. (2014) altered the stimuli and the testing paradigm of the QuickSIN to develop a short clinical test for measuring speech reception thresholds (SRTs) in eight different auditory scenes. Some listening conditions were expected to be easier for speech communication than the standard QuickSIN condition due to the ability of a normal-hearing listener to benefit from the binaural release from masking and from integrating audio and visual speech cues. Hearing loss can potentially degrade these abilities, resulting in worse-than-normal performance in these conditions. On the other hand, some listening conditions were expected to be more difficult than the standard condition. These particular conditions were designed to simulate challenging acoustic environments such as reverberant rooms and fast talkers, which often causes people with even mild to moderate hearing loss to complain of communication difficulties (Gordon-Salant & Fitzgibbons 1993). When combined with the standard QuickSIN condition, these conditions can provide a more comprehensive assessment of a person's operational speech communication ability. For occupations such as firefighting, aviation, law enforcement, military, etc., where communication is critical for mission success and survivability, such an assessment can be potentially used for predicting fitness-for-duty. This test battery is also useful for studying the effect of a hearing prosthesis or a hearing protection device on everyday speech communication.

---

[1]Audiology and Speech Pathology, Walter Reed National Military Medical Center, Bethesda, Maryland, USA; and [2]Army Hearing Division, United States Army Public Health Center, Aberdeen, Maryland, USA.

0196/0202/2018/XXXX-00/0 • Ear & Hearing • Copyright © 2018 Wolters Kluwer Health, Inc. All rights reserved • Printed in the U.S.A.

<zdoi; 10.1097/AUD.0000000000000567>

Brungart et al. (2014) reported normative speech-in-noise data on 49 normal-hearing listeners on their test battery. The present study used the same tests to measure speech-in-noise performance in participants with various degrees of hearing loss. The main goal of the present study was to identify the auditory test scenes that are most sensitive to speech-in-noise deficits and to measure the effect of hearing loss on speech-in-noise performance in those cases. Speech-in-noise performance in different listening conditions was also compared with the perceived hearing difficulties of participants.

METHODS

Subjects

Participants between 19 and 57 years of age, who were either civilian staff members or were eligible for health care, were recruited for this study at Walter Reed Army Medical Center (Washington, DC) or at Walter Reed National Military Medical Center (Bethesda, MD). Audiometric thresholds were either taken from the listener's health records, if measured within 6 months of the start of testing, or were measured clinically as part of the study. Subjects were divided into three groups: H1 (normal to near-normal); H2 (mild); and H3 (moderate to severe), according to the hearing profiles described in the U.S. Army standard for medical fitness (AR40-501) (U.S. Department of Army 1960). Those in the H1 group were subdivided into two groups: normal hearing (H1NH), having thresholds ≤20 dB HL from 250 to 8000 Hz, and borderline normal (H1BN), having at least one threshold >20 dB HL, but ≤30 dB HL from 250 to 3000 Hz and ≤45 dB HL from 4000 to 8000 Hz. Figure 1 shows average audiometric thresholds across subjects in each hearing profile group, and also lists gender and age statistics for each group. Data of the H1NH group were analyzed and presented in Brungart et al. (2014).

Stimuli

The testing method and the test conditions were identical to those described in Brungart et al. (2014). Subjects were required to recognize Institute of Electrical and Electronic Engineers (IEEE) sentences (Institute of Electrical and Electronic Engineers 1969) spoken by a female talker and presented in eight different conditions as follows:

1. *Babble*: standard QuickSIN condition with four-talker babble masker,
2. Audiovisual (*AV*): Babble condition with audiovisual IEEE sentences,
3. N_0S_π: Babble condition but phase of the target sentences shifted by 180° in one ear,
4. $AV{+}N_0S_\pi$: N_0S_π condition with audiovisual target sentences,
5. *Spatial*: using head-related transfer functions, target set in front, and two four-talker babble maskers (same four talkers as in all previous conditions, but different babble signals), one set to the left (−90°) and the other to the right (90°),
6. Reverberation (*Reverb*): Spatial condition with reverberation (RT60 = 0.25 sec) added to stimuli,
7. Time compression and reverberation (*TC+Rev*): Reverb condition with target speech rate increased by 50% using pitch-preserving time compression (i.e., time compressed to 66% of the original length) before applying reverberation, and
8. Speech-Spectrum Noise (*SSN*): standard QuickSIN condition with the babble replaced by continuous speech-spectrum noise masker.

The sentence lists, each with six IEEE sentences, were taken from the QuickSIN compact disc (CD), and the background noise level adjustment used in the CD for equalizing performance across lists was preserved to the greatest extent possible. In the four conditions that used four-talker babble (namely, Babble, AV, N_0S_π, and $AV{+}N_0S_\pi$), the sentences-masker pairs were identical to those on the CD. For the three conditions that used eight-talker babble (namely, Spatial, Reverb, and TC+Rev), the four-talker babble on the left was matched to the target list, while that on the right was taken from a different list.

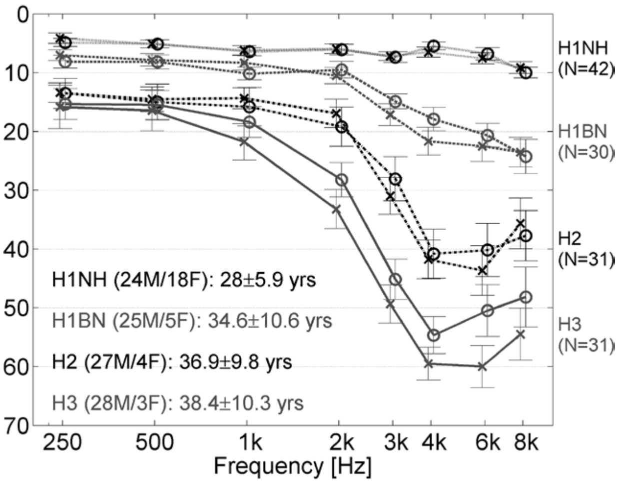

Fig. 1. Average audiometric thresholds in left (x) and right (o) ears for the four listener groups. Number of M and F subjects in each group, and the mean and ±1 SD of subject age in each group are also listed. BN indicates borderline normal; F, female; M, male; NH, normal hearing.

Test Procedure

In each listening condition, subjects first estimated the range of their psychometric curve using a method of adjustment. In this part, subjects were presented a continuous stream of IEEE sentences at 70 dB SPL, while they adjusted the background masker level using a slider. Subjects first increased the masker to a level where they could hear the speech but no longer understand any word [i.e., SRT corresponding to 0% speech intelligibility (SRT_0)] and then decreased it to a level where they began to understand all the words [i.e., SRT corresponding to 100% speech intelligibility (SRT_{100})].

The subjective SRT estimation was followed by a speech recognition task according to the standard QuickSIN procedure to measure the actual speech-in-noise threshold, denoted as SRT corresponding to 50% speech intelligibility (SRT_{50}), for that condition. In this task, subjects were required to hear and repeat back sentences from a QuickSIN list, in the presence of background noise. The signal to noise ratio (SNR) of the first sentence was set to be 12.5 dB above the average of the SRT_0 and SRT_{100} thresholds estimated for that condition. The SNR of each subsequent sentence was decreased in 5 dB steps, making it harder to recognize than the previous one. The speech signal was set at 70 dB SPL, and the masker level was adjusted according to the SNR. Subject performance was scored based on the

TABLE 1. Questions in the abbreviated SSQ questionnaire, along with their original SSQ number

SSQ No.	Original SSQ No.	Question
1	Speech 14	You are talking to someone on the telephone and someone next to you starts talking. Can you follow what's being said by both talkers?
2	Speech 12	You are in a group and the conversation switches from one person to another. Can you easily follow the conversation without missing the start of what each new speaker is saying?
3	Speech 5	You are talking to a person. There is continuous background noise, such as a fan or running water. Can you follow the conversation?
4	Spatial 8	In the street, can you tell how far away someone is, from the sound of their voice or footsteps?
5	Spatial 13	Can you tell from the sound whether a bus or truck (vehicle) is coming towards you or going away?
6	Spatial 17	Do you have the impression of sounds being where you would expect them?
7	Spatial 3	You are sitting in between two people. One of them starts to speak. Can you tell right away whether it is the person on your left or your right, without having to look?
8	Qualities 14	Do you have to concentrate very much when listening to someone or something?
9	Qualities 13	Can you easily judge another person`s mood by the sound of their voice?
10	Qualities 11	Do everyday sounds that you hear seem to have an artificial or unnatural quality?
11	Qualities 18	Can you easily ignore other sounds when trying to listen to something?
12	Qualities 5	Can you easily distinguish different pieces of music that you are familiar with?

SSQ, Speech, Spatial and Qualities of Hearing Scale.

number of key words recognized correctly in each sentence, and if no key word was recognized correctly in a sentence, then further sentences were not presented, assuming a score of zero for those sentences. On the basis of the SRT calculation methodology in the standard QuickSIN procedure, the SRT_{50} was calculated by adding 2.5 dB to the SNR of first sentence and then subtracting the total number of key words recognized correctly.

This procedure was repeated once for each condition, thus obtaining two estimates of each SRT. For the first set of SRT estimates, the eight listening conditions were presented in random order. In the second session, and for the second set, the order of the conditions from the first set was reversed, to counterbalance the presentation order effect. List numbers 1–16 from the QuickSIN CD were used in a pseudorandom order for estimating the SRT_{50} twice in each of the eight conditions. For each subject, list numbers 1–12 were randomly assigned across conditions, while list numbers 13–14 and 15–16 were always paired and were assigned to one of the six audio-only conditions because the video for these two list pairs was not available. The continuous stream of IEEE sentences for the subjective estimates (SRT_0 and SRT_{100}) was not from the QuickSIN CD, but rather from audiovisual recording of IEEE sentences by a different female talker, made using the same equipment and at the same time as the QuickSIN recordings. The total testing time per subject was about 25 to 30 min. This translates to about 3 to 4 min per condition, with about 2 mins required to obtain the subjective estimate (SRT_0 and SRT_{100}) and another 2 min to obtain the objective estimate (SRT_{50}) Table 1.

Speech, Spatial and Qualities of Hearing Scale Questionnaire

An abbreviated version of the Speech, Spatial and Qualities of Hearing Scale (SSQ) questionnaire (Gatehouse & Noble 2004) was administered to listeners to sample their perceived difficulties in everyday speech communication. This abbreviated version consisted of 12 questions that were selected from the three subareas of the SSQ to be highly correlated with functional hearing handicap, weakly correlated with each other, or having operational relevance for individuals in challenging occupations like firefighting, law enforcement, or the military [note that these questions were selected before the publication of the SSQ-12 abbreviated form of the questionnaire (Noble et al. 2013) and are unrelated to them]. The questions used in the abbreviated SSQ questionnaire are listed in Table 1 (Fig. 2).

RESULTS

Effect of Subject Group and Listening Condition

Figure 2 shows the average SRT_{50} values for the four listener groups in all listening conditions. As expected, thresholds for all subject groups significantly improved ($p < 0.001$) due to visual input (AV versus Babble and AV+N_0S_π versus N_0S_π) and binaural release from masking (N_0S_π versus Babble and AV+N_0S_π versus AV). There was no significant improvement or degradation for any group in the SSN ($p > 0.16$) condition, relative to the Babble condition. The three spatial conditions (Spatial, Rev, and TC+Rev) were not directly comparable to the babble condition because they had eight interfering talkers rather than four.

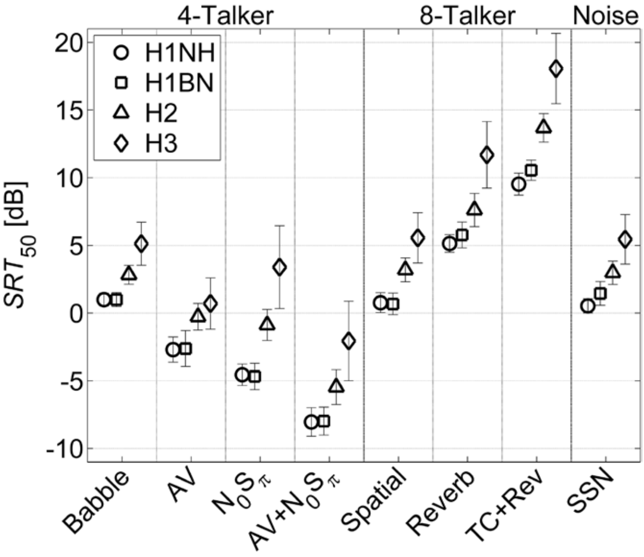

Fig. 2. Measured speech-in-noise thresholds (SRT_{50}) for each group in all eight listening conditions. Symbols and error bars denote means and 95% confidence intervals estimated across listeners within each group.

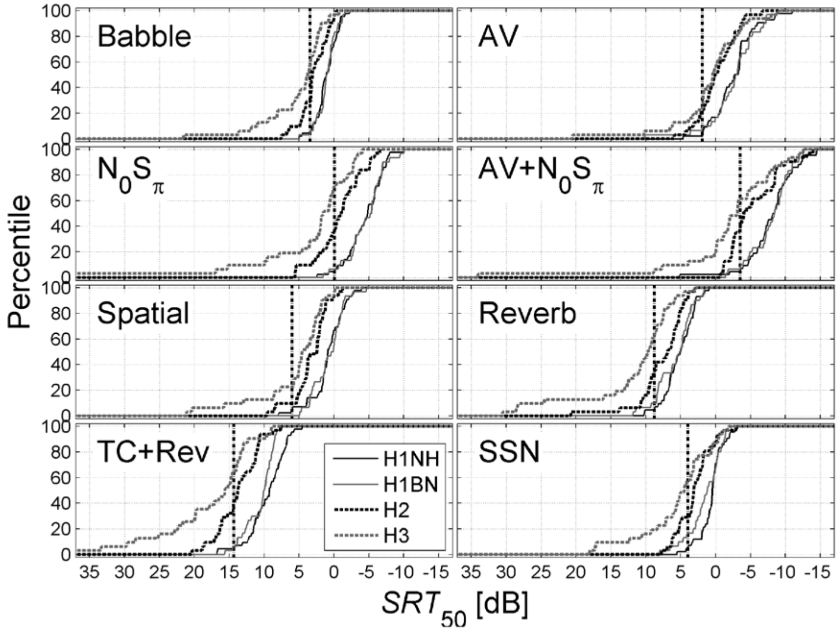

Fig. 3. SRT$_{50}$ cumulative distributions (in percentile). The vertical dash-dotted line in each panel represents the fifth percentile value for H1NH group in each condition.

However, relative to the baseline Spatial condition, performance degraded significantly ($p < 0.001$) due to Reverb and TC+Rev.

In all the conditions, the speech-in-noise performance degraded (i.e., SRT$_{50}$ increased) with moderate to severe hearing loss, but not with mild hearing loss. The average performances of H2 and H3 groups were worse than the control group (i.e., H1NH) in all listening conditions ($p \leq 0.004$ for H2 and $p \leq 0.003$ for H3), but the H1BN group did not show a significant deficit ($p > 0.075$) in any condition. Compared with the standard QuickSIN condition (i.e., Babble), the speech-in-noise deficit was noticeably greater in N$_0$S$_\pi$, AV+N$_0$S$_\pi$, Reverb, and TC+Rev conditions. Analysis of variance revealed that the listener group had a significant main effect ($p < 0.001$) on SRT$_{50}$ in all listening conditions, but the effect was stronger than the standard condition [$F(3,130) = 18.405$] only in N$_0$S$_\pi$ [$F(3,130) = 20.087$] and TC+Rev [$F(3,130) = 27.620$] condition.

Figure 3 shows cumulative distributions of SRT$_{50}$ values in all listening conditions. In N$_0$S$_\pi$, Reverb, and TC+Rev conditions, some H3 listeners had SRT$_{50}$ thresholds more than 20 dB higher than the average SRT$_{50}$ for the control group (i.e., H1NH). Consistent with the analysis of variance results, the highest separation between cumulative distributions of H1NH and H2 or H3 group was observed in N$_0$S$_\pi$ and TC+Rev conditions. To quantify the separation among groups, SRT$_{50}$ distribution of the H1NH group was compared with those of the other three groups in each listening condition using two-tailed student *t* tests, and the results are listed in Table 2.

No statistically significant difference was observed between the H1NH and H1BN groups in any listening condition ($p > 0.75$). H2 and H3 groups were significantly different ($p < 0.002$) than the control group in all but one listening condition (AV+N$_0$S$_\pi$ for H2 and AV for H3). But the highest separation, as indicated by the *t* statistics, was observed in the TC+Rev condition, followed by that in the N$_0$S$_\pi$ condition. The Babble, Spatial, Reverb, and SSN conditions also show high degree of separation, with *t* statistics exceeding 4.5 for H3 and, in some cases, for H2 Figure 4.

Another commonly used metric in clinical psychophysics for quantifying the performance of a patient is to compare it with the fifth percentile value of the control or the "normal" group (i.e., a performance level that 95% of the control group population achieves or exceeds). The fifth percentile level for the control group in each listening condition is denoted by the vertical dash-dotted line in Figure 3. Figure 4 shows that more than 58% of H3 subjects and more than 29% of H2 subjects performed worse than this level in Babble, N$_0$S$_\pi$, AV+N$_0$S$_\pi$, Reverb, TC+Rev, and SSN conditions.

Test–Retest Reliability

Despite having additional subjects with hearing loss, the test–retest reliability of the data in the present study was very

TABLE 2. Results of two-tailed *t* tests comparing the distribution of SRT$_{50}$ values of the control group (i.e., H1NH) with that of the other three groups

Group	Babble	AV	N$_0$S$_\pi$	AV+N$_0$S$_\pi$	Spatial	Reverb	TC+Rev	SSN
H1BN	–0.01	0.09	–0.19	0.08	–0.16	1.08	1.81	1.76
H2	**4.27**	**3.49**	**5.18**	3.03	**4.17**	**3.49**	**6.12**	**4.74**
H3	**4.88**	3.16	**4.92**	**3.75**	**4.71**	**5.05**	**6.14**	**5.07**

Values in bold denote cases with statistically significant difference after Bonferroni correction (p < 0.002).

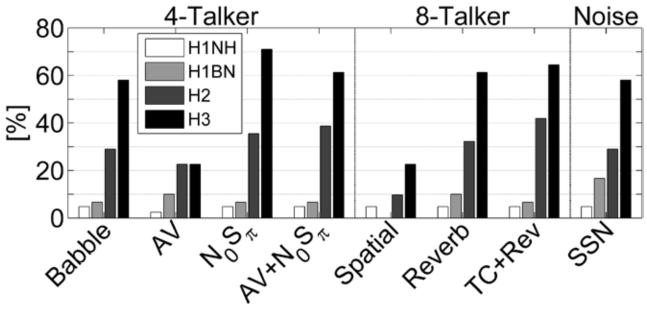

Fig. 4. The percentage of subjects in each group having an SRT$_{50}$ higher (i.e., worse performance) than the fifth percentile value for the control group in each listening condition.

close to that reported by Brungart et al. (2014). For example, the mean absolute difference in SRT$_{50}$ thresholds in the standard QuickSIN condition (i.e., Babble) in the present study was 2.36 dB, compared with 2.3 dB reported by Brungart et al. (2014). The mean absolute difference in SRT$_{50}$ varied between 2.24 dB (Reverb) and 3.71 (AV and AV+N$_0$S$_\pi$). The ranges for the mean absolute difference in SRT$_0$ and SRT$_{100}$ were 2.4 to 3.7 dB and 3.1 to 4.9 dB, respectively Table 3.

Another way to characterize the test–retest reliability is in terms of the Pearson correlation coefficient. The two estimates for the three SRTs were significantly correlated ($p < 0.001$) in all conditions. Table 3 lists the Pearson correlation coefficients. The two estimates were averaged for all subsequent analyses Table 4.

Cross-Correlations

The eight listening conditions, although designed to test different aspects of speech-in-noise performance, could have redundancies due to common factors such as speech stimuli and the testing paradigm. Table 4 shows Pearson correlation coefficients between the SNR$_{50}$ values. Strong correlations ($r > 0.8$) were observed between N$_0$S$_\pi$ and AV+N$_0$S$_\pi$, and among Babble, Reverb, and TC+Rev conditions. The AV condition was not strongly correlated ($r \leq 0.61$) with any other condition, while Spatial and SSN conditions showed moderate correlations with each other as well as with other conditions ($0.54 \leq r \leq 0.79$).

Audiometric Hearing Thresholds

The systematic degradation of the average speech-in-noise performance from H1NH group to H3 group in all conditions suggests that the speech-in-noise deficit could be related to the elevated audiometric thresholds. To test this hypothesis, SRT$_{50}$ performances in all listening conditions were modeled using linear multiregression of all audiometric thresholds from both ears. The variance in SRT$_{50}$ that was accounted for by the audiometric hearing thresholds varied across conditions, with the least variance accounted for in the AV condition ($R^2 = 0.26$; $p = 0.005$) and the most variance accounted for in the Reverb and TC+Rev conditions (both $R^2 = 0.56$; $p < 0.001$). Hearing thresholds also accounted for nearly 50% or more variance in other conditions that had significant speech-in-noise deficits, such as N$_0$S$_\pi$ ($R^2 = 0.49$; $p < 0.001$) and SSN ($R^2 = 0.54$; $p < 0.001$). Thus, the increased speech-in-noise deficits in certain conditions may be partially due to the loss of audibility resulting from elevated thresholds. Although the target speech level was the same in all conditions, the loss of high-frequency speech cues due to inaudibility would reduce the redundancy in the speech information received by listeners with hearing loss, making their speech communication ability less robust to noise-masking and environmental distortions. It might also represent a distortion component that systematically increased with increasing thresholds. Nevertheless, audiometric hearing thresholds fail to account for more than 40% of the variance in any condition Figure 5.

Perceived Speech-in-Noise Performance

Figure 5 compares the subjectively estimated range of speech-in-noise performance (i.e., SRT$_0$ to SRT$_{100}$) with the measured SRT$_{50}$ thresholds (+ symbols). In most conditions, the measured SRT$_{50}$ thresholds were not only within the perceived range of performance but were also very close to the center of the range, denoted by × symbols. The exceptions were the AV and AV+N$_0$S$_\pi$ conditions, where subjects tended to underestimate their thresholds resulting in SRT$_{50}$ values very close to SRT$_0$ values, and the TC+Rev condition, where subjects tended to overestimate their thresholds resulting in SRT$_{50}$ values closer to their perceived SRT$_{100}$. The results for the audiovisual conditions suggest that although listeners do not perceive the visual speech information to aid speech communication (possibly due to increased cognitive load required to process audiovisual stimuli), their speech-in-noise performance is improved by visual speech cues. On the other hand, the overestimated performance for the TC+Rev condition indicates that listeners, irrespective of their hearing loss, feel that they can understand rapid speech, but in reality, they make more errors than they think Figure 6.

Perceived Hearing Difficulties

To test the ability of the speech-in-noise performance in various listening conditions to capture the speech-in-noise communication problems, the correlations between SRT values (both subjective and measured) and SSQ questionnaire ratings were calculated. Figure 6 shows coefficient magnitudes for the correlations that were statistically significant after Bonferroni correction. SSQ questions 4, 7, 9, 11, and 12 had very weak or nonsignificant correlations with any SRT value in any condition. In general, SRT$_{50}$ values were more highly correlated with SSQ ratings than the subjective thresholds. Only questions 2, 3, 4, and 10 had relatively strong correlations ($r > 0.35$) with more than one condition. When compared by condition, only Babble, Reverb, TC+Rev, and SSN conditions had correlations

TABLE 3. Pearson correlation coefficient between two estimates of the three SRT values across all listeners

	Babble	AV	N$_0$S$_\pi$	AV+N$_0$S$_\pi$	Spatial	Reverb	TC+Rev	SSN
SRT$_0$	0.73	0.76	0.61	0.88	0.78	0.75	0.73	0.71
SRT$_{50}$	0.57	0.47	0.83	0.71	0.65	0.84	0.73	0.68
SRT$_{100}$	0.41	0.63	0.56	0.77	0.77	0.74	0.66	0.72

All correlations were statistically significant ($p < 0.001$).

TABLE 4. Cross-correlation coefficients among the eight listening conditions, calculated across all participants

	AV	N_0S_π	AV+N_0S_π	Spatial	Reverb	TC+Rev	SSN
Babble	0.61	0.70	0.62	0.72	**0.83**	**0.83**	0.77
AV	—	0.47	0.56	0.54	0.60	0.59	0.54
N_0S_π	—	—	**0.84**	0.64	0.62	0.71	0.63
AV+N_0S_π	—	—	—	0.56	0.52	0.64	0.56
Spatial	—	—	—	—	0.79	0.75	0.69
Reverb	—	—	—	—	—	**0.81**	0.76
TC+Rev	—	—	—	—	—	—	0.77

All correlations were statistically significant ($p < 0.001$). The correlation coefficients greater than 0.8 are written in bold.

with $r > 0.35$. TC+Rev and SSN were the only two conditions that yielded stronger correlations with SSQ questions than the standard QuickSIN condition (i.e., Babble). The AV condition was poorly correlated ($r < 0.3$) with SSQ questions.

The only two conditions that were correlated with SSQ question 6, which relates to the perceived location of the sound source, were N_0S_π and AV+N_0S_π. This suggests that the two N_0S_π conditions could be potentially useful in evaluating the role of binaural integration ability in speech-in-noise communication. Question number 8, which pertains to attention and effort required while listening to speech-in-noise, correlated more with TC+Reverb condition ($r > 0.3$) than any other condition. Therefore, one likely reason why all listeners found the TC+Rev condition to be the most challenging could be the higher cognitive load required for communicating in this condition.

DISCUSSION

In this study, a set of speech materials designed to evaluate speech perception in a variety of complex environments was used to examine how well listeners with different levels of hearing loss performed in each task. In designing the tests, it was hypothesized that some more complex speech environments might be more sensitive to the effects of hearing loss than the standard clinical version of the QuickSIN test, which presents a single talker monaurally or diotically in the presence of a four-talker babble in an anechoic space. Of the seven alternative listening conditions tested, two conditions stood out as being substantially more sensitive to the effect of hearing loss than the standard QuickSIN test (note that this increased sensitivity occurred despite the fact that the QuickSIN test was carefully normalized to reduce the variability in the intelligibility of the IEEE sentences in the Babble task, and this adjustment was not made for any of the other listening conditions. Potentially the differences across conditions might be even greater if these adjustments were made for the other listening conditions). The first was the TC+Rev condition, which was highly correlated with the standard babble test but was substantially more sensitive to the effects of hearing loss. This suggests that the performance in the TC+Rev condition is degraded by the same hearing deficits that interfere with performance in the standard QuickSIN task, but that these degradations are amplified by the temporal distortions caused by the increased speech rate and simulated Reverb. Notably, the TC+Rev condition was more correlated than any other condition with a survey question asking how much listeners needed to concentrate to understand speech, suggesting an interaction between cognitive effort and the extraction of fast speech from reverberant environments.

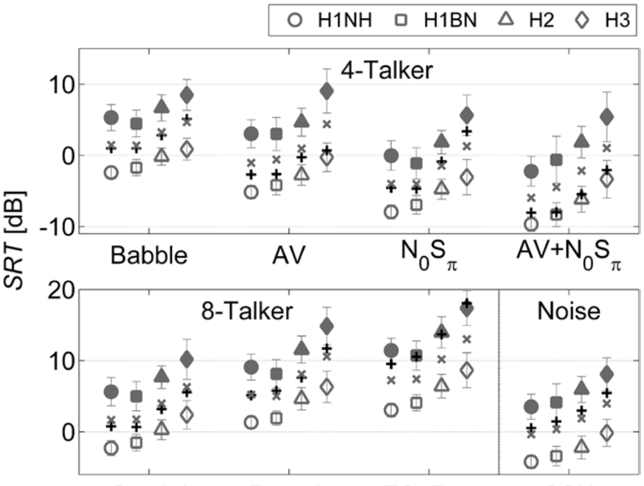

Fig. 5. Subjectively estimated speech-in-noise thresholds SRT_0 (open symbols) and SRT_{100} (filled symbols) for each group in all eight listening conditions. Symbols and error bars denote means and 95% confidence intervals estimated across listeners within each group. In each case, the gray "×" symbol denotes the mid-way point between SRT_0 and SRT_{100}, whereas the black "+" symbol represents the average SRT_{50} value.

The second condition that appeared to be more sensitive to the effects of hearing loss than the standard Babble condition was the N_0S_π condition. However, in contrast with the TC+Rev condition, the N_0S_π condition was very poorly correlated with performance in the standard babble condition. It was also one of the only two conditions (other being AV+N_0S_π) that were correlated with the SSQ question regarding the accuracy of sound localization. This suggests that it might be influenced by some aspect of binaural and spatial perception that is not addressed by the other conditions.

It is also notable that the AV conditions of the experiment showed relatively little difference between the subjects with different degrees of hearing loss. Comparing the AV and non-AV conditions with and without the N_0S_π manipulation, it appears that all listeners obtained approximately a 5 dB improvement in performance from the availability of visual cues. Also, in general, it seems that the AV cues tend to minimize the differences between listeners with and without hearing loss, presumably because the performance in those conditions is dominated more by differences in lipreading skills than in hearing loss. Nevertheless, the AV conditions represent an important aspect of the everyday face-to-face communication. Moreover, there may be some type of hearing loss or some hearing prostheses or hearing loss interventions

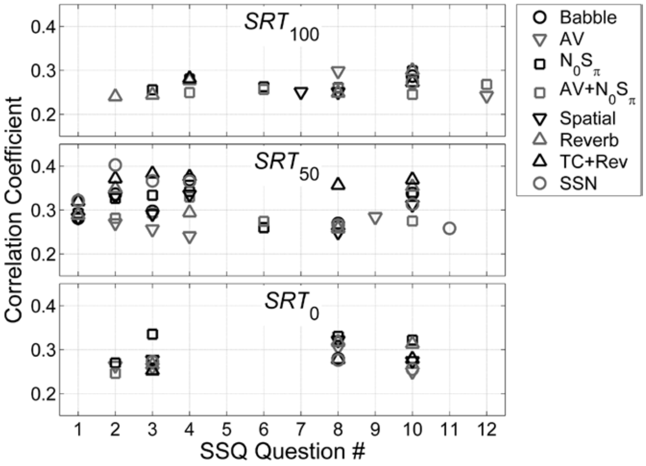

Fig. 6. Pearson correlation coefficient magnitudes between ratings on the 12 SSQ questions and subjective (SRT_0, SRT_{100}) and measured thresholds (SRT_{50}) in the eight listening conditions. Only those correlations that were statistically significant after Bonferroni correction are plotted. Correlation coefficients were positive for questions 8 and 10, and negative for the remaining questions.

that might be more sensitive to differences in visual cues, and the AV variants of the QuickSIN might be very useful in those situations.

A note should be made about the usefulness of a technique based on the method of adjustment to assess speech perception, rather than one based on objective speech intelligibility. In this experiment, listeners were fairly reliable at adjusting the level of masker to estimate the points of 100% and 0% speech intelligibility, and these subjective estimates were highly correlated with their objective SRT measurements ($p < 1 \times 10^{-8}$). This suggests that method of adjustment can be a very useful tool for assessing within-subject, across-masker differences in speech intelligibility in cases where a very rapid estimate is needed or where limitations in the amount of available speech materials (e.g., in the number of lists available in the QuickSIN) make it impractical to objectively test every condition. However, it does seem like these responses may also be influenced by other factors. For example, in this experiment, the SRT_{100} condition was the only one correlated with the subjective question asking about the clarity of music. However, there were differences in these measures across listening conditions (especially in the AV and TC+Rev conditions of this experiment), so care must be taken in using subjective measures to compare performance across different listening conditions.

Overall, these results suggest that a hearing test regimen that combines the TC+Rev task and the N_0S_π task might be an ideal combination for identifying individuals with hearing loss with speech-in-noise deficits but who score normally on traditional speech perception tests. Interestingly, the TC+Rev speech perception condition and N_0S_π tone detection (not speech perception) are also the two tests that other studies in our laboratory have found to be most sensitive to self-reported hearing deficits in blast-exposed individuals with normal or near-normal audiograms (Brungart et al. 2016). The authors are currently adapting these tests for testing in the free field, which could be used to evaluate devices like hearing aids, and exploring ways to standardize and distribute these modified versions of the QuickSIN for more widespread clinical use.

ACKNOWLEDGMENTS

This research was funded by the U.S. Army Center for Health Promotion and Preventive Medicine (USACHPPM) through a Health Promotion and Preventive Initiative (HPPI) program grant.

The views expressed in this article are those of the authors and do not reflect the official policy of the Department of Army/Navy/Air Force, Department of Defense, or U.S. Government.

The identification of specific products, scientific instrumentation, or organizations is considered an integral part of the scientific endeavor and does not constitute endorsement or implied endorsement on the part of the author, DoD, or any component agency.

The authors have no conflicts of interest to disclose.

Address for correspondence: Sandeep A. Phatak, Walter Reed National Military Medical Center, Building 19, Room 5504, 8901 Rockville Pike, Bethesda, MD 20889, USA. E-mail: s.a.phatak@gmail.com

Received July 25, 2017; accepted January 23, 2018.

REFERENCES

Brungart, D. S., Kubli, L., Phatak, S., et al. (2016). "Functional hearing impairments in normal-hearing listeners with a history of deployment-related blast exposure." *J Acoust Soc Am*, *140*, 3152.

Brungart, D. S., Sheffield, B. M., Kubli, L. R. (2014). Development of a test battery for evaluating speech perception in complex listening environments. *J Acoust Soc Am*, *136*, 777–790.

Buss, E., Whittle, L. N., Grose, J. H., et al. (2009). Masking release for words in amplitude-modulated noise as a function of modulation rate and task. *J Acoust Soc Am*, *126*, 269–280.

Cameron, S., Glyde, H., Dillon, H. (2011). Listening in spatialized noise-sentences test (LiSN-S): Normative and retest reliability data for adolescents and adults up to 60 years of age. *J Am Acad Audiol*, *22*, 697–709.

Carhart, R. (1965). Problems in the measurement of speech discrimination. *Arch Otolaryngol*, *82*, 253–260.

Dubno, J. R., Horwitz, A. R., Ahlstrom, J. B. (2002). Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *J Acoust Soc Am*, *111*, 2897–2907.

Freyman, R. L., Balakrishnan, U., Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *J Acoust Soc Am*, *109*(5 Pt 1), 2112–2122.

Garstecki, D. (1987). "Self-perceived hearing difficulty in aging adults with acquired hearing loss." *J Acad Rehabil Audiol*, *20*, 49–60.

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *Int J Audiol*, *43*, 85–99.

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *J Speech Hear Res*, *36*, 1276–1285.

Hannula, S., Bloigu, R., Majamaa, K., et al. (2011). Self-reported hearing problems among older adults: Prevalence and comparison to measured hearing impairment. *J Am Acad Audiol*, *22*, 550–559.

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

Jerger, J. (2011). Why do people without hearing loss have hearing complaints? *J Am Acad Audiol*, *22*, 490.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., et al. (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.

Litovsky, R. Y. (2012). "Spatial release from masking." *Acoust Today*, *8*, 18–23.

McArdle, R. A., Killion, M., Mennite, M. A., et al. (2012). Are two ears not better than one? *J Am Acad Audiol*, *23*, 171–181.

Nilsson, M., Soli, S. D., Sullivan, J. A. (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*, 1085–1099.

Noble, W., Jensen, N. S., Naylor, G., et al. (2013). A short form of the speech, spatial and qualities of hearing scale suitable for clinical use: The SSQ12. *Int J Audiol*, *52*, 409–412.

Pienkowski, M. (2016). "On the etiology of hearing difficulties in noise despite clinically normal audiograms." *Ear Hear*, *97*, 593–608.

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

Wilson, R. H., & Burks, C. A. (2005). "The use of 35 words to evaluate hearing loss in terms of signal-to-babble ratio: A clinic protocol." *J Rehabil Res Dev*, *42*, 839–852.

U.S. Department of Army (1960). *"Standards of Medical Fitness," Army Regulation 40–501, Rapid Action Revision Issue Date 4 August 2011*. Washington, DC: Headquarters Department of the Army. Accessed March 6, 2018. Retrieved from https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN3801_AR40-501_Web_FINAL.pdf.