Original Article

An examination of speech reception thresholds measured in a simulated reverberant cafeteria environment

Virginia Best*,†, Gitte Keidser*, Jörg M. Buchholz*,‡ & Katrina Freeston*

*National Acoustic Laboratories and the HEARing Cooperative Research Centre, Australian Hearing Hub, Macquarie University, Australia, †Department of Speech, Language and Hearing Sciences, Boston University, Boston, USA and ‡Audiology Section, Department of Linguistics, Australian Hearing Hub, Macquarie University, Australia

The British Society of Audiology

The International Society of Audiology

NAS NORDIC AUDIOLOGICAL SOCIETY

Abstract

Objective: There is increasing demand in the hearing research community for the creation of laboratory environments that better simulate challenging real-world listening environments. The hope is that the use of such environments for testing will lead to more meaningful assessments of listening ability, and better predictions about the performance of hearing devices. Here we present one approach for simulating a complex acoustic environment in the laboratory, and investigate the effect of transplanting a speech test into such an environment. *Design:* Speech reception thresholds were measured in a simulated reverberant cafeteria, and in a more typical anechoic laboratory environment containing background speech babble. *Study sample:* The participants were 46 listeners varying in age and hearing levels, including 25 hearing-aid wearers who were tested with and without their hearing aids. *Results:* Reliable SRTs were obtained in the complex environment, but led to different estimates of performance and hearing-aid benefit from those measured in the standard environment. *Conclusions:* The findings provide a starting point for future efforts to increase the real-world relevance of laboratory-based speech tests.

Key Words: Speech reception thresholds; real-world; hearing loss; hearing aids

Laboratory-based speech tests are routinely conducted to assess speech understanding in noise. However it is often noted that the performance of an individual in the laboratory does not necessarily correspond to their real-world listening ability. For example, several studies have noted a mismatch between the benefit of hearing aids or processing schemes as measured in the laboratory and how beneficial users report them to be in their everyday listening situations (e.g. Bentler et al, 1993; Walden et al, 2000; Cord et al, 2004; Wu, 2010).

Although there is no single way of measuring speech intelligibility in the laboratory, most approaches involve a relatively simple acoustic environment (audiometric booth or anechoic chamber) and a steady-state noise or unintelligible speech babble background. This kind of set-up offers good control and repeatability, but misses a number of ecologically relevant variables that may prove to be important (Cord et al, 2007; Jerger, 2009). Among these are dynamic variations in spatial and level characteristics of the acoustic environment, realistic levels of reverberation, and the presence of competing intelligible conversations that can be highly distracting. Many of these factors have been shown to affect speech perception, and in some cases more so for listeners with hearing loss, but their combined effect in the context of an ecological setting is not really known. Moreover, while some of these factors have been shown to influence hearing-aid processing, others have not been examined in detail. For example, we know very little about the impact of competing talkers on hearing-aid processing schemes, many of which are designed to *emphasize* speech sounds.

A few previous attempts have been made to conduct sentence-based speech testing under more realistic conditions. Killion and colleagues (1998) used a novel approach to test the real-world benefit of directional hearing aids. They placed real talkers in real environments (parties, restaurants, etc.), and had them produce sentences from a standard test at different vocal efforts to vary the signal-to-noise ratio (SNR). They recorded the signals arriving at the ears of a real listener in that same environment, and then used the recorded signals for speech testing over headphones. More recently, the multi-microphone/multi-speaker R-SPACE system was developed to achieve a similar goal with the HINT test but under controlled laboratory conditions (Revit et al, 2002; Compton-Conley et al, 2004; Gifford & Revit, 2010). To our knowledge there has been no systematic comparison of data obtained under more realistic conditions to those obtained in simpler conditions. This is an important step in

RIGHTS LINK

Abbreviations

|	
AIDI	Articulation-index weighted directivity index
4FAHL	Four-frequency average hearing loss
HI	Hearing-impaired
NH	Normally hearing
RIR	Room impulse response
SNR	Signal-to-noise ratio
SRT	Speech reception threshold

determining whether it is worth moving to these more complicated scenarios, and demonstrating where the benefits might lie.

Here we present a multi-loudspeaker approach for generating complex acoustic environments in the laboratory, and describe a preliminary attempt to conduct speech testing in one example environment based on a typical everyday listening situation. Our specific aim was to understand the effects of using a more complicated and realistic acoustic environment on the properties of speech reception thresholds (SRTs). The same sentence materials were used to measure SRTs in a 'standard' anechoic environment containing a speech babble background, and in a 'complex' simulation of a reverberant cafeteria containing multiple intelligible conversations. The same large group of listeners with a broad range of hearing losses completed testing in both environments so that correlations could be examined. Another advantage of using more complex environments that better resemble real-world listening situations is that they are more appropriate for evaluating hearing-aid processing strategies. Thus we also examined the ability of the two tests to capture changes in performance due to amplification in hearing-impaired listeners.

Methods

Participants

Forty-six listeners participated. Eighteen of these had normal hearing ('normally hearing', NH). Their age ranged from 18 to 57 years (mean 41 years) and their four-frequency average hearing loss (4FAHL, mean of left and right ear pure-tone thresholds at 500, 1000, 2000, and 4000 Hz) ranged from 1 to 15 dB (mean 6 dB). The other 28 listeners had bilateral sensorineural hearing losses ('hearing-impaired', HI). Their age ranged from 29 to 80 years (mean 70 years) and their 4FAHL ranged from 26 to 78 dB (mean 45 dB). Hearing levels did not differ between the ears by more than 25 dB at any audiometric frequency. Audiograms for each listener are plotted in Figure 1 along with group averages. Note that 4FAHL
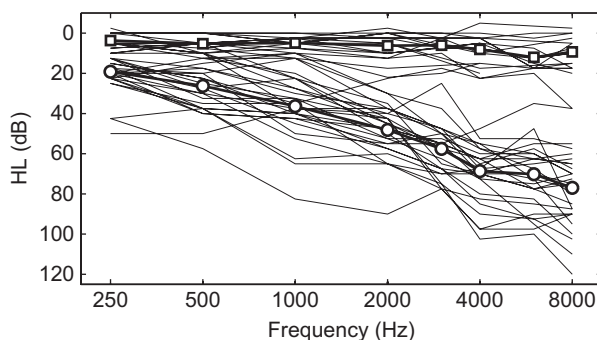
Figure 1. Audiograms for each listener (averaged across left and right ears), as well as group means for the NH group (squares) and the HI group (circles).

was correlated with age in the total pool (r = 0.71. p < 0.001) but not in the subgroup of HI listeners (r = − 0.08, p = 0.67) as a result of several young listeners with quite severe losses.

All listeners were asked to take home and complete a questionnaire addressing their hearing abilities[1]. The purpose was to determine how SRTs measured under standard and complex conditions relate to real-world experience as measured by self-report. Fifteen questions addressing disability under specific situations were taken from the Speech, Spatial and Qualities of Hearing questionnaire (Gatehouse & Noble, 2004). These questions included the 14 questions in the 'speech' subscale, as well as the question addressing listening effort from the 'qualities of hearing' subscale (question 18). Hearing-aid wearers answered all questions twice, based on listening unaided and aided.

All participants were paid a small gratuity for their participation. The treatment of participants was approved by the Australian Hearing Ethics Committee and conformed in all respects to the Australian government's National Statement on Ethical Conduct in Human Research. Note that a subset of these listeners also participated in another study described in a companion paper (Best et al, 2015, unpublished) and some of the data presented here (for the standard environment) also appear in that paper.

Hearing aids

Twenty five of the 28 HI listeners were regular hearing-aid wearers and participated in the experiment both with and without their own hearing aids. All hearing aids were behind-the-ear styles, seven with the receiver in the canal. They represented a variety of entry-level and high-end devices from Phonak, Resound, Siemens, Oticon, Bernafon, Unitron, and Rexton Day, and were set to the user's most common program for testing. We did not attempt to adjust the gain or compression settings to ensure uniformity across participants, but rather opted to use the settings each listener was accustomed to using in their daily lives.

The listening programs were most likely affected by different noise management technologies, of which directionality would have the most significant influence on the relative performance of devices in this task. In addition, the amount of directivity experienced by a listener can depend on the positioning of devices on the head and ears. Thus, to get an estimate of the directivity provided by each listener's hearing aids, unaided and aided directivity indices were measured in situ[2]. Extensive details of this measurement procedure can be found elsewhere (Keidser et al, 2013a). Briefly, the three-dimensional directivity index is calculated as the ratio of sensitivity to frontal sounds relative to sensitivity averaged across all directions. As the directivity index is frequency-dependent, a frequency weighting based on the articulation index can be used to derive a single directivity index in dB that is relevant for speech (the AIDI).

These measurements confirmed that different participants experienced different levels of directionality using their most common program. For each individual, the amount of directivity gained from wearing hearing aids (relative to that provided naturally by the listener's own head and ears) was captured as the difference in AIDI between the unaided and aided conditions. This value varied from around − 1 dB to 4 dB across participants.

Environment and stimuli

Testing took place in a large anechoic chamber fitted with a three-dimensional loudspeaker array of radius 1.8 m. The array was

configured with 41 equalized loudspeakers (Tannoy V8); 16 loud-speakers were equally spaced at 0° elevation, eight at ± 30° elevation, four at ± 60° elevation, and one loudspeaker was positioned directly above the center of the array. Stimulus playback was via a PC with a soundcard (RME MADI) connected to two D/A converters (RME M-32) and 11 four-channel amplifiers (Yamaha XM4180).

The listener was seated such that the head was in the center of the loudspeaker array, facing the frontal loudspeaker, and wore a small lapel microphone in order to be heard clearly by the experimenter who was seated outside the chamber wearing headphones. The experimenter monitored participants via webcam to ensure they maintained a relatively fixed head position, and could talk to them via an intercom as required.

Target sentences were Bamford-Kowal-Bench sentences spoken by an Australian male talker (Keidser et al, 2002). In the standard environment, targets were presented from the frontal loudspeaker (0°/0°, azimuth/elevation), and the background was multi-talker babble generated by presenting four independent samples of eight-talker speech babble from the four loudspeakers positioned at ± 45°/0° and ± 135°/0° (Figure 2, top). In the complex environment, a large reverberant room (dimensions: $15 \times 8.5 \times 2.8$ m) with a reverberation time (T_{30}) of approximately 0.5 s was simulated using ODEON software (Rindel, 2000). The room simulation incorporated a kitchen area, and 12 tables each surrounded by six chairs. The cupboards in the kitchen, all tables and one third of the chairs were simulated as wooden plates. The rest of the chairs were considered as "occupied with audience" and included the absorption characteristics of a person. The materials assigned to the walls, ceiling, and floor are indicated in Figure 2. To make the talker/listener configuration comparable to the standard environment, the listener was positioned amongst the tables and chairs, and the target talker was given a virtual position in the room in front of the listener at a distance of 2.0 m (Figure 2, bottom). This resulted in a direct-to-reverberation ratio (DRR) of about 3 dB. The background consisted of seven conversations between pairs of talkers seated at the tables and facing each other. This resulted in 14 masker talkers distributed around the listener at different horizontal directions, distances, and facing angles. Since the directional characteristic of the talker-sources was included in the simulation, the facing angle had a strong, frequency-dependent effect on the DRR, which varied between − 14 dB and − 1 dB. The dialogues spoken by the maskers were recorded in the anechoic chamber by a mix of Australian-accented talkers (six male, eight female), using transcripts taken from the listening comprehension component of the International English Language Testing System (Cambridge University Press).

Room impulse responses (RIRs) generated in ODEON were converted to loudspeaker signals using a loudspeaker-based auralization toolbox (LoRA; Favrot & Buchholz, 2010). Briefly, for each talker-listener setup this toolbox derives a multi-channel RIR (one channel per loudspeaker) and convolves it with the corresponding anechoic talker signal, resulting in a multi-channel sound file. In this process, the direct sound and early reflections (up to a reflection order of four) are realized by image sources and assigned to the loudspeakers that best match their true source direction. The diffuse part of each impulse response is realized by multiplying directional energy envelopes for each loudspeaker with uncorrelated samples of noise. Since the direct sound (and all early reflections) of the target and individual masking talkers were realized by individual loudspeakers, the accuracy of the reproduced sound field is not limited in frequency bandwidth as would be the case for sound-field reconstruction techniques such as higher-order Ambisonics or wave-field synthesis (e.g. Daniel
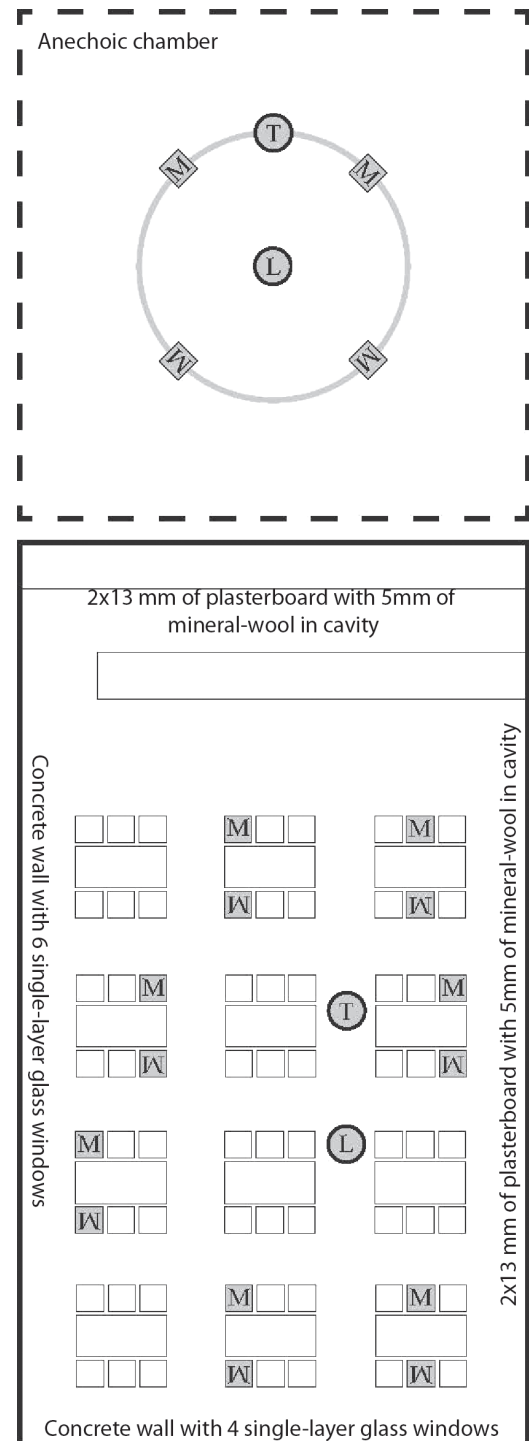
Figure 2. Schematic layout of the two environments. The standard environment (top) was an anechoic chamber, with a target (T) located directly in front of the listener (L), and four babble maskers (M) located at ± 45° and ± 135° azimuth. The complex environment (bottom) was a simulated reverberant cafeteria, including a kitchen area at one end of the room, and tables and chairs in the main area. The target was located directly in front of the listener and seven pairs of speech maskers were distributed in the room at different azimuths and distances. The floor was simulated as 6-mm pile carpet on closed-cell foam and the ceiling as 25 mm of mineral wool suspended by 200 mm from a concrete ceiling.

Table 1. Summary of room acoustic parameters of the complex and standard environments as predicted by the ODEON software. DRR: Direct-to-reverberation ratio; C50: Clarity; STI: Speech transmission index; SPL: Sound pressure level.

|	DRR (dB)			C50 (dB)			STI in quiet			SPL (dB)		
	min	*mean*	*max*	*min*	*mean*	*max*	*min*	*mean*	*max*	*min*	*mean*	*max*
Complex												
masker	−14.3	−6.9	−0.8	1.4	6.0	8.0	0.64	0.73	0.8	54.3	56.2	57.9
target		3.2			11.7			0.86			varied adaptively	
Standard												
masker		∞			∞			1			50	
target		∞			∞			1			varied adaptively	

et al, 2003). The accuracy of the applied ODEON room simulation software has been verified in different round robin studies (e.g. Bork, 2005) and the accuracy of the applied sound reproduction techniques by Favrot and Buchholz (2010) and Oreinos and Buchholz (2014). Hence, even though the applied cafeteria environment was not verified by direct comparison with a real room, it can be assumed that the cafeteria environment provides a realistic, or at least plausible, realization of an example reverberant environment.

In both environments, the background noise was presented continuously throughout a block by looping segments of about five minutes at a fixed level of 65 dB SPL (measured in the center of the array). This sound pressure level was estimated by the Odeon software and can be considered a realistic value. Although the overall sound pressure level of the background noise was matched, the properties of the two environments differed in many other ways. In order to highlight some relevant differences, a number of acoustic parameters are given in Table 1, which include the DRR (e.g. Zahorik, 2002), clarity (C50: ISO 3382) as a measure of the early-to-late reflection energy ratio, speech transmission index (STI: IEC 60268-16) measured in quiet, and sound pressure levels (SPL) for all the individual speech sources measured in isolation. The acoustic parameters for the cafeteria environment were calculated from the simulated room impulse responses measured in the center of the 3D loudspeaker array with an omni-directional microphone. The acoustic parameters for the standard environment are idealized and will have been modified slightly by the non-ideal playback environment. Additionally, the different target and masker signals were recorded at the ears of a Bruel & Kjaer 4128C Head and Torso Simulator (HATS). The resulting long-term power spectra calculated in auditory critical bands (i.e. critical band levels) are shown in Figure 3 (top and middle panels) for an example broadband SNR of 0 dB. Due to the similarity of the spectra at the left and right ear here only spectra averaged across the two ears are shown. The corresponding SNR as a function of frequency is shown in the bottom panel of Figure 3. Significant frequency-dependent differences of up to 7 dB can be observed in the different target and masker spectra as well as in the SNR. In order to illustrate the potential effect of hearing loss on target and masker audibility, the critical band level of threshold simulating noise (TSN) at the listeners' ears is shown in Figure 3 by the dotted lines. The TSN level was derived according to ANSI S3.5-1997 (Table 1), for the average hearing loss shown in Figure 1 (solid line, circles). Figure 4 illustrates the temporal behavior of the different maskers measured at the left ear of the HATS. The example envelopes (top and middle panels) were calculated at the output of an auditory bandpass filter with center frequency of 1000 Hz, normalized to its RMS value, and temporally smoothed with a 4th-order Butterworth lowpass filter with a cut-off

frequency of 32 Hz. The cafeteria noise exhibits significantly more low-frequency modulation ($\leq$ 32 Hz) than the standard noise (see modulation spectra in the bottom panel of Figure 4). This behavior is consistent across frequency. Although not shown here in detail, the squared interaural coherence c_{LR} was also calculated (Westermann et al, 2013), and revealed that both background noises are highly diffuse (i.e. exhibiting values $c_{12} < 0.2$ above about 200 Hz).

Procedures

SRT testing was conducted using automated software for the presentation and scoring of sentences (see Keidser et al, 2013b for details). On each trial, a sentence was presented and listeners spoke
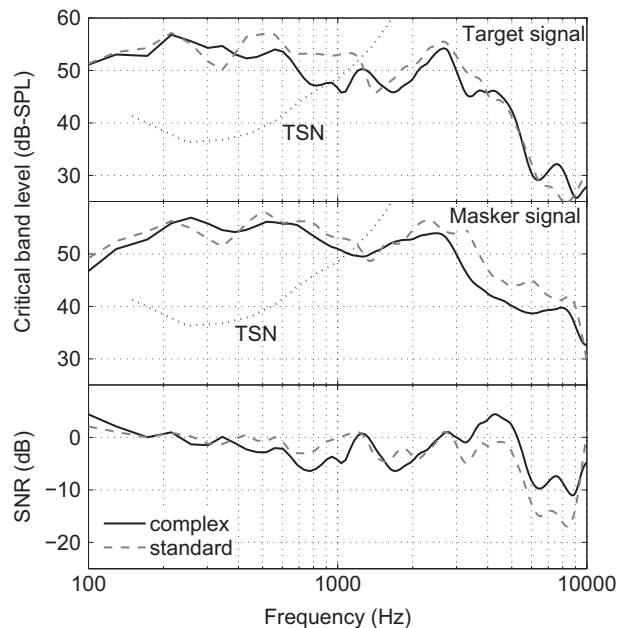
Figure 3. Critical band levels for the target (top panel) and masker (middle panel) signals for the standard (dashed gray lines) and complex (solid black lines) environment measured at the ears of a HATS placed inside the center of the 3D loudspeaker array. Due to the similarity of the ear signals the levels were averaged across the two ears. The critical band level of threshold simulating noise (TSN) is shown by the dotted lines. The bottom panel shows the SNR calculated as the difference between the above target and masker signals. The applied broadband SNR was 0 dB.
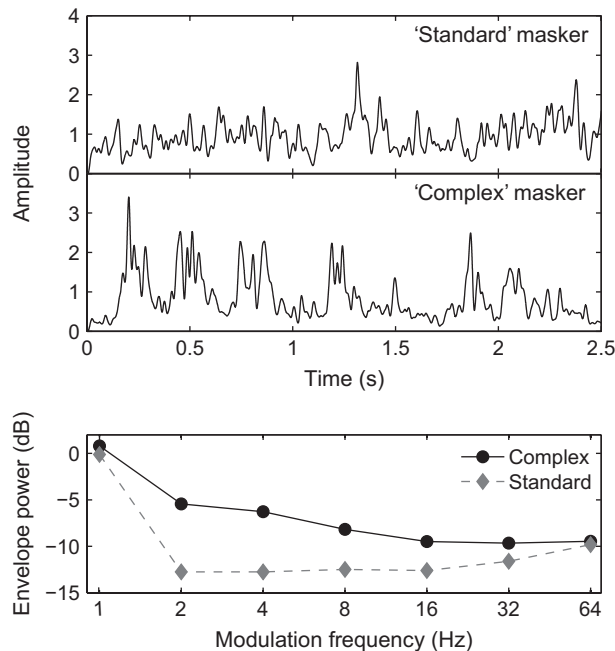
Figure 4. Temporal Hilbert envelopes of the masker in the standard (top panel) and complex (middle panel) condition and recorded at the left ear of a HATS placed in the center of the 3D loudspeaker array. The envelopes were calculated from the output of an auditory bandpass filter with a center frequency of 1000 Hz, and temporally smoothed using a 4th-order Butterworth lowpass filter with a cut-off frequency of 32 Hz. The corresponding amplitude modulation spectra are shown in the bottom panel. Signals were presented at 65 dB SPL.

aloud their responses. The experimenter entered the number of correct morphemes (out of a possible 3–8) into the software program, which triggered the next sentence.

For each environment (and hearing-aid condition, in the HI group) four blocks of trials were completed. In the first block, an adaptive procedure was used to estimate the 50% SRT within a standard error of 0.8 dB (see Keidser et al, 2013b for details). Three blocks were then completed using fixed target levels corresponding to fixed SNRs at the estimated SRT as well as at 2 dB above and below the estimated SRT. Each of these blocks consisted of 32 sentences. Sentences were paired with different SNRs for different subjects, and the order of presentation was randomized. No sentence was presented more than once to any listener. Each block took approximately five minutes to complete, for a total testing time of around 40 minutes. The order of testing of the two environments (and hearing-aid conditions, in the HI group) was counterbalanced across participants.

It is worth noting that the use of a fixed masker level means that different listeners were tested at different overall sensation levels depending on their hearing thresholds and the presence/absence of hearing aids. In addition, because the SNRs were chosen separately for each listener in each environment and hearing aid condition, the test SNRs also varied.

Percent correct scores at the three fixed SNRs were used to generate psychometric functions. Logistic functions were fit to the raw scores using the psignifit toolbox version 2.5.6 for MATLAB (see http://bootstrap-software.org/psignifit/) which implements the maximum-likelihood method described by Wichmann and Hill (2001).

Finally, SRTs (the SNR at 50% correct) were extracted. Note that the adaptively measured SRTs were only used to estimate the best SNRs for the psychometric function measurements and were not considered in the final data analysis.

Results

SRTs in standard and complex environments
As shown in Figure 5, unaided SRTs in the standard and complex environments were strongly correlated ($r = 0.93$, $p < 0.001$). However, SRTs were consistently higher in the complex environment (all points lie above the diagonal) especially for the poorer listeners (gradient of the least-squares fit is 1.5 dB/dB). Averaged across listeners within a group, the mean increase in SRT in the complex environment was 1.48 dB (NH) and 4.17 dB (HI).

Multiple regression analyses were conducted to determine whether age and/or hearing loss were reliable predictors of unaided SRTs in the two environments. In the standard environment, 4FAHL was a significant predictor but age was not (4FAHL: $\beta = 0.86$, $p < 0.001$; age: $\beta = -0.29$, $p = 0.77$; overall model fit: R^2adj $= 0.67$, $p < 0.001$). In other words, a better SNR was required by those with more severe hearing loss. The same was true in the complex environment (4FAHL: $\beta = 0.88$, $p < 0.001$; age: $\beta = -0.01$, $p = 0.91$; overall model fit: R^2adj $= 0.75$, $p < 0.001$).

When only the HI group was considered, 4FAHL (but not age) was again a significant predictor of unaided performance in the standard environment (4FAHL: $\beta = 0.72$, $p < 0.001$; age: $\beta = 0.00$, $p = 0.98$; overall model fit: R^2adj $= 0.47$, $p < 0.001$) as well as the complex environment (4FAHL: $\beta = 0.74$, $p < 0.001$; age: $\beta = 0.04$, $p = 0.79$; overall model fit: R^2adj $= 0.50$, $p < 0.001$). For aided performance in this group, 4FAHL was a much weaker predictor, in both the standard environment (4FAHL: $\beta = 0.42$, $p = 0.05$; age: $\beta = 0.23$, $p = 0.27$; overall model fit: R^2adj $= 0.10$, $p = 0.12$) and the complex environment (4FAHL: $\beta = 0.50$, $p = 0.02$; age: $\beta = 0.31$, $p = 0.12$;
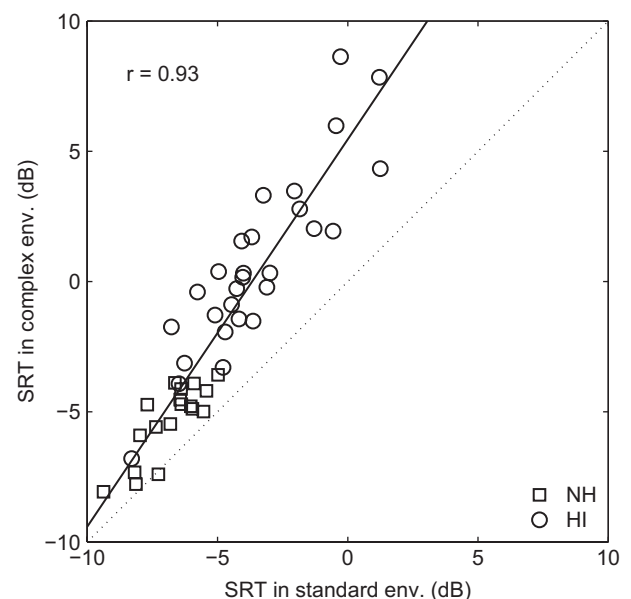
Figure 5. Scatterplot showing individual SRTs in the complex environment against SRTs in the standard environment. Different symbols indicate NH listeners (squares) and unaided HI listeners (circles). The solid line shows the least squares fit.

overall model fit: R^2adj = 0.20, p = 0.04), suggesting that amplification did not have the same effect across degree of hearing loss.

Hearing-aid benefits

To examine the effects of amplification in the two environments, a repeated-measures ANOVA was conducted on the SRTs for the 25 hearing-aid wearers. This analysis revealed significant main effects of hearing aid condition (unaided/aided; $F(1,24) = 10.53$, $p = 0.003$), and environment (standard/complex; $F(1,24) = 324.24$, $p < 0.001$), as well as a significant interaction ($F(1,24) = 6.73$, $p = 0.016$). The interaction reflects the fact that benefits were larger in the complex environment (mean 1.8 dB vs. 0.8 dB in the standard environment). As shown in Figure 6, benefits were correlated across the two environments ($r = 0.68$, $p < 0.001$). Quite strikingly, however, there was a large range of benefits across listeners on both environments (from −2.6 to 5.6 dB in the standard environment, and from −1.2 to 8.2 dB in the complex environment).

There are several factors that may contribute to the large individual differences in hearing-aid benefit we observed. First, listeners wore their own devices and thus prescriptions were not uniform. Secondly, our directivity index measurements indicated that different listeners experienced different levels of directionality (see sub-section "Hearing aids", in the Methods section). A third factor that may influence benefit is a listener's hearing loss; listeners with more severe losses received stimuli at lower sensation levels where amplification is most effective. Finally, the benefit of hearing aids was measured at different SNRs for different listeners (dictated by their unaided SRT). Given that amplification cannot improve the audibility of speech when noise is the limiting factor, the baseline SNR may be critically important in determining benefit.

To understand which of these relevant variables contributed most to the hearing-aid benefits observed in our two environments, multiple regression analyses were conducted. Given the wide variety of gain and compression settings across the hearing aids it was not possible to capture that variable in a single meaningful number. Thus the predictors included in the regression were AIDI difference, hearing loss (4FAHL), and SNR. In the standard environment, both AIDI difference and SNR were significant predictors (AIDI: $\beta = 0.37$, $p = 0.01$; SNR: $\beta = 0.60$, $p = 0.005$; overall model fit: R^2adj = 0.64, $p < 0.001$), suggesting that those fitted with stronger directionality and tested at higher SNRs showed larger benefits. In the complex environment, only the SNR was a significant predictor ($\beta = 0.68$, $p = 0.001$; overall model fit: R^2adj = 0.69, $p < 0.001$).

### *Relationship between objective and subjective measures*

As one of the broad motivations of this work is to provide more accurate predictions of real-world performance, it was of interest to examine how the SRTs measured in our two environments relate to the self-reported ratings of hearing ability collected in the questionnaire. A single score was calculated for each subject (separately for unaided and aided listening, where appropriate) by averaging over a subset of the SSQ scores. Specifically, eight questions were chosen that referred to situations involving selective attention to speech in the presence of noise, reverberation, or other talkers (speech items 1, 4, 5, 6, 7, 8, 9, 11).

Correlations between unaided self-report scores and unaided SRTs (including both NH and HI listeners) were highly significant in both environments (standard: $r = -0.78$, $p < 0.001$; complex: $r = -0.81$, $p < 0.001$). When only the HI group were considered, correlations were lower but still significant (standard: $r = -0.59$, $p = 0.001$; complex: $r = -0.63$, $p < 0.001$). Correlations for aided scores in this group were lower again (standard: $r = -0.49$, $p = 0.01$; complex: $r = -0.39$, $p < 0.05$). These correlations are comparable to those reported by Mendel et al (2007), who compared scores on various clinical speech in noise tests to ratings on the Hearing Aid Performance Inventory (their correlations ranged from 0.45–0.80).

Discussion

Results from this experiment indicate that the use of a more complex and realistic acoustic environment can change the psychometric properties of a simple sentence test. Using the same large group of listeners, we found that thresholds were higher on average in the complex environment. This increase was despite the fact that the cafeteria noise was more modulated (Figure 4) and thus provided more opportunity for clean glimpses of the target to be obtained ("listening in the dips"). The increase in threshold has several possible causes. Previous studies have demonstrated that complex maskers, especially those comprised of intelligible speech, can lead to increased thresholds by causing "central" or "informational" masking (e.g. Carhart et al, 1969; Brungart, 2001; Brungart et al, 2001; Best et al, 2012). Reverberation also tends to increase thresholds by degrading target speech information, temporally smearing targets and maskers, and reducing the ability to suppress spatially separated maskers (e.g. Culling et al, 2003; Lavandier & Culling, 2007, 2008; George et al, 2008). No doubt some combination of these factors (and possibly others) led to the increased SRTs we observed in the cafeteria environment. In general, modifications to speech tests that shift SRTs towards positive SNRs are useful for the goal of increasing real-world relevance, as environmental SNRs are generally above 0 dB (e.g. Pearsons et al, 1976; Smeds et al, 2012).

An interesting observation in this study was the interaction between hearing loss and environment, i.e. that the increase in unaided SRTs in the complex environment was particularly strong for the listeners
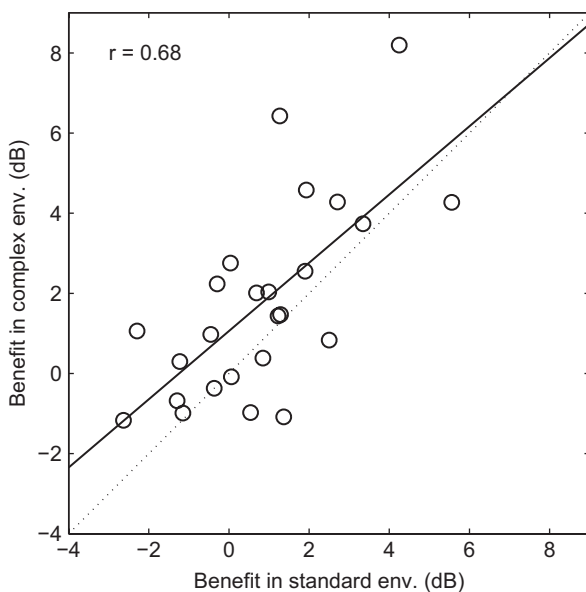
Figure 6. Scatterplot showing individual hearing-aid benefits in the complex environment against benefits in the standard environment (positive benefits indicate a reduction in the SRT, i.e. better aided performance). The solid line shows the least squares fit.

with the poorest hearing. There are several reasons why this might be the case.

First, it may be that NH listeners were better able to listen in the dips of the complex masker to offset the overall increase in difficulty discussed above. Previous studies have shown that HI listeners receive less benefit from listening in the dips largely because of reductions in audibility of the target and masker signals (e.g. Festen & Plomp, 1990; George et al, 2006; Bernstein & Grant, 2009; Christiansen & Dau, 2012; Rhebergen et al, 2014). To illustrate the potential effect of audibility on SRTs in the current study, unaided critical band levels are shown in Figure 3 (middle panel) for the complex and standard masker as well as for threshold simulating noise calculated for the average hearing loss shown in Figure 1. The "external" maskers are dominant only below about 1500 Hz, and above 1500 Hz the auditory internal, threshold simulating noise is dominant. Hence, the effect of the "external" maskers on the SRT is limited to rather low frequencies, and would be even further limited by a more severe hearing loss. Moreover, in modulated maskers such as the cafeteria noise, short-term levels in spectrotemporal dips can be more than 15 dB below the shown long-term levels. Similarly it can be deduced from Figure 3 (top panel) for the target signal that not only absolute target audibility but also the "audible bandwidth" is limited by the target level (or SNR) as well as hearing loss. Hence, the benefit received from listening in the dips is heavily affected by target and masker sensation level, SNR, and masker modulation spectrum (or type of masker) and is expected to decrease with increasing hearing loss.

Second, differences in the spectra of the target and masker signals at the listener's ears affect the frequency dependency of the SNR. According to Figure 3 (bottom panel) the SNR in the complex environment is reduced by about 4 dB at frequencies around 500–2000 Hz and increased by up to 6 dB at frequencies above about 3000 Hz. Hence, listeners with a more severe high-frequency hearing loss will not benefit from the improved SNR at high frequencies and will be penalized by the reduced SNR at mid frequencies.

Third, it is possible that HI listeners experience more informational masking in the presence of distracting maskers, although the literature on this issue is mixed (e.g. Helfer & Freyman, 2008; Agus et al, 2009; Woods et al, 2013).

Fourth, the presence of reverberation may exacerbate the effect of hearing loss on speech perception in noise, as has been reported previously (e.g. Harris & Reitz, 1985; Nabelek, 1988; Harris & Swenson, 1990).

Whatever the precise combination of factors at play, the important observation here is that testing in a more complex environment magnified the differences between SRTs, and thus may offer the practical advantage of better discriminating between individuals. For example, if two listeners differ by 1 dB in the standard environment, they would be expected to differ by 1.5 dB in the complex environment. However, this larger separation is useful only if the reliability of SRT measurements does not decrease by a similar or greater amount in the complex environment. If we assume that the threshold and slope of the fitted logistic functions are related to the mean and standard deviation of the underlying response distribution, then the inverse of the slope can be used as a surrogate measure of threshold reliability (e.g. see Strasburger, 2001). Slope values tended to be lower in the complex than in the standard environment (mean of 11 %/dB vs. 13 %/dB) and taking the rms average across listeners of slope values in each environment, we find that the reliability is 1.3 times lower in the complex environment. Thus there may be a small gain in sensitivity to individual differences in the complex

environment, although a more detailed investigation would be needed for a definitive answer.

Despite the observed changes when moving from the standard to the complex environment, the strong correlation between the two environments suggests that the complex environment did not substantially change the overall *ranking* of listeners; the "good listeners" in one environment were the "good listeners" in the other, etc. Thus we did not expect large improvements in our ability to predict the real-world performance of individual participants. Indeed self-reported abilities in speech-in-noise situations were only marginally better predicted by SRTs in the complex environment. It is worth noting also that self-report data are known to be variable and prone to individual biases, and thus it may be difficult to observe subtle improvements in predictions. Moreover, in this case participants answered the questionnaires rather broadly with reference to their general listening experiences. Improvements may be gained by obtaining self-report estimates in the specific environments being tested in the laboratory, and future work will explore this option.

Another question we asked was whether the use of a more complex environment would affect the estimated benefit of hearing aids when listening to speech in noise. Hearing-aid benefits tended to be larger in the complex environment, but for both environments we measured a wide range of hearing-aid benefits that included negative benefits. Analysis showed that larger hearing-aid benefits were strongly associated with higher test SNRs, with an additional contribution of directionality in the standard environment. The effect of SNR is illustrated further in Figure 7. Here, aided SRTs are plotted as a function of unaided SRTs for all HI listeners in both environments. It can be seen that the SRTs tended to diverge (i.e. hearing-aid benefits increased) as the unaided SRTs increased. Indeed it appears that large hearing-aid benefits only occurred when testing was in the positive SNR region, which mostly occurred for the more challenging complex environment (for the reasons discussed above). Our observation that baseline SNR affects hearing-aid benefits is consistent with the arguments of Plomp (1986), who has shown that
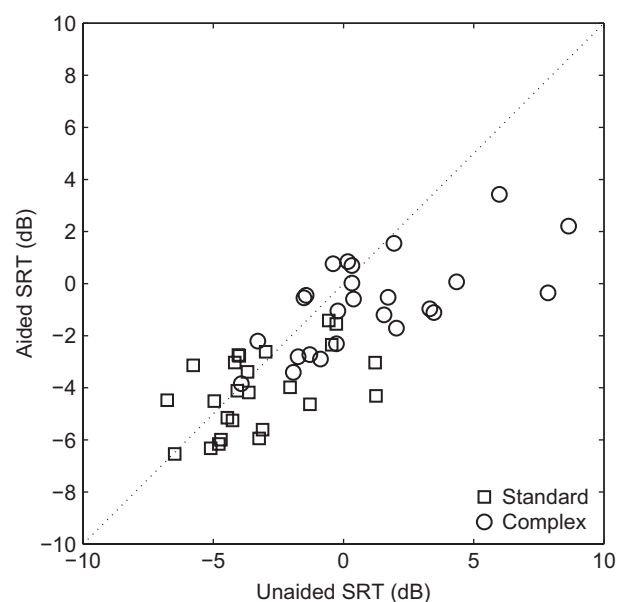
Figure 7. Scatterplot showing aided SRTs as a function of unaided SRTs in the standard (squares) and complex (circles) environments.

amplification mainly provides a benefit when audibility (and not external noise) is the limiting factor. Furthermore, the increased benefit measured for the more hearing-impaired listeners in the complex environment is in line with Rherbergen et al (2014), who have shown a nonlinear growth of masking for fluctuating noise maskers that is more pronounced for HI than NH listeners. It might be informative in future work to compare different environments at a fixed SNR, to remove such variables. It would also be of great interest to extend this work to more systematically examine the benefits provided by directional microphones, particularly given that the AIDI did not predict benefit in the complex environment, and the mounting evidence that current laboratory measures are poor predictors of self-reported real-world directional benefit (e.g. Cord et al, 2004, 2007; Wu, 2010).

In this study we examined the effects of just one manipulation designed to increase the realism of a speech in noise test. We observed modest changes in the psychometric properties of the test when conducted in a more complex acoustic environment, which translated into higher SRTs, larger difference between listeners, and larger estimated hearing-aid benefits. Even though these changes may be largely explained by the acoustic properties of the target and masker signals as well as the hearing losses of our listeners, it is important to note that by adding realism to the acoustic signals outcome measures can change significantly and may also provide improved ecological validity. Of course a range of other kinds of environments will need to be examined to determine how generalizable these results are. Moreover, manipulations of the target speech materials, e.g. by introducing talker-variability (Gilbert et al, 2013), varying syntactic complexity (Wingfield et al, 2006), varying sentence predictability (Wilson et al, 2007), or requiring comprehension of information (Tye-Murray et al, 2008) can also influence the characteristics of speech tests and may well interact with the environment. Ultimately a combination of ecologically motivated modifications might be the way to achieve more relevant outcome measures.

Conclusion

This study investigated the effects of measuring speech reception thresholds in a simulated reverberant cafeteria environment. The findings provide a starting point for future efforts to increase the real-world relevance of laboratory-based speech tests.

Notes

1. One NH listener was unable to complete the questionnaire and thus results from only 17 listeners are presented.
2. One listener did not have time to complete the directivity index measurement and thus only 24 measurements are analysed.

Acknowledgements

References

Agus T.R., Akeroyd M.A., Gatehouse S. & Warden D. 2009. Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise. *J Acoust Soc Am*, 126, 1926–1940.

ANSI S3.5-1997. Methods for calculation of the speech transmission index. A revision of ANSI S3.5-1969. *American National Standard*, 1997.

Bentler R.A., Niebuhr D.P., Getta J.P. & Anderson C.V. 1993. Longitudinal study of hearing-aid effectiveness II. Subjective measures. *J Sp Hear Res*, 36, 820–831.

Bernstein J.G. & Grant K.W. 2009. Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, 125, 2258–3372.

Best V., Marrone N., Mason C.R. & Kidd G., Jr. 2012. The influence of non-spatial factors on measures of spatial release from masking. *J Acoust Soc Am*, 13, 3103–3110.

Bork I. 2005. Report on the 3rd round robin on room acoustical computer simulation - Part II: Calculations. *Acta Acustica united with Acustica* 91, 753–763.

Brungart D.S. 2001. Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am*, 109, 1101–1109.

Brungart D.S., Simpson B.D., Ericson M.A. & Scott K.R. 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J Acoust Soc Am*, 110, 2527–2538.

Carhart R., Tillman T.W. & Greetis E.S. 1969. Perceptual masking in multiple sound backgrounds. *J Acoust Soc Am*, 45, 694–703.

Christiansen C. & Dau T. 2012. Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise. *J Acoust Soc Am*, 132, 1655–1666.

Compton-Conley C.L., Neuman A.C., Killion M. & Levitt H. 2004. Performance of directional microphones for hearing aids: Real-world versus simulation. *J Am Acad Audiol*, 15, 440–445.

Cord M., Baskent D., Kalluri S. & Moore B.C.J. 2007. Disparity between clinical assessment and real-world performance of hearing aids. *Hear Rev*, 14, 22–26.

Cord M.T., Surr R.K., Walden B.E. & Dyrlund O. 2004. Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids. *J Am Acad Audiol*, 15, 353–364.

Culling J.F., Hodder K.I. & Toh C.Y. 2003. Effects of reverberation on perceptual segregation of competing voices. *J Acoust Soc Am*, 114, 2871–2876.

Daniel J., Nicol R. & Moreau S. 2003. Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging *AES 114th Convention*. Amsterdam.

Favrot S. & Buchholz J.M. 2010. LoRA: A loudspeaker-based room auralization system. *Acta Acustica united with Acustica*, 96, 364–376.

Festen J.M. & Plomp R. 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am*, 88, 1725–1736.

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

George E.L.J., Festen J.M. & Houtgast T. 2006. Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, 120, 2295–2311.

George E.L.J., Festen J.M. & Houtgast T. 2008. The combined effects of reverberation and nonstationary noise on sentence intelligibility *J Acoust Soc Am*, 124, 1269–1277.

Gifford R.H. & Revit L.J. 2010. Speech perception for adult cochlear implant recipients in a realistic background noise: Effectiveness of preprocessing strategies and external options for improving speech recognition in noise. *J Am Acad Audiol*, 21, 441–451.

Gilbert J., Tamati T. & Pisoni D. 2013. Development, reliability and validity of PRESTO: A new high-variablity sentence recognition test. *J Am Acad Audiol*, 24, 26–36.

Harris R.W. & Reitz M.L. 1985. Effects of room reverberation and noise on speech discrimination by the elderly. *Audiology*, 24, 319–324.

Harris R.W. & Swenson D.W. 1990. Effects of reverberation and noise on speech recognition by adults with various amounts of sensorineural hearing impairment. *Audiology*, 29, 314–321.

Helfer K.S. & Freyman R.L. 2008. Aging and speech-on-speech masking. *Ear Hear*, 29, 87–98.

IEC 60268-16 Sound system equipment: Part 16: Objective rating of speech intelligibility by speech transmission index. *Edition 4.0 2011-06*, 2011.

ISO 3382 Acoustics: Measurement of the reverberation time of rooms with reference to other acoustical parameters. *Second Edition*, 1997.

Jerger J. 2009. Ecologically valid measures of hearing-aid performance. *Starkey Audiology Series*, 1, 1–4.

Keidser G., Ching T., Dillon H., Agung K., Brew C. et al. 2002. The National Acoustic Laboratories' (NAL) CDs of speech and noise for hearing-aid evaluation: Normative data and potential applications. *The Australian and New Zealand Journal of Audiology*, 24, 16–35.

Keidser G., Dillon H., Convery E. & Mejia J. 2013a. Factors influencing inter-individual variation in perceptual directional microphone benefit. *J Am Acad Audiol*, 24, 955–968.

Keidser G., Dillon H., Mejia J. & Nguyen C.V. 2013b. An algorithm that administers adaptive speech-in-noise testing to a specified reliability at selectable points on the psychometric function. *Int J Audiol*, 52, 795–800.

Killion M., Schulein R., Christensen L., Fabry D., Revit L. et al. 1998. Real-world performance of an ITE directional microphone. *Hearing Journal*, 51, 1–6.

Lavandier M. & Culling J.F. 2007. Speech segregation in rooms: Effects of reverberation on both target and interferer. *J Acoust Soc Am*, 122, 1713–1723.

Lavandier M. & Culling J.F. 2008. Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer. *J Acoust Soc Am*, 123, 2237–2248.

Mendel L.L. 2007. Objective and subjective hearing-aid assessment outcomes. *Am J Audiol*, 16, 118–129.

Nabelek A.K. 1988. Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. *J Acoust Soc Am*, 84, 476–484.

Oreinos C. & Buchholz J.M. 2014. Validation of realistic acoustical environments for listening tests using directional hearing aids *International Workshop on Acoustic Signal Enhancement (IWAENC 2014)*. Antibes, France.

Pearsons K.S., Bennett R.L. & Fidell S. 1976. Speech levels in various noise environments. *Environmental Health Effects Research Series, Office of Research and Development, U.S. Environmental Protection Agency*, EPA-600/1-77-025.

Plomp R. 1986. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *J Sp Hear Res*, 29, 146–154.

Revit L.J., Schulein R.B. & Julstrom S. 2002. Toward accurate assessment of real-world hearing aid benefit. *Hearing Review*, 9, 34–38.

Rhebergen K.S., Pool R.E. & Dreschler W.A. 2014. Characterizing the speech reception threshold in hearing-impaired listeners in relation to masker type and masker level. *J Acoust Soc Am*, 135, 1491–1505.

Rindel J.H. 2000. The use of computer modeling in room acoustics. *J Vibroengineering*, 3, 41–72.

Smeds K., Wolters F. & Rung M. 2012. Realistic signal-to-noise ratios *Proceedings of the International Hearing Aid Research Conference*. Lake Tahoe, pp. 93–94.

Strasburger H. 2001. Converting between measures of slope of the psychometric function. *Percept Psychophys*, 63, 1348–1355.

Tye-Murray N., Sommers M., Spehar B., Myerson J., Hale S. et al. 2008. Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *Int J Audiol*, 47, S31–S37.

Walden B.E., Surr R.K., Cord M.T., Edwards B. & Olson L. 2000. Comparison of benefits provided by different hearing aid technologies. *J Am Acad Audiol*, 11, 540–560.

Westermann A., Buchholz J.M. & Dau T. 2013. Binaural de-reverberation based on interaural coherence. *J Acoust Soc Am*, 133, 2767–2777.

Wichmann F.A. & Hill N.J. 2001. The psychometric function: I. Fitting, sampling and goodness-of-fit. *Percept Psychophys*, 63, 1293–1313.

Wilson R.H., McArdle R.A. & Smith S.L. 2007. An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J Sp Lang Hear Res*, 50, 844–856.

Wingfield A., McCoy S.L., Peele J.A., Tun P.A. & Cox L.C. 2006. Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *J Am Acad Audiol*, 17, 487–497.

Woods W.S., Kalluri S., Pentony S. & Nooraei N. 2013. Predicting the effect of hearing loss and audibility on amplified speech reception in a multi-talker listening scenario. *J Acoust Soc Am*, 133, 4268–4278.

Wu T.-H. 2010. Effect of age on directional microphone hearing-aid benefit and preference. *J Am Acad Audiol*, 21, 78–89.

Zahorik P. 2002. Assessing auditory distance perception using virtual acoustics. *J Acoust Soc Am*, 111, 1832–1846.