The effect of hearing loss on source-distance dependent speech intelligibility in rooms

A. Westermann and J. M. Buchholz

The effect of hearing loss on source-distance dependent speech intelligibility in rooms

A. Westermann[a),b)] and J. M. Buchholz[b)]
National Acoustic Laboratories, Australian Hearing, 16 University Avenue,
Macquarie University, New South Wales 2109, Australia
adwe@widex.com, jorg.buchholz@nal.gov.au

Abstract: Westermann and Buchholz [(2015). J. Acoust. Soc. Am. **137**(2), 757–767] found substantial improvements in speech reception thresholds (SRTs) for normal hearing listeners in a reverberant auditorium when the target talker was separated in distance from a two-talker masker. This study applied similar methodology, but tested listeners with a hearing impairment. On average, the participants received a 7 dB benefit in SRTs when the target was fixed at 0.5 m and the masker was moved from 0.5 to 10 m. But when the target was moved away, the SRTs increased by 5 dB. This indicates that hearing impaired listeners have difficulties suppressing nearby maskers while focusing attention on a far target.

1. Introduction

The auditory system employs multiple mechanisms to successfully understand speech in reverberant multi-talker environments. These auditory mechanisms are often disturbed by a hearing impairment (HI), which makes it hard (or even impossible) for HI listeners to communicate in such challenging "cocktail party scenarios" (e.g., Bronkhorst, 2000). Numerous studies have shown how normal hearing (NH) listeners, and to some degree also HI listeners, can take advantage of the angular separation as well as the voice characteristics of the individual talkers (e.g., Brungart *et al.*, 2001). Recently, Westermann and Buchholz (2015) showed how NH listeners can effectively use distance-related cues, especially those related to changes in the direct-to-reverberant ratio (DRR), to better understand a target talker in a background of masking talkers (for an overview on auditory distance perception, see Zahorik, 2005). Using binaural room impulse responses (BRIRs) measured in a reverberant auditorium, they presented a sentence test with the target and masker at different distances directly in front of the listener. To focus on reverberation cues, distance-dependent level and spectral changes were equalized. They investigated both a scenario in which the target was kept close (0.5 m) and the masker distance varied from 0.5 to 10 m, and a scenario in which the masker was kept close and the target distance varied. Measuring speech reception thresholds (SRTs), they found intelligibility improvements of up to 10 dB when the target was at 0.5 m distance and the masker was changed from 0.5 to 10 m. When the masker was kept close and the target was moved away the mean SRT still improved, but the individual SRTs varied largely. Some listeners received a substantial benefit from the spatial separation whereas other listeners performed even slightly worse than in the colocated condition. This large variability could not be explained by common objective measures. They therefore hypothesized it was caused by informational masking (IM), which was supported by the observation that all subjects received a substantial benefit when the speech masker was replaced by a mostly energetic, speech modulated noise masker. However, analyzing the masker errors in conditions with speech maskers showed that this behavior was not due to target-masker confusions as commonly observed in the IM dominated, colocated conditions (Ihlefeld and Shinn-Cunningham, 2008). From this they hypothesized that the nearby "clear" masker captured the attention of the listener over the "blurred" reverberant target and named this "distraction-based" IM, as in contrast to "confusion-based" IM. This segregation is in agreement with the conceptual framework provided by Shinn-Cunningham (2008),

---

[a)]Author to whom correspondence should be addressed.
[b)]Also at: Department of Linguistics, Macquarie University, Building C5A, Balaclava Road, North Ryde, NSW 2109, Australia.

which mention target-masker similarities (confusions) and drawing of exogenous attention (distractions) among the possible causes of IM.

Since previous studies have shown that HI listeners have severe deficits in utilizing reverberation cues for distance perception (Akeroyd *et al.*, 2007), it is important to investigate if HI listeners gain the same benefit as NH listeners when the distance between target and masker is varied. Furthermore, the effect of a HI on distraction-based IM is unknown.

2. Methods

2.1 Stimuli

As in Westermann and Buchholz (2015), two speech corpora were used in this experiment: the coordinate response measure (CRM; Bolia *et al.*, 2000) corpus and the speech material of the Listening in Spatialized Noise-Sentences test (LiSN-S; Cameron and Dillon, 2007). In the CRM corpus each sentence has the structure: "Ready [call sign] go to [color] [number] now," with eight call-signs, four colors (red, green, blue, and white) and eight numbers (1 through 8), resulting in 256 sentences for each of eight different talkers. SRTs were measured using only the four male talkers. Subjects were assigned the "Baron" call-sign, and asked to report the color/number corresponding to that speaker. Two different maskers were applied, a speech masker and a speech-modulated noise masker. The speech masker consisted of two randomly chosen CRM sentences with different talker and color/number combination from the target. The speech modulated noise masker was realized by applying the temporally smoothed broadband Hilbert envelope of each of the speech maskers to noise with the same long-term spectrum as all of the speech maskers (for details see Best *et al.*, 2013).

To allow conclusions with a more general validity, the target sentences and maskers from the LiSN-S were also tested. With the LiSN-S, SRTs are measured with a continuous two-talker masker (Cameron and Dillon, 2007) and thus, the target and masker speech signals are less synchronized than in the CRM corpus. The target and masker signals were realized by different female talkers.

All (anechoic) target and masker signals were convolved with BRIRs measured at different distances in an auditorium using a Brüel & Kjær (B&K, Denmark) Head and Torso Simulator (HATS). The auditorium had a reverberation time of $T_{30} = 1.9$ s at 2 kHz and a volume of approximately 1150 m^3. These were the same BRIRs as used in Westermann and Buchholz (2015). Three different spatial configurations were tested here; one where target and masker were colocated at a distance of 0.5 m ($M_{s0.5}T_{s0.5}$), one where the target talker was at 0.5 m distance and the masker at 10 m distance ($M_{s10}T_{s0.5}$) and one with the opposite of the latter ($M_{s0.5}T_{s10}$). Note that the labels of the spatial conditions show the applied distances as well as the masker type, i.e., M_{s10} indicates a speech masker at 10 m distance and $M_{n0.5}$ is a (speech modulated) noise masker at 0.5 m distance. At 0.5 m distance the direct sound provided the main energy and at 10 m the room reverberation, with a DRR of +15.1 dB and −7.7 dB, respectively. The room reverberation was always perceptually fused with the direct sound without any noticeable echoes.

To maintain the time-alignment of target and masker, which is only critical for the CRM corpus, the propagation delay introduced from the distance between sound sources was removed by time-aligning the direct sound component of the measured BRIRs. Furthermore, in order to minimize intelligibility improvements directly resulting from distance-dependent changes in long-term spectrum and overall level the maskers were equalized. The equalization, or spectral matching, was designed so that the equalized long-term spectra of the masker, analyzed in critical bands, always equaled the corresponding spectrum of the masker colocated with the target (i.e., either $T_{s0.5}M_{s0.5}$ or $T_{s10}M_{s10}$). Even though this process removed cues that are relevant for distance perception in general, it did not affect the short-term or time-varying spectral and spatial details of the masker that carry the main information on the DRR. However, the equalization might have slightly increased the involvement of IM as well as energetic masking (EM). Finite Impulse Response (FIR) equalization filters with a length of 512 taps (at a sampling frequency of 44.1 kHz) were designed and applied using MATLAB. The equalization procedure was applied to both the CRM and LiSN-S speech corpora.

2.2 Procedures

Experiments were carried out in a double-walled booth, using equalized, Sennheiser HD-215 circumaural headphones driven by a RME Hammerfall HDSPe AIO soundcard and a computer running a MATLAB GUI. Preceding each experiment, both air and

bone conduction audiometric thresholds were measured in octave bands from 250 to 8000 Hz. For both the CRM and LiSN-S the masker level was kept at a root-mean-square level of 60 dB sound pressure level (SPL), measured in a B&K type 4153 artificial ear before compensation for hearing loss. The target level was initially set to 67 dB SPL and varied relative to the masker following a one-up one-down rule to adaptively estimate the SRT. In order to (partly) compensate for audibility, linear amplification was applied according to the National Acoustic Laboratories-revised profound (NAL-RP) scheme (Dillon, 2001). The individually prescribed insertion gains were realized using 512-tap long FIR filters designed in MATLAB.

First the LiSN-S test was measured and then the CRM. Within each test the order of presentation was randomized and all conditions measured with the CRM corpus were repeated once. After the initial hearing screening and audiometry, the subjects were given verbal instruction read by the experimenter. Before the CRM experiment started, training was performed using one random condition to ensure familiarity with the GUI and understanding of the task. No training was applied before the LiSN-S test. Testing each subject required a single session of 1.5 h.

2.3 Subjects

Nine HI subjects (3 females and 6 males) participated, aged 44–77 years (mean 67). All subjects had symmetrical sloping sensorineural hearing losses, and all were native Australian English speakers and experienced hearing aid users. The individual audiograms are shown in Fig. 1 (solid lines). All subjects were active participants from the National Acoustic Laboratories database, and had significant experience with speech intelligibility tests. In order to allow a direct comparison between the derived HI and NH data, results from 16 NH listeners (<20 dB hearing level) were taken from Westermann and Buchholz, 2015. As the $M_{n0.5}T_{s10}$ condition was not measured for the CRM with the initially recruited HI listeners, four additional participants (three male and one female) with similar hearing losses (Fig. 1, dashed lines) as well as age (68–79 years; mean 74.75) as the original group were recruited to give an indication of performance in this condition. These participants were also tested on the $M_{s0.5}T_{s0.5}$, $M_{s0.5}T_{s10}$, and $M_{n10}T_{s0.5}$ conditions for reference purposes. As a consequence, different numbers of subjects participated in the different conditions, which is further described in Fig. 2.

3. Results

The left and right panels of Fig. 2(a) show the SRTs measured for the NH and HI listeners, respectively, using the CRM speech corpus. The filled black symbols denote the speech masker and the open symbols the speech-modulated noise masker. The lower panels show the corresponding spatial advantage, calculated as the difference between the individual per subject SRT in the colocated condition and the individual SRT per subject in the separated condition. When moving the speech masker from 0.5 m (i.e., colocated condition $M_{s0.5}T_{s0.5}$) to 10 m (i.e., spatially separated condition $M_{s10}T_{s0.5}$) the SRT decreased on average by about 7 dB for the HI listeners, i.e., the listener's performance is strongly improved. However, this improvement is smaller than the average
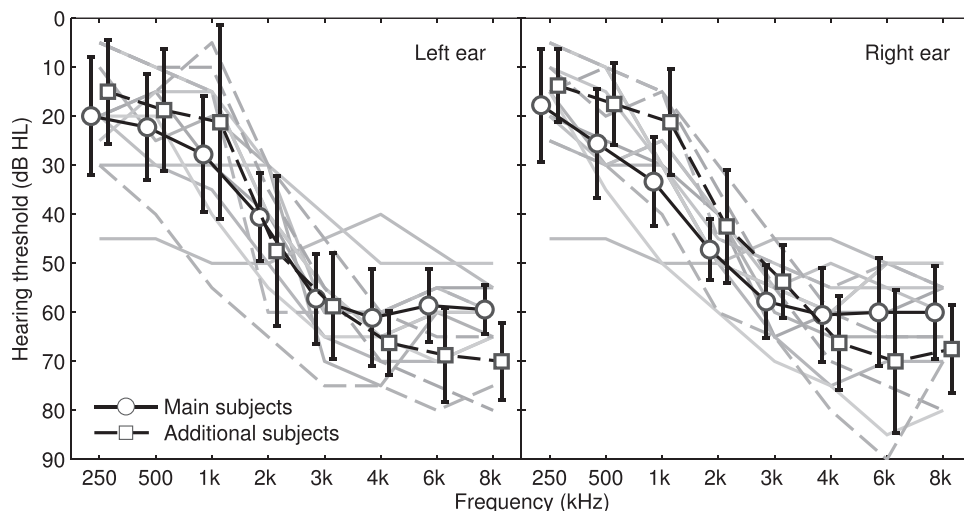

Fig. 1. Mean and standard deviation of hearing thresholds separated into nine main subjects (solid lines) and four additional subjects (dashed lines).
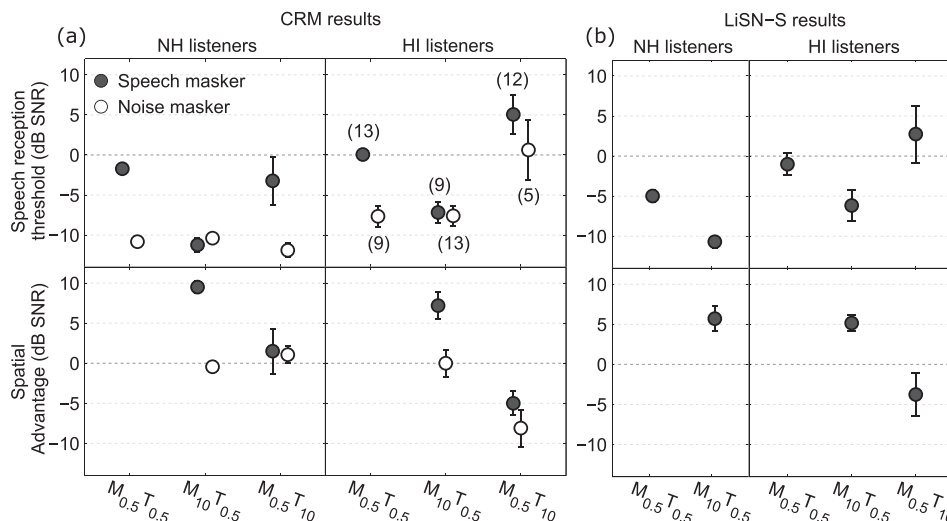
Fig. 2. Top panels: Mean and across-subject 95% confidence intervals of SRTs (a) measured with the CRM corpus and (b) measured with the LiSN-S corpus. Bottom panels: Mean and 95% confidence intervals of the spatial benefit. Note that 16 NH and 9 HI subjects participated in the different conditions unless indicated by a number in ().

improvement of 10 dB observed with the NH listeners. For the speech-modulated noise masker, the SRT in the colocated condition decreased to the same value as in the far-masker ($M_{n10}T_{s0.5}$) condition and thus, no spatial advantage was observed. This was the same for NH and HI subjects. When the masker was kept at 0.5 m and the target distance was increased from 0.5 to 10 m (i.e., from $M_{s0.5}T_{s0.5}$ to $M_{s0.5}T_{s10}$) the mean SRT increased (intelligibility decreased) by 5 dB for the HI listeners. This decrease in performance is in qualitative agreement with some of the NH subjects, but a significant number of NH subjects still showed a clear improvement as illustrated by the large spread of the NH data and discussed in Westermann and Buchholz (2015). It should be noted that the adaptive SRT lead to loudness discomfort in one subject in the $M_{s0.5}T_{s10}$ condition and could therefore not be retrieved.

The difference between the NH and HI listeners in the $M_{n0.5}T_{s10}$ condition, of about 12 dB, indicates that the performance of the HI listeners was greatly affected by the additional reverberation with the target at 10 m distance, where the NH listeners were largely unaffected (i.e., comparing $M_{n0.5}T_{s10}$ and $M_{n10}T_{s0.5}$).

Due to the different numbers of subjects that participated in the different CRM conditions, statistically treated here as data "missing at random," a linear mixed-effects model was applied to analyze the SRT data using the software R 3.1.3 with the packages nlme 3.1–120 and multcomp 1.4–0. Fitting the model to the HI data with condition as a fixed effect and a subject-specific intercept as the random effect showed a significant effect of condition ($F = 167.1$; $p < 0.001$). An estimate of the mean difference between pairs of conditions adjusted for multiple comparisons according to Hothorn et al. (2008) showed significant differences ($p < 0.001$) between all conditions except ($p \geq 0.97$) for $M_{n0.5}T_{s0.5}$, $M_{n10}T_{s0.5}$, and $M_{s10}T_{s0.5}$ as well as $M_{s0.5}T_{s0.5}$ and $M_{n0.5}T_{s10}$. An additional linear mixed-effects model with group (NH or HI), condition, and their interaction as fixed effects and a subject-specific intercept as a random effect was fitted to the data. It showed significance for all fixed effects ($p < 0.001$). An estimate of the mean difference between pairs of conditions with adjustment for multiple comparisons showed significant differences ($p \leq 0.005$) between groups (NH or HI) for all conditions except for $M_{s0.5}T_{s0.5}$ ($p = 0.18$). The error bars in Fig. 2(a) indicate 95% confidence intervals, which in the case of the SRTs were derived from the corresponding standard error of the mean for each condition separately and in the case of the spatial advantage from the described linear mixed-effects models.

Figure 2(b) shows the measured SRTs and the corresponding spatial advantage using the LiSN-S corpus. Similar to the CRM data for the HI listeners as well as to the NH data in the LiSN-S, SRTs decreased by approximately 5 dB when the speech masker was moved to 10 m (i.e., from $M_{s0.5}T_{s0.5}$ to $M_{s10}T_{s0.5}$) and SRTs increased by about 4 dB when the target was moved to a 10 m distance ($M_{s0.5}T_{s10}$). Applying the same statistical analysis as with the CRM data showed that also here the effect of condition was significant ($p < 0.001$) and that all pairs of mean SRT estimates were significantly different from each other. Similarly, fitting an additional model

comparing the NH and HI data, the effect of condition was not significant ($p = 0.126$) but the effect of group and the interaction was significant ($p < 0.001$). Finally, the estimated NH SRTs were all significantly different from the estimated HI SRTs ($p < 0.001$).

4. Discussion

Generally, both the CRM and LiSN-S results show that increasing the distance to a speech masker (i.e., comparing $M_{s0.5}T_{s0.5}$ and $M_{s10}T_{s0.5}$) results in an improvement of mean SRTs of about 7 dB in HI subjects (and about 10.5 dB in NH subjects), whereas the SRTs masked by speech modulated noise are unaffected by the spatial separation. This difference can be explained by considering the concepts of EM and IM (for a review, see Kidd *et al.*, 2007 or Shinn-Cunningham, 2008). The similarity between SRTs for the speech masker ($M_{s10}T_{s0.5}$) and the speech-modulated noise masker ($M_{n10}T_{s0.5}$) in the spatially separated condition suggests that the spatial separation aids the perceptual segregation of target and masker, which removes target-masker confusions, and thus, reduces IM. This is supported by the data in Table 1, where masker errors observed with the CRM are significantly lowered (almost halved) from 13.2% to 7.5% when the speech masker is moved further away.

Comparing the data for the NH and HI listeners illustrates that most SRTs are increased for the HI subjects, which is more pronounced in the spatially separated conditions. The increase in colocated thresholds is usually explained by decreased sensitivity to loudness cues, which typically provides the main segregation cue for resolving IM (Brungart *et al.*, 2001). The increase in SRTs in the spatially separated condition may be explained by increased EM in the HI listeners (see Best *et al.*, 2013) due to reduced target audibility as well as reduced temporal and spectral resolution, and maybe distorted spatial cues (Glyde *et al.*, 2015). This is supported by the observation that the SRTs for the purely energetic, speech-modulated noise masker are increased by the same amount as the SRTs for the speech masker.

When the speech masker was kept at 0.5 m and the target was moved further away ($M_{s0.5}T_{s10}$), SRTs increased for all HI subjects. This was different from the NH group, which also showed a large inter-subject variability. Since the NH subjects showed a consistent benefit with the speech-modulated noise masker in this condition, the subject-dependent behavior observed with the speech masker is likely linked to IM effects. However, the SRTs measured with the HI participants with the speech-modulated noise masker indicate that the HI listeners did perform substantially worse than the NH listeners when the target was further away. The detrimental effects of reverberation on speech are well documented for HI listeners (Helfer and Wilber, 1990) and this difference in SRTs is likely a direct cause thereof. Hence, when the target is further away the HI listeners are both affected by the IM effects of the nearby maskers and by EM due to the increased reverberation in the target speech component itself.

Considering the IM encountered in condition $M_{s0.5}T_{s10}$, Westermann and Buchholz (2015) argued that a distraction-based rather than a confusion-based nature of IM was encountered. According to Table 1 similar conclusions can be drawn for the HI listeners, as this condition provides the lowest number of masker errors (where the target and masker are confused) for all speech-masker conditions (i.e., 3.9% for HI and 5.7% for NH subjects). Westermann and Buchholz (2015) argued that the close and clear masker captures the attention of the listeners and distracts from the far and blurred target. If this is the case, then the results would indicate that the ability to selectively attend to the target, and thereby to suppress the distractors, are highly subject dependent and reduced in HI subjects. This might be linked to cognitive factors as well as auditory factors, which, due to the hearing loss as well as the increased age, may be both reduced in HI subjects. However, since the SRT in the $M_{s0.5}T_{s10}$ condition is highly positive (about 5 dB) it may have already reached an upper limit of IM above which the SRT does not increase any further (i.e., IM resolved by loudness cues). Independent of this alternative explanation, future studies should consider if, in

Table 1. Percentage of masker errors for the measured CRM results.

Near target			Far target		
Condition	HI	NH	Condition	HI	NH
$M_{s05}T_{s05}$	13.2%	14.8%	$M_{s05}T_{s05}$	13.2%	14.8%
$M_{s10}T_{s05}$	7.5%	7.0%	$M_{s05}T_{s10}$	3.9%	5.7%

particular, cognitive factors or abilities (such as the executive function) can explain the large differences between subjects for the far-target and close-masker condition with speech maskers.

References and links

Akeroyd, M., Gatehouse, S., and Blaschke, J. (**2007**). "The detection of differences in the cues to distance by elderly hearing-impaired listeners," J. Acoust. Soc. Am. **121**(2), 1077–1089.

Best, V., Thompson, E. R., Mason, C. R., and Kidd, G. (**2013**). "An energetic limit on spatial release from masking," J. Assoc. Res. Otolaryngol. **14**(4), 603–610.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**(2), 1065–1066.

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acta Acust. Acust. **86**(1), 117–128.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**(5), 2527–2538.

Cameron, S., and Dillon, H. (**2007**). "Development of the listening in spatialized Noise-Sentences Test (LiSN-S)," Ear Hear. **28**(2), 196–211.

Dillon, H. (**2001**). *Hearing Aids* (Thieme, New York), pp. 239–242.

Glyde, H., Buchholz, J. M., Nielsen, L., Best, V., Dillon, H., Cameron, S., and Hickson, L. (**2015**). "Effect of audibility on spatial release from speech-on-speech masking," J. Acoust. Soc. Am. **138**(5), 3311–3319.

Helfer, K. S., and Wilber, L. A. (**1990**). "Hearing loss, aging, and speech perception in reverberation and noise," J. Speech, Lang., Hear. Res. **33**(6), 149–155.

Hothorn, T., Bretz, F., and Westfall, P. (**2008**). "Simultaneous inference in general parametric models," Biometr. J. **50**(6), 346–363.

Ihlefeld, A., and Shinn-Cunningham, B. G. (**2008**). "Spatial release from energetic and informational masking in a selective speech identification task," J. Acoust. Soc. Am. **123**(6), 4369–4379.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2007**). "Informational masking," in *Auditory Perception of Sound Sources* (Springer, New York), pp. 143–189.

Shinn-Cunningham, B. G. (**2008**). "Object-based auditory and visual attention," Trends Cognit. Sci. **12**(5), 182–186.

Westermann, A., and Buchholz, J. M. (**2015**). "The effect of spatial separation in distance on the intelligibility of speech in rooms," J. Acoust. Soc. Am. **137**(2015), 757–767.

Zahorik, P. (**2005**). "Auditory distance perception in humans: A summary of past and present research," Acta Acust. Acust. **91**, 409–420.