View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Queensland eSpace

The importance of interaural time differences and level differences in spatial release from masking

Helen Glyde[a),b)] **and Jörg M. Buchholz**[b)]
The HEARing Cooperative Research Centre,
550 Swanston Street, Carlton, Victoria 3053, Australia
helen.glyde@nal.gov.au, jorg.buchholz@nal.gov

Harvey Dillon[c)] **and Sharon Cameron**[c)]
National Acoustic Laboratories, 16 University Avenue, Macquarie University,
New South Wales 2067, Australia
harvey.dillon@nal.gov, sharon.cameron@nal.gov

Louise Hickson[c)]
School of Health and Rehabilitation Sciences, University of Queensland,
St Lucia, Queensland 4067, Australia
lhickson@uq.edu.au

Abstract: Numerous studies have described improvements in speech understanding when interaural time differences (ITDs) and interaural level differences (ILDs) are present. The present study aimed to investigate whether either cue in isolation can elicit spatial release from masking (SRM) in a speech-on-speech masking paradigm with maskers positioned symmetrically around the listener. Twelve adults were tested using three presentations of the Listening in Spatialized Noise–Sentences Test, with each presentation modified to contain different interaural cues in the stimuli. Results suggest that ILDs provide a similar amount of SRM as ITDs and ILDs combined. ITDs alone provide significantly less benefit.

© 2013 Acoustical Society of America

PACS numbers: 43.66.Pn, 43.66.Dc [QJF]
Date Received: April 21, 2013 **Date Accepted:** June 14, 2013

1. Introduction

The role of spatial release from masking (SRM) in successful speech understanding in noise has been well-established in the literature with many researchers demonstrating that normal-hearing adults can understand speech at significantly poorer signal-to-noise ratios when speech is spatially separated from the noise rather than co-located (e.g., Refs. 1–5). The ability of hearing-impaired adults to achieve SRM has been shown to be deficient using a wide variety of experimental paradigms (e.g., Refs. 1, 6–10). Despite the large body of research into SRM, surprisingly little research exists regarding which acoustic cues underpin this ability. Further research in this area is critical if progress is to be made regarding how best to address the SRM deficits seen in hearing-impaired individuals. This paper reports a follow-up experiment to the Glyde *et al.*[6] study, aimed at identifying whether interaural time differences (ITDs) and interaural level differences (ILDs) in isolation can produce SRM using a specific experimental paradigm which incorporates two speech maskers symmetrically placed around the listener.

[a)] Author to whom correspondence should be addressed. Also at: The University of Queensland, St. Lucia, QLD 4067, Australia.
[b)] Also at: The National Acoustic Laboratories, Macquarie University, NSW 2104, Australia.
[c)] Also at: The HEARing Cooperative Research Centre, Carlton, VIC 3053, Australia.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 130.102.158.13 On: Thu, 29 Oct 2015 06:08:17

Previous research that examined the role of ITDs and ILDs in SRM utilized experimental procedures in which only one masking sound source was present. Under these conditions the contribution of ITDs was found to be smaller than that of ILDs.[11–13] Though Bronkhorst and Plomp[12] demonstrated that SRM was largest when both interaural cues were present, the better signal-to-noise ratio (SNR) received by one ear due to the existence of only one masker did allow ILDs to be used to achieve SRM through essentially monaural listening. Therefore, one could assume that the addition of an extra masker on the opposite side of the head would reduce the importance of ILDs as it would prevent one ear from having a consistently better SNR. However, if the maskers are speech based, as was the case in the Bronkhorst and Plomp experiment, it is still possible that ILDs could be used to exploit momentary differences in SNR between the ears.

Culling *et al.*[14] used head-related transfer functions (HRTFs) to examine the contribution of ITDs and ILDs to SRM when three distracting sound sources were present. They tested three different spatially separated, non-symmetrical conditions which varied based on the position of the distracting speech in the horizontal plane (either $-30°$, $+60°$, $+90°$ or $+30°$, $+60°$, $+90°$ or $+90°$, $+90°$, $+90°$) while the target was always at 0° azimuth. Using sentences voiced by a male speaker as both the target and the distracting material, Culling *et al.* observed no significant difference in benefit gained from ITDs or ILDs, when collapsed across spatial positions. However, when distracters were present at both negative and positive azimuth positions, no SRM was observed in the ILD alone condition.

Kidd *et al.*[15] conducted one of the few studies that have directly investigated the importance of ITDs and ILDs when two symmetrically placed maskers are used. High pass and low pass filtering was applied to speech stimuli to alter access to ITDs and ILDs. Contrary to the results of Culling *et al.*,[14] Kidd and colleagues found that both cues could provide SRM. However, as both ITDs and ILDs are present to some extent across the frequency range of speech, it is possible that the SRM obtained in either the high pass or low pass filtering condition could have depended, to some degree, on either cue. Therefore, more information is needed regarding the role of interaural cues when symmetrically placed maskers exist.

2. Method

This study was conducted under the ethical oversight of the Australian Hearing Ethics Committee and the University of Queensland Behavioral and Social Sciences Ethical Review Committee.

2.1 Stimuli

The raw speech materials used here are the same as those described in detail in Glyde *et al.*[6] Briefly, the Listening in Spatialized Noise–Sentences Test (LiSN-S) has 120 target sentences which are voiced by a female speaker. In each condition, the target sentences are presented at 0° azimuth. The LiSN-S conditions differ from each other based on the spatial location (0° vs $\pm 90°$ azimuth) or vocal identity (same as, or different female speakers to, the speaker of the target sentences) of the two looped children's stories which form the distracting speech. The four listening conditions are: same voice at 0° (SV0), same voice at $\pm 90°$ (SV90), different voices at 0° (DV0), and different voices at $\pm 90°$ (DV90). The LiSN-S speech materials are considered to be high in informational masking, particularly in the same voice conditions because of the high confusability between target and maskers. The different spatial conditions were realized using non-individualized HRTFs.

To isolate the SRM achieved with each interaural cue, the speech materials were modified by processing the original HRTFs in MATLAB before they were convolved with the raw speech material. Minimum-phase versions of the original HRTFs were derived via the Hilbert transform and used to realize HRTFs without ITDs. That is, a filter with the same magnitude response as the original HRTFs, but with the same

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 130.102.158.13 On: Thu, 29 Oct 2015 06:08:17

delay at all frequencies irrespective of direction was applied. To create HRTFs without ILDs the original $\pm 90°$ HRTFs were modified using a minimum-phase filter whose magnitude response was equal to the $0°$ HRTF minus (in dB) each of the $\pm 90°$ HRTFs. The resulting HRTFs had the ITDs appropriate to the $\pm 90°$ directions, but the same ILDs as the average of the original left and right HRTFs for the $0°$ condition. The magnitude spectra of the resulting HRTFs are shown in Fig. 1. It should be noted that even in the $0°$ conditions small ILDs are present in the high frequencies due to the slight asymmetry of the Kemar manikin which was used to derive the original HRTFs. For more information regarding this type of processing, please see Kulkarni *et al.*[16]

This processing resulted in three versions of the LiSN-S test which differed from each other based on the interaural cues available. These versions are referred to as: (i) the reference condition (which contained both ITDs and ILDs), (ii) the ITDs only condition, and (iii) the ILDs only condition.

2.2 Participants

Twelve native English speakers with normal hearing (range 24–53 yr, mean 33.6 yr) participated in the research. Hearing thresholds did not exceed 20 dB HL at any octave frequency between 250 and 8000 Hz.

2.3 Procedure

Testing for each participant was conducted over three 15 min sessions. In each session, testing for one of the three cue conditions was completed. Test sessions were between one and two weeks apart and the order of the test sessions was counterbalanced to minimize learning effects.

Testing was conducted in a sound-attenuated booth. The LiSN-S speech materials were presented over equalized HD215 headphones (Wennebostel, Germany) connected to the computer through a buddy 6G Universal Serial Bus soundcard (Port Colborne, Ontario). The LiSN-S was presented according to the test procedure described in Cameron and Dillon.[17] The target sentences were initially presented at a SNR of +7 dB. The participant was tasked with repeating each target sentence and the

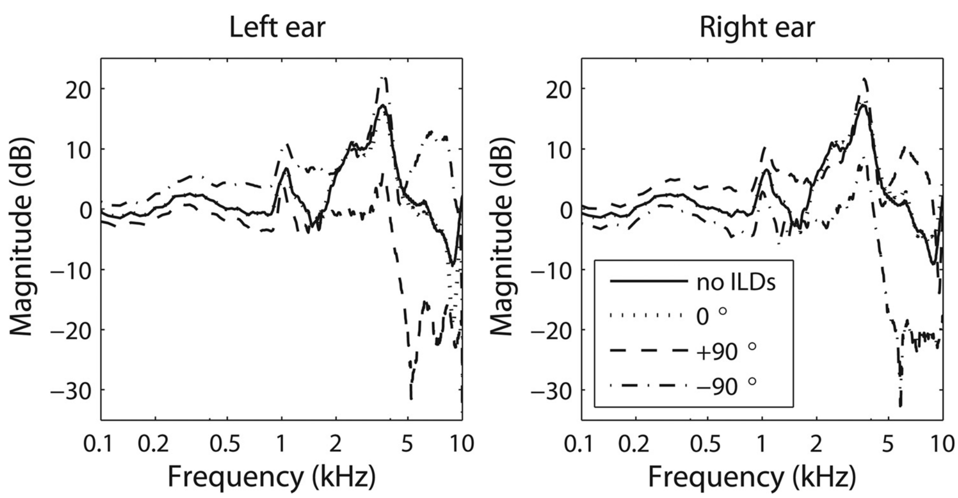

Fig. 1. Magnitude spectra of the different HRTFs used in the three different LiSN-S versions for the left (left panel) and right ear (right panel). The magnitude spectra for the condition "no ILDs" is identical for all spatial conditions and is indicated by the solid line. The magnitude spectra for the original condition and the condition without ITDs are identical and are shown for the different spatial conditions: $0°$ dotted lines, $+90°$ dashed lines, $-90°$ dashed-dotted lines.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 130.102.158.13 On: Thu, 29 Oct 2015 06:08:17

participant's speech reception threshold (SRT) was determined by varying the target level adaptively in 2 dB steps. All words in the target sentences were scored. Testing in each condition continued until at least 17 scored sentences, plus practice, had been completed and the participant's standard error was less than 1 dB, or until 30 sentences had been presented.

3. Results

Statistical analysis was conducted using Statistica (Tulsa, OK) version 10. Figure 2 illustrates the mean SRT and 95% confidence intervals in each of the four LiSN-S conditions for the three cue conditions. For both the different and same voice scenarios, the benefit gained from spatial separation was smallest when only ITD cues were present. There was also smaller benefit, regardless of cue condition, in the different voice conditions than in the same voice conditions.

A three-way analysis of variance, with cue condition, voice, and location included as factors revealed a significant interaction of cue condition and location [$F(2,22) = 49.5$, $p < 0.0001$) indicating that the effect of spatially separating the distracters from the target differs significantly depending on the interaural cues available in the condition. Planned comparisons, with a Bonferroni correction for the number of comparisons made ($\alpha = 0.033$), were conducted to evaluate differences within the interaction. Performance in the ITD only condition was significantly poorer than the reference condition (ITD and ILD) for the same voice scenario [$F(1,11) = 54.6$, $p < 0.0001$] and different voice scenario [$F(1,11) = 7.8$, $p = 0.018$]. The ITD only condition also resulted in significantly poorer SRM than the ILD only condition when the distracting speech and target sentences were voiced by the same female speaker [$F(1,11) = 60.3$, $p < 0.0001$). The difference between the amount of SRM achieved in the ITD only and ILD only conditions for the different voices approached, but did not reach, significance [$F(1,11) = 5.0$, $p = 0.046$]. No significant differences were found between the ILD only condition and the reference condition for either same or different voice [$F(1,11) = 0.1$, $p = 0.938$ and $F(1,11) = 0.7$, $p = 0.432$].

A small but insignificant difference in mean SRT can be observed between the reference condition and the ILDs only condition in the 0° configuration.

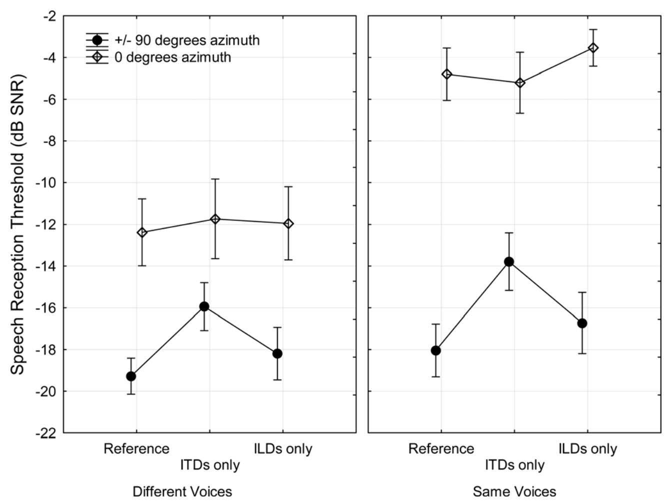

Fig. 2. Speech reception thresholds obtained for each spatial location (+/−90° azimuth and 0° azimuth) as a function of cue condition. The panel on the left shows the results when different female speakers voiced the distracting stories and the panel on the right shows the results when the same female speaker voiced the distracting stories.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 130.102.158.13 On: Thu, 29 Oct 2015 06:08:17

4. Discussion

The aim of this study was to examine whether ITDs alone or ILDs alone can derive SRM when the target speech is frontal and speech maskers are symmetrically located on each side of the head. Consistent with the findings of Kidd *et al.*,[15] the study demonstrated that the presence of either ITDs or ILDs is sufficient to achieve SRM to some extent.

Interestingly, in the current study, the SRM achieved when only ILDs were available was greater than that achieved through the use of ITDs. Unlike in the earlier research, such as that of Bronkhorst and Plomp[12] or Carhart *et al.*,[11] this finding cannot be considered a consequence of a static head shadow effect as the distracters were symmetrically placed around the target affording neither ear a consistently better SNR. It could, however, still be the result of a more dynamic head shadow where the listener attends to the ear with the better SNR at any given moment in time and frequency. It may also have been facilitated by the level differences between the two distractors that head diffraction causes in the signal at each ear considered separately.

No significant difference was found between the ILD-only condition and the reference condition regardless of the distracters used. This lack of significant difference could be interpreted as indicating that, if ILDs are present, the addition of ITDs will not materially change the amount of SRM achieved. This suggests that either an amount of redundancy exists in the information provided by the two types of cues or an absolute performance limit exists within this test paradigm (e.g., Ref. 18), and that the SRM achieved with ILDs takes the participant to this limit preventing ITDs from helping further.

Regardless of the reason behind the larger SRM found with ILDs alone than with ITDs alone, it is important to note the observed ILD dominance is contrary to the results of Kidd *et al.*,[15] who employed the same positioning of maskers, and Culling *et al.*,[14] who employed a similar method of separating ITDs from ILDs. The difference in findings between this research and Kidd *et al.*[15] may be attributable to the different methods used to isolate ITDs and ILDs. Using high pass and low pass filtering isolates the frequency ranges in which each cue is most effective but does not remove all traces of the other cue. In addition, filtering removes components of the speech information, which may confound results.

Determining the reason for the difference in results between the current study and the work of Culling *et al.* will require additional experimentation as numerous methodological dissimilarities exist between the studies. Factors that may have contributed to the discrepancy include the density and positioning of the masking speech materials. When one considers that the SRT in the ILD only condition of the Culling *et al.* experiment is far higher than the ITD only condition for the case where maskers are present on both sides of the head (approximately -0.5 dB compared to -3 dB), it seems likely that the differing results may be due to the density of the masker preventing the use of any dynamic head shadow.

Acknowledgments

The authors acknowledge the financial support of the HEARing CRC, established and supported under the Cooperative Research Centres Program—an initiative of the Australian Government, and the financial support of the Commonwealth Department of Health and Ageing. The authors would also like to thank Nicky Chong-White for her assistance preparing the required speech materials and Virginia Best for her helpful insights and feedback.

[1]S. A. Gelfand, L. Ross, and S. Miller, "Sentence reception in noise from one versus two sources: Effects of aging and hearing loss," J. Acoust. Soc. Am. **83**, 248–256 (1988).
[2]K. Allen, S. Carlile, and D. Alais, "Contributions of talker characteristics and spatial location to auditory streaming," J. Acoust. Soc. Am. **123**, 1562–1570 (2008).
[3]E. C. Cherry, "Some experiments on the recognition of speech with one and two ears," J. Acoust. Soc. Am. **25**, 975–979 (1953).

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 130.102.158.13 On: Thu, 29 Oct 2015 06:08:17

[4]A. S. Bregman, *Auditory Scene Analysis* (MIT Press, Cambridge, MA, 1990).
[5]S. Cameron, H. Glyde, and H. Dillon, "Listening in spatialized noise–sentences test: Normative and retest reliability data for adolescents and adults up to 60 years of age," J. Am. Acad. Audiol. **22**, 697–709 (2011).
[6]H. Glyde, S. Cameron, H. Dillon, L. Hickson, and M. Seeto, "The effects of hearing impairment and aging on spatial processing," Ear. Hear. **34**, 15–28 (2013).
[7]J. R. Dubno, J. B. Ahlstrom, and A. R. Horwitz, "Binaural advantage for younger and older adults with normal hearing," J. Am. Acad. Audiol. **51**, 539–556 (2008).
[8]N. Marrone, C. R. Mason, and G. Kidd, Jr., "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," J. Acoust. Soc. Am. **124**(5), 3064–3075 (2008).
[9]P. L. Divenyi and K. M. Haupt, "Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. III. Factor representation," Ear. Hear. **18**(3), 189–201 (1997).
[10]T. L. Arbogast, C. R. Mason, and G. Kidd, Jr., "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098 (2002).
[11]R. Carhart, T. W. Tillman, and K. R. Johnson, "Release of masking for speech through interaural time delay," J. Acoust. Soc. Am. **42**(1), 124–138 (1967).
[12]A. W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**(4), 1508–1516 (1988).
[13]J. R. Dubno, J. B. Ahlstrom, and A. R. Horwitz, "Spectral contributions to the benefit from spatial separation of speech and noise," J. Speech Lang. Hear. Res. **45**, 1297–1310 (2002).
[14]J. F. Culling, M. L. Hawley, and R. Y. Litovsky, "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**(2), 1057–1065 (2004).
[15]G. Kidd, C. R. Mason, V. Best, and N. Marrone, "Stimulus factors influencing spatial release from speech-on-speech masking," J. Acoust. Soc. Am. **128**(4), 1965–1978 (2010).
[16]A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," J. Acoust. Soc. Am. **105**(5), 2821–2840 (1999).
[17]S. Cameron and H. Dillon, "Development of the Listening in Spatialized Noise-Sentences Test (LISN-S)," Ear. Hear. **28**, 196–211 (2007).
[18]V. Best, E. R. Thompson, C. R. Mason, and G. Kidd, "An energetic limit on spatial release from masking," J. Assoc. Res. Otolaryngol. In Press.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 130.102.158.13 On: Thu, 29 Oct 2015 06:08:17