The influence of non-spatial factors on measures of spatial release from masking

Virginia Best^{a)}

Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts 02215

Nicole Marrone

Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

Christine R. Mason and Gerald Kidd, Jr.

Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts 02215

(Received 29 April 2011; revised 17 February 2012; accepted 19 February 2012)

This study tested the hypothesis that the reduction in spatial release from masking (SRM) resulting from sensorineural hearing loss in competing speech mixtures is influenced by the characteristics of the interfering speech. A frontal speech target was presented simultaneously with two intelligible or two time-reversed (unintelligible) speech maskers that were either colocated with the target or were symmetrically separated from the target in the horizontal plane. The difference in SRM between listeners with hearing impairment and listeners with normal hearing was substantially larger for the forward maskers (deficit of 5.8 dB) than for the reversed maskers (deficit of 1.6 dB). This was driven by the fact that all listeners, regardless of hearing abilities, performed similarly (and poorly) in the colocated condition with intelligible maskers. The same conditions were then tested in listeners with normal hearing using headphone stimuli that were degraded by noise vocoding. Reducing the number of available spectral channels systematically reduced the measured SRM, and again, more so for forward (reduction of 3.8 dB) than for reversed speech maskers (reduction of 1.8 dB). The results suggest that non-spatial factors can strongly influence both the magnitude of SRM and the apparent deficit in SRM for listeners with impaired hearing.

© 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.3693656]

PACS number(s): 43.71.Ky, 43.66.Dc, 43.66.Pn [MAA]

Pages: 3103–3110

I. INTRODUCTION

The intelligibility of speech under masked conditions is improved when the masking sounds are spatially separated from the target sound relative to the case where all of the sounds arise from the same location. This improvement stems in part from a simple acoustical factor, in which the signal-to-noise ratio (SNR) at one ear is improved by the acoustic filtering of the head for asymmetric configurations (Zurek, 1993; Edmonds and Culling, 2006). In addition, spatially separated sounds produce different interaural time differences (ITDs), which provide further effective improvements in SNR within each auditory filter (Zurek, 1993; Edmonds and Culling, 2005). Beyond these effects, spatial separation of sounds may also produce important perceptual cues that can lead to large advantages. For example, spatial separation reduces uncertainty about how to sort the acoustic mixture, particularly when there are few other distinguishing features among competing sounds. The magnitude of the improvement seen in the latter case may often be much larger than that related to simple increases in SNR. For example, Marrone *et al.* (2008a) presented listeners with a mixture of three similar, simultaneous talkers, and measured the benefit of separating the maskers symmetrically about the target in the horizontal plane. Despite the fact that

long-term head shadow advantages were largely negated in this configuration, they found a very large spatial release from masking (SRM) with a group mean improvement in threshold of ~12 dB.

One deficit often demonstrated by listeners with sensorineural hearing impairment (HI) when compared to listeners with normal hearing (NH) is reduced SRM. For steady-state noise maskers the difference in SRM between NH and HI groups is relatively small (on the order of a few decibels) and can, to a large extent, be explained by the loss of a useful head-shadow advantage due to reduced high-frequency audibility in HI listeners (Duquesnoy, 1983; Bronkhorst and Plomp, 1989; Bronkhorst and Plomp, 1992; Dubno *et al.*, 2002). However, for speech maskers, the disparity between groups can be very large and does not appear to be related solely to a loss of head-shadow advantages (Arbogast *et al.*, 2005; Marrone *et al.*, 2008b; Neher *et al.*, 2009). For example, Marrone *et al.* (2008b) used the same three-talker mixture described previously and found a difference in SRM between NH and HI groups that was substantially larger than that reported for noise maskers (about 9 dB on average).

These past results suggest that the difference in SRM between NH and HI groups may depend on the amount of “informational masking” (IM) produced by the stimulus. Specifically, when the maskers are highly similar to and/or confusable with the target, SRM tends to be large in NH listeners, and the difference between NH and HI listener groups also tends to be large. However, the comparison between the

^{a)}Author to whom correspondence should be addressed. Electronic mail: ginbest@bu.edu

findings using noise and speech maskers is complicated by the fact that noise maskers have more acoustic overlap (i.e., overlap in time and/or frequency) with the target than do speech maskers and, thus, they generate more “energetic masking” (EM; for a recent review of EM and IM see Kidd *et al.*, 2008). This is particularly important when comparing NH and HI groups, because HI listeners appear to be more susceptible to EM than NH listeners but equally susceptible to IM (Helfer and Freyman, 2008; Agus *et al.*, 2009; Best *et al.*, 2011).

Experiment 1 aimed to examine the effect of the amount of IM on SRM in NH and HI listeners without the confound of changes in EM. The experiment was an extension of the work of Marrone and colleagues (2008a,b) and the relevant data from those studies are included here. SRM was measured for a speech target in the presence of two symmetrically placed speech maskers that were either highly intelligible and confusable with the target (high in IM) or time-reversed and unintelligible, and thus less confusable with the target (lower in IM). The comparison of these two kinds of maskers has the advantage that they have very similar spectrotemporal structures and thus generate approximately equivalent amounts of EM (e.g., Dirks and Bower, 1969; Freyman *et al.*, 2001; Brungart and Simpson, 2002; Summers and Molis, 2004; Rhebergen *et al.*, 2005; Brungart and Simpson, 2007; Hornsby and Ricketts, 2007; Kidd *et al.*, 2010).

Experiment 2 was conducted to examine whether the interaction between IM and hearing status could be replicated in NH listeners when the signal quality was varied. The speech stimuli were systematically degraded using noise vocoding. Vocoding adversely affects speech understanding and interferes with the segregation of competing voices by reducing spectral resolution and disrupting temporal fine structure cues (Shannon *et al.*, 1995; Qin and Oxenham, 2003). It thus provides a useful qualitative parallel in NH listeners to some changes in the stimulus produced by hearing impairment (Moore, 2007). However, it was *not* intended to be a plausible simulation of the underlying physiological mechanisms causing hearing loss. The basic idea was to reduce overall performance in NH listeners (to be more like HI listeners) and examine the interaction of this impairment with the intelligibility of the maskers in terms of the effect on SRM. Importantly, under the conditions tested in experiment 2, vocoding was implemented prior to the introduction of the spatial cues used to separate the sound sources. This meant that the spatial cues that normally provide the basis for the perception of different speech source locations were available in both the fine structure of the noise carriers and the envelopes, despite the degradation in speech quality. If vocoding were done post-spatialization, the fine structure ITDs would be lost and reductions in localizability and SRM would be expected (Drennan *et al.*, 2007). The issue of order-of-processing (vocoding vs spatialization) was examined explicitly in a recent study by Garadat and colleagues (2009). Somewhat surprisingly, they found no difference in SRM due to the order of processing; i.e., SRM was the same regardless of whether the stimuli were vocoded before or after spatialization. Drennan *et al.* (2007) also reported that the loss of fine structure ITDs had less of an impact on the SRM

for speech presented against speech maskers than on speech presented against noise maskers. These results suggest that envelope cues may be sufficient to produce spatial benefits with competing speech sounds.¹ In the present study, preserving the spatial differences in the stimuli and using NH listeners with presumably normal spatial hearing capabilities allowed us to assess the influence of signal quality (and overall intelligibility) on SRM in the absence of any specific deficits in spatial processing.

II. EXPERIMENT 1: LISTENERS WITH HEARING LOSS

A. Participants

Participants in experiment 1 were six young HI listeners (25–43 years, mean 31 years) and six young NH listeners (20–43 years, mean 31 years). The HI group had bilateral, symmetric, moderate-to-severe sensorineural hearing losses. The group mean audiogram is shown in Fig. 1. Each listener in the NH group had audiometric thresholds of 20 dB or better at octave frequencies from 250 Hz to 8 kHz. Data for five of the six NH listeners in the *forward* and *reversed* conditions have been published previously (BARE room, Marrone *et al.*, 2008a). Data for the six HI listeners in the *forward* condition only have also appeared elsewhere (YHL group/BARE room, Marrone *et al.*, 2008b). One of the HI listeners and three of the NH listeners had participated in previous experiments using similar speech materials.

B. Methods

Listeners were seated in a large single-walled sound-attenuating booth (Industrial Acoustics Corporation). The booth had perforated metal walls and ceiling with carpet on the floor and produced little reverberation (Marrone *et al.*, 2008a). Stimuli were presented over loudspeakers located at a distance of 1.5 m from and level with the listener’s head position. Digital stimuli were generated on a PC outside the booth and then fed through separate channels of Tucker-Davis Technologies System II hardware. This included 16-

FIG. 1. Mean audiograms for the better and worse ears in the HI group. Error bars show across-listener standard deviations.

bit digital-to-analog conversion at 40 kHz (DA8), low-pass filtering at 20 kHz (FT6), and programmable attenuation (PA4). Stimuli were amplified (Tascam) before being routed to the loudspeakers (Acoustic Research 215PS).

On each trial, the stimulus was composed of three sentences from the Coordinate Response Measure corpus (Bolia *et al.*, 2000), presented simultaneously. The sentences all had the format “Ready <callsign> go to <color> <number> now.” The three sentences were spoken by three different female talkers selected at random on each trial from the four available talkers. One sentence was designated as the target and always contained the call-sign “Baron” with the color and number keywords being randomly selected from the choices available. The masker sentences contained randomly selected call-signs and keywords that differed from the target and from each other. Subjects were instructed to identify the keywords that were uttered by the target talker. Responses were registered by pressing one colored button and one numbered button on a hand-held keypad. The statement “correct” or “incorrect” and the target keywords were provided on the display of the keypad by way of feedback. Responses were counted as correct only if both color and number were correct. The three sentences were presented from a single loudspeaker directly in front of the listener (*colocated* configuration) or the target was presented from the frontal loudspeaker and the masker sentences were presented symmetrically at 90° to either side of the midline (*separated* configuration). The masker sentences were either played naturally (*forward* condition) or were time-reversed to render them unintelligible (*reversed* condition).

For NH listeners, the target level was fixed at 60 dB SPL² and the level of the maskers was adapted to vary the target-to-masker ratio (TMR). For HI listeners, a presentation level of 30 dB sensation level (SL) was used where possible (i.e., 30 dB above an individual’s quiet identification threshold). In cases where 30 dB SL was too high for the range of masker levels to be accommodated within the limits of the playback system, a lower SL was used. For all listeners at least 10 dB SL was achieved. Thresholds were estimated using a one-down one-up procedure that tracked the 50% correct point on the psychometric function. The two maskers were equal in level to each other, and this level was varied adaptively in 4 dB steps initially and then in 2 dB steps following the fourth reversal. Each run consisted of at least 30 trials and at least nine reversals, and thresholds were calculated by averaging the TMR at an even number of reversal points after the first three (or four). A session consisted of one run of each of the different conditions presented a random order, and each listener completed four such sessions. The mean thresholds presented in the following are based on the arithmetic mean across sessions.

C. Results

In the *forward* condition [Fig. 2(A)], mean masked speech reception thresholds for the NH and HI groups were similar in the *colocated* configuration (NH: 2.8 dB; HI: 4.6 dB), but NH listeners achieved much lower thresholds

FIG. 2. (A) Mean TMR at 50% threshold for NH (black) and HI (white) listeners in the *forward* condition of experiment 1. (B) Mean TMR at 50% threshold in the *reversed* condition of experiment 1. (C) Mean spatial release from masking (*colocated* – *separated* thresholds) in experiment 1. Error bars in all panels show across-subject standard deviations.

in the *separated* configuration (NH: – 8.6 dB; HI: – 1.0 dB). A two-way mixed-design analysis of variance (ANOVA) on the *forward* thresholds revealed a significant effect of configuration [$F(1,10) = 148.2$, $p < 0.001$], group [$F(1,10) = 22.0$, $p = 0.001$], and a significant interaction [$F(1,10) = 17.5$, $p = 0.002$]. In the *reversed* condition [Fig. 2(B)], NH thresholds were lower than HI thresholds in both the *colocated* (NH: – 7.4 dB; HI: 1.7 dB) and *separated* (NH: – 12.3 dB; HI: – 1.6 dB) configurations. A two-way mixed-design ANOVA on the *reversed* thresholds revealed a significant effect of configuration [$F(1,10) = 47.2$, $p < 0.001$], group [$F(1,10) = 56.8$, $p < 0.001$], and no significant interaction

[$F(1,10) = 1.8$, $p > 0.05$]. The simple subtraction of the thresholds in the two configurations indicates a larger group difference in SRM for *forward* than *reversed* maskers [Fig. 2(C)]. Bonferroni-corrected t -tests confirmed that the difference in SRM between groups was significant for *forward* [$t(10) = 4.2$, $p = 0.002$], but not *reversed* [$t(10) = 1.3$, $p > 0.05$] maskers.

To examine patterns of individual performance across different conditions, Fig. 3 shows scatter plots of thresholds for *forward* and *reversed* maskers (NH: black symbols, HI: white symbols). In the *colocated* configuration [Fig. 3(A)], although the thresholds are correlated [$r(10) = 0.74$, $p = 0.006$], the pool of listeners shows a much larger range of performance for *reversed* maskers than for *forward* maskers (particularly the NH listeners). In fact, it can be seen that all of the thresholds are positive for the *forward* maskers (to the right of vertical dotted line), whereas the bet-

ter (NH) listeners were able to achieve negative thresholds with *reversed* maskers.

In the *separated* configuration [Fig. 3(B)], the thresholds across the two masker types are highly correlated [$r(10) = 0.93$, $p < 0.001$] and the variation across listeners is similar for the two masker types with the better listeners achieving negative threshold values for both types of masker. Consistent with a greater equivalence in performance across masker types in the *separated* configuration, the slope of the least-squares fit is close to unity (1.2 vs 2.9 in the *colocated* configuration).

The results of experiment 1 support the idea that the amount of IM present can affect the SRM deficit measured in listeners with hearing loss. Experiment 2 was conducted to examine whether a similar interaction is observed in NH listeners presented with systematically degraded signals.

III. EXPERIMENT 2: NOISE VOCODING IN LISTENERS WITH NORMAL HEARING

A. Participants

Participants in experiment 2 were six young adult NH listeners (18–22 years, mean 19 years). All listeners had audiometric thresholds of 20 dB or better at octave frequencies from 250 Hz to 8 kHz. None of the listeners had previous experience in experiments of this kind.

B. Methods

Stimuli in experiment 2 were delivered via headphones³ (Sennheiser HD 580) to listeners seated in a double-walled audiometric booth (Industrial Acoustics Company). The sentence materials and the stimulus delivery system, other than the transducers and amplifiers, were identical to those used in experiment 1. The target was presented diotically and the maskers were presented either diotically (*colocated* configuration) or were lateralized to opposite sides by introducing a whole-waveform interaural delay of $\pm 600 \mu\text{s}$ (*separated* configuration). As in experiment 1, the masker sentences were either played naturally (*forward* condition) or were time-reversed to render them unintelligible (*reversed* condition). The booth was fitted with a monitor and mouse, and subjects responded by clicking on 1 of 32 colored and numbered buttons on a custom graphical user interface. No feedback was provided.

Before the target and masker sentences were combined, each sentence was processed using noise vocoding (cf. [Shannon et al., 1995](#)). Sixth-order Butterworth filters were used to filter the sentences into 8, 16, or 32 contiguous log-spaced bands in the range 80–8000 Hz. Hilbert envelopes were extracted from each band, low-pass filtered at 500 Hz, and used to modulate narrow bands of noise with the same bandwidth as the analysis bands. Independent white noise tokens were used for the target and masker sentences. Vocoding was done prior to the introduction of ITDs to the masker sentences, so that the binaural cues (ITDs imposed on the fine structure and envelope) were preserved in the vocoded stimulus. An unprocessed condition was also included, giving four processing conditions in total.

FIG. 3. Scatter plots showing thresholds for *forward* vs *reversed* maskers in the (A) *colocated* and (B) *separated* configurations of experiment 1. Black and white symbols show NH and HI listeners, respectively. Solid lines show least-squares fits to the data points.

On each trial, the masker level was fixed at 60 dB SPL⁴ and the target level was varied randomly to produce one of seven TMRs equally spaced between -15 and $+15$ dB. The different TMRs were tested four times each to give a block of 28 trials. The 16 experimental conditions (two spatial configurations, two masker conditions, and four processing conditions) were tested in separate blocks. These 16 blocks were tested in a random order within one experimental session. Eight sessions were completed by each listener, for a total of 32 trials per unique condition/TMR. Psychometric functions were generated from the mean percent correct scores for each listener, and thresholds were extracted from logistic fits to these functions at levels corresponding to 50% correct. At least one full session, covering all of the conditions, was completed by each listener by way of training prior to the eight sessions that were scored and included in the following analyses.

C. Results

Figure 4(A) shows thresholds in the *forward* condition for each processing condition. For the *colocated* configuration, thresholds were clustered around 5 dB, with little effect seen for the different degrees of processing (total range of 1.4 dB). Thresholds diverged, however, in the *separated* configuration (total range of 5.2 dB). A two-way repeated-measures ANOVA on the *forward* thresholds found a significant effect of configuration [$F(1,5) = 20.5$, $p = 0.006$], processing condition [$F(3,15) = 12.3$, $p < 0.001$], and a significant interaction [$F(3,15) = 10.5$, $p = 0.001$]. In the *reversed* condition [Fig. 4(B)], an effect of processing condition was evident for both spatial configurations, with ranges of 7.1 and 8.9 dB in the *colocated* and *separated* configurations, respectively. A two-way repeated-measures ANOVA on the *reversed* thresholds revealed a significant effect of configuration [$F(1,5) = 48.6$, $p = 0.001$], processing condition [$F(3,15) = 101.7$, $p < 0.001$], but no significant interaction [$F(3,15) = 2.0$, $p > 0.05$].

SRM for the different conditions is shown in Fig. 4(C). Vocoding led to a systematic reduction in SRM in all cases, but this decline was more marked in the *forward* condition than in the *reversed* condition (decrease from unprocessed to eight-band vocoding of 3.8 dB in *forward* as compared to 1.8 dB in *reversed*). One-way repeated-measures ANOVAs confirmed that the effect of processing condition was significant for *forward* [$F(3,15) = 10.2$, $p = 0.001$], but not *reversed* [$F(3,15) = 1.9$, $p > 0.05$] maskers. As in experiment 1, this is largely a consequence of the limited variation in thresholds for the *forward colocated* condition.

To examine individual differences in performance across the processing conditions, Fig. 5 shows scatter plots of thresholds for *forward* and *reversed* maskers. Each listener is indicated by a different symbol, with the color of the symbol distinguishing the different processing conditions (from unprocessed, black, to eight-band vocoding, white). In the *colocated* configuration [Fig. 5(A)], although the points were strongly correlated [$r(22) = 0.87$, $p < 0.001$], the different listeners under the different processing conditions showed a much larger range of thresholds for *reversed*

FIG. 4. (A) Mean TMR at 50% threshold in the *forward* condition of experiment 2 for the different processing conditions (black, unprocessed; dark gray, 32 bands; light gray, 16 bands; white, 8 bands). (B) Mean TMR at 50% threshold in the *reversed* condition of experiment 2. (C) Mean spatial release from masking in experiment 2. Error bars in all panels show across-subject standard deviations.

maskers than for *forward* masker. Similar to what was observed in experiment 1, the lack of variation in the *forward* condition reflected a clustering of thresholds at positive TMRs (to the right of the vertical dotted line). In the *separated* configuration [Fig. 5(B)], the thresholds were again highly correlated across masker types [$r(22) = 0.89$, $p < 0.001$]. In this configuration, no clustering of the thresholds at positive TMRs was evident, and the variation across listeners and processing conditions was similar for the two masker types. The greater correspondence between scores for the two masker types in the *separated* configuration is further supported by the fact that the least-squares fit to the

FIG. 5. Scatter plots showing thresholds for *forward* vs *reversed* maskers in the (A) *colocated* and (B) *separated* configurations of experiment 2. Different colors indicate the different processing conditions, from unprocessed (black) to eight-band vocoded (white), and different symbol types represent different individuals. Solid lines show least-squares fits to the data points.

data has a slope closer to unity (1.0 vs 1.9 in the *colocated* configuration).

IV. DISCUSSION

In this pair of experiments examining SRM in competing speech mixtures, we found that differences in SRM between listeners with normal vs impaired hearing are determined in part by the characteristics of the competing speech. This interaction was replicated in NH listeners given systematically degraded speech signals, suggesting that variations in SRM do not necessarily imply variations in spatial abilities.

In experiment 1, SRM for *forward* maskers was found to be significantly smaller for HI listeners than NH listeners, as reported previously (Duquesnoy, 1983; Bronkhorst and

Plomp, 1989; Bronkhorst and Plomp, 1992; Marrone *et al.*, 2008b). However, SRM for *reversed* maskers was comparable across listener groups. In other words, the measured group difference in SRM depended critically on the characteristics of the stimulus. The large difference in SRM for the *forward* maskers appears to be related to the fact that all listeners had particularly high thresholds in the *colocated* configuration, being clustered at a positive TMR. It appears that the difficult baseline condition forces all listeners into a region where the target is the loudest sound in the mixture, and in this region NH and HI groups perform similarly. Spatial separation takes listeners out of this region, presumably by resolving confusion about which sounds belong to which talker, i.e., by helping with the “selection” problem (cf. Shinn-Cunningham, 2008). This enables listeners to achieve lower thresholds, which are influenced by their individual hearing status and listening abilities. In the *forward* condition of the current study, this lead to particularly large SRM values for NH listeners (11.4 dB on average), who were able to achieve very low thresholds in the *separated* configuration. HI listeners, who were unable to perform as well in the *separated* configuration, showed less SRM (5.6 dB on average). With *reversed* maskers, this clustering of *colocated* thresholds did not occur, and the SRM was smaller and more comparable across listener groups (4.9 and 3.3 dB).

These results demonstrate that variations in baseline performance for tasks varying in IM can have a strong influence on the measured SRM. This is an idea that has recently been discussed elsewhere (Freyman *et al.*, 2008; Kidd *et al.*, 2010) and suggests that there are important non-spatial factors contributing to “spatial” benefits. Another important implication of the present results is that the amount of IM in a speech intelligibility task can affect the *difference* in SRM observed between NH and HI listener groups. Specifically, the difference in SRM between NH and HI groups in experiment 1 was 5.8 dB when the maskers were naturally spoken and 1.6 dB when the maskers were time-reversed. Thus conclusions regarding “spatial deficits” in HI listeners, and comparisons of such deficits across studies, must be made with caution and with consideration of the type of masking that is present.

The ideas raised in experiment 1 were elaborated in experiment 2 using listeners with normal hearing sensitivity and presumably also with normal spatial hearing capabilities. By systematically degrading the speech stimuli, variations in SRM were obtained within NH listeners that paralleled the variations seen across NH and HI groups in experiment 1. SRM was reduced with vocoding in both masker conditions, but the reduction was greatest for the *forward* maskers. This highlights again that the amount of IM in a task can influence the SRM measured and the interaction with other parameters, such as hearing status or signal quality. Garadat and colleagues (2009) also came to the conclusion that the magnitude of SRM depends on the degree of IM present. However, their results are in the opposite direction to those reported here—they found that progressively coarser vocoding *increased* SRM in their task, which involved the identification of spondaic words against sentence maskers. This discrepancy between the

two studies may reflect the fact that the targets used by Garadat *et al.* (2009) were not confusable with the maskers, and thus task performance was not limited by IM. Consistent with this speculation, their listeners achieved extremely low thresholds even in the colocated configuration, with the group mean thresholds reported to be lower than -20 dB TMR for unprocessed stimuli.

Despite the issues discussed previously regarding the *colocated* reference point, a consistent finding in the current study was that thresholds in the *separated* configurations, where IM is reduced and listeners are presumably approaching their best performance, were higher in HI listeners than in NH listeners. This deficit is likely a consequence of the many factors that have been discussed in previous studies, such as reduced audibility, reduced spectral resolution, and/or a loss of sensitivity to temporal fine structure (Moore, 2007).

V. CONCLUSION

SRM was reduced by hearing impairment (and by the disruption of speech via noise vocoding in NH listeners), but the size of this reduction was dependent on the characteristics of the speech maskers. Closer examination of thresholds suggested that the particular difficulty of segregating three colocated talkers lead to inflated SRM values in NH listeners, and thus larger apparent deficits in the HI group (or the NH group given degraded signals). Reversing the maskers alleviated this issue to some extent. These results suggest that the specific characteristics of competing signals can strongly influence SRM, and while impaired spatial processing may be a factor in the reduced SRM observed in HI listeners it need not be the primary explanation.

ACKNOWLEDGMENTS

Work supported by NIDCD (DC004545, DC00100 and DC004663) and AFOSR (FA9550-08-1-0424). V.B. was supported in part by a University of Sydney Postdoctoral Research Fellowship.

¹Informal listening suggested to us that the perception of spatial separation of the simultaneous vocoded talkers was completely lost when spatialization was done prior to vocoding. Thus, we strongly doubt that the measured SRM would be similar, as observed by Garadat *et al.* (2009) and Drennan *et al.* (2007) under their experimental conditions.

²The level of the speech from each loudspeaker was calibrated with reference to a microphone suspended at the position of the listener's head with the listener absent.

³The reason for choosing headphones for Experiment 2 was to simplify the interpretation of the results by avoiding head-shadow induced changes in SNR at each ear.

⁴Speech levels for headphone testing were calibrated using an acoustic coupler and a sound level meter.

Agus, T. R., Akeroyd, M. A., Gatehouse, S., and 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.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (2005). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **117**, 2169–2180.

Best, V., Mason, C. R., and Kidd, G., Jr. (2011). "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing talkers," *J. Acoust. Soc. Am.* **129**, 1616–1625.

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**, 1065–1066.

Bronkhorst, A. W., and Plomp, R. (1989). "Binaural speech intelligibility in noise for hearing-impaired listeners," *J. Acoust. Soc. Am.* **86**, 1374–1383.

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

Brungart, D. S., and Simpson, B. D. (2002). "Within-ear and across-ear interference in a cocktail-party listening task," *J. Acoust. Soc. Am.* **112**, 2958–2995.

Brungart, D. S., and Simpson, B. D. (2007). "Effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task," *J. Acoust. Soc. Am.* **122**, 1724–1734.

Dirks, D. D., and Bower, D. R. (1969). "Masking effects of speech competing messages," *J. Speech Hear. Res.* **12**, 229–245.

Drennan, W. R., Won, J. H., Dasika, V. K., and Rubinstein, J. T. (2007). "Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise," *J. Assoc. Res. Otolaryng.* **8**, 373–378.

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (2002). "Spectral contributions to the benefit from spatial separation of speech and noise," *J. Speech Lang. Hear. Res.* **45**, 1297–1310.

Duquesnoy, A. J. (1983). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," *J. Acoust. Soc. Am.* **74**, 739–743.

Edmonds, B. A., and Culling, J. F. (2005). "The spatial unmasking of speech: evidence for within-channel processing of interaural time delay," *J. Acoust. Soc. Am.* **117**, 3069–3078.

Edmonds, B. A., and Culling, J. F. (2006). "The spatial unmasking of speech: evidence for better-ear listening," *J. Acoust. Soc. Am.* **120**, 1539–1545.

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

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2008). "Spatial release from masking with noise-vocoded speech," *J. Acoust. Soc. Am.* **124**, 1627–1637.

Garadat, S. N., Litovsky, R. Y., Yu, G., and Zeng, F. G. (2009). "Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech," *J. Acoust. Soc. Am.* **126**, 2522–2535.

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

Hornsby, B. W. Y., and Ricketts, T. A. (2007). "Directional benefit in the presence of speech and speechlike maskers," *J. Am. Acad. Audiol.* **18**, 5–16.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay, Springer Handbook of Auditory Research (Springer, New York), pp. 143–190.

Kidd, G., Jr., Mason, C. R., Best, V., and Marrone, N. (2010). "Stimulus factors influencing spatial release from speech-on-speech masking," *J. Acoust. Soc. Am.* **128**, 1965–1978.

Marrone, N., Mason, C. R., and Kidd, G., Jr. (2008a). "Tuning in the spatial dimension: evidence from a masked speech identification task," *J. Acoust. Soc. Am.* **124**, 1146–1158.

Marrone, N., Mason, C. R., and Kidd, G., Jr. (2008b). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," *J. Acoust. Soc. Am.* **124**, 3064–3075.

Moore, B. C. J. (2007). *Cochlear Hearing Loss: Psychological, Psychological, and Technical Issues* (Wiley, Chichester, UK).

Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., and van Schaik, A. (2009). "Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: Effects of hearing loss, age, and cognition," *Int. J. Audiol.* **48**, 758–774.

Qin, M. K., and Oxenham, A. J. (2003). "Effects of simulated cochlear implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am.* **114**, 446–454.

- Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2005). "Release from informational masking by time reversal of native and non-native interfering speech," *J. Acoust. Soc. Am.* **118**, 1274–1277.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Shinn-Cunningham, B. G. (2008). "Object-based auditory and visual attention," *Trends Cogn. Sci.* **12**, 182–186.
- Summers, V., and Molis, M. R. (2004). "Speech recognition in fluctuating and continuous maskers: effects of hearing loss and presentation level," *J. Speech Lang. Hear. Res.* **47**, 245–256.
- Zurek, P. M. (1993). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.