Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing talkers[a)]

Virginia Best,[b)] Christine R. Mason, and Gerald Kidd, Jr.

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

Listeners with sensorineural hearing loss are poorer than listeners with normal hearing at understanding one talker in the presence of another. This deficit is more pronounced when competing talkers are spatially separated, implying a reduced "spatial benefit" in hearing-impaired listeners. This study tested the hypothesis that this deficit is due to increased masking specifically during the simultaneous portions of competing speech signals. Monosyllabic words were compressed to a uniform duration and concatenated to create target and masker sentences with three levels of temporal overlap: 0% (non-overlapping in time), 50% (partially overlapping), or 100% (completely overlapping). Listeners with hearing loss performed particularly poorly in the 100% overlap condition, consistent with the idea that simultaneous speech sounds are most problematic for these listeners. However, spatial release from masking was reduced in *all* overlap conditions, suggesting that increased masking during periods of temporal overlap is only one factor limiting spatial unmasking in hearing-impaired listeners. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3533733]

I. INTRODUCTION

Understanding what someone is saying in the presence of a competing talker can be difficult as a result of at least two distinct kinds of masking. Energetic masking (EM) describes the reduction in the audibility of a target signal that occurs when the target and masking signals compete for representation in the same peripheral neural units. Peripheral overlap occurs for signals that overlap in time and frequency, as well as for those that are proximal in time or frequency (e.g., in the case of forward masking and upward spread of masking). Informational masking (IM, also called perceptual or central masking; cf. Kidd *et al.*, 2008b) describes the interference that cannot be explained easily by interactions in the auditory periphery. It is generally associated with confusion between the target and masker(s) or an inability to focus attention on the target. In the case of speech mixtures, the extent of each kind of masking can vary. The competing signals are typically broadband but vary in the degree to which they overlap in time and frequency and thus in the amount of EM that is present. Furthermore, a variety of more complex factors may influence the amount of IM, including the similarity of competing voices (Brungart, 2001), and uncertainty about where and when to direct attention (Kidd *et al.*, 2005; Best *et al.*, 2007; Brungart and Simpson, 2007; Kitterick *et al.*, 2010).

The spatial separation of competing speech sources is known to improve the intelligibility of a target talker. While this "spatial release from masking" (SRM) likely reflects a release from both EM and IM, it is often difficult to determine the contributions of the two kinds of masking to the SRM observed in a specific listening situation. One estimate of the amount of SRM that is attributable to a reduction in EM comes from studies that have measured SRM for a speech target presented against a speech-shaped (and often speech-envelope-modulated) noise. The SRM observed in this stimulus configuration results from improvements in the target-to-masker ratio at one ear caused by the acoustic shadowing of the head as well as by improvements in audibility due to binaural processing of differences in interaural delays. The SRM afforded by these cues is typically on the order of 5–10 dB (Zurek, 1993; Yost, 1997; Bronkhorst, 2000).

Stimulus configurations that are dominated by IM (e.g., two or three simultaneous, similar talkers) may produce an SRM that is considerably larger than that found in corresponding conditions for EM (e.g., Allen *et al.*, 2008; Marrone *et al.*, 2008b). Arbogast *et al.* (2002) created an experimental stimulus in which two competing speech signals contained minimal spectral overlap and hence minimal EM. They did this by filtering the target and masker sentences into narrow frequency bands and then presenting mutually exclusive sets of bands for each. This listening situation was thought to comprise predominantly IM, and thus the SRM observed when the sources were spatially separated was taken to be primarily a release from IM (around 18 dB in that study). Freyman *et al.* (1999, 2001) also found a large release from IM (6–10 dB) using a stimulus paradigm in which the perception of spatial separation between a target talker and a masker talker was induced using the precedence effect. This release occurred despite a small *increase* in EM resulting from the addition of a delayed masker. When a noise masker was used instead of the speech masker in the same paradigm, the resulting SRM was negligible. These

---

findings support the idea that SRM is related to the EM/IM content of the listening situation (see also Kidd *et al.*, 1998 for a similar finding in a nonspeech task).

It is well known that listeners with sensorineural hearing impairment (HI) have trouble understanding one talker in the presence of noise or competing talkers (Carhart and Tillman, 1970; Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Bronkhorst, 2000). For a talker presented against broadband noise, hearing loss (HL) raises speech reception thresholds relative to those seen in normally hearing (NH) listeners. This appears to be both due to the reduced audibility and the reduced frequency resolution that accompanies sensorineural HL (Moore, 1985, 2007). In temporally fluctuating maskers, recent evidence has suggested that the deficit may also be related to an inability to use temporal fine structure cues (Lorenzi *et al.*, 2006; Hopkins *et al.*, 2008). These effects are consistent with increased EM in HI listeners, which may explain part of the difficulty in competing talker situations. Whether HL also influences IM, however, is unclear.

Only a handful of studies have attempted to explicitly separate the effects of HL on EM and IM in the context of speech segregation. Arbogast *et al.* (2005) measured masked speech reception thresholds in NH and HI listeners but used spectrally interleaved target and masker bands in an attempt to reduce EM (as described above). HI listeners showed poorer thresholds than NH listeners for a co-located target and masker, with a deficit of 13 dB in noise and 4 dB in competing speech. They suggested that spectral smearing of the acoustically narrow speech and noise bands had the effect of producing more EM in the HI listeners. In a recent study, Agus *et al.* (2009) used a similar approach in conjunction with a "speech-in-speech-in-noise" design to reduce EM and isolate IM in younger listeners and older listeners with various degrees of HL. They reported similar estimates of IM in younger and older listeners, suggesting that the critical difference between these groups for speech tasks is the amount of EM. Helfer and Freyman (2008) came to a similar conclusion based on the finding that young listeners and older listeners with HL showed the same drop in performance when a low-IM masker (noise) was replaced with a relatively high-IM masker (speech).

For a variety of tasks, HI listeners show reduced SRM compared to listeners with normal hearing (e.g., Duquesnoy, 1983; Bronkhorst and Plomp, 1992; Marrone *et al.*, 2008a). One factor that may limit SRM in HI listeners is a reduced ability to use high-frequency head shadow cues because of elevated high-frequency thresholds (Dubno *et al.*, 2002). However, there may be other factors that contribute as well. For example, as described above, situations that are dominated by EM tend to exhibit less SRM than situations dominated by IM. Given the evidence that HI listeners are more susceptible to EM than NH listeners, it is plausible that this is a factor in their reduced SRM.

This experiment tested two related hypotheses. First, we hypothesized that it is the *simultaneous* portions of competing speech waveforms that give rise to the deficits seen in HI listeners, probably as a result of increased EM. Second, we hypothesized that the reduced spatial release from speech-on-speech masking seen in HI listeners is due to this increased EM during the simultaneous portions of the speech. These hypotheses were tested using an approach in which the temporal overlap of a pair of competing sentences was systematically varied. While previous studies have removed temporal overlap by presenting entire messages sequentially (e.g., Webster and Thompson, 1954), we varied overlap on a word by word basis such that the competing messages were still temporally coincident at the sentence level. To enable precise control over the amount of overlap, all words were compressed to the same duration and the gaps between the target and masker words were held constant.[1] If these hypotheses proved true, then it follows that both overall performance and SRM should be compromised the most in HI listeners relative to NH listeners in the case of complete temporal overlap, where the potential for EM is maximized. Moreover, we expected that performance and SRM would not be compromised at all in the case of zero overlap, because EM should not be a significant factor.

The current experiments were conducted in young listeners to avoid the difficulties associated with disentangling the effects of HL from possible effects of aging as addressed in many of the studies mentioned above (e.g., Helfer and Freyman, 2008; Marrone *et al.*, 2008a; Agus *et al.*, 2009). In addition, a control experiment was conducted to ensure that any differences observed between the listener groups could not be attributed to simple differences in audibility, especially in the high frequencies where the groups differed most in terms of their audiometric thresholds. A subset of the NH listeners repeated the experiment with the stimuli spectrally shaped to approximate the audibility curve of the HI group. This condition was not designed to simulate HL, but rather to eliminate one factor, reduced sensation level, from the list of factors that might disrupt performance in the HI group. The remaining deficits in performance in the HI group, beyond those exhibited by this control group, could then be attributed to suprathreshold aspects of HL, such as reduced frequency resolution that might increase susceptibility to EM.

II. METHODS

A. Participants

Nine NH listeners (18–25 years of age, mean 21) and nine HI listeners (18–44 years of age, mean 26) participated. The NH listeners were screened to ensure that their pure-tone thresholds were in the normal range (no greater than 20 dB HL) for octave frequencies from 250 to 8000 Hz. The HI listeners had mild to moderately severe, bilateral, symmetric, sloping, sensorineural HLs. Six of the nine HI listeners had losses of unknown etiology, two were associated with hereditary syndromes and one was hereditary but non-syndromic. Audiograms for each HI listener are shown in Fig. 1. Seven of the nine HI listeners were regular bilateral hearing aid users but their aids were removed for testing. Six listeners from the NH group participated in a control study, in which the experiment was repeated under a condition simulating the audibility experienced by the HI group. All listeners were paid for their participation.

B. Stimuli

A corpus of monosyllabic words recorded at Boston University's Hearing Research Center (for details see Kidd *et al.*,

J. Acoust. Soc. Am., Vol. 129, No. 3, March 2011

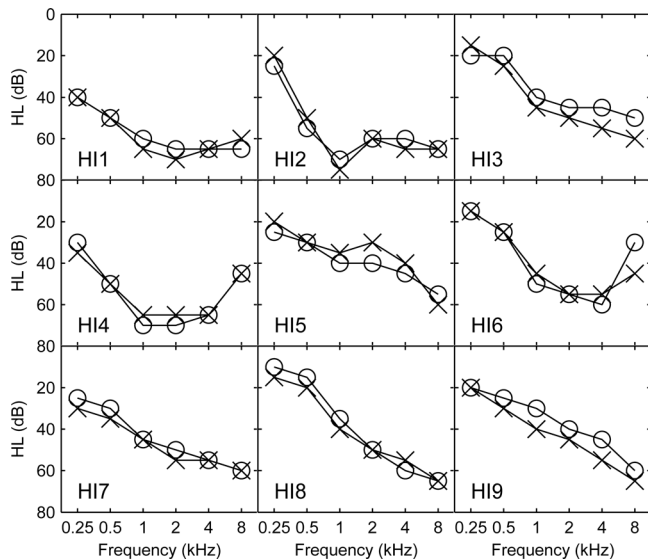Best *et al.*: Spatial release and temporal overlap   1617

FIG. 1. Left- and right-ear audiograms (crosses and circles, respectively) for each listener in the HI group.

2008a) was modified for use in the experiment. This corpus comprises 40 words organized into five different categories (names, verbs, numbers, adjectives, and nouns) spoken by eight male and eight female talkers. Each word in the corpus was temporally compressed to have a duration of exactly 200 ms using PRAAT software (Boersma and Weenink, 2009). The compression algorithm uses a pitch-synchronous overlap and add (PSOLA) algorithm that maintains the voice pitch information. Note that the unprocessed stimuli had durations ranging between 385 and 1051 ms, so in proportional terms, the compression ranged from 19% to 52%.

A target sentence was created by concatenating five words from the same talker with 200-ms intervening silent gaps (see Fig. 2). A masker sentence was created in the same way using a different talker of the same gender. The first word of the target was always the name BOB to identify it as the target sentence, but the other four target words and all five of the masker
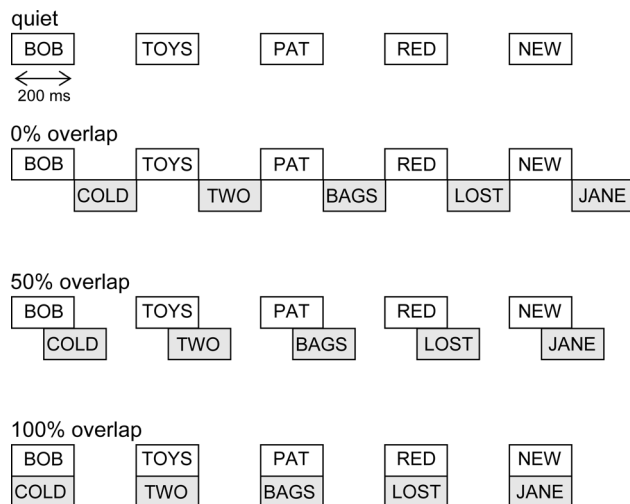

FIG. 2. Schematic illustration showing the temporal structure of the target (white) and masker (gray) sentences in the quiet training condition (top) as well as in the masked conditions with 0%, 50%, and 100% overlap.

words were randomly chosen without replacement from the remaining 39 words in the corpus with no attempt to use correct syntax. For example, target and masker sentences could be BOB TOYS PAT RED NEW and COLD TWO BAGS LOST JANE. All words were scaled to equal root-mean-square before concatenating, and the target and masker sentences were presented at the same level, chosen separately for each of the two listener groups to be comfortably loud [NH: 76 dB, HI: 96 dB sound pressure level (SPL)].

For the control study, an attempt was made to simulate in NH listeners the overall audibility of the speech stimuli in the HI group. The stimuli were filtered using a custom filter that had attenuation characteristics matching the mean audiogram of the HI group. Although this filter introduced attenuation across the spectrum, the attenuation was greatest in the high-frequency region and thus is referred to throughout the text as a "low-pass filter" and the control group as the normally hearing low-pass ("NHLP") group. The stimulus level was set to 96 dB SPL before filtering such that the resulting sensation level in the NHLP group would be similar to that of a listener with the average HI audiogram.

C. Procedures

Digital stimuli were generated on a personal computer (PC) using MATLAB software (MathWorks Inc, Natick, MA). The stimuli were digital-to-analog (D/A) converted and attenuated using System II hardware (Tucker-Davis Technologies, Alachua, FL) and presented over HD580 headphones (Sennheiser, Wedemark, Germany). Listeners were seated in a sound-treated booth fitted with a monitor and mouse and indicated their responses by clicking with the mouse on a graphical user interface. The interface contained a 5×8 grid of buttons, where each button was labeled with one word from the corpus. To assist with navigation around the grid, the five columns represented the five word categories, and the eight options in each category were arranged alphabetically within a column (except for the numbers, where numerals were used and arranged in order from lowest to highest).

Listeners responded to the five target words by clicking five different buttons on the interface (in the order in which they heard the words, or remembered hearing the words).[2] Each word in the response was scored separately and was only judged correct if it was the correct word in the correct position. Incorrect words were classified into one of three error categories. "Masker errors" were errors in which the response word corresponded to any of the words in the masker sentence. "Order errors" were errors in which the response word corresponded to one of the other target words in the sentence. "Random errors" were the remaining errors, in which the response word did not correspond to any of the presented words. Take the example used above, where the target sentence was BOB TOYS PAT RED NEW and the masker was COLD TWO BAGS LOST JANE. If the response was BOB TOYS FOUND NEW JANE, the words BOB and TOYS would be scored as correct, FOUND would be classified as a random error, NEW as an order error, and JANE as a masker error. Because the first word was always the same, performance was essentially perfect and was not

included in the presented results. For the remaining words, listeners showed some evidence of a serial order position effect with performance being superior at either end. This effect did not interact with any of the manipulations of interest in the study, and thus performance is presented as a mean score across the last four words in a response.

D. Conditions

Before commencing the main experiment, listeners completed one or more training blocks in which the target was presented diotically with no masker sentence present. These blocks gave listeners a chance to learn the response grid and also provided baseline intelligibility measures for the compressed words. Each listener completed between one and six blocks of 30 trials, depending on previous experience with the stimuli and response grid and how quickly they felt confident about using the response grid effectively. In cases where more than one block was completed, only the score for the final block is reported.

In the main experiment, target and masker sentences were presented with one of three levels of temporal overlap (see Fig. 2). Temporal overlap was either 0% (target and masker were perfectly interleaved and did not overlap at all), 50% (target and masker overlapped for half of the word duration), or 100% (target and masker were simultaneous and overlapped completely). In the 0% and 50% conditions, the temporally leading sentence was randomly the target or masker sentence, with equal numbers of trials of each over the course of a block.

While the target sentence was always presented diotically, the masker sentence was presented with one of two spatial configurations. In the "co-located" configuration, the masker sentence was also diotic. In the "separated" configuration, the masker sentence was lateralized to the right side by introducing an interaural time difference of 0.6 ms.

An experimental session consisted of six 30-trial blocks, one in each of the six conditions (two spatial configurations and three levels of temporal overlap). Each listener completed four such sessions, and the order of the blocks was randomized within each session. Because there were four words scored per trial and 120 trials per condition, percent correct scores for each listener were based on 480 response words in total.

III. RESULTS

A. Individual results

The percent correct scores for the nine NH listeners are shown in Fig. 3. The dashed lines at the top of each panel show quiet performance. The three lines in each panel show performance for the three overlap conditions in the co-located (left) and separated (right) configurations. All listeners performed well in quiet, with scores between 82% and 97% correct. Most listeners performed poorly (50% or below) in the co-located configuration for all overlap conditions. An exception to this is listener NH1 who performed relatively well in the 0% and 50% overlap conditions. In all cases, there was a clear benefit of spatial separation. For each individual listener, performance in the separated configuration was similar across
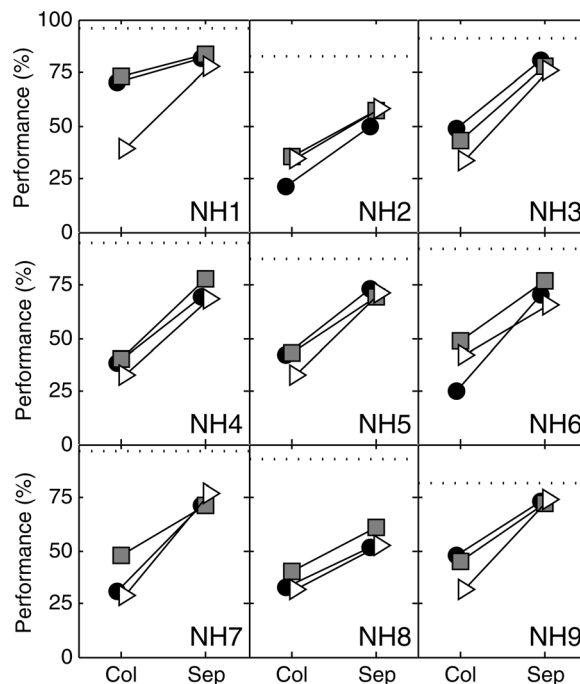

FIG. 3. Individual data for the nine listeners in the NH group. Each panel shows performance (in percent correct) for a different listener as a function of spatial configuration. The three lines in each panel represent the three overlap conditions (0%, black circles; 50%, gray squares; 100%, white triangles). Dashed lines at the top of each panel show quiet performance.

the three overlap conditions and either approached quiet performance (e.g., NH3 and NH9) or was somewhat poorer than the quiet performance (e.g., NH2 and NH8).

The percent correct scores for the nine HI listeners are shown in Fig. 4. Again the dashed lines at the top of each panel show quiet performance, and it is clear that some
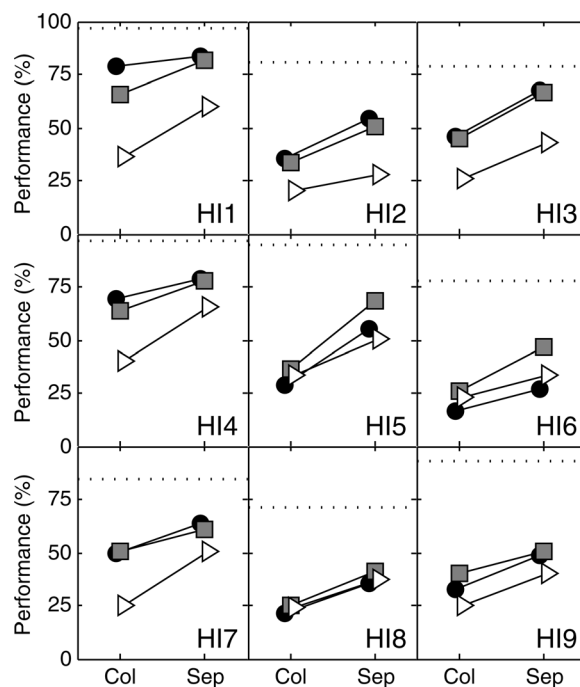

FIG. 4. Individual data for the nine listeners in the HI group. Other details as per Fig. 3.

listeners had trouble identifying these words even in quiet (e.g., HI3 and HI8 with scores around 75%). Most listeners performed poorly (50% or below) in the co-located configuration for all overlap conditions. Two exceptions are listeners HI1 and HI4 who performed relatively well in the 0% and 50% overlap conditions. Again, in all cases, there was a benefit of spatial separation. For some listeners (e.g., HI1 and HI3), performance in the separated configuration for 0% and 50% overlap approached performance in quiet, but 100% overlap scores were substantially worse. For other listeners (e.g., HI8 and HI9) scores in the separated configuration were poor in all overlap conditions.

The percent correct scores for the six NH listeners in the control experiment are shown in Fig. 5. These panels can be compared to the corresponding panels in Fig. 3, which show data from the same listeners. Quiet scores with low-pass filtered stimuli were similar to scores with unfiltered stimuli. Moreover, in the masked conditions, each listener showed similar performance across the two experiments. One exception is NH6, who showed some drop in performance with low-pass filtering in the 100% overlap condition.

B. Mean results

Figure 6 shows group mean scores for the co-located (top row) and separated (middle row) configurations. The bottom row shows the subtraction of these scores (SRM), which will be discussed below. The left column compares results for the NH and HI groups. The right column compares results for the NH subgroup who participated in the main experiment as well as in the control experiment (NH and NHLP). The three clusters of bars in each panel show the three overlap conditions as labeled.

The first comparison of interest is between the NH and HI groups. In the co-located configuration (top left panel), NH listeners achieved scores of 40%, 46%, and 34% in the 0%, 50%, and 100% overlap conditions, respectively. The HI group performed similarly, with scores of 42%, 43%, and 28%. Separation of the masker (middle left panel) provided a benefit in all overlap conditions for the NH group and the HI group. However, the separated scores were poorer in the HI
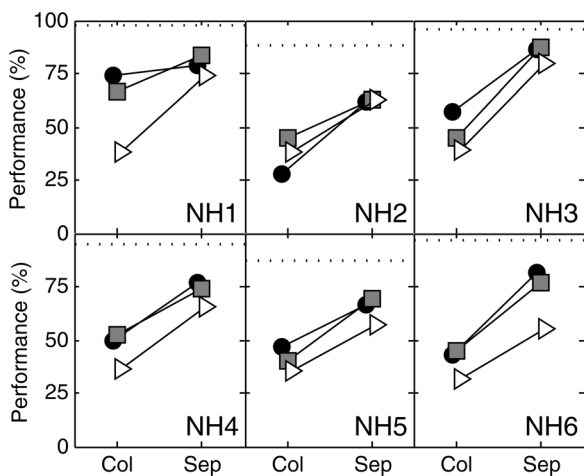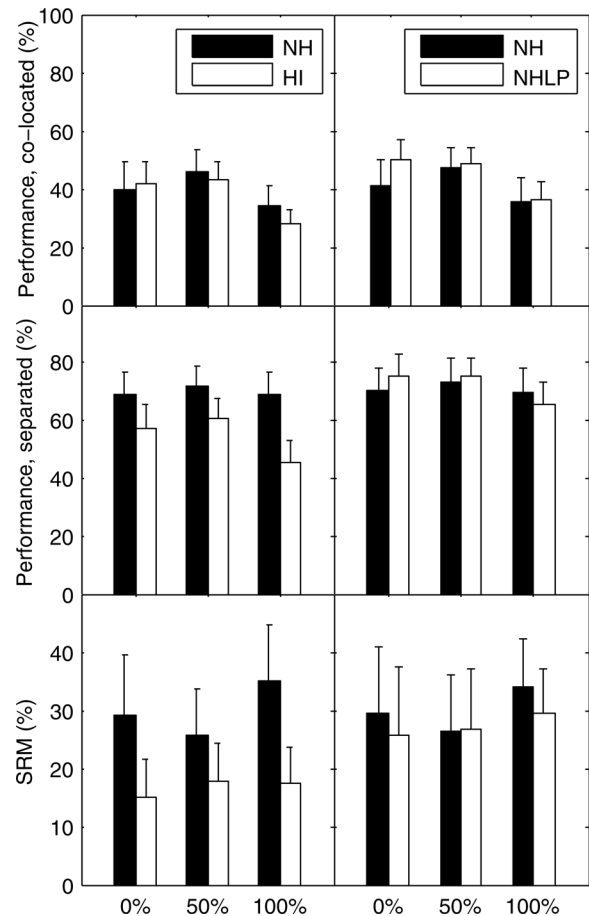

FIG. 6. Mean performance and SRM for the NH and HI groups (left column) and the NH subgroup without and with low-pass filtering (right column). The top and middle panels show performance (in percent correct) for the co-located and separated configurations, respectively. The bottom panel shows SRM (in percentage points—note the ordinate scale is different to upper panels). The three clusters of bars show results for the three overlap conditions. Bars represent mean results pooled across listeners and the error bars show ±1 standard deviation of the mean.

group than in the NH group, particularly for the 100% overlap condition (57%, 61%, and 45% vs 69%, 72%, and 69%). A three-way analysis of variance (ANOVA) (Table I) showed significant main effects of spatial configuration and overlap condition, but no significant main effect of listener group.[3] Listener group interacted significantly with spatial configuration and with overlap condition, but the interaction between spatial configuration and overlap condition was not significant nor was the three-way interaction. Planned comparisons (t-tests, $p < 0.05$) indicated that scores in the HI group were significantly poorer than those in the NH group in the 100% overlap condition for both co-located and separated configurations, but did not differ significantly in any other condition.

The second comparison of interest is between the broadband and low-pass filtered conditions for the subset of six NH listeners who completed both. The black bars in the right column of Fig. 6 show mean performance for this subset in the main experiment. Comparison to mean performance for the full group of nine listeners (black bars in the left column of Fig. 6) indicates that this subset was a good representative sample of the larger group. In the co-located configuration (top right panel), these subjects performed slightly better in

FIG. 5. Individual data for the six listeners in the NHLP group. Other details as per Fig. 3.

TABLE I. ANOVA results and effect sizes (partial η-squared) for analysis of overall performance. The top half of the table compares NH and HI groups. A three-way ANOVA was conducted with spatial configuration and overlap condition as within-subjects factors, and listener group as a between-subjects factor. The bottom half of the table compares NH and NHLP groups. A three-way ANOVA was conducted with spatial configuration, overlap condition, and filtering as within-subjects factors.

Groups	Factor	df	F	p	Effect size
NH vs HI	Spatial	1, 16	328.6	0.000[a]	0.95
	Overlap	2, 32	16.8	0.000[a]	0.51
	Group	1, 16	2.8	0.12	0.15
	Spatial × overlap	2, 32	2.1	0.14	0.11
	Spatial × group	1, 16	26.8	0.000[a]	0.63
	Overlap × group	2, 32	3.5	0.042[b]	0.18
	Spatial × overlap × group	2, 32	2.0	0.15	0.11
NH vs NHLP	Spatial	1, 5	134.3	0.000[a]	0.96
	Overlap	2, 10	6.1	0.018[b]	0.55
	Filter	1, 5	2.0	0.22	0.28
	Spatial × overlap	2, 10	0.6	0.56	0.11
	Spatial × filter	1, 5	2.1	0.21	0.29
	Overlap × filter	2, 10	4.8	0.035[b]	0.49
	Spatial × overlap × filter	2, 10	1.0	0.42	0.16

[a]Significant at $p < 0.005$.
[b]Significant at $p < 0.05$.

the control experiment than in the main experiment for all overlap conditions (scores of 50%, 49%, 36%). In the separated configuration (middle right panel), scores were again slightly higher in the NHLP group for 0% and 50% overlap but slightly worse for 100% overlap (scores of 75%, 76%, and 66%). A three-way ANOVA (Table I) showed significant main effects of spatial configuration and overlap condition, but no significant main effect of filtering. Listener group interacted significantly with overlap condition, but the other two-way interactions and the three-way interactions were not significant. Planned comparisons (paired t-tests, $p < 0.05$) found a significant effect of filtering only for the co-located configuration with 0% overlap, where listeners performed slightly better with low-pass filtered stimuli. This may reflect a small learning effect, as the low-pass filtered experiment was completed after the broadband experiment.

The results reported here confirm that the HI group performed more poorly than the NH group for completely overlapping stimuli. Low-pass filtering the stimuli for the NH group did not reduce performance under any condition, suggesting that the deficit in the HI group relative to the NH group cannot be attributed primarily to reduced high-frequency audibility.

C. SRM

The bottom panels of Fig. 6 show the SRM (difference in percentage points between performance in the separated and co-located configurations) for the different listener groups and overlap conditions. Taking first the NH and HI comparison (bottom left panel), the NH group obtained benefits of 29, 26, and 35 percentage points for the 0%, 50%, and 100% overlap conditions. The HI group showed less SRM (15, 18, and 17 percentage points). A two-way ANOVA (Table II) conducted on the benefits revealed a significant main effect of group but

no significant main effect of overlap and no interaction.[4] Planned comparisons (t-test, $p < 0.05$) confirmed that the effect of HL was significant in each of the three overlap conditions. Comparison of the NH and NHLP groups (bottom right panel) reveals similar amounts of SRM in the three overlap conditions (NH: 30, 26, and 34 percentage points; NHLP: 26, 27, and 29 percentage points). A two-way ANOVA (Table II) found no significant main effect of filtering or overlap and no significant interaction. In sum, these analyses confirm that SRM was significantly reduced by HL in all overlap conditions but was not significantly reduced by low-pass filtering.

D. Error patterns

To explore the basis of the differences between the NH and HI groups further, incorrect responses for each listener were classified into the three different error types as described in Sec. II C. The types of errors made by listeners can often be revealing as to the nature of the interference they experience under different listening conditions. For example, in closed-set speech identification experiments, the occurrence of errors in which listeners report words from the competing sentences is often interpreted as evidence for IM (Brungart, 2001; Kidd et al., 2005). Conversely, responses that are not present in the stimulus are associated with EM.

Masker errors were the most common error type overall, occurring on 23% of all trials (pooled across overlap conditions, spatial configurations, and listeners). Order errors occurred on 12% of trials and random errors on 15% of trials. Note that masker errors were much more frequent than would be expected if listeners chose from the grid randomly when they erred. Random guessing predicts that masker/order/random errors should occur in proportions of 0.13/0.08/0.79, respectively. Instead, the observed proportions were 0.48/0.24/0.28.

Figure 7 illustrates the frequency of the different error types with a breakdown by overlap condition and listener group. The left and right columns show the co-located and separated configurations, and the top, middle, and bottom panels in each column show rates of masker errors, order errors, and random errors. In each panel, the bars represent mean error rates as a percentage of total trials for the NH and HI groups in the different overlap conditions.

TABLE II. ANOVA results and effect sizes (partial η-squared) for analysis of SRM. The top half of the table compares NH and HI groups. A two-way ANOVA was conducted with listener group as a between-subjects factor and overlap condition as a within-subjects factor. The bottom half of the table compares NH and NHLP groups. A two-way ANOVA was conducted with filtering and overlap condition as within-subjects factors.

Groups	Factor	df	F	p	Effect size
NH vs HI	Overlap	2, 32	2.1	0.14	0.11
	Group	1, 16	26.7	0.000[a]	0.63
	Overlap × group	2, 32	2.1	0.15	0.11
NH vs NHLP	Overlap	1, 5	2.1	0.21	0.29
	Filter	2, 10	0.6	0.56	0.11
	Overlap × filter	2, 10	1.0	0.42	0.16

[a]Significant at $p < 0.005$.

J. Acoust. Soc. Am., Vol. 129, No. 3, March 2011

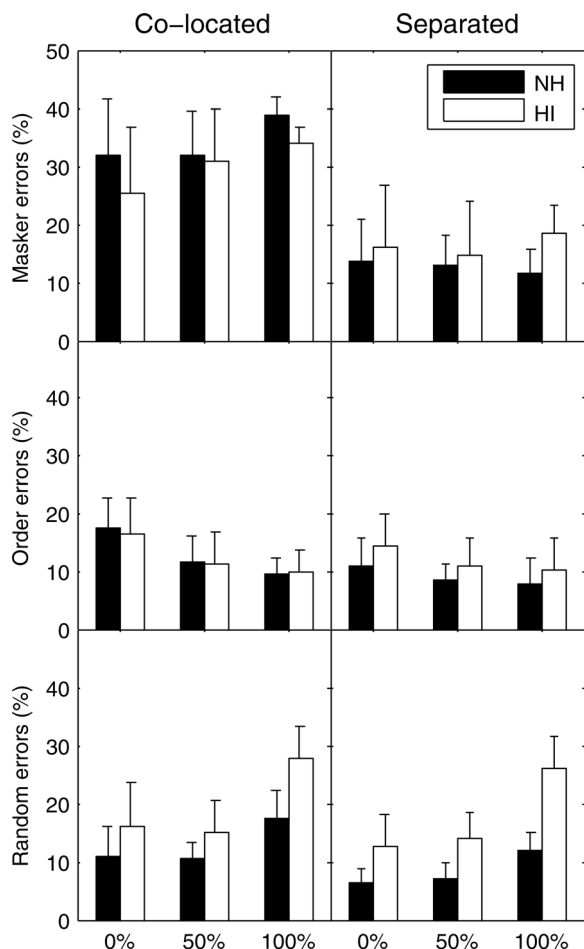Best et al.: Spatial release and temporal overlap   1621

FIG. 7. Error rates for the NH and HI groups. The left and right columns show data for the co-located and separated configurations, and the three rows show masker errors, order errors, and random errors. The three clusters of bars in each panel show the three overlap conditions. Bars show the mean error rates in percent of total trials, such that the different error rates add up to the total percentage of incorrect trials. The error bars show ±1 standard deviation of the mean.

TABLE III. ANOVA results and effect sizes (partial η-squared) for analysis of error rates in NH and HI listener groups. The top third of the table shows results for masker errors, the middle third shows results for order errors, and the bottom third shows results for random errors. For each error type, a three-way ANOVA was conducted with spatial configuration and overlap condition as within-subjects factors and listener group as a between-subjects factor.

Error type	Factor	df	F	p	Effect size
Masker	Spatial	1, 16	399.4	0.000[a]	0.96
	Overlap	2, 32	4.9	0.014[b]	0.23
	Group	1, 16	0.0	0.95	0.00
	Spatial × overlap	2, 32	13.5	0.000[a]	0.46
	Spatial × group	1, 16	20.1	0.000[b]	0.56
	Overlap × group	2, 32	0.8	0.45	0.05
	Spatial × overlap × group	2, 32	4.3	0.023[b]	0.21
Order	Spatial	1, 16	32.6	0.000[a]	0.67
	Overlap	2, 32	33.4	0.000[a]	0.68
	Group	1, 16	0.4	0.54	0.02
	Spatial × overlap	2, 32	8.5	0.001[a]	0.35
	Spatial × group	1, 16	13.5	0.002[a]	0.46
	Overlap × group	2, 32	0.0	0.96	0.00
	Spatial × overlap × group	2, 32	0.7	0.49	0.04
Random	Spatial	1, 16	34.0	0.000[a]	0.68
	Overlap	2, 32	76.8	0.000[a]	0.83
	Group	1, 16	18.0	0.001[a]	0.53
	Spatial × overlap	2, 32	1.7	0.20	0.10
	Spatial × group	1, 16	4.3	0.056	0.21
	Overlap × group	2, 32	9.4	0.001[a]	0.37
	Spatial × overlap × group	2, 32	0.8	0.45	0.05

[a]Significant at $p < 0.005$.
[b]Significant at $p < 0.05$.

In the co-located configuration, masker errors increased with increasing temporal overlap and were slightly more common in the NH group (top left panel). In the separated configuration, masker errors were reduced in all conditions and groups, the effect of overlap was not apparent, and masker errors were slightly more common in the HI group (top right panel). A three-way ANOVA (Table III) showed a significant main effect of spatial configuration and overlap condition but no significant main effect of group. Spatial location interacted significantly with overlap condition and with group, and the three-way interaction was also significant. Planned comparisons (t-tests, $p < 0.05$) suggested that the NH group made more masker errors in the co-located configuration for 100% overlap, and the HI group made more masker errors in the separated configuration for 100% overlap.

Order errors decreased with increasing temporal overlap in both co-located and separated configurations and were similar across the two listener groups (middle row). A three-way ANOVA (Table III) revealed significant main effects of spatial configuration and overlap condition but no significant main effect of group. Spatial configuration interacted signifi-cantly with overlap condition and with group but the three-way interaction was not significant. Planned comparisons (t-tests, $p < 0.05$) revealed no difference between listener groups in any condition.

As expected, random errors were most frequent in the 100% overlap condition. Moreover, the HI group consistently made more random errors than the NH group. A three-way ANOVA (Table III) showed significant main effects of spatial configuration, overlap condition, and listener group, as well as a significant interaction between overlap and group. The three-way interaction was not significant. Planned comparisons (t-tests, $p < 0.05$) indicated that the HI group made more random errors in the 50% and 100% overlap conditions for both spatial configurations and in the 0% overlap condition for the separated configuration only.

IV. DISCUSSION

This study tested the hypothesis that the increased speech-on-speech masking seen in HI listeners and the reduced spatial release from such masking are due to increased EM in the simultaneous portions of the speech waveforms. A novel two-talker paradigm was used that gave rise to a large amount of IM. By varying the temporal overlap of the competing talkers, the amount of EM was systematically varied. We predicted that HI listeners would show

the largest deficit relative to NH listeners in the case of complete temporal overlap (where EM was maximized) with little or no deficit in the case of zero overlap (where EM was negligible). It was also hypothesized that increased EM would lead to reduced SRM in HI listeners under conditions involving temporal overlap, but that zero overlap would give rise to comparable SRM in NH and HI groups. While the first hypothesis was confirmed, the second hypothesis was not.

Both NH and HI groups performed most poorly in the 100% overlap condition, reflecting the fact that temporal overlap in the competing sentences caused detrimental EM (or "mutual masking," Darwin, 2007) in addition to the strong IM caused by the presence of masking words from the same set as the target words. It was interesting to find that scores in the 50% overlap condition tended to follow scores in the 0% overlap condition rather than representing an intermediate case. It seems that these word tokens were redundant enough that presenting half of a word in isolation was similar to presenting the whole word in isolation with respect to conveying word identity. The deficit for the 100% overlap condition was larger in the HI group than in the NH group, supporting our hypothesis that it is the simultaneous portions of competing speech mixtures that pose the most difficulty for listeners with HL.

Our finding implies that the primary factor reducing performance in HI listeners for competing speech tasks is increased EM and not increased IM. This complements previous studies that have reported similar amounts of IM in younger and older listeners despite differences in the amount of EM (Li et al., 2004; Helfer and Freyman, 2008; Agus et al., 2009). Together these results indicate that HL in younger and older listeners increases EM and reduces performance in realistic listening situations.

In another related study, Arehart and colleagues showed that increased masking between concurrent vowel stimuli in listeners with HL degraded the individual representation of the vowels (Arehart et al., 2005; Rossi-Katz and Arehart, 2005). They also showed that presenting the vowel pair to different ears, and thus removing the peripheral overlap, remediated this deficit to some extent (Arehart et al., 2005). Presenting the simultaneous talkers with different ITDs did not provide a comparable release given that the scores for the HI group remained low (and in particular random errors remained high) in the ITD-separated configuration. It seems likely that the separation by ITD was less effective than separation by ear because of the peripheral overlap (and EM) that remained.

The prediction that removing temporal overlap in the competing sentences would allow HI listeners to achieve SRM equivalent to that seen in NH listeners was not borne out. Rather, spatial release was reduced in HI listeners in *all* overlap conditions. This arose because scores in the separated configuration were worse in the HI group (although not significantly) even for the 0% overlap condition. Despite this group result, there were some listeners for whom performance in the separated configuration for 0% and 50% overlap approached quiet levels (e.g., HI1 and HI3), and overall the recovery toward quiet performance in those conditions was better than in the 100% overlap condition. This suggests that, at least in some listeners, there is no deficit in spatial perception *per se* but only in the ability to use perceived location to separate simultaneous signals.

For the listeners who did show a deficit in all masked conditions, regardless of temporal overlap, there are several possible explanations. One explanation is that HI listeners cannot use spatial separation to perceptually segregate sound sources—and therefore reduce IM—to the same extent that NH listeners can. This seems unlikely given the large and salient ITD difference that was used in the current study; past studies have found that HI listeners are often able to use much subtler spatial cues to discriminate sound source locations than those employed here (Colburn, 1982; Häusler et al., 1983; Smith-Olinde et al., 1998). Another possibility is that distortions of the stimulus related to HL caused increased confusions between the target and the masker that persisted in the HI group even with the aid of spatial separation. Both of these explanations, suggesting that IM is not released as effectively in HI listeners, would predict an increase in masker errors in the separated configuration. However, this was not observed. A third explanation is that EM is increased in HI listeners even in cases of temporally non-overlapping sounds, perhaps due to increased forward masking (Nelson and Freyman, 1987; Festen, 1993; Jin and Nelson, 2006). This explanation receives some indirect support from the error analysis, which indicated that the deficit in the HI group relative to the NH group in the 0% overlap condition was driven by increased random errors. An experiment that delivered target and masker sentences to the two ears separately (as per Arehart et al., 2005) might help to test this possibility. Finally, a fourth possible explanation is that the degraded representation of speech caused by HL interacts with task difficulty, impairing performance on challenging speech tasks even when the signals are acoustically perfectly segregated as in the non-overlapping condition of this study. It has been shown previously, for example, that simulated HL exacerbates the effects of time compression of words such as that used in this study (Stuart and Phillips, 1998), and that HL increases the processing load during listening such that the subsequent recall of speech is adversely affected (see for e.g., McCoy et al., 2005). These additional costs in HI listeners under such conditions may place a hard "limit" on performance so that the full benefit of manipulations like spatial separation cannot be achieved.

V. CONCLUSIONS

This experiment examined performance in a speech-on-speech task as a function of the temporal overlap of the competing signals. Listeners with HL performed particularly poorly relative to listeners with normal hearing in the 100% overlap condition, consistent with the idea that simultaneous speech sounds are most problematic for these listeners. However, their SRM was reduced in *all* overlap conditions, suggesting that increased masking during periods of temporal overlap is only one factor limiting spatial unmasking in HI listeners for competing speech situations. Reduced audibility did not appear to account for the current results.

[1]Note that the compression of individual words to a fixed duration is different to the more typical approach of compressing naturally spoken sentence materials by a constant factor (e.g., Adank and Janse, 2009; Feng et al., 2010).

[2]The method used here may seem more demanding than some other methods for obtaining responses to speech (such as free free recall), because of the requirement to visually search for the target words on the grid. However, our experience with this task in previous studies (Kidd et al., 2008a; Kidd et al., 2010) has indicated that young listeners pick it up very quickly and after a small amount of practice find it to be a rapid and convenient means of registering a response.

[3]All of the statistical comparisons were also done after applying a rationalized arcsine transform (Studebaker, 1985) to the percent correct scores. This transformation did not change any of the outcomes and thus only the results for the untransformed data are plotted and reported.

[4]The statistical outcomes were the same if the SRM was normalized to quiet performance, i.e., if the SRM was stated as a proportion of the theoretical maximum release such that an SRM of 1 corresponds to a return to quiet performance.

Adank, P., and Janse, E. (**2009**). "Perceptual learning of time-compressed and natural fast speech," J. Acoust. Soc. Am. **126**, 2649–2659.

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.

Allen, K., Carlile, S., and Alais, D. (**2008**). "Contributions of talker characteristics and spatial location to auditory streaming," J. Acoust. Soc. Am. **123**, 1562–1570.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

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.

Arehart, K. H., Rossi-Katz, J., and Swensson-Prutsman, J. (**2005**). "Double-vowel perception in listeners with cochlear hearing loss: Differences in fundamental frequency, ear of presentation, and relative amplitude," J. Speech Lang. Hear. Res. **48**, 236–252.

Best, V., Ozmeral, E., and Shinn-Cunningham, B. G. (**2007**). "Visually-guided attention enhances target identification in a complex auditory scene," J. Assoc. Res. Otolaryngol. **8**, 294–304.

Boersma, P., and Weenink, D. (**2009**). "Praat: Doing phonetics by computer," http://www.praat.org/ (Last viewed May 08, 2010).

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

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

Brungart, D. S., and Simpson, B. D. (**2007**). "Cocktail party listening in a dynamic multitalker environment," Percept. Psychophys. 79–91.

Carhart, R., and Tillman, T. W. (**1970**). "Interaction of competing speech signals with hearing losses," Arch. Otolaryngol. **91**, 273–279.

Colburn, H. S. (**1982**). "Binaural interaction and localization with various hearing impairments," Scand. Audiol. Suppl. **15**, 27–45.

Darwin, C. J. (**2007**). "Spatial hearing and perceiving sources," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 215–232.

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.

Feng, Y., Yin, S., Kiefte, M., and Wang, J. (**2010**). "Temporal resolution in regions of normal hearing and speech perception in noise for adults with sloping high-frequency hearing loss," Ear Hear. **31**, 115–125.

Festen, J. M. (**1993**). "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice," J. Acoust. Soc. Am. **94**, 1295–1300.

Festen, J. M., and 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.

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., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Häusler, R., Colburn, H. S., and Marr, E. (**1983**). "Sound localization in subjects with impaired hearing. Spatial-discrimination and interaural-discrimination tests," Acta Oto-Laryngol., Suppl. **400**, 1–62.

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

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Jin, S. H., and Nelson, P. B. (**2006**). "Speech perception in gated noise: the effects of temporal resolution," J. Acoust. Soc. Am. **119**, 3097–3108.

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.

Kidd, G., Jr., Best, V., and Mason, C. R. (**2008a**). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," J. Acoust. Soc. Am. **124**, 3793–3802.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2008b**). "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, New York), pp. 143–190.

Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.

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.

Kitterick, P. T., Bailey, P. J., and Summerfield, A. Q. (**2010**). "Benefits of knowing who, where, and when in multi-talker listening," J. Acoust. Soc. Am. **127**, 2498–2508.

Li, L., Daneman, M., Qi, J., and Schneider, B. (**2004**). "Does the information content of an irrelevant source differentially affect speech recognition in younger and older adults?," J. Exp. Psychol. Hum. Percept. Perform. **30**, 1077–1091.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. **103**, 18866–18869.

Marrone, N., Mason, C. R., and Kidd, G., Jr. (**2008a**). "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.

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

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., and Wingfield, A. (**2005**). "Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech," Q. J. Exp. Psychol. A **58**, 22–33.

Moore, B. C. J. (**1985**). "Frequency selectivity and temporal resolution in normal and hearing-impaired listeners," Br. J. Audiol. **19**, 189–201.

Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological, and Technical Issues* (Wiley, Chichester, UK), pp. 1–91.

Nelson, D. A., and Freyman, R. L. (**1987**). "Temporal resolution in sensorineural hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 709–720.

Rossi-Katz, J. A., and Arehart, K. H. (**2005**). "Effects of cochlear hearing loss on perceptual grouping cues in competing-vowel perception," J. Acoust. Soc. Am. **118**, 2588–2598.

Smith-Olinde, L., Koehnke, J., and Besing, J. (**1998**). "Effects of sensorineural hearing loss on interaural discrimination and virtual localization," J. Acoust. Soc. Am. **103**, 2084–2099.

Stuart, A., and Phillips, D. P. (**1998**). "Recognition of temporally distorted words by listeners with and without a simulated hearing loss," J. Am. Acad. Audiol. **9**, 199–208.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Webster, J. C., and Thompson, P. O. (**1954**). "Responding to both of two overlapping messages," J. Acoust. Soc. Am. **26**, 396–402.

Yost, W. A. (**1997**). "The cocktail party problem: Forty years later," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. A. Gilkey and T. R. Anderson (Erlbaum, New Jersey), pp. 329–348.

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.