

Tuning in the spatial dimension: Evidence from a masked speech identification task

Nicole Marrone,^{a)} Christine R. Mason, and Gerald Kidd, Jr.

*Department of Speech, Language, and Hearing Sciences and the Hearing Research Center,
Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215*

(Received 28 February 2007; revised 23 May 2008; accepted 28 May 2008)

Spatial release from masking was studied in a three-talker soundfield listening experiment. The target talker was presented at 0° azimuth and the maskers were either colocated or symmetrically positioned around the target, with a different masker talker on each side. The symmetric placement greatly reduced any “better ear” listening advantage. When the maskers were separated from the target by ±15°, the average spatial release from masking was 8 dB. Wider separations increased the release to more than 12 dB. This large effect was eliminated when binaural cues and perceived spatial separation were degraded by covering one ear with an earplug and earmuff. Increasing reverberation in the room increased the target-to-masker ratio (T/M) for the separated, but not colocated, conditions reducing the release from masking, although a significant advantage of spatial separation remained. Time reversing the masker speech improved performance in both the colocated and spatially separated cases but lowered T/M the most for the colocated condition, also resulting in a reduction in the spatial release from masking. Overall, the spatial tuning observed appears to depend on the presence of interaural differences that improve the perceptual segregation of sources and facilitate the focus of attention at a point in space. © 2008 Acoustical Society of America.

[DOI: 10.1121/1.2945710]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.66.Lj [RLF]

Pages: 1146–1158

I. INTRODUCTION

There are a number of examples in the auditory system of selective responses along a simple stimulus dimension. Perhaps the most obvious and best understood example is stimulus frequency where the tuned responses of the peripheral transduction mechanism have been thoroughly mapped out and examined. Beyond peripheral filtering, however, are instances in which the actions of higher level processes lead to performance that reveals an enhanced degree of selectivity. Greenberg and Larkin (1968), for example, used the probe-signal method to demonstrate tuning in the frequency domain that was not attributable to peripheral filtering but rather to the focus of attention at an expected signal frequency. Similar findings revealing the role of expectation and selectivity have been reported for other dimensions, such as duration (e.g., Wright and Dai, 1994), spectral shape (Hill *et al.*, 1998), and modulation frequency (e.g., Wright and Dai, 1998).

The behavioral evidence for tuning along the spatial (azimuthal) dimension is less compelling and surprisingly limited (cf. Scharf, 1998). Clearly, certain neurons in the brainstem or cortex exhibit selective responses to the primary binaural cues of interaural time and level differences (e.g., Goldberg and Brown, 1968; Middlebrooks and Pettigrew, 1981; Yin and Kuwada, 1983; Tsuchitani, 1988; Yin and Chan, 1990; Sterbing *et al.*, 2003; Stecker *et al.*, 2003; King *et al.*, 2007) so that, akin to the example of frequency noted previously, there is a basis in auditory physiology for

expecting tuned responses to spatial location. Further, there is electrophysiological evidence suggesting that higher-level processes affect selectivity in azimuth. Teder-Salejarvi and Hillyard (1998) and Teder-Salejarvi *et al.* (1999) have demonstrated changes in event-related potentials (ERPs) in humans based on attended versus unattended locations of a sound stimulus. When a test stimulus was presented at an attended location, larger ERPs were obtained than when presented at unattended locations. Filter bandwidths inferred from their data were often sharply tuned, in some cases less than 5°. Psychophysical evidence for the important role of attentional focus in location, based on presentation of speech stimuli at likely versus unlikely locations, has also been presented recently by Kidd *et al.* (2005b) and Brungart *et al.* (2006).

Studies of masking in sound fields also provide examples of a selective response to spatial location. More masking occurs when target and masker(s) are colocated than when they are separated in azimuth. This selective response, a progressive reduction in masking as source separation increases, could be interpreted as evidence for tuned spatial channels analogous to frequency channels. However, the release from masking that occurs due to spatial separation of sound sources is a complex phenomenon comprising both lower-level and higher-level components (e.g., Kidd *et al.*, 1998; Freyman *et al.*, 1999; Arbogast *et al.*, 2002; Drennan *et al.*, 2003, 2007; Hawley *et al.*, 2004; Best *et al.*, 2005, 2006) and it is not always apparent which factor drives release from masking. Thus, the mechanism(s) responsible for the apparent tuned response is unclear and may be different in different conditions. The main lower-level components that are thought to provide a basis for spatial release from

^{a)}Electronic mail: n-marrone@northwestern.edu

masking are the “better ear advantage” [attending to the ear with the more favorable target-to-masker ratio (T/M) caused by the acoustic filtering of the head]¹ and “binaural analysis” [defined here as within-channel improvement in T/M due to a masking-level-difference (MLD) type of mechanism, where the terms within and across channel refer to frequency channels]. The relevant higher-level components are less clearly understood but are thought to reflect (at least) the combined actions of perceptual segregation of sources and the focus of attention at a point in space (cf. Kidd *et al.*, 2005b). The type of masking that is present may have a profound effect on the pattern of results as source separation is varied. For example, Kidd *et al.* (1998) found small (less than 5 dB, on average, for mid- and low-frequency targets) amounts of spatial release from masking in a nonspeech pattern identification task for a noise masker producing primarily energetic masking. For the same listeners, targets, and task the spatial release from a primarily informational masker (sequences of random-frequency tone complexes producing little if any energetic masking) was, on average, as much as 20 dB at low and high target frequencies. The Kidd *et al.* (1998) results revealed a progressive and pronounced decline in T/M for the informational masking condition as spatial separation of target and masker varied from collocated to 180°. However, although the results for both energetic and informational maskers were consistent with a tuned response (although of very different magnitudes), and both clearly depended on the binaural cues of interaural time (ITD) and level (ILD) differences, the underlying mechanisms were thought to be fundamentally different. For the energetic masker, by far the largest spatial release occurred for the high frequency target and followed the published values for head shadow (e.g., Shaw, 1974) fairly closely after taking into account the mildly reverberant room. Predictions for spatial release from energetic masking for speech in noise are based on the same two factors and are generally successful in accounting for the empirical results (cf. Zurek, 1993; Bronkhorst, 2000). For the informational masker used by Kidd *et al.* (1998), head shadow may have been one factor, although the across-channel nature of the masking complicates the interpretation of the effect. Binaural analysis almost certainly was not a factor because the masking was not within-channel masking leading to the conclusion that higher-level factors were primarily responsible.

Another example of a spatially tuned response obtained perceptually that appeared to be due largely to higher level processes was reported by Arbogast and Kidd (2000). They used a variant on the probe-signal method to examine tuning in azimuth. The task was to identify the upward or downward trajectory of a sequence of tone pulses presented through a loudspeaker. Sequences of tonal masking sounds were presented from several other loudspeakers roughly concurrently with the target. The target was most likely to be presented at one location but on a small proportion of the trials it was randomly presented from other locations. The masker frequencies were randomized and remote from the target, limiting energetic masking while emphasizing informational masking. Both accuracy and response times were better at the more likely target location than at less likely

locations, which Arbogast and Kidd (2000) interpreted as evidence for spatial filtering. The filter-like responses they observed appeared to be fairly sharply tuned although the effects were relatively small and the attenuation characteristics of the “filter” could not be accurately determined.

Thus, there is some psychophysical evidence that suggests tuning in azimuth. However, the evidence is incomplete and in some cases inconsistent. For example, in contrast to the (apparently) sharply tuned pattern of masked responses reported by Arbogast and Kidd (2000), Boehnke and Phillips (1999) found evidence indicating much broader perceptual tuning based on the results of a gap detection task. Based on their results, they proposed two broadly tuned, overlapping spatial channels located in the left and right auditory hemifields. However, that interpretation also has been questioned (Oxenham, 2000). Other studies using different paradigms have also suggested the presence of spatial channels or filters. For example, Carlile *et al.* (2001) used a procedure in which localization judgments following exposure to noise at one location were altered in a manner consistent with the presence of a set of spatially arranged (in azimuth) filters.

For the masking results, at least, the role of head shadow, in which the magnitude of the acoustic effect varies with spatial position and frequency, complicates determining how the other factors contribute to spatial tuning. One approach to minimizing the role of head shadow is the symmetric placement of two maskers around a target. This approach has been used by several investigators (e.g., Helfer, 1992; Peissig and Kollmeier, 1997; Bronkhorst and Plomp, 1992; Noble and Perrett, 2002; Li *et al.*, 2004) to examine spatial release from masking that occurs beyond any acoustical “better ear listening” effects. However, only the study by Noble and Perrett (2002) provides an indication of spatial tuning when the target is one talker and the maskers are other independent talkers, symmetrically placed around the target. They tested two spatial separations, $\pm 30^\circ$ and $\pm 90^\circ$ (in addition to collocated) and found that the relatively small amount of spatial release (approximately 4 dB) for the $\pm 30^\circ$ condition increased by about 1 dB for the $\pm 90^\circ$ condition. Somewhat like the Arbogast and Kidd (2000) study, it is possible to conclude that the selectivity of a perceptual filter was sharper than the narrowest spacing tested but the precise properties of the filter could not be determined.

In recent years, interest has turned to speech-on-speech masking, in part because it represents a common everyday listening situation, but also because higher level processes seem to be more of a factor in determining spatial release. In two studies (Arbogast *et al.*, 2002; Kidd *et al.*, 2005a), much larger amounts of spatial release were found for speech-on-speech masking than in noise-masked control conditions producing primarily energetic masking. Further, in the Kidd *et al.* (2005a) study, there was also a differential effect of reverberation depending on masker type. Increased reverberation increased the T/M at threshold for both collocated and spatially separated conditions, with a large amount of spatial release still apparent. Other studies, however, (cf. Culling *et al.*, 2003) have reported that increased reverbera-

tion completely eliminates spatial release for speech. Thus, currently, the role of reverberation in spatial release from speech-on-speech masking is uncertain.

In the present study our goal was to examine spatial release from masking under conditions producing different amounts of energetic and informational masking of speech and in rooms with varying degrees of reverberation. The energetic/informational masking distinction was varied by using different types of masking stimuli and reverberation was varied by changing the sound absorption characteristics of the listening environment. This study attempts to answer the question of whether or not a filter-like function can be measured in these situations when the listener's task is to selectively attend to one talker at a particular location and the interfering sources are progressively separated (symmetrically) from the target.

II. METHODS

A. Listeners

Six normal hearing adult volunteers (5 female, 1 male) between 25 and 43 years of age participated in this study. The listeners had audiometric thresholds of 20 dB HL (hearing level re: ANSI 3.6–2004) or better in each ear for octave frequencies from 250 to 8000 Hz. Listeners participated in the experiment in four to six sessions that were approximately two hours each (including short breaks) and were paid for their participation. One additional listener with an anacusic ear after the auditory nerve was severed surgically was tested on a subset of the conditions (see Sec. II D 2).

B. Stimuli

The four female talkers from the Coordinate Response Measure (CRM) speech identification test (Bolia *et al.*, 2000) were used for the target and masker sentences. Each sentence in this corpus has the following structure: "Ready [callsign] go to [color] [number] now." The corpus contains all possible combinations of eight callsigns, four colors, and eight numbers. On each trial, the listener heard three sentences spoken by three different randomly chosen talkers. Each sentence had a different callsign, color, and number. One of the sentences was the target, denoted by the callsign "Baron." The target and masker talkers varied from trial to trial. Once the target sentence was selected, the two masker sentences were then chosen without replacement from the remaining set of possible talkers, callsigns, colors, and numbers.

In addition, a subset of listeners was tested in a condition where the speech of the two masker talkers was temporally reversed (see Sec. III D). The purpose of this manipulation was to test a condition under which the masker had spectrotemporal properties similar to natural speech (and therefore produces about the same amount of energetic masking) but was expected to produce less informational masking because it was not intelligible.

C. Room characteristics

The experiment was conducted in a single-walled Industrial Acoustics Company (IAC) sound booth (12 ft 4 in. long,

13 ft wide, and 7 ft 6 in. high) with stimuli presented through loudspeakers (Acoustic Research 215 PS). This room was designed to allow for changing the reverberation characteristics with panels of different acoustic reflectivity, such as acoustic foam or Plexiglas®. Two room conditions were used for this experiment. In one low-reverberation condition, the surfaces were untreated and had the typical perforated metal surface that is standard in IAC booths and a carpeted floor (the "BARE" room). In the second high-reverberation condition, all surfaces were covered with Plexiglas® panels to increase the acoustic reflections in the room (the "PLEX" room). Various measurements of the room acoustics for these conditions were described by Kidd *et al.* (2005a), documenting the increase in reverberation as a result of adding Plexiglas® panels to the room. For example, there was a greater than 7 dB decrease in the direct to reverberant energy ratio (from 6.3 to -0.9 dB for BARE versus PLEX, respectively) as measured at the approximate position of the listener's head 5 ft from the loudspeaker. Also, the reverberation time (as measured using pulse trains with the loudspeakers in the configuration used in the present study) increased by approximately a factor of 4 (from 0.06 s to just over 0.25 s when the surfaces were covered with Plexiglas®).

D. Procedures

1. General

Listeners were seated in the sound booth in the center of an array of seven loudspeakers arranged in a semicircle with a 5 ft radius in the horizontal plane at a height approximately level with the listener's ears. The loudspeakers were located at 0° (directly in front of the listener), ±15°, ±45°, and ±90° (directly to the right and left of the listener).

The computer used to control the experiment was located outside the booth along with the Tucker-Davis Technologies (TDT) hardware used to present the stimuli. Stimuli were played at a 40-kHz rate via a 16-bit, 8-channel digital-to-analog converter (DA3-8), low-pass filtered at 20 kHz (FT-6), and attenuated (PA-4). The target sentence was routed through a programmable switch (SS-1). On trials when the target and maskers were colocated (0° spatial separation), the two masker sentences were digitally added, then routed through separate digital-to-analog converter channels, filters, and attenuators before being combined in a mixer (SM3), passed through a power amplifier (Tascam), and routed to the loudspeaker. On trials when the maskers were spatially separated from the target, each sentence was routed through separate digital-to-analog converter channels, filters, attenuators, and power amplifiers (Tascam), and played through separate loudspeakers.

The system was calibrated before each session so that the loudspeakers were correctly positioned (5 ft from listener at head height) and the output level measured with a Brüel & Kjær microphone suspended in that position for a given input was verified and the same from each loudspeaker. For a flat-spectrum Gaussian noise of the same level at the input to the loudspeakers measured at the position of the listener, the

overall SPL was approximately 3 dB higher in the more reverberant room (PLEX). This correction was made when the results are reported in dB SPL.

The task was 1-interval 4×8 -alternative forced-choice (four colors: red, white, blue, and green and the numbers 1–8) in which the listeners were asked to identify the color and number from the sentence with the call sign Baron. Listeners were instructed to keep their head facing forward (toward the target loudspeaker at 0° azimuth) but were not restrained. Responses were entered on a handheld keypad with a liquid crystal display (Q-term-II). The word “Listen” appeared on the display at the beginning of each trial. After stimulus presentation, listeners responded to the prompts, “Color [B R W G]?” and “Number [1–8]?” on the keypad. For a response to be scored as correct the listener had to identify both the color and number accurately. Feedback was given on each trial (e.g., “Correct, it was red six”). To familiarize listeners with these procedures, they initially completed two 30-trial blocks with sentences presented in quiet at 60 dB SPL. Every listener had 100% correct speech identification in that condition.

Threshold for target identification was measured in quiet (no maskers) in each room condition using a one-up, one-down adaptive procedure to estimate the 50% correct point on the psychometric function (Levitt, 1971). Each track had a minimum of 30 trials and 9 reversals (typically many more than 9 were obtained) to estimate threshold. The threshold estimate was computed after discarding the first three or four reversals (whichever produced an even number) and thus was based on at least the last six reversals. The initial step size was 4 dB and was reduced to 2 dB after three reversals. Two estimates of threshold in quiet were measured and averaged. If the threshold estimates were more than 5 dB different from one another, an additional two estimates were collected and used in the average.

In all other blocks of trials, the target was presented simultaneously with two maskers. The target level was fixed at 60 dB SPL and the masker level was varied adaptively using the same procedure as for quiet threshold measurements. The two maskers always had the same rms level. At the beginning of each adaptive track, the target was clearly audible above the maskers (+20 dB T/M). The masker level was varied adaptively in 4 dB steps initially and then in 2 dB steps following the third reversal. Each track had a minimum of 30 trials and at least 9 reversals (typically many more than 9 were obtained). The threshold estimate was computed after discarding the first three or four reversals (whichever produced an even number) and thus was based on at least the last six reversals. Threshold estimates were averaged over eight adaptive tracks per condition. Trials were blocked by spatial configuration of the maskers. The maskers were either colocated (0°), near ($\pm 15^\circ$), intermediate ($\pm 45^\circ$), or far ($\pm 90^\circ$). The blocks were presented in random order, such that the masker location changed after every two adaptive tracks. To facilitate comparisons across conditions, masked thresholds will all be expressed as T/M 's in decibels. The T/M is calculated as the fixed target level minus the level of the individual maskers at adaptive threshold.

2. “Monaural” control condition

Although there is no acoustically better ear when the masker talkers are symmetric about the target talker based on the long-term rms value, there is the possibility that the head shadowed representations in the two ears are different enough in the spatially separated cases from the colocated case to provide some benefit of spatial separation. For example, in the $\pm 90^\circ$ case because each masker talker is essentially low-pass filtered by the head before it is received by the far ear, the sum of the two maskers will be frequency-dependent and different from what the spectra are for the colocated case. Thus, it is conceivable that the benefit of spatial separation could be due to monaural cues. In addition, there is presumably a complicated pattern of interaural differences that would lead to moments of improved T/M at each ear. To test the possibility that such information could be useful in performing the task, listeners wore commercially available hearing protectors (an earplug and earmuff) on one ear (left) and repeated a subset of the spatial conditions. This will be referred to as the monaural condition even though it is not strictly monaural listening, but a means of reducing binaural cues (and controlling for the potential acoustic benefit of spatial separation). All six listeners were tested with an earplug and earmuff in a subset of spatial configurations (0° and $\pm 90^\circ$). To compare to true monaural listening, one additional listener with an anacusic ear was tested.

The hearing protectors used were disposable E-A-R® plugs and the AOSafety® Economy Earmuff, both manufactured by the Aearo Company. In order to estimate the amount of attenuation obtained, listeners wore earplugs in both ears and both earmuffs while speech threshold estimates were obtained in quiet. If they did not achieve at least 35 dB of attenuation (± 5 dB) relative to their unoccluded speech thresholds, the earplugs were reinserted, the earmuffs were repositioned, and new threshold estimates were obtained. The earplug and earmuff from the right ear were then removed from the headband, which had been modified so that it could be positioned comfortably on the listener's head. The monaural earmuff was held in place by the tightness of the headband.

3. Time-reversed speech maskers

The speech stimuli and procedures used in the current experiment were chosen to emphasize informational masking. Time-reversed speech maskers were also tested to determine the effect of decreasing the amount of informational masking. Time-reversed speech maskers may be considered less effective informational maskers than time-forward speech because of their lack of meaning, but preserve the spectrotemporal complexity of natural speech. This point is discussed in more detail in later sections and in the Appendix. Five of the six listeners were tested with time-reversed speech maskers for a subset of the spatial configurations (masker locations at 0° and $\pm 90^\circ$). The masker sentences were selected from trial to trial in the same manner as for the forward speech but were played backwards. The procedures for these conditions were otherwise identical to those described earlier.

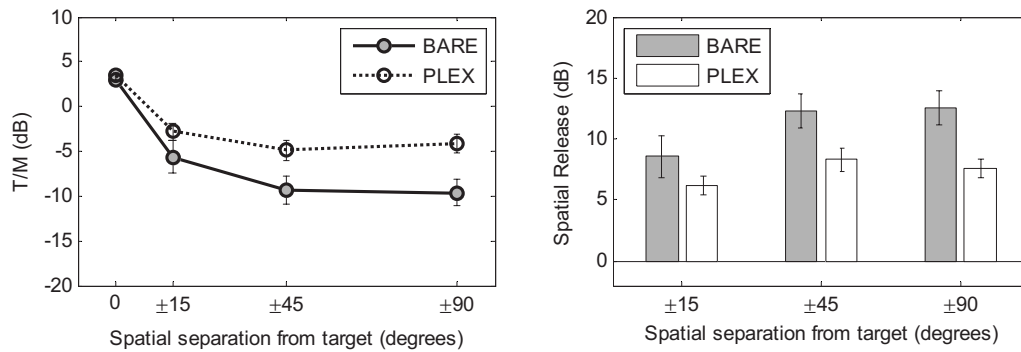


FIG. 1. (Left) Group mean target-to-masker ratios (T/M ; expressed in decibels and computed relative to the level of individual maskers) at masked threshold as a function of the horizontal separation of two masker talkers for the BARE (filled circles) and PLEX (open circles) room conditions. The BARE room is a large IAC booth with 0.06 s reverberation time. The addition of Plexiglas® panels increases the reverberation time by a factor of 4 to 0.25 s in the PLEX room. The error bars are ± 1 standard error of the mean. (Right) Average spatial release from masking (dB), for each listener, the difference in T/M at threshold between collocated and separated conditions, at three masker separations.

III. RESULTS

The unmasked adaptive thresholds (50% correct points) were within two decibels of one another in the two room conditions. The mean threshold in quiet for target speech at 0° azimuth was 13.5 dB SPL (with a standard error, SE, of 1.3 dB) in the room condition with low-reverberation and 15.5 dB SPL (SE=1.2 dB) in the more reverberant room (after correcting for the 3 dB difference in overall SPL).

The left-hand panel of Fig. 1 displays the group mean thresholds (and ± 1 standard error about the mean) for target color and number identification in the presence of two competing talkers as a function of the angular separation between target and maskers. It illustrates the main effects of both spatial separation and reverberation. The abscissa is the degree of separation between the target and maskers in azimuth, from 0° (no separation, or collocated) to $\pm 90^\circ$. The ordinate is T/M at threshold in decibels. There are two functions displayed: one for the results obtained in the low-reverberation room (BARE: filled circles connected by solid lines) and the other for the results obtained in the room with more reverberation (PLEX: open circles connected by dotted lines). A repeated-measures analysis of variance (ANOVA) confirmed that there were significant main effects of spatial separation [$F(3, 15)=70.1$, $p<0.001$] and room reverberation [$F(1, 5)=33.6$, $p=0.002$] on T/M at threshold. Post-hoc analyses (pairwise comparisons) within each room condition indicated that the spatial separations were significantly different from one another ($p<0.05$) with the exception of $\pm 45^\circ$ versus $\pm 90^\circ$. There was also a significant interaction between spatial separation and reverberation [$F(3, 15)=11.5$, $p<0.001$] confirming the result that is apparent from the nonparallel functions plotted in the left-hand panel of Fig. 1.

The highest average T/M 's were found for the collocated case, regardless of room condition. When the target and masker talkers were collocated, there was essentially no difference between the T/M at threshold in the two room conditions (3.0 dB, SE=0.3 dB for BARE and 3.4 dB, SE=0.4 dB for PLEX). Also, thresholds in the collocated condition were consistent across listeners in both rooms as evident from the small error bars.

Overall, thresholds decreased as the amount of spatial separation between the target and maskers increased for both rooms. In the low-reverberation room (BARE), when the maskers were spatially separated from the target by $\pm 15^\circ$, thresholds decreased to -5.7 dB (SE=1.8 dB). Thresholds decreased further with greater spatial separation, but the average results were essentially the same at $\pm 45^\circ$ and $\pm 90^\circ$ [thresholds of -9.3 dB (SE=1.6) and -9.6 dB (SE=1.5), respectively]. The same pattern was true in the more reverberant room (PLEX), although the decrease in thresholds was less pronounced. At $\pm 15^\circ$, thresholds decreased to -2.7 dB (SE=1 dB). For the greater spatial separations, thresholds decreased further and the values were again comparable to one another [-4.9 dB (SE=1.2) at $\pm 45^\circ$ and -4.2 dB (SE=1.0) at $\pm 90^\circ$]. Individual differences were noted in the maximum decrease in threshold with increasing spatial separation, although the overall pattern of results was consistent across all listeners (see Fig. 2) with most of the benefit occurring in the first $\pm 15^\circ$ separation.

A. Spatial release from masking

These results can also be evaluated in terms of the amount of spatial release from masking. For each listener this value was calculated as the T/M at threshold for the collocated condition minus the T/M at threshold in the spatially separated conditions. The results of this computation are displayed in the right-hand panel of Fig. 1. The groups of bars are for the 3 spatial separations while the ordinate is the amount of spatial release from masking in decibels. The values shown are group mean differences and standard errors. In the low-reverberation room (BARE), there was 8.6 dB of spatial release from masking (SE=1.7 dB) when the two masker talkers were presented from $\pm 15^\circ$. There was nearly 4 dB additional benefit of moving the maskers to either of the wider spatial separations [total spatial release of 12.3 dB (SE=1.4) at $\pm 45^\circ$ and 12.6 dB (SE=1.4) at $\pm 90^\circ$]. In the more reverberant room, there was approximately 2 dB less release from masking when the maskers were at $\pm 15^\circ$ [PLEX: 6.2 dB (SE=0.8)]. The average amount of spatial release across listeners in the more reverberant room at $\pm 45^\circ$

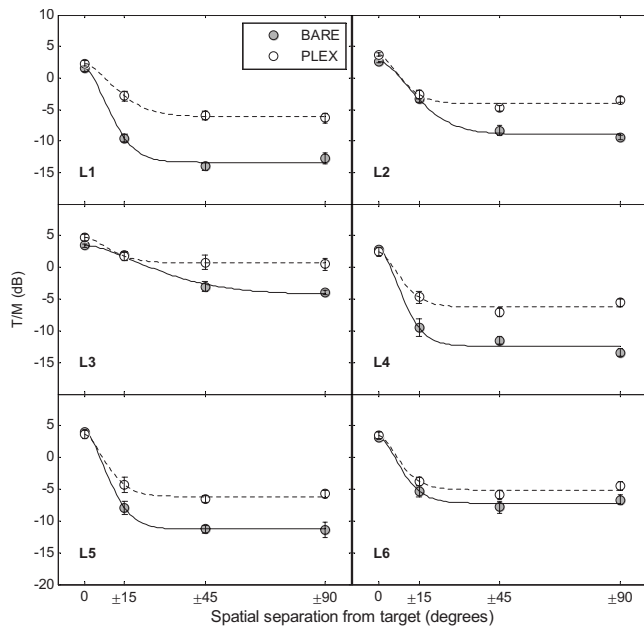


FIG. 2. Individual results plotted for the same conditions as shown in Fig. 1. Error bars are ± 1 standard error of the mean. The lines are best-fitting filter functions (see the text and Table I).

and $\pm 90^\circ$ was still substantial [8.3 dB (SE=0.9) and 7.6 dB (SE=0.7), respectively], although it was less than in the low-reverberation room.

B. Spatial tuning

A useful way of summarizing the relationship between spatial separation of the symmetrically placed maskers and the reduction in masking is to compute best-fitting filter functions based on the masked results. The interpretation of these filter functions as evidence of spatial tuning is considered in the Discussion. The form of the filter applied to the data was the familiar $\text{roex}(p, r)$ filter commonly used to characterize auditory filters in the frequency domain (e.g., Patterson *et al.*, 1982; Glasberg and Moore, 1986). This filter was chosen both for computational convenience and because it provides estimates of not only the bandwidth but also the range of the filter (maximum amount of “attenuation”). How-

ever, this choice was not meant to suggest that it is the “correct” filter shape for the processes involved. The filters were computed on the results from individual listeners—partly to highlight the differences among listeners—and are shown in Fig. 2. The abscissa is the angular separation between target and maskers in degrees, and the ordinate is the T/M at threshold in decibels. Thus, the curves all begin at the point denoting 0° separation and then display an attenuation characteristic as spatial separation increased and thresholds decreased. In order to show the individual listener T/M 's, the curves were *not* normalized to 0 dB attenuation; therefore the attenuation values are equivalent to the amount of spatial release, which is obtained by subtracting the T/M for a spatially separated condition from the T/M for the colocated condition. The two functions and associated data points represent the two room conditions (BARE: filled circles, PLEX: open circles). The bandwidth of the filters (assuming a symmetric filter² computed from the 3 dB down point), the maximum attenuation, or range, of the filter (representing the value r in decibels in the roex filter expression) and the angular separation at which the maximum release was achieved are presented in Table I. The angular separation was defined as the smallest separation that was equal to the maximum release in dB on the fitted function. For 5 of the 6 listeners, the -3 dB bandwidth occurred at less than $\pm 10^\circ$ separation for both rooms' functions with the average value near $\pm 6^\circ$. The remaining listener, L3, appears to be something of an outlier with respect to these measurements. The range of the filter/maximum masking release on the fitted functions varied across listeners from roughly 8 to 15 dB for the low-reverberation room (BARE) and 4–10 dB in the more reverberant room (PLEX). The angular separations at which the maximum release was first reached ranged from 31° to 55° with the exception of the function for Listener 3 in the low-reverberation room, which did not show asymptotic behavior in the range tested.

C. Monaural listening

As mentioned in Sec. II, the amount of attenuation obtained with the monaural hearing protectors was estimated by measuring the speech identification threshold in quiet as

TABLE I. Filter characteristics from best-fitting $\text{roex}(p, r)$ filter functions computed on masked thresholds from individual listener data displayed in Fig. 2.

| Listener | BARE | | | PLEX | | |
|--|------------------------------|----------------------|------------------------------|------------------------------|----------------------|------------------------------|
| | 3 dB down point (\pm deg) | Maximum release (dB) | Maximum release (\pm deg) | 3 dB down point (\pm deg) | Maximum release (dB) | Maximum release (\pm deg) |
| L1 | 5 | 15 | 37 | 9 | 8.5 | 47 |
| L2 | 9 | 11.6 | 55 | 7 | 7.8 | 33 |
| L3 | 22 | 7.7 | 90 | 15 | 4.2 | 37 |
| L4 | 5 | 15.3 | 33 | 6 | 8.8 | 33 |
| L5 | 5 | 15.5 | 35 | 6 | 9.9 | 33 |
| L6 | 6 | 10.4 | 33 | 6 | 8.6 | 31 |
| Intersubject mean (standard deviation) | ± 8.7 (6.7) | 12.6 (3.2) | ± 47.2 (22.6) | ± 8.2 (3.5) | 8.0 (3.5) | ± 35.7 (5.9) |

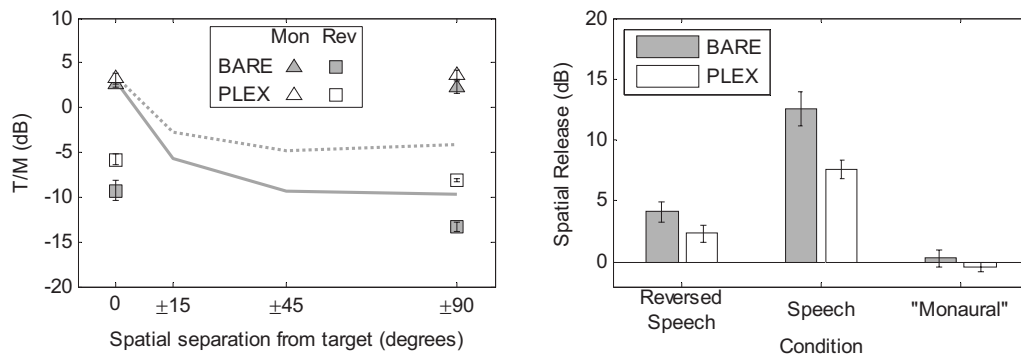


FIG. 3. (Left) Mean T/M at threshold at 0° and $\pm 90^\circ$ in both rooms for the monaural condition (“Mon,” where one ear was occluded with an earplug and muff) and for the condition where participants listened binaurally but the speech maskers were time-reversed (“Rev”). Error bars represent ± 1 standard error of the mean. The group mean T/M ’s from Fig. 1 (binaural listening, forward speech maskers) are replotted for comparison (BARE: solid line, PLEX: dashed line). (Right) Mean spatial release from masking (dB), calculated on an individual listener basis and then averaged across listeners, for the conditions plotted in the left-hand panel.

the listener wore hearing protectors on both ears and comparing that value to the one measured in the unoccluded case. Averaged across all listeners and sessions, the amount of attenuation achieved was 38.1 dB (SE=1.7 dB), relative to each listener’s unoccluded speech threshold.

The left-hand panel of Fig. 3 displays the T/M ’s at threshold in the monaural condition for the colocated and $\pm 90^\circ$ separation conditions in the two room conditions (BARE: filled triangles, PLEX: open triangles). Also shown for reference are the mean data curves from Fig. 1 for binaural (unoccluded) listening for the same spatial conditions (a solid gray line for the BARE room result and a dashed gray line for the PLEX room result). The average threshold for speech identification with colocated target and maskers was essentially the same regardless of whether the listeners were performing the task monaurally (listening with the earplug and earmuff) or binaurally (unoccluded) for both room conditions. The average T/M at 0° for the monaural listening condition in the low-reverberation room was 2.5 dB (SE=0.2 dB) and was 3.3 dB (SE=0.6) in the more reverberant room; both T/M ’s were within one decibel of the results for binaural listening.

The most striking (although not unexpected) finding in the monaural data was that wearing the hearing protectors on one ear nearly eliminated the benefit of spatial separation. When the maskers were presented at $\pm 90^\circ$, the average T/M ’s for the masked speech thresholds were approximately equivalent to the colocated thresholds in both room conditions. In the low-reverberation room, the average T/M at $\pm 90^\circ$ was 2.2 dB (SE=0.7 dB) and in the more reverberant room, it was 3.7 dB (SE=0.5 dB). For both room conditions, the average monaural T/M ’s at threshold in the colocated and spatially separated conditions were not statistically different from the average T/M for the binaural colocated thresholds in a repeated-measures ANOVA [BARE: $F(2,4) = 1.3$, $p=0.36$; PLEX: $F(2,4)=0.410$, $p=0.69$]. In addition, the one truly monaural listener (complete unilateral deafness following vestibular schwannoma removal) who was tested showed no difference in the T/M at threshold for the colocated versus separated conditions in both rooms. The severe reduction in spatial release from masking in the monaural condition is shown clearly in the right-hand panel of Fig. 3,

which displays spatial release for the binaural listening condition in the center pair of bars and the monaural listening condition in the right-most pair of bars for both room conditions. By inference, any acoustic differences between the colocated and separated conditions that were present monaurally were insufficient to explain the spatial release from masking found in the binaural condition.

D. Effect of masker type

Figure 3 (left-hand panel) also shows the group means and standard errors for the colocated and $\pm 90^\circ$ spatially separated conditions using time-reversed speech maskers (squares). The corresponding time-forward mean results are again provided for comparison (BARE: solid gray line; PLEX: dashed gray line). In the colocated condition, the average T/M at threshold was lower with the time-reversed speech maskers than when the maskers were intelligible in both room conditions. In the low-reverberation room, mean threshold T/M at 0° was -9.3 dB (SE=1.2 dB), which was 12.3 dB lower than with the forward-speech maskers. A similar result was found in the more reverberant room, where group-mean threshold at 0° was -5.8 dB (SE=0.6 dB). This was 9.2 dB lower than the corresponding threshold with forward-speech maskers. When the maskers were presented at $\pm 90^\circ$ there was a modest benefit of spatial separation in both room conditions. In the low-reverberation room, mean threshold at $\pm 90^\circ$ was -13.4 dB (SE=0.5 dB), corresponding to 4.1 dB spatial release. In the more reverberant room, there was 2.3 dB spatial release, as mean threshold at $\pm 90^\circ$ was -8.1 dB (SE=0.1 dB). As compared to the forward-speech maskers, all of the threshold T/M ’s were lower for the reversed speech maskers, but the greater reductions occurred for the colocated condition. Further, the amount of reduction in T/M was similar for both room conditions. This effect reduced the mean spatial release from masking observed with the reversed speech maskers, although the benefit remained statistically significant. Spatial release for the reversed speech condition is plotted in the left-most pair of bars in the right-hand panel of Fig. 3. A two-factor repeated-measures ANOVA on T/M at threshold for the reversed speech maskers with room and spatial condition as between-

subject factors confirmed there were significant main effects for room [$F(1,4)=63.3$, $p=0.001$] and spatial conditions [$F(1,4)=19.2$, $p=0.01$]. The interaction between room and spatial condition was also significant [$F(1,4)=11.7$, $p=0.03$].

IV. DISCUSSION

The current findings support the notion of a tuned response in azimuth resulting in a filter-like pattern of masking results. This conclusion is based on the progressive decrease in T/M at masked threshold with increasing spatial separation of symmetrically placed speech maskers over a narrow range of azimuths. The magnitude of the effect varied across listeners but, on average, was more than 12 dB in the low-reverberation room. This effect appears to be mediated by a variety of aspects of the listening environment, stimuli, and task. However, this benefit did not appear to increase in proportion to increasing differences in target-masker azimuth across the range of values tested (see Fig. 1). Instead, most of this effect was obtained within the first 15° of spatial separation of sources with the full benefit for most listeners realized by about 45°. The -3 dB bandwidths from the fitted filter functions are thus quite narrow with the average value less than $\pm 10^\circ$. The bandwidths computed here are similar to the narrow bandwidths found by Teder-Salejarvi *et al.* (1999).

The tuning primarily reflects, we believe, a reduction in informational masking as the target and masker are spatially separated. Although it is difficult to determine the relative amounts of energetic and informational masking that are present in a speech-on-speech masking task, there are several factors that suggest that this is the case. The large release from masking in the collocated condition when the masker speech is reversed is consistent with a reduction in informational masking. However, this technique only provides a rough approximation of the amount of informational masking that occurs and is based on the assumption that energetic masking is equivalent when speech maskers are played backwards or forwards—an assumption that may not hold in some cases (e.g., Rhebergen *et al.*, 2005). Unlike broadband noise, reversed speech has similar temporal fluctuations to normal speech, making it intermediate to noise and speech along a hypothetical continuum of target-masker similarity. Further, there is the possibility that the similar and highly structured nature of target and masker sentences used in the current study differentially affects the amount of energetic masking present in time-forward and reversed maskers. In order to examine this possibility, a control experiment using speech-shaped speech-envelope-modulated noise that took its modulation pattern from forward and time-reversed speech was conducted with four additional listeners. This experiment is described in the Appendix. The conclusion drawn from those data is that the reduction in masking for the time-reversed maskers cannot be attributed to a decrease in energetic masking but must be due to a decrease in informational masking. Further, the spatial release found in both conditions was much less than was found here for speech maskers and was in good agreement with past results in which noise maskers were presented from symmetric loca-

tions around a speech target (e.g., Bronkhorst and Plomp, 1992; Helfer, 1992; Peissing and Kollmeier, 1997; Noble and Perrett, 2002). There is also other work, notably that of Brungart *et al.* (2006), that has suggested that the majority of masking that occurs in the coordinate response measure task employing speech maskers is informational in nature.

The type of masking present in the collocated condition of this study has implications for the mechanisms responsible for spatial release. Clearly, the basis for the reduction in thresholds in the spatially separated conditions is the presence of interaural differences. This conclusion is supported by the results of the monaural condition in which no significant spatial release was found. In the symmetric masker configuration, there was no overall acoustical better ear advantage. Measurements made on a Knowles Electronics Manikin for Acoustic Research showed that the overall level (root mean square computed over entire sentence) of either of the maskers alone or target plus maskers differed by less than 1 dB across ears from all of the different spatial configurations tested in the experiment. However, this does not rule out the possibility that there could be short-term fluctuations in level or overall spectral differences due to head shadow effects, resulting in epochs of better T/M when the maskers are spatially separated from the target. Unlike the asymmetry leading to a “better ear” typical with spatial separation of a target and a single masker, the two ears receive roughly equal fluctuations of T/M but either ear may be “better” in the spatially separate case as compared to the collocated case³ at certain moments in time. If this could explain some of the large spatial release from masking observed, then listeners should exhibit some spatial release when listening monaurally. However, the benefit of spatial separation was eliminated in the condition simulating monaural listening. As a further control, we tested a subset of the listeners while both ears were covered with hearing protectors. After increasing the target level to compensate for the attenuation (i.e., presented at the same sensation level), the spatial release was restored. This assured that the loss of normal pinna cues was not the reason for the difference. Thus, the advantages found here clearly depend on binaural listening.

Previous findings using speech maskers placed symmetrically around a speech target are most relevant to the current results. Noble and Perrett (2002) studied the spatial release from masking for a target speech source masked by symmetrically placed speech, speechlike, or noise maskers. In a soundfield experiment with maskers placed at $\pm 30^\circ$, they found about a 4 dB spatial release for two speech maskers. This was smaller than the spatial release found in the current study for $\pm 15^\circ$ of separation. They found even smaller releases when the maskers were noise. In one experiment, they also measured performance for speech masker placements of $\pm 90^\circ$ and found only about a 1 dB further increase (relative to $\pm 30^\circ$) in spatial release. Given the filter widths computed here (attenuation maxima at about 38° excluding Listener 3 who had atypically large bandwidths), their results appear to be consistent with ours except that the magnitude of their spatial release was substantially less. It seems likely that their smaller spatial release from speech

masking is a consequence of their stimuli and tasks producing smaller amounts of informational masking than those in the present study.

Spatial release from masking in single noise-masker conditions increases with increasing target-masker separation over a range of values (e.g., Plomp, 1976; Plomp and Mimpfen, 1981), a result that is consistent with the predictions of Zurek's (1993) model. Also, a speech target masked by a single speech masker follows a similar pattern of spatial release. Asymmetric, single masker placement could result in a function like those shown in Fig. 2 for either type of masker as it is progressively separated from a speech target. However, although both noise-masking-speech and speech-masking-speech produce declining threshold functions in the asymmetric placement condition, the fact that the two differ in the symmetric placement condition is, we believe, quite significant and is consistent with the conclusion that the underlying mechanisms are different as well. This interpretation is in agreement with Noble and Perrett's conclusions despite the discrepancy in the size of the effect. Additionally, the recent work of Brungart *et al.* (2006) suggesting that energetic masking is not a major factor for speech materials and procedures similar to those used here calls into question any strong role of binaural analysis (i.e., within-channel improvements in T/M , or "masking level differences"). Bronkhorst (2000) has proposed a straight-forward approach to the prediction of spatial release for a speech target presented from the front in the multiple masker case. Based on his equation (Bronkhorst, 2000, p. 123) and the parameters he estimated by fitting data from several studies, the prediction for the amount of spatial release for two maskers placed symmetrically at $\pm 90^\circ$ is about 2 dB. That value is consistent with empirical reports of spatial release from noise maskers (e.g., Noble and Perrett, 2002), although it is somewhat smaller than the 4.6 dB effect found by Bronkhorst and Plomp (1992) for a modulated noise masker. For the procedures and target stimuli used in the current study, we found a value of about 1.5 dB in the low-reverberation room when measured empirically using broadband noise maskers⁴ and about 3.5 dB for either time-forward or -reversed speech-shaped speech-envelope-modulated noise (refer to the Appendix). Thus, when the primary limitation on performance is energetic masking, the conditions tested in this study yield comparatively small spatial advantages. It appears to be possible that the presence of a high degree of informational masking is necessary (but not necessarily sufficient) to observe the large and sharply tuned effects found here. It is also clear that models of spatial release that only take into account better-ear listening and binaural analysis cannot predict these large effects.

The interpretation of the current results is that the greatest amount of informational masking was present in the colocated condition for the forward-speech maskers. Either spatial separation of the stimuli or time reversing the maskers caused a large reduction in the informational masking. Once the informational masking was reduced by means of decreasing target-masker similarity by time-reversing the masker, further reductions due to spatial separation were minimal and possibly indicate that performance had reached a limit im-

posed by the remaining energetic masking. The conclusion then is that the amount of informational masking in a task influences the degree of spatial benefit observed after eliminating the better-ear advantage. This interpretation is consistent with the relatively small amounts of spatial release from masking observed in previous studies that used various energetic maskers or presented the stimuli in low-uncertainty conditions or provided other strong perceptual segregation cues (cf. Kidd *et al.*, 1998; Arbogast *et al.*, 2002; Noble and Perrett, 2002; Hawley *et al.*, 2004; Culling *et al.*, 2004; Best *et al.*, 2005).

One of the factors influencing the magnitude of spatial release observed was the amount of reverberation in the listening environment. There was less spatial release from masking in the more reverberant room (the PLEX condition). Although the T/M 's at threshold were stable across room conditions for the colocated target and maskers, the T/M 's at threshold were higher in the more reverberant room when the maskers were spatially separated.

Increasing reverberation did not affect the colocated thresholds in the current three-talker experiment, but did in the two-talker experiments of Plomp (1976), Culling *et al.* (2003), and Kidd *et al.* (2005a). This might be explained by several factors, including differences between studies in the stimuli and procedures used and in the amount of reverberation present. In the two studies using rooms or simulations with longer reverberation times than those of the current experiment (Plomp, 1976; Culling *et al.*, 2003), increasing reverberation increased T/M by 2–4 dB for the colocated condition. However, in both studies the T/M 's in the colocated conditions were generally lower (better) than those found here, likely because there was only a single masker talker. Importantly, the effect of increasing reverberation on the benefit of spatial separation found here is somewhat different than that reported by Kidd *et al.* (2005a) in the same room conditions for a single masker either colocated with the target at 0° or spatially separated at 90° . In that study, when the masker was another speech signal (targets and maskers were comprised of multiple narrow and mutually exclusive frequency bands), a large release from masking was preserved in the more reverberant room (16.0 dB in PLEX versus 16.7 dB in BARE). The main difference between the results of the two studies is that the group mean T/M 's at threshold in the colocated condition were more variable and lower in the earlier study. As noted previously, the colocated thresholds found here for two speech maskers were remarkably constant across listeners and room conditions including the monaural control. In the earlier study, the range of performance across listeners in the colocated case for a single masker was much larger. The lower mean thresholds in the earlier study, combined with the larger intersubject differences, suggest that the source segregation cues were different. Although we do not know for certain, it seems possible that thresholds in the colocated, two-talker masker used here are determined simply by relative level. When three talkers of the same sex uttering similar sentences are colocated, there may be insufficient cues (e.g., F0 differences, timbre differences, etc.) to segregate the target until it is the loudest of the three voices. The values of T/M at

threshold were around 2–3 dB meaning that the target was that much higher in level than either masker alone or approximately equal to the sum of the maskers. These values are comparable to performance in a study by Brungart *et al.* (2001) for monaural segregation of three same-sex talkers using the same stimuli. In the Kidd *et al.* (2005a) study, other segregation cues—notably timbre differences resulting from the narrowband processing of the speech targets and maskers (see also Arbogast *et al.*, 2002)—provided another means for segregating the sounds. It is possible that the timbre cue was disrupted by reverberation, causing thresholds to increase in the collocated condition about as much as in the spatially separated condition. Future work is necessary to determine whether longer reverberation times than those used here would disrupt the relative level segregation cue and further increase the collocated thresholds for two or more masker talkers.

A related issue is that the large individual differences in spatial release found here are almost entirely attributable to differences in performance in the spatially separated conditions. The level cue that we speculate forms the basis for segregation in the collocated case seems to be one that most listeners were able to use quite effectively and is robust with respect to this amount of reverberation. However, in the spatially separated conditions, the ability to use interaural differences to segregate and selectively attend to the target—and ignore the maskers—varied widely across listeners. The large individual differences found in the spatially separated condition are not unlike the large individual differences found in the ability to use various segregation cues to overcome informational masking (e.g., Neff and Dethlefs, 1995; Durlach *et al.*, 2003; Richards *et al.*, 2002).

Threshold T/M increased in the spatially separated condition with increasing reverberation. This finding supports earlier work by Helfer (1992) using nonsense syllables masked by symmetrically positioned “cafeteria noise” and by Kidd *et al.* (2005a) and Culling *et al.* (2003) for asymmetric speech maskers. However, the conclusion that can be drawn from the current data differs from that of Culling *et al.* (2003) for several reasons. They observed that the benefit of spatial separation was eliminated with increasing reverberation. Several differences in the design of their experiment and the current experiment could help to explain this apparent discrepancy. First, because they were interested in the interaction of F0 contours (intonated, monotonous, or inverted) with spatial separation and reverberation, the task they used consisted of identifying key words from a male talker in the presence of a single female talker. This potentially involved less informational masking than in the present study as a strong segregation cue was present in all conditions. Their thresholds for the collocated case were better and the release from masking with spatial separation in the simulated anechoic space was smaller than in the current results. In addition, the elimination of the effect in reverberation, as well as the overall increase in thresholds in reverberation, may have been due to the greater amount of reverberation present in their experiment.

Considered across these various studies, reverberation clearly adversely affects performance in multitalker situa-

tions. Culling *et al.* (2003), Kidd *et al.* (2005a) and the current results have all demonstrated that the T/M at threshold increases with increasing reverberation when the competing talkers are spatially separated from the target. The most likely reason for this effect, we believe, is that the temporal “smearing” caused by reverberation reduces the interaural time and level differences that normally help the listener segregate the different sound sources and permit the listener to focus attention on the target. Performance in the absence of spatial cues appears to depend on the availability of other cues (e.g., overall level differences) for segregating the target from the collocated maskers that are less sensitive to the temporal smearing caused by this amount of increased reverberation. Therefore, whether spatial release per se is reduced, and, if so, to what extent, appears to depend on the specific conditions tested in the experiment.

V. SUMMARY AND CONCLUSIONS

The motivation for the current study was to better understand the processes that allow listeners to selectively attend to one source while ignoring competing sources in realistic room conditions. Listeners gave a selective response to a talker located straight ahead in the presence of two collocated or symmetrically positioned interfering talkers. Overall, filter-like properties were observed in the pattern of responses and it seems likely that this effect depends on the specifics of the stimuli and task. The spatial advantage observed appears to be a consequence of the listener using interaural differences to improve perceptual segregation of the target from the maskers and to focus attention at a point in space in order to overcome informational masking.

Spatial release from masking increased in these experiments as a function of increasing target–masker separation in azimuth from 0° to $\pm 45^\circ$ with negligible improvement for increasing spatial separation further to $\pm 90^\circ$. Most of this release occurred for the initial separation of $\pm 15^\circ$, suggesting that even small spatial separations provide large perceptual benefits. The spatial filters that were computed from the results were quite narrow (generally less than $\pm 10^\circ$ at the 3 dB down points) and maximum values of attenuation were mostly achieved in the range of $\pm 30^\circ$ to $\pm 45^\circ$. A relatively large amount of spatial release was observed in this task, in excess of 12 dB for the larger separations, when listening binaurally in the low-reverberation room. When participants listened with one ear occluded by an earplug and earmuff to simulate monaural listening, spatial release was nearly eliminated; the average T/M 's in the collocated and spatially separated conditions were not statistically different from the average T/M for the binaural collocated condition. The control conditions of time-reversed speech maskers, modulated noise maskers, and Gaussian noise maskers all produced much less spatial release than was obtained for the time-forward speech maskers. The large effects produced when informational masking was emphasized allow us to highlight the importance of the perceptual aspects of a process that has traditionally been thought of as explainable by fairly low level processes as in the model proposed by Zurek (1993), to account for spatial release from masking for speech in noise.

When room reverberation was increased, a fairly large spatial release from speech-on-speech masking (over 8 dB) was still present. As in the low-reverberation room, most of the spatial release occurred for the initial $\pm 15^\circ$ separation. In the more reverberant room, listeners needed a more favorable T/M to identify the target when the talkers were spatially separated, but the same T/M when the talkers were colocated. Thus, increasing reverberation reduced the spatial advantage. We speculate that the segregation cues in the colocated condition resulted from stimulus level differences that were unchanged by increasing reverberation. In the spatially separated conditions, the segregation cues were likely differences in perceived location based on interaural timing and level differences that were disrupted by the increase in reverberation.

Time reversing the maskers resulted in a greater change in performance for the colocated condition than in the spatially separated condition and this reduced the spatial release from masking. The reduction in similarity between target and masker provided a large segregation benefit. When the talkers were spatially separated, there was less improvement due to time-reversing the maskers possibly because of the large improvement already achieved through spatial separation. This suggests that informational masking is greater when there are fewer segregation cues.

ACKNOWLEDGMENTS

This work was supported by Grant No. FA9950-05-1-2005 from the Air Force Office of Scientific Research and by Grant Nos. DC008440, DC00100, DC04545 and DC04663 from the National Institute on Deafness and other Communication Disorders (NIDCD). This work partly fulfills the requirements for the doctor of philosophy degree at Boston University for the first author. The authors wish to thank the listeners that participated in the experiment, as well as Steven Colburn, Melanie Matthies, Barbara Shinn-Cunningham, Frederick Gallun, Virginia Best, and Nathaniel Durlach for helpful discussions of the results. Deborah Corliss and Jacqueline Therieau provided assistance with the Plexiglas® in the Sound Field Lab. They also thank Richard Freyman and two anonymous reviewers for useful comments on earlier versions of this manuscript.

APPENDIX

A potential explanation advanced for the reversed speech result (reduced T/M at threshold especially for the colocated condition) is that the forward speech produced more energetic masking than the reversed speech because of the coherence of the target and masker envelopes. Because all of the sentences—both targets and maskers—have the same structure and are to some degree time aligned, the envelopes of the targets and maskers are positively correlated (near 0.5 by our estimates). Reversed speech, however, diminishes the correlation between target and masker (near zero or slightly negative). If the peaks in the envelopes are more highly correlated in the time-forward case, then it is possible that the spectral overlap is also greater, whereas a lower envelope correlation with time-reversed maskers could

provide a better opportunity for extracting target information in envelope minima. Because this alternative explanation for the reversed-speech result affects the extent to which we can attribute the colocated findings to informational masking, an additional experiment was conducted to examine this possibility in more detail. A new group of four listeners was recruited and tested using the envelopes of the forward and time-reversed speech maskers to modulate speech-shaped noise (derived from the corpus of speech used in the experiment). Following some preliminary analysis and listening experience, the envelope low-pass cutoff was set at 10 Hz. Speech-shaped speech-modulated noise was used because it is generally considered to produce primarily energetic masking of speech with little concomitant informational masking. Because the envelopes are derived from forward versus reversed speech, the extent to which the target and masker envelopes overlap is about the same as for actual speech. The test procedures used were identical to those employed in the main experiment with samples of the time-forward or -reversed noises replacing the speech maskers. If the difference between forward and reversed speech maskers was due to greater energetic masking in the forward speech maskers, or to a better opportunity to extract target information in the masker envelope minima for the reversed speech, then there should be less masking obtained for the reversed speech-modulated noise masker than for the forward speech-modulated noise masker.

The results indicated no significant difference between forward and reversed noise at either spatial condition: colocated and $\pm 90^\circ$. For the colocated condition, the group mean T/M 's were about -6.5 dB, whereas the corresponding values for the spatially separated condition were about -10 dB. These values, including the approximately 3.5 dB spatial release from masking, are in good agreement with similar values reported by Bronkhorst and Plomp (1992).

Our interpretation of the results of this control experiment is that, despite the greater temporal overlap of the target and masker envelopes in the time-forward condition than in the time-reversed condition, the amount of energetic masking was about the same. Therefore, the previous conclusions regarding the large improvements in T/M at threshold for reversed speech maskers compared to forward speech maskers being attributable to a release from informational masking appears to be supported by these results. It seems likely that the generally sparse spectral overlap of the CRM materials, as reported by Brungart *et al.* (2006), causes only small amounts of energetic masking even when the target and masker envelopes are somewhat coherent. The generalization of this finding to other speech intelligibility results should be made with caution, if at all, because of the closed-set highly structured nature of the CRM test and stimuli.

¹What happens to the information from the “poorer ear” when considering the advantage of better ear listening is not always stated explicitly. If the two ears are considered separate channels, then the best strategy should be to optimally combine information from each. So, the poorer ear would contribute to the overall percept. If combining the inputs to the two channels results in a single representation or image that is, in some sense, noisier than the better input, then summation would be disadvantageous.

However, if the listener is able to select only a single channel, then obviously attending to the better ear is advantageous.

²Although we do not have any direct evidence regarding the (a) symmetry of the filter, there is some evidence from closely related conditions (Kidd *et al.*, 2005b) suggesting a preference for attending to stimuli presented from the right-hand side in highly uncertain listening conditions. Whether this would affect filter symmetry is not known and because of the symmetric placement of the maskers, there is not a way to evaluate filter asymmetry in the current design.

³Apart from the complex spectrotemporal patterns that may be different for the colocated and spatially separated conditions, imagine this extremely over simplified example to illustrate the point. The target presented from 0° arrives roughly equally at the two ears however, each of the two maskers when presented from ±90° are essentially low-pass filtered by the head when received in the opposite ears. Therefore, the high frequency masker energy from one of the two sources is reduced in each ear relative to the case of both maskers arriving equally at the two ears when colocated with the target. Whether or not these effects should actually aid in identifying the target is not obvious but the results of the monaural control condition would appear to suggest that they do not.

⁴It has been demonstrated previously by Bronkhorst and Plomp (1992) and Noble and Perrett (2002) that the spatial release for two symmetrically placed independent Gaussian noise maskers is quite small—on the order of one or two decibels. We double checked this finding for the stimuli and procedures used in the current experiment by replacing the speech maskers with two independent broadband noises. The average spatial release from masking for a group of four listeners (two who participated in the speech masking experiments and two new listeners) for the symmetrically placed noises was essentially the same in both room reverberation conditions (BARE: 1.5 dB and PLEX: 0.9 dB). This is consistent with the previous results referenced earlier.

American National Standards Institute. (2004). American national standard specification for audiometers. ANSI 3.6–2004

Arbogast, T. L., and Kidd, G., Jr. (2000). "Evidence for spatial tuning in informational masking using the probe-signal method," *J. Acoust. Soc. Am.* **108**, 1803–1810.

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.

Best, V., Gallun, F. J., Ihlefeld, A., and Shinn-Cunningham, B. G. (2006). "The influence of spatial separation on divided listening," *J. Acoust. Soc. Am.* **120**, 1506–1516.

Best, V., Ozmeral, E., Gallun, F. J., Sen, K., and Shinn-Cunningham, B. G. (2005). "Spatial unmasking of birdsong in human listeners: Energetic and informational factors," *J. Acoust. Soc. Am.* **118**, 3766–3773.

Boehnke, S. E., and Phillips, D. P. (1999). "Azimuthal tuning of human perceptual channels for sound location," *J. Acoust. Soc. Am.* **106**, 1948–1955.

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. (2000). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acust. Acta Acust.* **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., Chang, P. S., Simpson, B. D., and Wang, D. (2006). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," *J. Acoust. Soc. Am.* **120**, 4007–4018.

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

Carlile, S., Hyams, S., and Delaney, S. (2001). "Systematic distortions of auditory space perception following prolonged exposure to broadband noise," *J. Acoust. Soc. Am.* **110**, 416–424.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (2004). "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**, 1057–1065.

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

Drennan, W. R., Gatehouse, S., and Lever, C. (2003). "Perceptual segregation of competing speech sounds: The role of spatial location," *J. Acoust. Soc. Am.* **114**, 2178–2189.

Drennan, W. R., Won, J. H., Dasika, V. K., and Rubenstein, J. T. (2007). "Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise," *J. Assoc. Res. Otolaryngol.* **8**, 373–383.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (2003). "Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," *J. Acoust. Soc. Am.* **114**, 368–379.

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.

Glasberg, B. R., and Moore, B. C. J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," *J. Acoust. Soc. Am.* **79**, 1020–1033.

Goldberg, J. M., and Brown, P. B. (1968). "Responses of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Physiological mechanisms of sound localization," *J. Neurophysiol.* **32**, 613–636.

Greenberg, G. Z., and Larkin, W. D. (1968). "Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method," *J. Acoust. Soc. Am.* **44**, 1513–1523.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Am.* **115**, 833–843.

Helfer, K. S. (1992). "Aging and the binaural advantage in reverberation and noise," *J. Speech Hear. Res.* **35**, 1394–1401.

Hill, N. I., Bailey, P. J., and Hodgson, P. (1998). "A probe-signal study of auditory discrimination of complex sounds," *J. Acoust. Soc. Am.* **102**, 2291–2296.

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

Kidd, G., Jr., Mason, C. R., Brughera, A., and Hartmann, W. M. (2005a). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," *Acust. Acta Acust.* **91**, 526–536.

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.

King, A. J., Bajo, V. M., Bizley, J. K., Campbell, R. A. A., Nodal, F. R., Schulz, J., and Schnupp, J. W. H. (2007). "Physiological and behavioral studies of spatial coding in the auditory cortex," *Hear. Res.* **229**, 106–115.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.

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

Middlebrooks, J. C., and Pettigrew, J. D. (1981). "Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location," *J. Neurosci.* **1**, 107–120.

Neff, D. L., and Dethlefs, T. M. (1995). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," *J. Acoust. Soc. Am.* **98**, 125–134.

Noble, W., and Perrett, S. (2002). "Hearing speech against spatially separate competing speech versus competing noise," *Percept. Psychophys.* **64**, 1325–1336.

Oxenham, A. (2000). "Influence of spatial and temporal coding on gap detection," *J. Acoust. Soc. Am.* **107**, 2215–2223.

Patterson R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.* **72**, 1788–1803.

Peissig, J., and Kollmeier, B. (1997). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Am.* **110**, 1660–1670.

Plomp, R. (1976). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," *Acustica* **34**, 200–211.

Plomp, R., and Mimpen, A. M. (1981). "Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences," *Acustica* **48**, 325–328.

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.
- Richards, V. M., Tang, Z., and Kidd, G., Jr. (2002). "Informational masking with small set sizes," *J. Acoust. Soc. Am.* **111**, 1359–1366.
- Scharf, B. (1998). "Auditory attention: The psychoacoustical approach," in *Attention*, edited by H. Pashler (Psychology Press Ltd., Hove, East Sussex) pp. 75–117.
- Shaw, E. A. G. (1974). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *J. Acoust. Soc. Am.* **58**, 1848–1861.
- Stecker, G. C., Mickey, B. J., Macpherson, E. A., and Middlebrooks, J. C. (2003). "Spatial sensitivity in field PAF of cat auditory cortex," *J. Neurophysiol.* **89**, 2889–2903.
- Sterbing, S. J., Hartung, K., and Hoffmann, K. P. (2003). "Spatial tuning to virtual sounds in the inferior colliculus of the guinea pig," *J. Neurophysiol.* **90**, 2648–2659.
- Teder-Salejarvi, W. A., and Hillyard, S. A. (1998). "The gradient of spatial auditory attention in free-field. An event-related potential study," *Percept. Psychophys.* **60**, 1228–1242.
- Teder-Salejarvi, W. A., Hillyard, S. A., Roder, B., and Neville, H. J. (1999). "Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials," *Brain Res. Cognit. Brain Res.* **8**, 213–227.
- Tsuchitani, C. (1988). "The inhibition of cat lateral superior olivary unit excitatory responses to binaural tone bursts. I. The transient chopper discharges," *J. Neurophysiol.* **59**, 164–183.
- Wright, B. A., and Dai, H. (1994). "Detection of unexpected tones with short and long durations," *J. Acoust. Soc. Am.* **95**, 931–938.
- Wright, B. A., and Dai, H. (1998). "Detection of sinusoidal amplitude modulation at unexpected rates," *J. Acoust. Soc. Am.* **104**, 2991–2997.
- Yin, T. C., and Chan, J. C. (1990). "Interaural time sensitivity in medial superior olive of cat," *J. Neurophysiol.* **64**, 465–488.
- Yin, T. C., and Kuwada, S. (1983). "Binaural interaction in low frequency neurons in the inferior colliculus of the cat. III. Effects of changing frequency," *J. Neurophysiol.* **50**, 1020–1042.
- Zurek, P. M. (1993). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, Needham Heights, MA), 255–276.