ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 91 (2005) 967 – 979

Bottom-up and Top-down Influences on Spatial Unmasking

Barbara G. Shinn-Cunningham, Antje Ihlefeld, Satyavarta, Eric Larson
Boston University Hearing Research Center, Departments of Cognitive and Neural Systems and Biomedical Engineering, 677 Beacon St., Boston, Massachusetts 02215

Summary
The ability to detect and understand a source of interest (a "target") in the presence of a competing source (a "masker") is better when the sources are spatially separated than when they are at the same location, an effect known as "spatial unmasking". Many models account for spatial unmasking by predicting reduction of within-frequency-band masking; however, recent studies report significant spatial unmasking even when within-band "energetic masking" is minimal. The current study examines whether spatial unmasking depends on the veracity and kinds of spatial cues present when the target and masker are similar and processed to have little spectral overlap. For the tested stimuli, traditional within-band models predict (at most) a modest amount of spatial unmasking that varies with condition. Instead, we observe large spatial unmasking effects. Moreover, after accounting for the broadband target and masker intensities at the acoustically better ear, the amount of spatial unmasking is essentially independent of the kind of spatial cues that cause the target and masker the be perceived at different locations. Only at the lowest target-to-masker ratios (when within-band masking becomes significant) does the amount of spatial unmasking depend on the interaural phase differences in target and masker. These results emphasize that the relative overall intensities of the masker and target are critical for predicting how much perceptual interference the masker causes, even when "energetic masking" is minimal. We believe that in everyday settings, both traditional bottom-up factors and a higher- level mechanism depending on spatial perception contribute to spatial unmasking. We argue that the latter mechanism is a form of spatial attention, critical for meditating competition between similar, simultaneous sources in everyday settings.

PACS no. 43.66.Pn, 43.66.Pn, 43.55.Br

1. Introduction

There are many studies of "spatial unmasking" that show that a listener's ability to detect and/or understand a sound source of interest (a "target") improves when it comes from a different direction than a competing source (a "masker"). Spatial unmasking has been demonstrated in both detection tasks (e.g., see [1]) as well as in supra-threshold tasks such as comprehending speech (e.g., see [2]).

In traditional models of spatial unmasking, the masking source is assumed to reduce the audibility of the target signal within each critical frequency band. Two distinct spatial factors can influence the audibility of the target (e.g., see [3, 4]): 1) energy effects, whereby the relative energy of the target and masker at the ears changes with target and masker location, altering target audibility in each frequency band; and 2) binaural processing, which allows listeners to detect the presence of target energy in a particular band if the target and masker contain different interaural time and / or level differences.

Received 8 December 2004, Revised 23 June 2005, accepted 26 September 2005.

If a target is directly in front of the listener and a masker to the right of the listener, the target-to-masker energy ratio (TMR) in each frequency will generally be larger at the left ear than the right ear due to head-shadow effects. Listening only to the left-ear signal would yield better performance in this condition than when sources are co-located. This "better ear" acoustic effect can account for much of the spatial unmasking that is observed in many studies (particularly for stimuli that have frequencies above 2 kHz, where head-shadow effects play a large role; e.g., see [5]).

Binaural processing assumed to improve performance by allowing a listener to detect target energy that would be masked otherwise. In conditions where the TMR within a critical band is low, the masker signal dominates the total signal reaching the ears of the listener. At very low within-band TMRs, the pattern of activity in neurons sensitive to interaural time differences (ITDs) will also be dominated by the interaural time difference in the masker. However, at slightly higher TMRs, before the monaural response at either ear is noticeably altered by the presence of a target with a different ITD than the masker, the target alters the neural response of ITD-sensitive neurons. In particular, the relatively quiet target decreases the overall firing rate of brainstem neurons that would respond well to the masker-

alone stimulus, because the target causes interaural decorrelation. This binaural mechanism appears to operate separately in each frequency channel. Specifically, across frequency, the target ITDs need not be consistent with any single location in space to produce the entire benefit, as long as the target and masker ITDs within each frequency channel differ appropriately [6, 7]. Such results suggest that spatial unmasking does not rely upon differences in perceived location, but rather on more peripheral, processing mechanisms working independently in each frequency channel (e.g., [7]).

In most conditions where target and masker are distinct in their spectro-temporal characteristics (including detecting a tone in noise background or understanding a talker in steady-state noise), traditional models that take into account better-ear acoustic effects and binaural processing give relatively accurate predictions of how much spatial unmasking occurs [8, 4]. However, more recent studies show that perceived spatial location sometimes influences spatial unmasking in ways that are not predicted by such analysis. In general, such spatial unmasking arises in conditions in which there is "informational masking," i.e., masking that is not caused by interference in the peripheral representation of the target by the masker (e.g., see [9]).

Freyman and colleagues [10, 11] found that listeners were better at understanding a target talker from straight ahead when presented with two copies of a masking signal (one from the right, one slightly delayed from straight ahead) than when presented with the masker signal only from straight ahead. Freyman et al. attributed these results to differences in the perceived locations of target and masker. Because of the precedence effect [12], the temporally leading copy of the masker resulted in a shift of the perceived masker location away from the target. However, this difference in perceived location had little effect on target understanding when the masker was a noise source, easy to distinguish from the target speech. Only when the masker was speech, similar to the target, did adding a leading copy of the masker from the side of the listener improve the ability to understand the target source.

Traditional binaural models predict that in reverberant conditions, less binaural masking should occur because the left- and right-ear signals are interaurally decorrelated by reverberant energy [13, 4]. Consistent with this observation, spatial unmasking often decreases with increasing levels of reverberation [13, 4]. However, Kidd and colleagues [2] recently demonstrated that, even in a reverberant setting, perceiving a masker and target at different locations can improve performance when target and masker are similar. In their study, a small number of narrowband signals derived from an original speech signal were summed to produce an intelligible speech-like target. Sources were manipulated to control the degree to which they overlapped in their spectral content as a way of teasing apart contributions from "energetic masking" (reduction in the reliability of the peripheral representation of the portions of the target due to the presence of the masker). When presented with a noise masker that did not significantly overlap the target in frequency, there was very little masking or spatial unmasking, regardless of the environment. When presented with a noise masker that overlapped the spectral content of the target, there was a large amount of masking as well as significant spatial unmasking, but the amount of spatial unmasking decreased with increasing reverberant energy. Finally, when presented with another spectrally sparse speech-like masker that did not overlap the spectral content of the target, there was substantial masking, large amounts of spatial unmasking, and robust spatial unmasking even when there was significant reverberant energy.

These results suggest that there is another factor contributing to spatial unmasking in cases where energetic masking is not the main limitation on performance and the target and masker have similar spectro-temporal content (e.g., for speech masked by speech or by time-reversed speech; however, not for speech in steady-state noise or temporally-modulated broadband noise; e.g., see [14, 15]). When target and masker are easily confused with one another (i.e., in the presence of this specific form of "informational masking;" [9, 16]), differences in the *perceived locations* of target and masker appear to produce spatial unmasking.

These past studies were conducted with loudspeakers in anechoic space or a room, making it difficult to quantify exactly how the TMR at each ear varied with experimental condition. The current study used binaural stimuli presented over headphones rather than speakers in a sound field or in anechoic space. In addition to verifying that similar effects occur using headphone stimuli, the current experimental approach allows us to quantify any differences in the target and masker energies in the signals available to the listener in each spatial configuration and condition (i.e., to account for the acoustic better-ear effect).

We wished to verify the hypothesis that when target and masker are similar and therefore easy to confuse with one another, differences in the perceived locations of the target and masker allow a listener to focus on the signal at a desired location, producing spatial unmasking (e.g., see [17, 18]). Moreover, we wondered whether the amount of spatial unmasking that we obtained would depend on the veracity of the spatial cues that caused the perceived locations of target and masker to differ. On one hand, simulating sources with realistic spatial cues (e.g., using full head-related impulse responses, HRIRs) produce more compact spatial percepts than when only a subset of cues is present (e.g., using only interaural time differences; see [19]), and the spatial extent of the perceived sources might influence how helpful perceived spatial separation is for segregating target from masker. On the other hand, if all that is necessary to attend to the target is any difference that is salient, any combination of spatial cues may be sufficient for the full effect.

In order to reduce the contributions of within-band masking, we used spectrally sparse target and masker signals that had very little spectral overlap (e.g., see [20]) but

that were similar in their spectro-temporal characteristics (and easily confused with each other). From the results of Kidd and his colleagues [2], such conditions produce significant spatial unmasking even when there is very little energetic masking, in contrast to what happens with dissimilar target and masker with similar spectral overlap, where essentially no masking (or spatial unmasking) is obtained. To test whether *any* manipulation of the target and masker signals that produced differences in perceived location produces similar spatial unmasking, we manipulated the kinds of spatial cues used to control the perceived locations of target and masker. Amongst the conditions was one in which interaural phase differences (IPDs) in the target and masker were identically zero, a case that traditional binaural models predict would provide little or no spatial unmasking, even if within-band masking plays some role (despite our efforts to minimize these effects). Results suggest that in complex settings, spatial unmasking occurs through energy effects at the ears, bottom-up binaural processing, and some higher-level mechanism, possibly a form of top-down spatial attention that helps mediate competition for central resources caused by simultaneous sources.

2. Methods

2.1. Subjects

Four college-aged students participated in the study. All subjects had normal hearing as confirmed by an audiometric screening. Subjects were paid for their participation in the experiment. Most subjects were naive, having no prior experience in psychophysical studies, although all received training to familiarize them with the stimuli and procedures before formal data collection began (see Procedures, below).

2.2. Stimuli

Raw speech stimuli were taken from the Coordinate Response Measure corpus (e.g., see [21]), which consists of sentences of the form "Ready <*call sign*>, go to <*color*> <*number*> now." In the sentences we used, the call sign was one of the set ["Baron," "Eagle," "Tiger," and "Arrow"]; the color was one of the set [white, red, blue, green]; and the number was one of the digits between one and eight, excluding the number seven (as it is the only two-syllable digit and is therefore relatively easy to identify). For each session, one of the four male talkers was randomly selected as the target talker.

In each trial, two different sentences were used as sources. One utterance, designated as the *target*, always contained a particular call sign that was selected randomly for each experimental session. The second utterance (designated the *masker*) was chosen to contain one of the other three call signs, chosen randomly from trial to trial. In all cases, the numbers and colors in the competing utterances were randomly chosen, although target and masker colors and numbers were constrained to differ from each other within each trial, as were the target and masker talkers.

In order to reduce the energetic interference between target and masker, each speech signal was processed to produce intelligible speech-like signals that had little overlap in the frequency domain (see also [20]). Each target and masker speech signal was band-pass filtered into 16 frequency bands of 1/3 octave width, with center frequencies spaced evenly on a logarithmic scale between 175 Hz and 5.6 kHz, every one-third octave. On each individual trial, eight of the 16 target bands were chosen randomly to construct the target. To ensure that the spectral content was somewhat balanced from trial to trial, the eight random bands were chosen such that four were selected from the lower eight bands (175–882 Hz) and four were selected from the upper eight bands (1.1–5.6 kHz). The remaining eight bands were used to construct the masker, using otherwise identical processing. In all cases, the target and masker signals were constructed by summing eight modulated sinusoids whose center frequencies equaled the center frequencies of the randomly selected bands. The modulation envelope multiplying each sinusoidal carrier was derived from the envelope of the corresponding third-octave band of the original target or masker speech signal, calculated using the Hilbert transform. The left- and right-ear signals were constructed by summing modulated sinusoids that differed for the two ears in order to produce different aspects of the spatial cues contained in pseudo-anechoic head-HRIRs measured on a KEMAR manikin at a distance of one meter (for details of HRIR measurements, see [22]).

Three different processing schemes, differing in what spatial cues were generated in the binaural target and masker, were used (producing *full-cue*, *envelope-only*, and *carrier-only* stimuli) in order to measure the degree to which the amount of spatial unmasking depended on the kinds of spatial cues present in the stimuli. This processing was applied separately, but identically, to target and masker stimuli.

Figure 1 shows flow charts illustrating the three processing schemes described here. *Full-cue* stimuli (Figure 1a) were generated by 1) extracting the Hilbert-transform envelopes within each of the selected eight third-octave bands, 2) multiplying the eight envelopes with appropriate sinusoidal carriers at each of the band center frequencies, 3) convolving each narrowband signal by the HRIR from the desired location, and 4) summing the eight resulting binaural signals to produce the final binaural signal. *Envelope-only* stimuli (Figure 1b) were generated by 1) band-pass filtering the HRIR-processed speech stimuli, 2) extracting the Hilbert-transform envelopes from each of the eight narrowband binaural signals to produce left- and right-ear envelope signals for each band, 3) multiplying the left- and right-ear envelopes for each band by the same sinusoidal carrier, generating eight left-right pairs of modulated sinusoids, and 4) summing the eight binaural modulated sinusoids to produce the final binaural signal. *Carrier-only* stimuli (Figure 1c) were generated by

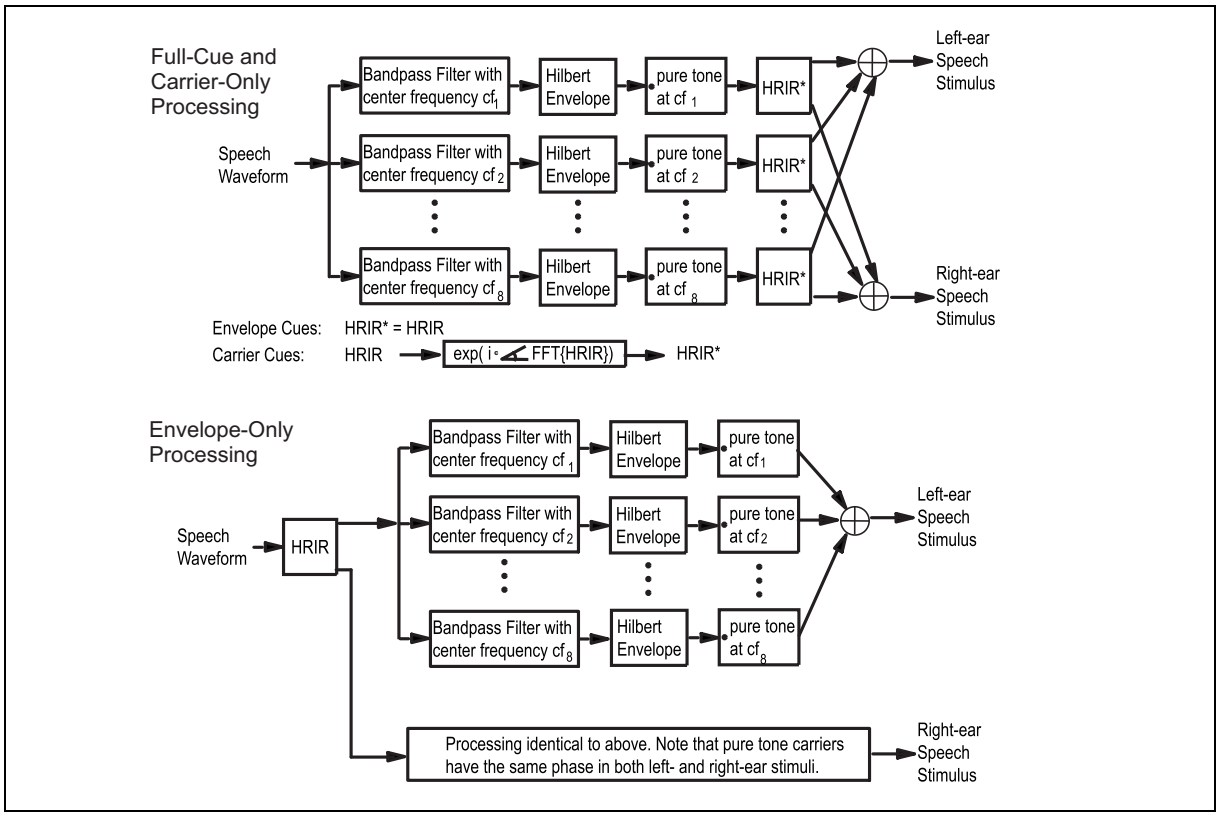

Figure 1. Three different processing schemes were used to generate stimuli, as shown in the above block diagrams. The three processing schemes differed in how the left- and right-ear head-related impulse responses (HRIRs), $h_l(t,\theta)$ and $h_r(t,\theta)$, were used to impart spatial cues in the stimuli, as shown. In the *carrier-only* stimuli, modified HRIRs [$h'_l(t,\theta)$ and $h'_r(t,\theta)$] were used that had the same frequency-dependent phase shift as the original HRIRs, but which had unity gain at all frequencies. In all cases, the input speech was filtered into eight randomly selected third-octave bands, the envelope in each band was extracted and used to modulate appropriate sinusoidal carriers, and the results summed.

1) convolving the appropriate eight sinusoidal carriers by the HRIR from the desired location to produce left- and right-ear carrier signals for each band, 2) equalizing the level of the resulting left- and right-ear sinusoids so that they differed only in their phase, 3) extracting the Hilbert-transform envelopes within each of the eight third-octave bands of the original speech signal, 4) multiplying the left- and right-ear sinusoidal carriers for each band by the appropriate modulation envelope, generating eight left-right pairs of modulated sinusoids, and 5) summing the eight binaural modulated sinusoids to produce the final binaural signal.

As a result of this processing, the *full-cue* stimuli simulated all of the spatial cues that would be present if a source consisting of modulated sinusoids had been presented from the appropriate location in space. *Envelope-only* stimuli had interaural level differences that were comparable to those in the *full-cue* stimuli as well as interaural time differences in the modulation envelopes of each carrier; however, the interaural phase differences in the fine structure of the stimuli were zero. Finally, in the *carrier-only* stimuli, the carrier sinusoids had fine-time interaural phase differences (but no interaural level differences) appropriate for a sinusoidal source of that frequency at the HRIR location. However, the *carrier-only* stimuli contained no spatial cues in their modulation envelopes, which were identical in each band at the two ears.

Despite these differences, all three spatial processing approaches produced signals whose perceived lateral location was straight ahead when processed with a 0° HRIR and was far to the right of midline when processed with a 90° HRIR. Furthermore, the resulting stimuli, although qualitatively unlike natural speech, were all intelligible in quiet for experienced subjects. In fact, all subjects were screened to ensure that they were able to identify target words from eight-band (non-spatialized) sentences presented in quiet. All four subjects achieved scores equal to or greater than 95% correct on this screening.

The original target and masker raw speech signals were first normalized to have the same root-mean-square (RMS) energy prior to any spatial or envelope processing. Following this normalization, the level of the original target was decreased as necessary to set the desired target- to-masker energy ratio (TMR), holding the masker level constant. The TMR between the unprocessed target and masker signals (henceforth referred to as $\mathrm{TMR_{nom}}$) was set to one of six values chosen randomly for each trial, spanning the range from −40 dB to +10 dB, every 10 dB.

Table I. Average and standard deviation of the absolute level of the target (in root-mean-square dB SPL) for the full-cue condition. For each spatial configuration and ear, the mean (and standard deviation) absolute level of the target was computed by averaging over the stimuli used in the experiments.

		−40 dB	−30 dB	−20 dB	−10 dB	0 dB	10 dB
Target 0°, Masker 0°	left ear	35.1 (1.9)	45.6 (1.6)	55.0 (1.9)	65.9 (1.7)	75.8 (1.4)	84.9 (1.9)
	right ear	33.4 (1.9)	44.0 (1.6)	53.3 (1.8)	64.1 (1.5)	74.2 (1.3)	83.2 (1.9)
Target 90°, Masker 90°	left ear	28.7 (2.1)	38.8 (1.9)	47.6 (2.1)	57.9 (2.4)	68.3 (2.4)	77.6 (2.3)
	right ear	33.2 (1.8)	43.4 (1.3)	52.4 (1.6)	62.8 (1.6)	73.0 (1.9)	82.3 (1.9)
Target 0°, Masker 90°	left ear	35.7 (1.3)	45.4 (1.0)	54.8 (1.9)	64.9 (1.7)	75.2 (1.7)	84.6 (1.9)
	right ear	34.0 (1.3)	43.6 (1.1)	53.1 (1.9)	63.3 (1.7)	73.7 (1.7)	82.9 (2.0)
Target 90°, Masker 0°	left ear	27.9 (2.5)	38.7 (1.9)	47.7 (1.8)	58.3 (1.8)	68.4 (1.9)	78.4 (2.1)
	right ear	32.5 (2.2)	43.2 (1.8)	52.1 (1.4)	62.7 (1.7)	73.0 (1.8)	83.1 (1.7)

Table II. Average target-to-masker energy ratio (TMR) at the left and right ears, computed over 1152 target / masker pairs, when the nominal TMR of the stimuli (prior to spatial processing) is zero.

	Target 0°, Masker 0°		Target 90°, Masker 90°		Target 0°, Masker 90°		Target 90°, Masker 0°	
	left ear	right ear	left ear	right ear	left ear	right ear	left ear	right ear
Full cue	0.2 (2.8)	0.3 (2.8)	-0.4 (4.3)	-0.2 (3.3)	7.3 (3.9)	1.0 (3.7)	-7.0 (3.4)	-0.5 (3.1)
Envelope only	-0.5 (2.9)	-0.5 (2.9)	-0.3 (4.4)	-0.4 (3.5)	7.5 (1.1)	1.1 (2.9)	-7.6 (3.4)	-1.0 (3.1)
Carrier only	-0.4 (2.4)	-0.4 (2.4)	0.3 (2.7)	0.3 (2.7)	-0.2 (2.5)	-0.2 (2.5)	0.1 (2.3)	0.1 (2.3)

After scaling the target as necessary, the target and masker signals were separately processed as shown in Figure 1 to generate appropriate binaural target and masker signals. These binaural target and masker signals were then summed and scaled to have the same peak amplitude (in order to maximize the useful dynamic range at playback). This processing resulted in TMRs at the left and right ears that could differ from TMR_{nom}, as well as a small random overall level rove (approximately 1–2 dB from trial to trial; see Table I). Moreover, the average TMR in the two ears varied systematically as a result of the spatial processing (varying from −7.0 to +7.3 dB, depending on which ear is considered, which spatial configuration simulated, and which spatial processing scheme employed; see Table II). The TMR at the ear with the better acoustic TMR will be denoted as TMR_{be} to distinguish it from the TMR set in the pre-processed target and masker, TMR_{nom}. The presentation levels of the target and masker are analyzed further in the section below entitled *Absolute and Relative Target Presentation Levels*, with results reported in Tables I and II. As discussed below, the target was fully intelligible in quiet at even the lowest level, which had an average absolute root-mean-square level of 28 dB SPL in the left ear and 33 dB SPL in the right ear.

The main rationale for using three different forms of spatial cues in the stimuli was to test whether the amount of spatial unmasking depends on the realism of the spatial cues when the main source of masking is due to difficulty in separating target and masker, not inaudibility of the target. However, the particular choices of spatial processing were also guided by our desire to separate out spatial unmasking effects that are due to traditional within-band binaural unmasking mechanisms and those due to spatial perception. Although the stimuli were processed to have little spectral overlap, it is impossible to completely eradicate energetic masking. However, for the chosen stimulus conditions, traditional interaural cross-correlation models (e.g., [23, 24, 25, 26, 27]) that can account for spatial release of within-band masking predict that the amount of spatial unmasking should be smaller for *envelope-only* stimuli than for *full-cue* or *carrier-only* stimuli. Figure 2 illustrates why this is the case.

Figure 2 shows the response of a cross-correlation-based binaural model as a function of time for combinations of a target (a 551-Hz channel of processed speech as in Figure 1 with spatial cues for 90°) and a masker (broadband noise at 0°). Each panel shows the model response to target alone (level set to 38 dB SPL), masker alone (with level set to 50 dB SPL), or target plus masker (producing a within-band target-to-masker energy ratio of -12 dB, roughly the magnitude of the smallest within-band target-to-masker ratio that arose in the current experiments). Total left- and right-ear signals were first processed through a computational ANF model [28]. The running, normalized cross-correlation of the left- and right-ear 551-Hz channel ANF responses was then computed in a rectangular time window (7.2 ms).

The top left panel of Figure 2 shows the auditory nerve fiber response to the target speech signal (note that at this time scale, the ANF inputs look very similar for all conditions, so only the full-cue response is shown). The top right-hand panel shows the response of the model to the (full-cue) broadband noise at 0° (note that for a masker at 0°, the model response would be essentially the

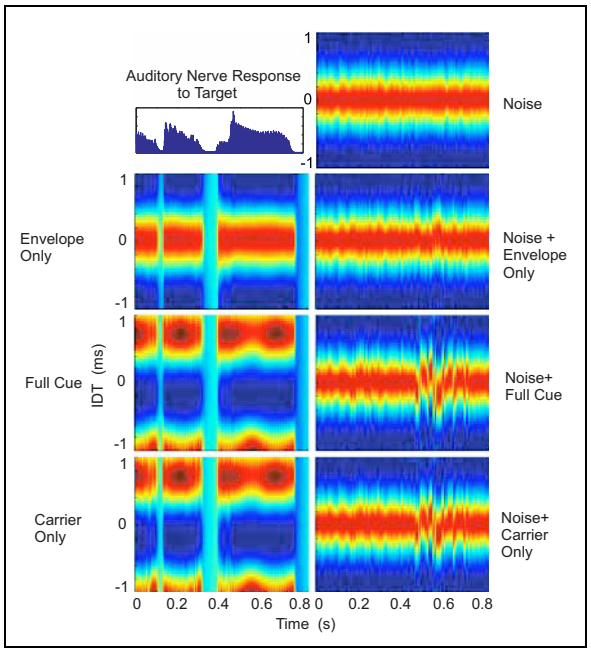

Figure 2. A binaural cross-correlation model predicts that *envelope-only* processing of a target stimulus should produce less spatial unmasking than *full-cue* or *carrier-only* processing. The top left panel shows the auditory nerve response as a function of time at one ear for the target stimulus. Each of the remaining panels shows the time-varying output of a binaural cross-correlation model tuned to 551 Hz (see text for details). The vertical axis represents different interaural time differences in the cross-correlation function. Color represents the value of the cross-correlation function, with red representing large values and blue representing low values. The top right panel shows model responses for a third-octave wide noise centered at 551 Hz. The left column shows model responses for a sample band of speech processed using the *envelope-only*, *full-cue*, and *carrier-only* processing schemes (see text). The right-hand panels in rows 2–4 show the model response for the sum of the corresponding target speech and the noise sample when the target-to-masker energy ratio is −12 dB.

same for envelope-only and carrier-only stimuli). Rows 2, 3, and 4 in the left column shows the model response for the envelope-only (second row), full-cue (third row), and carrier-only (bottom row) targets. All target-only responses (left column) show slow modulations of on-off activity correlated with the envelope energy fluctuations in the speech band (see top left panel). Because the envelope-only processed stimulus has zero ITD in the 551-Hz tone carrier, the model output for this stimulus (second row) shows activity peaks at zero ITD. In contrast, the peaks in the full-cue and carrier-only conditions (third and bottom rows) produce activity peaks near 0.8 ms. The right column of Figure 2 shows the model responses for the sum of the broadband noise and the corresponding targets. The responses to noise plus full-cue target and noise plus carrier-only target show large fluctuations when the instantaneous energy in the target is high, enabling detection of these portions of the target. In contrast, the response to the noise plus envelope-only target shows almost no fluctuations, illustrating why traditional models of spatial unmasking will predict less unmasking for envelope-only stimuli than for full-cue or carrier-only stimuli. To the extent that all spatial-cue conditions yield similar spatial unmasking, any traditional, within-band binaural model will not be able to account for these results.

2.3. Procedures

Each subject performed the initial screening described above to verify that they could understand the target in quiet. Following this screening, subjects each performed one training session followed by 12 formal experimental sessions, each of which lasted roughly one hour. The training session and each formal experimental session were identical in their construct, consisting of multiple blocks. The target call sign was fixed throughout a session, but varied randomly from session to session. The spatial configuration of the target and masker and the spatial cue condition (*full-cue*, *envelope-only*, and *carrier-only*) were held fixed in each block, but varied randomly from block to block. There were four spatial configurations tested: target and masker co-located at 0°, target and masker co-located at 90°, target at 0° and masker at 90°, and target at 90° and masker at 0°. Within each session, each subject performed one block of each combination of four spatial configurations and three spatial cue conditions, for a total of 12 blocks per session.

Each block consisted of 48 trials, corresponding to 8 trials at each of the six $\mathrm{TMRs}_{\mathrm{nom}}$, in random order. Thus, within each session, subjects completed 8 trials at each $\mathrm{TMR}_{\mathrm{nom}}$, spatial configuration, and spatial cue condition for a total of 576 trials per session. Across the 12 sessions, subjects performed 96 trials for each $\mathrm{TMR}_{\mathrm{nom}}$ in each condition.

Throughout each session, a computer screen displayed a message reminding the subject which call sign was the target and the location (front or side) from which the target would appear. At the end of each trial, subjects indicated the target color and number through a graphical user interface (GUI). The computer controlling the experiment recorded subject responses. After each trial, correct-answer feedback was provided by a written message indicating the correct color and number.

3. Results

Results were very consistent from subject to subject so only across-subject averages and standard deviations are reported. In general, performance increased with increasing target to masker ratio across the range of $\mathrm{TMR}_{\mathrm{snom}}$ tested.

3.1. Percent Correct

For all tasks, the percentage of correct trials was calculated as a function of the $\mathrm{TMR}_{\mathrm{nom}}$. Trials were considered "correct" only if both the color and number of the target were

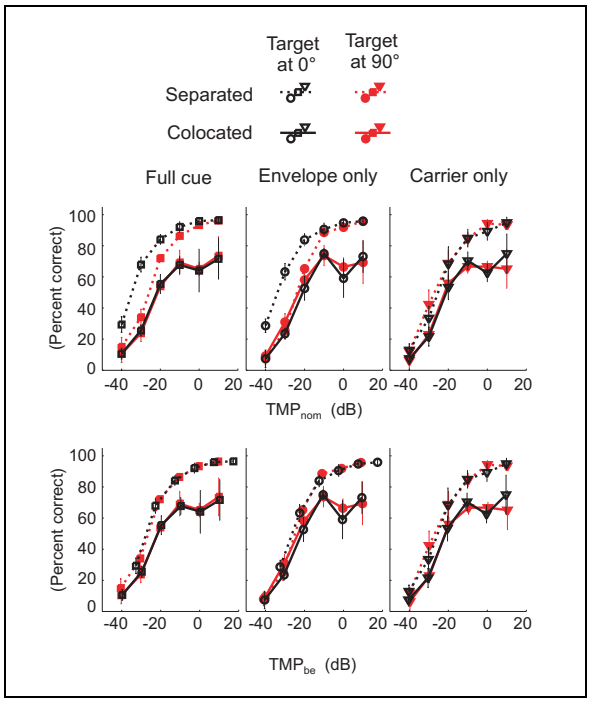

Figure 3. Performance improves with increasing target-to-masker level and with perceived spatial separation between target and masker for all spatial cue conditions. The top row shows percent correct performance as a function of TMR_{nom}, averaged across subject (error bars show across-subject standard deviation). Panels show results for *full-cue* processing, *envelope-only* processing, and *carrier-only* processing, from left to right. Each panel shows results for the four spatial configurations of target and masker, with target and masker spatially separated (dashed lines) or co-located (solid lines); and for target at 0° (open dark symbols) or target at 90° (filled light symbols). The bottom row shows the same results as a function of TMR_{be} (correcting for differences in the acoustic TMR at the better ear).

correctly identified. Given that there are four possible colors and seven possible numbers, chance performance was 1/4×1/7 or roughly 4%.

The top row of Figure 3 shows the average percent correct as a function of TMR_{nom} in each spatial configuration and condition, averaged across subjects (results in the bottom row are discussed below). The error bars show the across-subject standard deviation. Results for the *full-cue*, *envelope-only*, and *carrier-only* spatial cue conditions, are shown in the first, second, and third columns of the figure, respectively.

Results are essentially identical when the target and masker are at the same spatial location, independent of spatial cue condition and whether the sources are at 0° or 90°. For these co-located target and masker conditions (solid lines), performance is near 10% key words correct at the lowest TMR_{nom} (approaching chance performance), increases roughly linearly with increasing TMR_{nom} up to -10 dB, and is roughly equal for TMR_{nom} between -10 and 10 dB. For some conditions (most notably, for the 0° target and masker in the *envelope-only* and *carrier-only* conditions, shown by open symbols and solid lines in the middle and right panels), there is some indication that performance is actually worse at 0 dB TMR_{nom} than at −10 dB TMR_{nom}. This effect has been seen in previous studies and has been attributed to the listener making additional confusions between target and masker when they are at the same levels [29]. When the target is slightly less intense than the masker (but still audible and easy to hear, for instance at −10 dB TMR_{nom}) performance is assumed to improve because listeners can use the target-masker level difference to focus attention on the quieter talker. Although not explicitly tested here, we believe that performance in the spatially co-located conditions would improve if TMR_{nom} was increased beyond +10 dB.

For all three spatial cue conditions, performance generally improves when target and masker are at different locations (in all panels, dashed lines generally are above solid lines); however, the exact size and nature of the improvement depends on spatial configuration as well as on which spatial cues are presented. For all three spatial-cue conditions, performance asymptotes near 100% correct at large TMR_{nom} when target and masker are spatially separated. At lower TMR_{nom}, however, there are differences between the three spatial-cue conditions. For *full-cue* and *envelope-only* stimuli (left and center panels of Figure 3), performance is best when the target is straight ahead and the masker is to the right of the listener (dashed lines with open symbols) and is intermediate when the target is at 90° and the masker is straight ahead (dashed lines and filled symbols; although at large TMR_{nom}, there is no discernable difference between the two spatially-separated target and masker configurations because performance asymptotes at high performance levels). Furthermore, results for the *full-cue* and *envelope-only* cues are very similar (compare left and center panels in Figure 3). For the *carrier-only* stimuli, performance is equal for the two spatially separated spatial configurations for all conditions (compare the two dashed lines in the right-most panel), and generally lower than performance for comparable *full-cue* and *envelope-only* stimuli (for target at 0° and masker at 90°). The extent to which these differences across spatial cue condition can be attributed to differences in the TMR at the ears is considered further in subsequent sections.

3.2. Absolute and relative target presentation levels

Because we changed the target level to achieve the desired TMR_{nom} and then scaled each stimulus independently, the absolute level of the target varied systematically with TMR_{nom}, spatial configuration, and spatial cue condition as well as randomly from trial to trial. Informal tests confirmed that the target was always intelligible (performance was near 100% correct), even at the lowest absolute levels used. Thus, any changes in the ability to identify the target across TMR_{nom}, spatial configuration, and / or spatial cue condition can be attributed to the presence of the masking signal. We also computed the absolute presentation levels (in dB SPL) of the target in the *full-cue* condition, the condition for which the absolute presentation level of the

target varied the most. Table I shows the mean and standard deviation of the RMS levels of a *full-cue* target in dB SPL as a function of TMR_{nom} for each spatial cue configuration, for the left and right ears. These results show that the minimum mean absolute level of the target was 27.9 dB SPL (in the left ear for the case when TMR_{nom} was -40 dB and the target was at 90°), roughly akin to listening to whispered speech (e.g., see [30, p. 11]).

Just as the absolute target level varied systematically with spatial configuration and spatial cue condition, the average TMRs in the signals at the ears varied and did not always match TMR_{nom}, the ratio of the raw target energy to the raw masker energy. Even for a fixed spatial configuration and spatial-cue condition, there was variation in the TMR in the left- and right-ear stimuli across different trials. However, these random fluctuations were independent of TMR_{nom}.

We computed the mean and variability in the root-mean-square TMR in the final left- and right-ear stimuli for all spatial configurations and spatial cue conditions. Table II summarizes the results of this analysis assuming that TMR_{nom} is zero (for other values of TMR_{nom}, the actual TMR at the ears is equal to TMR_{nom} plus the value observed when TMR_{nom} is zero). Results in the first two columns show the TMRs in the left- and right-ear signals when target and masker are co-located. These results confirm that, on average, the left- and right-ear signals have essentially the same TMR when target and masker are simulated from the same spatial location, equal to TMR_{nom} (zero for the condition shown in Table II). For all configurations, the TMR in the near, right ear does not vary systematically with spatial configuration or spatial cue condition and is essentially equal to TMR_{nom} (i.e., given the trial-to-trial variability in the stimuli), primarily because the acoustic better-ear advantage relies on the acoustic head shadow, which is negligible in the near ear when the target moves from 0° to 90°. Focusing on the full-cue configurations (top two rows in Table II), there are systematic deviations between TMR_{nom} and the average observed TMRs in the left ear when target and masker are at different locations. When the target is at 0° and the masker is at 90°, the average *full-cue* left-ear TMR is 7.3 dB larger than TMR_{nom}. When the target is at 90° and the masker is at 0°, the average TMR in the left ear is −7.0 dB compared to TMR_{nom}. Results for the *envelope-only* TMRs (middle rows in Table I) are roughly the same as those of the *full-cue* TMRs. Finally, because carrier-only stimuli contained no head-shadow effects (only fine-time interaural phase difference cues), their actual TMR was always roughly equal to TMR_{nom} (bottom rows in Table II).

3.3. Performance after accounting for better-ear effects

The bottom panels in Figure 3 show the same results appearing in the top panels of the figure, but as a function of TMR_{be} rather than TMR_{nom}. In order to correct for the energy differences discussed above, each of the psychometric functions shown in the top of the figure were shifted by the average change in TMR at the better ear due to the spatial processing, given in Table II.

Results in the bottom row of Figure 3 show that changes in TMR_{be} due to spatial processing account for many of the effects observed in the top of the figure. Within each of the lower panels, the curves for the two spatial configurations where the target and masker were at the same location lie on top of each other (solid lines in the bottom panels of Figure 3); similarly, the curves for the two configurations where target and masker are spatially separated also lie on top of each other (dashed lines in each panel). However, for a given TMR_{be} the conditions in which the sources were perceived to come from different locations produce better performance than the conditions in which both sources were in the same direction (in all three panels in the bottom row of Figure 3, the dashed lines lie above the solid lines). For $TMRs_{be}$ of −20 dB or greater, performance is essentially identical for all three kinds of spatial processing. For instance, for performance at the 50% correct level, the benefit of spatial separation is equivalent to a roughly 5 dB increase in TMR_{be} (consider the horizontal separation between solid and dashed lines in each of the bottom panels of Figure 3 at the vertical position corresponding to 50% correct). However, there is a hint that performance is slightly worse for the *envelope-only* condition than for both *full-cue* and *carrier-only* conditions at the lowest TMR_{be}, an effect discussed further in the next section.

3.4. Spatial unmasking after accounting for better-ear effects

To quantify spatial unmasking, we wished to directly compare performance for spatially separated and spatially coincident configurations, after taking into account TMR_{be}. However, because we controlled TMR_{nom}, not TMR_{be}, we first had to interpolate results in order to compare performance in spatially separated and spatially coincident conditions.

For each subject in each condition, we used the psychometric function relating performance to TMR_{be} (see bottom panels of Figure 3) and interpolated to estimate performance at every TMR_{be} from -40 dB to +10 dB, in five dB steps. We generated two average psychometric functions, one for spatially separated and one for co-located sources. For the two co-located source configurations, we averaged percent correct estimates at each value of TMR_{be} from -40 dB to +10 dB. For the two spatially separated configurations, we averaged the interpolated percent correct estimates at each value of TMR_{be} for which both curves were defined. However, because the values of TMR_{be} that were presented varied with processing condition and spatial configuration, there were some values of TMR_{be} where only one of the functions was defined. At these points (e.g., $TMR_{be} = -40$ dB), we estimated the percent correct in the combined psychometric function from the single defined value from one spatial configuration.

This interpolation and averaging process resulted in two psychometric functions for each listener in each spatial cue

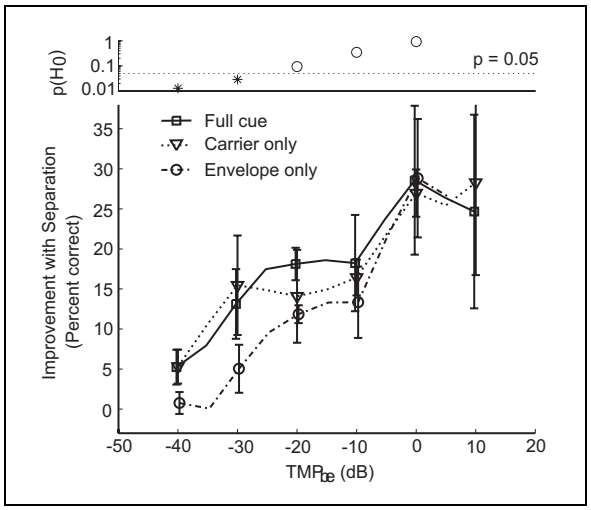

Figure 4. After accounting for TMR_{be}, the amount of spatial unmasking is similar for *full-cue*, *envelope-only*, and *carrier-only* stimuli when the target is loud; however, at the lowest target intensities, there is less spatial unmasking for the *envelope-only* stimuli than for *full-cue* or *carrier-only* stimuli. The improvement (amount of spatial unmasking in percent correct) is plotted as a function of TMR_{be} averaged across subject (error bars show across-subject standard deviation). In each condition, the psychometric functions as a function of TMR_{be} were linearly interpolated to predict performance. The two curves for spatially separated target and masker were averaged to estimate performance when sources are separate; similarly, the curves when target and masker were co-located were averaged. The difference of these curves was computed for each subject, and then averaged. At each TMR_{be}, a single-way ANOVAs tested the hypothesis that spatial unmasking varied with stimulus type. The inset, showing the probability that the null hypothesis is true, disproves the null hypothesis when TMR_{be} is less than -20 dB.

condition, one for co-located stimuli and one for separated stimuli, with points every five dB in TMR_{be}. Individually for each spatial cue condition and listener, we subtracted the estimated percent correct for co-located sources from the estimated percent correct when the sources were separated as a function of TMR_{be}. The resulting curves estimate how much spatial separation of the target and masker improves performance (in units of percent correct) for each condition and each listener as a function of TMR_{be}.

Figure 4 shows the average curves, collapsed across subject (error bars show across subject standard deviation). For all TMR_{be} greater than about -20 dB, the improvement is independent of spatial cue condition. However, at the lowest TMR_{be}, there is less improvement in performance with spatial separation in the envelope-only condition than for the full-cue and carrier-only conditions (which produce essentially the same improvement).

The null hypothesis, that stimulus type did not affect spatial unmasking, was tested using one-way ANOVA analysis on the data at each TMR_{be} for which all three curves were defined. Results showed that the null hypothesis could be rejected for all TMR_{be} <-20 dB ($p < 0.05$). This is illustrated in the inset at the top of Figure 4, which shows the probability that the null hypothesis is true as a function of TMR_{be}.

This analysis shows that above and beyond any spatial unmasking caused by changes in TMR_{be}, performance is better when target and masker are perceived at different spatial locations than when they are heard at the same location. At favorable target levels, the improvement is the same, independent of what spatial cue processing produces the differences in perceived location. However, at less favorable TMR_{be}, there is slightly more spatial unmasking for the *full-cue* and *carrier-only* processed stimuli than for the *envelope-only* stimuli.

3.5. Within-band spectral overlap

Although the target and masker stimuli were created to have little spectral overlap, there was, nonetheless, some overlap between target and masker spectra. Because the target and masker bands and tokens were chosen randomly, the degree of overlap varied from trial to trial. In order to estimate how much spectral overlap occurred, we analyzed how much masker energy fell within each target band 1) when both adjacent frequency bands contained a masker (a worst case analysis), 2) when the band above was masker and the band below was target, and 3) when the band below was masker and the band above was target. TMR_{band}, the average within-band TMR, was calculated for each third-octave band, spatial processing condition, and talker when TMR_{nom} was zero and the target and masker were both at 0°(the spatial configuration creating the lowest TMR for a given value of TMR_{nom}). Of course, the average within-band TMR depends on spatial configuration. Similarly, our simple analyses ignore upward spread of masking (and thus undoubtedly underestimate within-band effects). However, these subtleties are ignored here, as the goal of this analysis was to roughly estimate the TMR_{nom} at which within-band interference between target and masker might begin to affect a listener's ability to hear the target in a particular band and prove that such interference cannot account for the observed interference caused by the masker across the broad range of TMRs we used (not to quantify in detail, for each condition, exactly how large such interference was).

Figure 5 plots TMR_{band} for the three cases (masker above the target band, masker below the target band, and masker above and below the target band) for each of the individual talkers used in the study and for each of the spatial-cue processing conditions. The thick black line shows what the within-band TMR would be in the worst case, when both adjacent bands are masker; the two dashed lines show the within-band TMR for the cases when only only the band above (dotted line) or the band below is masker (dot-dashed line).

The worst-case (masker above and below) average TMR_{band} varies between approximately +30 dB and +60 dB. TMR_{band} depends only weakly on spatial-cue condition, but does vary with band center frequency and talker. The lowest within-band TMR is approximately

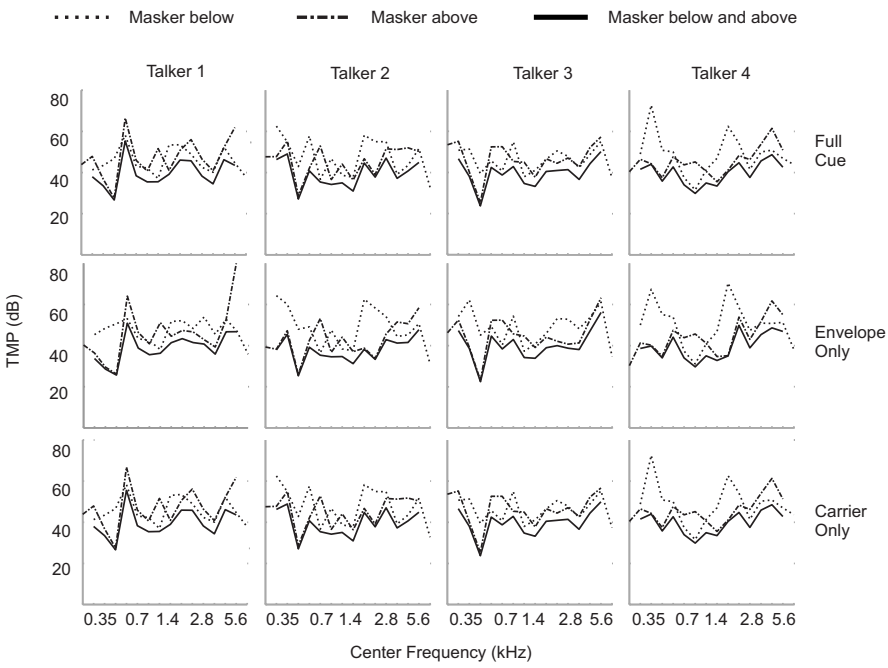

Figure 5. Some masker energy did fall within a critical band of the target, although the overlap was insignificant for TMR_{nom} greater than −20 dB. Each panel shows the within-band TMR for each possible target band when the target and masker are co-located at 0° and TMR_{nom} was 0 dB. Panels show results for each talker and for each spatial-cue processing scheme. Results were computed by averaging the within-band TMR for all stimuli for each target band for three different cases: 1) when the next-lower frequency band contained masker and the next-higher band contained target (dotted line), 2) when the next-higher frequency band contained masker and the next-lower band contained target (dot-dashed line), and 3) when both the next-lower and the next-higher frequency bands contained masker (solid line). Note that both the next-lower and next-higher frequency band could also contain target, in which case the within-band TMR would be significantly greater than the values shown.

+25 dB, occurring for a target in band four when there is a masker in band five for talkers 1–3.

This analysis suggests that the masker will not cause any significant within-band masking of the target for TMR_{nom} greater than −20 dB. However, spectral overlap will become increasingly influential as TMR_{nom} decreases below −20 dB.

4. Summary and discussion

Previous results suggest that perceived differences in target and masker location, rather than release from within-band energetic masking, contribute to spatial unmasking when target and masker are statistically similar [17, 16, 2]. Our results are consistent with these past studies in showing that spatial unmasking occurs in conditions where there should be little within-frequency-band interaction of target and masker. However, in addition, because the current study was performed using virtual auditory stimuli, we were able to extend these previous results by 1) analyzing the relative levels of target and masker at the ears of the listener and 2) controlling the form of the spatial cues present in the stimuli.

In the current study, energetic masking was minimized in order to emphasize spatial unmasking that arises through differences in perceived location. Target and masker were identically generated from eight randomly selected, non-overlapping narrow bands of speech taken from the same corpus. As a result, target and masker are easily confused, but cause little within-band interference of one another. We found that target intelligibility improves with spatial separation because: 1) intelligibility increases as the broadband TMR at the acoustically better ear increases, and 2) target intelligibility improves when target and masker are perceived in different locations. For the current task and stimuli, the first factor (changes in the broadband TMR_{be}) can account for nearly 7.5 dB of spatial unmasking when the target is straight ahead and the masker at 90° (see Table II, where for envelope-only processing, the TMR is +7.5 dB when target is at 0° and the masker is at 90°). The second factor (differences in perceived location) accounts for an additional 5 dB of spatial unmasking near 50% threshold, regardless of the spatial cue condition or which spatially separated configuration is considered (in the lower panels of Figure 3, the psychometric functions for spatially separated sources, shown by dashed lines, are displaced to the left of results when sources are co-located, shown with solid lines). In fact, whenever TMR_{nom} is greater than −20 dB and the spectral overlap between target and masker is negligible, this

second factor (perceiving target and masker at different locations) produces the same amount of spatial unmasking, independent of spatial cue condition. Thus, when energetic masking is not a significant factor in the interference caused by the masker, 1) realistic cues are no more effective in producing spatial unmasking than reduced spatial cues, and 2) interaural phase differences in the target and masker are not necessary to produce spatial unmasking.

Traditional models of spatial unmasking are based on detecting otherwise inaudible parts of the target through within-frequency-band binaural analysis, often performed independently in each channel (e.g., see [20, 7]). In the current experiment, within-band interference is significant only when $\mathrm{TMR_{nom}}$ is less than -20 dB, yet spatial unmasking is observed across all the TMRs tested. Furthermore, traditional binaural models predict that *full-cue* and *carrier-only* conditions should produce more spatial unmasking than *envelope-only* conditions. However, when $\mathrm{TMR_{nom}}$ is large in the current experiments, the amount of spatial unmasking that occurs due to differences in perceived location is the same for all spatial cue conditions. Unlike what has been concluded in some previous studies (e.g., see [15]), better-ear acoustic effects and interaural time differences in target and masker are not sufficient to account for the spatial unmasking we find. In order to account for the current results (and the results of previous studies implicating perceived spatial location in spatial unmasking), models must look across frequency and must also explicitly incorporate computation of source location.

Consistent with some past results, our analysis suggests that the broadband TMR at the better ear accounts for a portion of the spatial unmasking we observe (e.g., see [8, 15]); however, some other studies have found conditions in which binaural performance is worse than performance with the acoustically better ear, suggesting that listeners cannot truly "turn off" the bad acoustic ear when listening to binaural stimuli (e.g., see [31, 32, 33]). Those past studies that show worse binaural performance than better-ear monaural performance differ from the current study in that 1) the dominant masking effects are within band, rather than more central in origin and 2) the signal level in the worse acoustic ear is relatively intense. Either of these factors may explain why, in the current study, one can predict performance by considering the TMR in the better ear, whereas one cannot ignore what happens in the worse acoustic ear in these other studies.

In the current experiments, $\mathrm{TMR_{nom}}$ was controlled by decreasing the target level. It is theoretically possible that some of the decrease in target intelligibility with decreasing $\mathrm{TMR_{nom}}$ might be due to the absolute target level becoming too low. However, informal listening tests confirmed that the target was fully intelligible in quiet at the most extreme target attenuations used. Thus, the decreases in performance with decreasing $\mathrm{TMR_{nom}}$ can be attributed to interference from the masker. At the lowest values of $\mathrm{TMR_{nom}}$ it is likely that some of the masker interference on target intelligibility is from energetic overlap of target and masker within a critical band. When there is energetic overlap of target and masker, the amount of spatial unmasking should depend on the spatial-cue condition. Indeed, there was less spatial unmasking for *envelope-only* conditions than the other spatial-cue conditions at low signal levels. However, for larger $\mathrm{TMR_{be}}$, the amount of unmasking attributed to differences in perceived location was independent of spatial-cue condition. These results support the conclusion that for most of the current conditions the masker does not degrade target audibility in a traditional sense, but causes some kind of central interference.

We believe that when the target and masker are perceived at different spatial locations, the listener can reduce the central interference caused by the more salient (more intense, louder) masker by focusing attention on the target. This idea is consistent with recent studies that have suggested that differences between target and masker in many different physical dimensions can aid target understanding: differences in fundamental frequency, talker gender, voice timbre, sound level, and perceived location have all been suggested as possible cues that can be used to reduce interference between otherwise similar targets and maskers (e.g., see [34, 35, 36, 37, 38, 39, 16]).

In vision, attention to a particular spatial position facilitates neural responses to visual stimuli at the attended location and suppresses responses from other locations (see [40] for a recent review). Models of visual attention postulate that every visual stimulus has some inherent salience. When multiple objects compete for limited neural processing resources, the relative salience of target and masker determines how well observers can process each of the competing stimuli; moreover, attention to some stimulus attribute (such as object color, location, etc.) can enhance the relative salience of objects with that attribute [40, 41]. Similar explanations may apply to the current results, in which the relative auditory salience of target and masker depends directly on $\mathrm{TMR_{be}}$. Increasing $\mathrm{TMR_{be}}$ increases how effective the target is at competing for neural resources by increasing the relative salience of the target. Moreover, perceived differences between the spatial locations of the target and masker allow the listener to engage spatial attention, increasing the salience of a target at the attended location and reducing the interference caused by a masker at a different location. In this view, attention to a source at a particular location effectively "turns down" the masker level and "turns up" the target level, leading to less interference of the masker on target perception.

These results are consistent with the idea that top-down attention can reduce central interference when the target and masker are qualitatively different in one or more dimensions. By construct, central interference is the primary type of masker interference in the current experiments, and target and masker differ only in their perceived location. Thus, differences in perceived location make a significant contribution to spatial unmasking in the current study where there is little contribution from traditional within-channel factors and no other perceptual dimensions along which target and masker differ (no other cues to guide top-down attention). Of course, in many everyday circum-

stances, target and masker do overlap in their spectral content. In these cases, target audibility within each critical band probably plays a large role in limiting performance, and traditional bottom-up factors are likely to dominate how much spatial unmasking will be observed. Similarly, in most everyday settings, target and masker will differ in multiple dimensions in addition to spatial location, so that differences in perceived location play a less critical role in reducing interference.

While the current results help to define the different ways in which perceived spatial separation between target and masker can lead to spatial unmasking, much work remains. Additional experiments are necessary to tease apart the roles and relative importance of peripheral, bottom-up and central, top-down factors in spatial unmasking, and to begin to develop a comprehensive model explaining the many types of interference that can occur between a target and a masker.

5. Conclusions

We conclude that there are at least four mechanisms by which spatial separation of a target and a competing masker source leads to spatial unmasking:

1) Spatial separation of a target and masker generally increases $\mathrm{TMR}_{\mathrm{band}}$, the target-to-masker energy ratio within a critical band at the acoustically better ear. Within each frequency channel, the narrowband TMR directly affects audibility of the target energy in a given channel: improvements in the narrowband TMR at the better ear can increase how much of the target is audible in that channel (a bottom-up, stimulus driven mechanism contributing to spatial unmasking).

2) When target and masker overlap in their spectro-temporal content and are at different lateral angles, the addition of a target can produce interaural decorrelation of the total signal at the ears that can allow a listener to detect a low-level target in a particular time and frequency through bottom-up binaural processing. However, this mechanism cannot produce spatial unmasking if the target and masker do not overlap in their spectral content (compared to the effective bandwidth of the cross-correlation computation). In the current study, this factor probably accounts for the small differences in the amount of spatial unmasking observed for the three kinds of spatial-cue processing at the lowest TMRs tested.

3) Spatial separation of a target and masker generally increases $\mathrm{TMR}_{\mathrm{be}}$, the RMS target-to-masker energy ratio at the better ear computed across all frequencies. The current results show that even in conditions where there is little spectral overlap between target and masker, increasing the broadband TMR at the better ear improves speech intelligibility (as much as 7.5 dB in the current experiment).

4) Differences in perceived location allow a listener to focus attention on a target signal and reduce interference from a concurrent masker. In the current experiment, this benefit is roughly the same (equivalent of a 5 dB improvement in broadband TMR at threshold) regardless of what spatial cues lead to differences in perceived location. However, this top-down contribution to spatial unmasking is only likely to influence results when target and masker are easily confused, and may be redundant when target and masker differ in other, non-spatial dimensions.

Acknowledgement

This work was supported in part by grants from the Office of Naval Research (N00014-04-1-0131) and the National Institutes of Health (R01 N00014-04-1-0131). This work was motivated by helpful discussions with many colleagues, including Simon Carlile, Steve Colburn, Nat Durlach, Erick Gallun, Gerald Kidd, and Chris Mason. Chris Mason, Gerald Kidd, and two anonymous reviewers provided many detailed, insightful comments that greatly improved this manuscript. Erol Ozmeral's astute listening identified an error in the randomization of the stimuli we used during pilot testing, saving us untold time and grief. Tim Streeter assisted with gathering the results reported here.

References

[1] R. H. Gilkey: Effects of frequency and masker duration on free-field masking. J. Acoust. Soc. Am. **92** (1992) 2334.

[2] J. Kidd, G., C. R. Mason, A. Brughera, W. M. Hartmann: The role of reverberation in release from masking due to spatial separation of sources for speech identification. Acustica united with Acta Acustica **91** (2005) 526–536.

[3] R. M. Stern, C. Trahiotis: Models of binaural perception. – In: Binaural and Spatial Hearing in Real and Virtual Environments. R. Gilkey, T. Anderson (eds.). Erlbaum, New York, 1997, 499–532.

[4] P. M. Zurek, R. L. Freyman, U. Balakrishnan: Auditory target detection in reverberation. J. Acoust. Soc. Am. **115** (2004) 1609–20.

[5] V. Best, E. Ozmeral, F. J. Gallun, K. Sen, B. G. Shinn-Cunningham: Spatial unmasking of birdsong in human listeners: Energetic and informational factors. J. Acoust. Soc. Am. (2005) (in press).

[6] M. A. Akeroyd: The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking. J. Acoust. Soc. Am. **116** (2004) 1135–48.

[7] B. A. Edmonds, J. F. Culling: The spatial unmasking of speech: evidence for within-channel processing of interaural time delay. J. Acoust. Soc. Am. **117** (2005) 3069–78.

[8] P. M. Zurek: Binaural advantages and directional effects in speech intelligibility. – In: Acoustical Factors Affecting Hearing Aid Performance. G. Studebaker, I. Hochberg (eds.). College-Hill Press, Boston, MA, 1993.

[9] N. I. Durlach, C. R. Mason, J. Kidd, G., T. L. Arbogast, H. S. Colburn, B. G. Shinn-Cunningham: Note on informational masking. J. Acoust. Soc. Am. **113** (2003) 2984–7.

[10] R. L. Freyman, K. S. Helfer, D. D. McCall, R. K. Clifton: The role of perceived spatial separation in the unmasking of speech. J. Acoust. Soc. Am. **106** (1999) 3578–3588.

[11] R. L. Freyman, U. Balakrishnan, K. Helfer: Release from informational masking in speech recognition. MidWinter

Meeting of the Association for Research in Otolaryngology, St. Petersberg Beach, FL, 2000.

[12] R. Y. Litovsky, H. S. Colburn, W. A. Yost, S. J. Guzman: The precedence effect. J. Acoust. Soc. Am. **106** (1999) 1633–1654.

[13] R. Plomp: 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** (1976) 200–211.

[14] D. S. Brungart, B. D. Simpson: Within-ear and across-ear interference in a cocktail-party listening task. J. Acoust. Soc. Am. **112** (2002) 2985–2995.

[15] J. F. Culling, M. L. Hawley, R. Y. Litovsky: The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. J. Acoust. Soc. Am. **116** (2004) 1057–65.

[16] N. I. Durlach, C. R. Mason, B. G. Shinn-Cunningham, T. L. Arbogast, H. S. Colburn, J. G. Kidd: Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. J. Acoust. Soc. Am. **114** (2003) 368–79.

[17] R. L. Freyman, U. Balakrishnan, K. Helfer: Spatial release from informational masking in speech recognition. J. Acoust. Soc. Am. **109** (2000) 2112–2122.

[18] R. L. Freyman, U. Balakrishnan, K. S. Helfer: Spatial release from informational masking in speech recognition. J. Acoust. Soc. Am. **109** (2001) 2112–22.

[19] S. Carlile: Virtual auditory space: Generation and applications. RG Landes, New York, 1996.

[20] T. L. Arbogast, C. R. Mason, G. Jr. Kidd: The effect of spatial separation on informational and energetic masking of speech. J. Acoust. Soc. Am. **112** (2002) 2086–98.

[21] D. S. Brungart: Evaluation of speech intelligibility with the coordinate response method. J. Acoust. Soc. Am. **109** (2001) 2276–9.

[22] B. G. Shinn-Cunningham, N. Kopco, T. J. Martin: Localizing nearby sound sources in a classroom: Binaural room impulse responses. J. Acoust. Soc. Am. **117** (2005) 3100–3115.

[23] H. S. Colburn: Theory of binaural interaction based on auditory-nerve data. I: General strategy and preliminary results on interaural discrimination. J. Acoust. Soc. Am. **54** (1973) 1458–1470.

[24] H. S. Colburn: Theory of binaural interaction based on auditory-nerve data. II: Detection of tones in noise. J. Acoust. Soc. Am. **64** (1977) 525–533.

[25] R. M. Stern, C. Trahiotis, A. M. Ripepi: Some conditions under which interaural delays foster identification of speech-like stimuli. – In: Dynamics of Speech Production and Perception. P. Divenyi, G. Meyer (eds.). IOP Press, Amsterdam, 2004.

[26] M. A. Akeroyd, A. Q. Summerfield: Integration of monaural and binaural evidence of vowel formants. J. Acoust. Soc. Am. **107** (2000) 3394–406.

[27] J. F. Culling, H. S. Colburn: Binaural sluggishness in the perception of tone sequences and speech in noise. J. Acoust. Soc. Am. **107** (2000) 517–27.

[28] Z. Zhang, M. G. Heinz, I. C. Bruce, L. H. Carney: A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. J. Acoust. Soc. Am. **109** (2001) 648–670.

[29] D. S. Brungart: Informational and energetic masking effects in the perception of two simultaneous talkers. J. Acoust. Soc. Am. **109** (2001) 1101–9.

[30] B. C. J. Moore: An introduction to the psychology of hearing (5e). Academic Press, San Diego, CA, 2003.

[31] A. W. Bronkhorst, R. Plomp: The effect of head-induced interaural time and level differences on speech intelligibility in noise. J. Acoust. Soc. Am. **83** (1988) 1508–1516.

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

[33] B. G. Shinn-Cunningham, J. Schickler, N. Kopco, R. Litovsky: Spatial unmasking of nearby speech sources in a simulated anechoic environment. J. Acoust. Soc. Am. **110** (2001) 1118–29.

[34] G. Kidd, C. R. Mason, P. S. Deliwala, W. S. Woods, H. S. Colburn: Reducing informational masking by sound segregation. J. Acoust. Soc. Am. **95** (1994) 3475–3480.

[35] J. Bird, C. J. Darwin: Effects of a difference in fundamental frequency in separating two sentences. 11th International Symposium on Hearing: Auditory Physiology and Perception, Grantham, UK, 1997.

[36] C. J. Darwin, R. W. Hukin: Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. J. Acoust. Soc. Am. **107** (2000) 970–7.

[37] D. S. Brungart, B. D. Simpson, M. A. Ericson, K. R. Scott: Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. **110** (2001) 2527–38.

[38] C. J. Darwin, D. S. Brungart, B. D. Simpson: Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. J. Acoust. Soc. Am. **114** (2003) 2913–22.

[39] W. R. Drennan, S. Gatehouse, C. Lever: Perceptual segregation of competing speech sounds: the role of spatial location. J. Acoust. Soc. Am. **114** (2003) 2178–89.

[40] R. Desimone, J. Duncan: Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. **18** (1995) 193–222.

[41] J. H. Reynolds, T. Pasternak, R. Desimone: Attention increases sensitivity of V4 neurons. Neuron **26** (2000) 703–14.