Influences of modulation and spatial separation on detection of a masked broadband target[a]

Norbert Kopčo[b] and Barbara G. Shinn-Cunningham[c]
Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 22 June 2007; revised 24 June 2008; accepted 9 July 2008)

Experiments explored the influence of amplitude modulation and spatial separation on detectability of a broadband noise target masked by an independent broadband noise. Thresholds were measured for all combinations of six spatial configurations of target and masker and five modulation conditions. Masker level was either fixed (Experiment 1) or roved between intervals within a trial to reduce the utility of overall intensity as a cue (Experiment 2). After accounting for acoustic changes, thresholds depended on whether a target and a masker were colocated or spatially separated, but not on the exact spatial configuration. Moreover, spatial unmasking exceeded that predicted by better-ear acoustics only when modulation cues for detection were weak. Roving increased the colocated but not the spatially separated thresholds, resulting in an increase in spatial release from masking. Differences in both how performance changed over time and the influence of spatial separation support the idea that the cues underlying performance depend on the modulation characteristics of the target and masker. Analysis suggests that detection is based on overall intensity when target and masker modulation and spatial cues are the same, on spatial attributes when sources are separated and modulation provides no target glimpses, and on modulation discrimination in the remaining conditions. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2967891]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.66.Rq, 43.66.Mk [RLF] Pages: 2236–2250

I. INTRODUCTION

The extent to which one sound source masks another depends to a large degree on how similar the two sources are in characteristics such as their spectral profile, temporal structure, and spatial location. While a fair amount is known about how these individual characteristics affect the ability to detect and understand a masked target, relatively little is known about how these characteristics interact. In everyday situations, listeners often are faced with the task of understanding one complex, fluctuating signal in the presence of similar, complex signals from different locations, such as understanding one talker in the presence of competing talkers. If we are ever to understand perception in everyday situations, we must explore how source characteristics such as spectral content, amplitude fluctuations over time (modulation), and spatial location jointly affect perception.

This paper considers the individual and combined effects of two stimulus characteristics: modulation structure and spatial location. *A priori*, one might imagine that the two variables are redundant with one other, so that there is no added benefit when spatial cues in a target and a masker differ if they already differ in their modulation structure (and vice versa). Alternatively, it is possible that masking effects related to temporal modulation and spatial location are largely independent of one another and that effects of the two attributes are additive. Finally, it is possible that differences in temporal modulations actually facilitate the effectiveness of spatial cues in releasing masking, or vice versa, resulting in superadditivity of their individual effects. This study investigates these alternative possibilities using a detection task with simple broadband noise targets and maskers by manipulating both temporal and spatial characteristics independently and jointly.

Several previous studies looked at spatial release from masking (SRM) for nonspeech stimuli that fluctuated over time. The target stimuli in these studies ranged widely, including click trains (Saberi *et al.*, 1991; Gilkey and Good, 1995; Good *et al.*, 1997), chirp trains (Lane *et al.*, 2004; Kopco, 2005), and pulsed 1/3-octave bands of noise (Zurek *et al.*, 2004). However, none of these studies looked at how modulation influences SRM.

Other studies examining the relationship between modulation and spatial processing in masked detection tasks differed substantially in approach and the specific questions addressed, making it difficult to compare results across studies. For example, some explored comodulation and binaural masking release (van de Par and Kohlrausch, 1998; Hall *et al.*, 2006) while others looked at monaural and interaural level discrimination (Stellmack *et al.*, 2005), the interaction between modulation detection interference and spatial processing (Sheft and Yost, 1997), or the equivalence of binaural processing of low-frequency fine time structure versus high-frequency envelope structure (Bernstein and Trahiotis, 1994; van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002). Physiological data from the cat inferior colliculus (IC) suggest that binaural cues in the temporal envelope contribute to SRM (Sterbing *et al.*, 2003; Lane and Delgutte, 2005). However, some psychophysical studies sug-

---

[a] Portions of this work were presented at the 149th and 151st meetings of the Acoustical Society of America.

[b] Permanent Address: Department of Cybernetics and AI, Technical University, Košice, Slovakia. Electronic mail: kopco@bu.edu

[c] Electronic mail: shinn@bu.edu

gest that the stimulus temporal envelope does not affect SRM. For example, binaural detection thresholds obtained for a harmonic tone complex and broadband noise targets are very similar, despite dramatic differences in their envelopes (van de Par *et al.*, 2004). Overall, these studies do not provide a consistent account of how spatial cues and modulation jointly affect detection of a target embedded in noise.

Some work suggests that the influence of modulation on masked target detection depends on whether the target or the masker is modulated. For example, when listeners must detect a target embedded in maskers, reaction times depend less strongly on the number of distractors when the target is amplitude modulated and the maskers are unmodulated than when the target is a pure tone and the maskers are amplitude modulated (Asemi *et al.*, 2003). This asymmetry suggests that the modulated target is more likely to "pop out" of the background of unmodulated maskers than the reverse, making detection of a modulated target robust to the addition of interferers. In comodulation masking release (CMR) studies, adding off-target-frequency components that are modulated identically with the on-frequency masker improves the detectability of an unmodulated target (Hall *et al.*, 1984; van de Par and Kohlrausch, 1998; Winter *et al.*, 2004). However, we know of no studies reporting a corresponding benefit of increasing masker bandwidth when the target, rather than the masker, is modulated, so it is possible that there is a perceptual asymmetry between modulating the target versus modulating the masker in such situations, as well.

II. EXPERIMENTS AND HYPOTHESES

Two experiments were performed to study how modulation and spatial location of the target and masker affect target detection. Both target and masker were broadband noises that were either unmodulated or sinusoidally amplitude modulated (SAM). As a result, across-channel processing and across-frequency grouping were likely to contribute to performance. Moreover, for these broadband targets and maskers, listeners could not detect the target by using spectral sidebands (as might be the case when the target is a SAM tone; Dau and Ewert, 2004) and the opportunity to use profile analysis (Green, 1988) was minimized (because of the similarity of the target and masker spectral profiles).

A single modulation frequency (40 Hz) was used throughout the study, chosen both because humans are fairly sensitive to modulation at this frequency (Viemeister, 1979) and because responses of space-sensitive IC neurons are affected by modulation at this frequency (Lane and Delgutte, 2005).

Spatial separation of a broadband target from a broadband masker results in a frequency-dependent change in the target-to-masker energy ratio (TMR) at the ears. The resulting TMR profile as a function of frequency varies from one target/masker configuration to another, so that TMR should affect performance differently for different spatial configurations of the target and masker. The contribution of binaural processing to target detection should therefore depend on spatial configuration. In particular, if the TMR profile is such that the most favorable TMRs are at low frequencies, then interaural time difference (ITD) processing is likely to contribute to detection (Kopco and Shinn-Cunningham, 2003). On the other hand, if the most favorable TMRs are at high frequencies, then the contribution of ITD processing to performance is likely to be smaller. Finally, the contribution of across-frequency integration to detection, if any, is likely to be larger when the TMR is similar across frequency than when the TMR is very large in one band and small in others. As a result, the relative contribution of different detection cues (e.g., changes in overall energy and interaural decorrelation) also is likely to vary from one target/masker configuration to another.

Three different spatially separated configurations were included in this study to evaluate whether the interaction of modulation and spatial cues depends on the specific target/masker configuration. Specifically, in one of the chosen configurations the maximum in the TMR profile was in a low-frequency region, while in the remaining configurations it was at high frequencies.

As described above, the way in which modulation and spatial configuration interact is poorly understood. The current experiments were designed to explore how these cues jointly affect performance. If the processing of the two cues is strictly serial then the effects of the cues should be additive. This would occur if (1) spatial processing improves the effective TMR of the signal prior to any modulation processing, (2) modulation processing operates on the output of the spatial processing stage, and (3) detection is based on the output of the modulation processing. If the two cues both work to help listeners perceptually segregate the target from the masker, then the cues may be redundant. Specifically, if differences in modulation of the target and masker are sufficient to segregate the target and masker, then providing additional spatial cue differences in the target and masker might not improve performance. In this case, the benefits of modulation and spatial cue differences would be less than additive. Alternatively, if spatial cue differences are necessary for modulation differences to be useful (or vice versa), then the effects of differences in the two cues may be superadditive.

In addition to exploring whether the two cues are additive, subadditive, or superadditive, we tested two specific hypotheses about how source modulation structure and source location affect detection for broadband signals.

H1. The effect of modulation on SRM will depend on whether the target, the masker, or both target and masker are modulated (e.g., see the results of Asemi *et al.*, 2003).

H2. The effect of modulation on detection threshold will depend on spatial configuration because the relative importance of individual cues changes with spatial configuration. (1) When the best TMR occurs in low frequencies, ITD processing will be relatively influential on performance. (2) If perceived location rather than ITD processing is the critical factor in determining how spatial cues contribute to detection, performance will depend on whether or not the target and masker are spatially separated, but not on the exact spatial configuration. (3) When TMR is relatively constant with frequency, across-frequency integration is likely to contribute to detection.

Experiment 1 was performed with the masker noise presented at a fixed level. However, overall stimulus level may be the primary cue for detection when the target and masker are similar in their spectrotemporal structure and spatial cues, and therefore likely to be perceived as one unitary object from a particular location. To reduce the efficacy of overall level, Experiment 2 roved the masker level from interval to interval within each trial.

III. METHODS

A. Subjects

Seven subjects (four female and three male, including author N.K.) participated in Experiment 1. Seven subjects (three female and four male, two of whom participated in Experiment 1) participated in Experiment 2 (Experiment 2 was conducted almost a year after Experiment 1, so it is unlikely that learning from Experiment 1 transferred to Experiment 2 for the two subjects who performed both experiments). All subjects had normal hearing (confirmed by an audiometric screening), with ages ranging from 23–32 years.

B. Stimuli

The target and masker stimuli were both broadband noises with flat spectrum between either 0.3 and 8 kHz (target) or 0.2 and 12 kHz (masker), generated using a MATLAB implementation of the Butterworth bandpass filter (39th order for target and 33rd order for masker) with a stopband attenuation of 60 dB and stopband frequencies of 0.2–10.05 kHz (target) and 0.1–14 kHz (masker). The 200-ms-long target $s_T(t)$ was temporally centered on the masker $s_M(t)$, which had a duration of 300 ms. Both target and masker were ramped at onset and offset by 30 ms $\cos^2$ ramps. Modulation, if present, was sinusoidal with a frequency of 40 Hz and depth $m=0.5$ and had a random initial phase ϕ chosen from ten possible phases ($\phi=2\pi j/10$, $j=1,\ldots,10$). The stimuli were of the form

$$s_{i,k}(t) = A_i[1 + m_i \cos(2\pi 40t + \phi_{i,k})]n_{i,k}(t),$$

where $i=T$ for the target and $i=M$ for the masker, k is the trial number, $n_{i,k}(t)$ is a random bandpass-filtered noise token, and A_i is a scaling factor that determines the stimulus presentation level. The same five modulation conditions were explored in both experiments: no modulation ($m_T=m_M=0$), in-phase comodulation ($m_T=m_M=0.5$; $\phi_{M,k}=\phi_{T,k}$), target-only modulation ($m_T=0.5$; $m_M=0$), masker-only modulation ($m_T=0$; $m_M=0.5$), and pi-out-of-phase modulation ($m_T=m_M=0.5$; $\phi_{M,k}=\phi_{T,k}+\pi$).

Modulation increases the long-term rms energy of a signal by a factor of $(1+m^2)^{-0.5}$. For the modulation depth and form used here, modulation increases the rms energy of the modulated signal by approximately 0.5 dB. All results were corrected for this rms energy effect by scaling the measured thresholds and reporting thresholds in units of TMR.

Space was simulated using pseudoanechoic nonindividualized head-related impulse responses (HRIRs) recorded at four locations (−45°, 0°, 45°, and 90°, left to right) at a distance of 120 cm from the center of the head, using miniature microphones placed at the entrance of the ear canals of a female listener who did not participate as a subject in this study (see Shinn-Cunningham *et al.*, 2005, for a full description of these HRIRs). Five spatial configurations were explored in Experiment 1: two with the sources colocated at 0° or −45° and three with the sources spatially separated [(T at 90°, M at 0°), (T at 0°, M at 90°), and (T at 45°, M at −45°)]. An additional colocated condition (90°) was added in Experiment 2 to create three matching pairs of colocated and separated spatial configurations.

In both experiments, the average level of the masker was the same in all trials, prior to processing by the HRIRs (which altered the level of the signals reaching the ears). Therefore, because of HRIR processing, there were frequency-dependent variations in the signals reaching the ears across the different masker locations (graphs in Fig. 2 can be used to estimate how the received masker level changed at the two ears). For the masker at 0°, the maximum masker level received at the ears was 61 dB sound pressure level (SPL). In Experiment 1, the masker level was constant across the three intervals within a trial, while in Experiment 2 the masker level was roved independently in each interval by a value uniformly distributed between ±5 dB (the target, if present, was roved with the masker, which kept constant the TMR measured prior to HRIR processing).

Stimulus files, generated off-line at a sampling rate of 50 kHz, were stored on the hard disk of a control computer (IBM PC compatible). Ten random noise tokens were pregenerated to be used as targets and another ten tokens were produced to be used as maskers in this study (i.e., target and masker were always independent samples of noise). These 20 tokens were bandpass filtered (10 by the target filter and 10 by the masker filter, which had a slightly wider passband), modulated (by 1 of 10 modulation envelopes, differing in initial phase), and HRIR filtered (by an HRIR corresponding to locations of −45°, 0°, 45°, or 90°) to produce 440 target stimuli [10 tokens × (10 modulation envelopes + no modulation) × 4 locations] and 440 similar masker stimuli. On each trial, three different masker tokens and one target token were randomly selected, scaled, and concatenated into a stimulus file that contained three masker intervals with the target randomly added to the second or the third interval.

TDT System 3 hardware was used for D/A conversion. The result was amplified through a TDT headphone buffer and presented via Etymotic Research ER-1 insert earphones (with approximately flat frequency response in the range 100 Hz–15 kHz). No filtering was done to compensate for the transfer characteristics of the playback system. A simple alphanumeric interface in MATLAB was used to give instructions to subjects, gather responses, and provide feedback. The subject indicated the perceived target interval by hitting the appropriate numeric key (“2” or “3”) on the computer keyboard. Experiments were performed in a single-walled sound-treated booth.

C. Experimental procedure

Each trial consisted of three intervals, each of which contained a masker. Either the second or the third interval (randomly chosen with equal probability on each trial) also contained the target. The intervals were separated by 50-ms-long silent gaps. Subjects performed a two-alternative, forced-choice task in which they were asked to identify which interval, the second or the third, contained the target. Correct-answer feedback was provided at the end of each trial.

A three-down-one-up adaptive procedure was used to estimate detection thresholds (Levitt, 1971), defined as the 79.4% correct point on the psychometric function. Each run started with a description of the measurement condition of the run (e.g., written instructions might read “In this run, the target is modulated and the distractor is not modulated, the target comes from an azimuth of 0° and the distractor from 90°. Next, you will hear a sample of the noise distractor that you should ignore, followed by the target that you should identify. Hit RETURN to hear the sample.”). The subject could listen to the sample repeatedly until he/she was confident that he/she understood the task.

The staircase measurement procedure started with the target presented at a clearly detectable level and continued until 11 “reversals” occurred. The target level was changed by 4 dB on the first reversal, 2 dB on the second reversal, and 1 dB on all subsequent reversals. For each adaptive run, detection threshold was estimated by taking the average target presentation level over the last six reversals.

Each of the two experiments consisted of six 1 h sessions performed on different days (the first session of each experiment was a practice session, serving to familiarize the subjects with the experimental procedure). In each session, the thresholds were measured for all combinations of spatial and modulation conditions (25 thresholds in Experiment 1 and 30 in Experiment 2), with the order of conditions randomized between sessions and between subjects. One adaptive run took approximately 2–3 min to complete.

Informal interviews of the listeners confirmed that at moderate to high TMRs, listeners found it very easy to interpret the two simulated stimuli as a target noise and a distractor noise coming from the indicated locations with the described modulation characteristics (as opposed to hearing them as one combined noise). This was likely the case because of the following: (1) at the beginning of the experiment, the subjects were given a detailed description of the stimulus combinations they should expect; (2) prior to each adaptive run, listeners had the opportunity to familiarize themselves with the target and masker stimuli presented separately before they heard them combined; and (3) the procedure started with both the target and the masker clearly audible. It is difficult to know whether or not the listeners perceived the two stimuli as separate objects when the target level was near the threshold. However, none of the subjects reported any difficulty performing the task (for example, none of them reported being confused about what to listen for in order to detect the target).

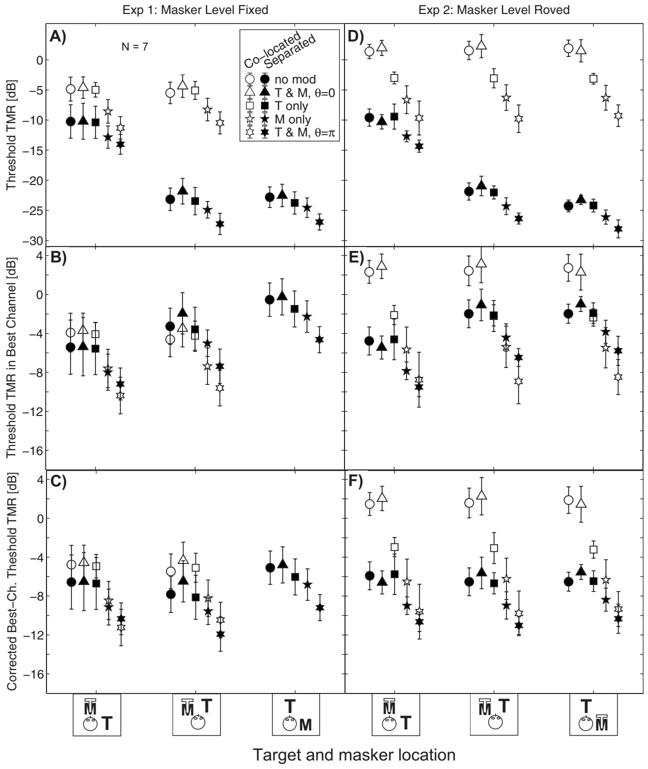

FIG. 1. Raw data plotted as a function of the masker location measured with the masker level fixed (Experiment 1; panels A, B, and C) and roved (Experiment 2; panels D, E, and F). All graphs show the across-subject mean and standard deviations in measured threshold TMRs: panels A and D show the raw threshold TMR energy ratios, panels B and E show the threshold TMRs in the best channel, and panels C and F show the threshold TMRs in the best channel after correcting for the frequency-dependence of the threshold TMR sensitivity.

IV. RESULTS AND ANALYSIS

A. Experiment 1: Fixed masker level

1. Overall results

Panels A, B, and C in Fig. 1 present the data collected in Experiment 1, with the masker level fixed (Panels D, E, and F show the data from Experiment 2, discussed in Sec. IV B). The data are plotted as a function of the masker location (indicated by the position of the letter “M” in the icons along the abscissa). Two spatial configurations are plotted for each masker location, one with the target and masker colocated (open symbols) and one with the target displaced from the masker (filled symbols).[1] The spatially separated target was at the location indicated by the filled letter “T” in the icons along the abscissa. The thresholds for different modulation conditions are represented by different symbols.

Figure 1(a) shows the across-subject mean and standard deviation of the TMR at detection threshold (lower values correspond to better performance). Thresholds varied by more than 20 dB, depending on the spatial configuration and modulation condition. For a given modulation condition and masker location, performance when the target and masker were spatially separated (filled symbols) was always better than when they were colocated (open symbols), revealing robust SRM. The colocated thresholds for target and masker at 0° and −45° were nearly identical, suggesting that the exact spatial configuration of the target and masker was not important as long as the sources were colocated (this observation, based on the two configurations in Experiment 1, is further supported by the results of Experiment 2 in which three colocated thresholds were measured). In contrast, the spatially separated thresholds were strongly influenced by the specific target and masker locations: performance was worse with the masker at 0° than with the masker at −45° or 90° [compare the leftmost group of filled symbols in Fig. 1(a) to the center or the rightmost groups].

Within each spatial configuration, the no-modulation, in-phase comodulation, and target-only modulation (circles, triangles, and squares, respectively) thresholds were generally comparable, and these thresholds were higher (performance was worse) than the remaining thresholds. Masker-only modulation yielded improvements in performance (pentagrams fall below circles), while out-of-phase modulation of the target gave the lowest thresholds (hexagrams tend to fall below pentagrams).

A three-way repeated-measures analysis of variance (ANOVA) was performed with factors of modulation, spatial separation (colocated versus separated), and masker location (0, −45°), paralleling the layout of Fig. 1(a). The (M 90°, T 0°) configuration was omitted because it had no corresponding colocated measurement. This statistical analysis found a significant modulation × separation interaction ($F_{4,24}=7.63$, $p=0.0004$), a significant separation × masker location interaction ($F_{1,6}=950$, $p<0.0001$), and significant effects of all three main factors ($p<0.0001$). Notably though, neither the interaction between modulation and masker location nor the three-way interaction was significant ($p>0.1$). These results suggest that, although overall performance and the effect of separation depend on spatial configuration, at least for the spatial configurations explored in this study, the effect of modulation on the thresholds is similar within each spatial configuration rather than varying with target and masker locations.

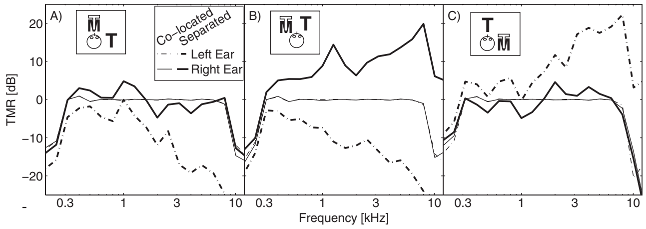

FIG. 2. TMR in 1/3-octave frequency bands in the six target/masker spatial configurations used in this study. Each panel shows the left- and right-ear TMRs in the colocated and separated spatial configurations for one masker location (indicated by the inset) as a function of the center frequency of the third-octave filters.

2. Energy effects in 1/3*-octave bands*

One factor contributing to the large spatial benefits and to the dependence of these improvements on spatial configuration is the better-ear advantage, arising from the changes in the level at which the stimuli are received at the left and right ears when target and masker are spatially separated. In general, spatial separation of the target and masker sources produces a larger TMR at one of the ears (the "better ear"), and a smaller TMR at the other ear, compared to when the sources are colocated (where the TMR is equal at the two ears). To explore the extent to which changes in TMR at the acoustically better ear could account for the observed spatial unmasking, we calculated the TMR in each of the signals reaching the listeners' two ears as a function of frequency.

For each spatial configuration, we selected a target and a masker processed by the appropriate HRIRs and filtered both target and masker into 22 log-spaced 1/3-octave signals per ear (ANSI, 1986). In this analysis, the target and masker were set to have the same level prior to spatial processing. (Note that the effects of spatial processing on the TMR at the ears are identical for all modulation conditions.) The resulting frequency-dependent TMRs show the proper correction needed, at each frequency, to calculate the TMR at detection threshold in each of the 22 frequency bands.[2] The results of this analysis are plotted in Fig. 2.

Each panel in Fig. 2 shows the TMRs for one fixed masker location (indicated by the inset icon), with each combination of the ear (solid versus dashed lines for right versus left ear, respectively) and the spatial configuration (thin versus thick lines for colocated versus spatially separated) plotted separately. (Note that the dashed and solid thin lines lie nearly on top of each other, so only the solid thin lines are easily visible.)

TMRs for the colocated configurations (thin lines) were approximately zero (or less than zero at the edges where no target energy was present), independent of the masker location (across panels) or the ear (solid versus dashed thin lines). The spatially separated TMRs were frequency dependent and varied both with the ear (solid versus dashed thick lines within each panel) and with the masker location (panel A versus panel B versus panel C). The largest improvement in TMR with spatial separation was approximately 5 dB in panel A (right-ear channel centered at 1 kHz), approximately 20 dB in panel B (right-ear channel centered at 8 kHz), and approximately 22 dB in panel C (left-ear channel centered at 8 kHz). Assuming that the listeners detect the target by detecting its presence due to the energy effects in the frequency channel with the most favorable TMR, detection performance with spatial separation is expected to improve due to the spatial configuration by an amount equal to the maximum TMR shown in each panel of Fig. 2. Note that this analysis assumes that, in each condition, performance is determined solely by the single frequency channel with the most favorable TMR and that the threshold TMR calculated in 1/3-octave band is the same for all frequency channels. Therefore, this analysis ignores possible contributions of across-frequency integration and binaural processing. Moreover, the exact TMRs computed in this way will depend on the detailed shapes of the peripheral auditory filters used, as well as how they change with center frequency, so that slightly different corrections would be found with different filter assumptions. However, this analysis provides a first-order correction for the wide variation in TMR with frequency caused by HRIR processing.

Figure 1(b) shows the threshold TMRs in the best frequency channel, determined by adding the best-channel correction (i.e., the peak values from Fig. 2) to the respective thresholds in Fig. 1(a). Colocated thresholds [open symbols in Fig. 1(b)] were essentially unchanged, as the TMR correction was near zero at all frequencies. However, correction of the spatially separated configurations reduced the effect of spatial separation to the point that many spatially separated thresholds (e.g., all thresholds with masker at −45°) were actually higher (performance was worse) than the corresponding colocated thresholds. Although this correction removed a good portion of the spatial effects on performance, ANOVA performed on the better-ear, best-frequency corrected thresholds found the same significant main factors and interactions as did the uncorrected thresholds [Fig. 1(a)] suggesting that the correction, while reducing the dependence of thresholds on the masker location, did not account for all of the variation in performance with spatial configurations.

3. Additional correction for frequency dependence of threshold TMR

The better-ear best-frequency correction yielded threshold TMRs that were much more similar than the uncorrected TMRs. To the extent that this correction was sufficient to account for the behavioral results, it suggests that (a) the threshold TMR is the same in all channels independent of frequency, (b) a simple 1/3-octave filter is an adequate representation of auditory filtering for the current analysis, and (c) there is no contribution of across-frequency integration or binaural processing to performance. The effect of any deviation from these assumptions is likely to depend on the spectral profiles of the target and masker signals, which differ with spatial configuration (see Fig. 2).

We now examine the assumption that threshold TMR in 1/3-octave band is constant as a function of frequency. In a previous study that measured SRM for broadband chirp-train signals masked by noise, threshold TMRs for narrowband targets were not constant as a function of frequency; instead, threshold TMRs were lower for higher-frequency targets (Kopco, 2005). When listening in a 9 kHz channel, best-channel analysis based on 1/3-octave filtering yielded thresholds that were nearly 4 dB lower than threshold TMRs using a 1 kHz channel. A simple frequency-dependent linear correction fit these earlier results relatively well (Kopco, 2005). The same correction, derived from the empirical fit to the data in this previous study, was applied to the current results:[3]

$$\mathrm{TMR}_{\mathrm{corrected}} = \mathrm{TMR}_{\mathrm{uncorrected}} + k_l \mathrm{CF} + k_2 . \qquad (1)$$

Here, $\mathrm{TMR}_{\mathrm{uncorrected}}$ are the data from Fig. 1(b), CF is the center frequency of the best-TMR filter in Hz, the constant k_1 was fitted to Kopco's (2005) data (k_1 was estimated to be -4.9×10^{-4} dB/Hz), and the constant k_2 was arbitrarily set to 1.34 dB to minimize the offset of the corrected data from the raw colocated data. (Note that the constant k_2 does not influence relative comparisons, as it shifts all data points by the same amount, but simply accounts for the absolute value of the TMR threshold). The frequency-corrected best-TMR model uses the same assumptions as the best-channel TMR correction shown in Fig. 1(b), except that it relaxes the assumption of a constant frequency-independent threshold TMR sensitivity. Instead, threshold TMR is assumed to decrease linearly with increasing center frequency.

Figure 1(c) shows the thresholds corrected by Eq. (1). Compared to the graphs in Fig. 1(b), the corrected spatially separated thresholds [filled symbols in Fig. 1(c)] were always better than or equal to the corresponding colocated thresholds (open symbols). Thresholds were roughly equal across all masker locations [in Fig. 1(c), the M 0°, T 90° thresholds were approximately equal to the corresponding M −45°, T 45° thresholds, as well as to the M 90°, T 0° thresholds; the trend was confirmed by data shown in Fig. 1(f) from Experiment 2]. Because the same correction was applied to all thresholds for a given spatial configuration, independent of the modulation condition, colocated thresholds still changed more as a function of the modulation condition than did the spatially separated thresholds. (Supporting these observations, ANOVA performed on the corrected data only found one significant interaction, modulation × separation, $F_{4,24}=7.65$, $p<0.0005$; all three main effects were significant, with $p<0.05$.) With these corrections, the spatially separated thresholds were only consistently lower than colocated thresholds in the no-modulation, in-phase modulation, and target-only modulation conditions [filled versus open circles, triangles, and squares in Fig. 1(c)]. Colocated and spatially separated thresholds were statistically indistinguishable in the masker-only modulation and out-of-phase modulation conditions for all spatial configurations.

Given the similarity of the corrected best-channel threshold TMRs at different masker locations [Fig. 1(c)], there only appears to be a modest effect of across-frequency integration in this study (i.e., there are no large differences across different spatial configurations, even though the best frequency and the overall shape of the better-ear TMR as a function of frequency vary dramatically with spatial configuration). Similarly, spatial processing only appears to contribute when the masker is modulated in a way that does not provide glimpses of the target (in the no modulation, in-phase modulation, and target-only modulation conditions).

In all of the following sections, the frequency-corrected best-channel TMR thresholds [from Figs. 1(c) and 1(f)] are used because (1) this correction accounts for the dependence of the thresholds on the masker location; (2) even though consideration of binaural processing and across-frequency integration could also produce corrections that explain some of the variability as a function of the masker location,[4] parsimony argues that these factors played only minor roles in this experiment; and (3) the fact that spatially separated configurations produce thresholds that depend less on the modulation condition than do colocated configurations is independent of the method used to account for energy effects or of the masker location. (However, note that it is currently not clear what causes the frequency dependence of the 1/3-octave filtered threshold TMRs.)

4. Results collapsed across the masker location

To better assess the interaction between modulation and separation, Fig. 3 shows the data collapsed across masker location. Figure 3(a) plots the across-subject mean threshold TMRs in the best 1/3-octave channel (and within-subject standard deviation, chosen here because it removes the between-subject differences from the computation of stan-

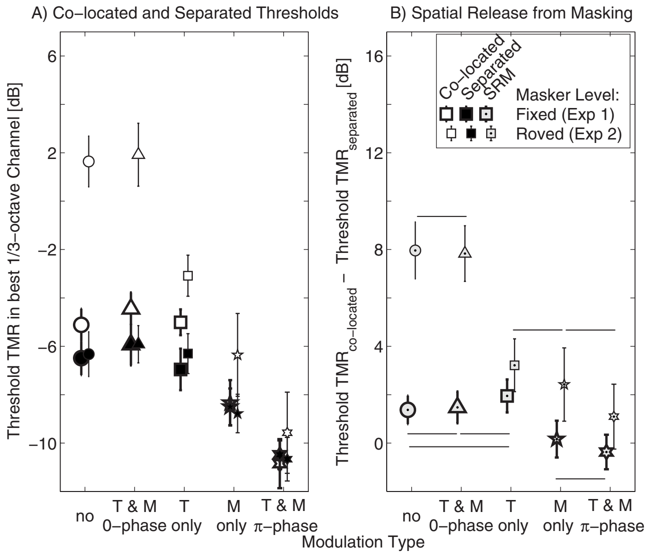

FIG. 3. Threshold TMRs in the frequency-corrected best 1/3-octave channel (panel A) and SRM (panel B) as a function of the modulation type, averaged across the masker locations (error bars give the within-subject standard deviation). The horizontal lines in panel B indicate SRMs that were not significantly different at the 0.01 level in a t-test after correcting for multiple comparisons (lines below large symbols for Experiment 1; lines above small symbols for Experiment 2). Different symbols are used to identify the modulation type, as in Fig. 1. The legend in panel B applies to both panels and all modulation conditions.

dard deviation[5]) as a function of the modulation type. The large filled and open symbols represent the spatially separated and colocated thresholds, respectively (the small symbols represent the results of Experiment 2, discussed in Sec. IV B).

The effect of modulation on performance was similar for colocated and separated spatial configurations. Thresholds were essentially the same for the no-modulation, in-phase comodulation, and target-only modulation conditions [compare large open and filled circles, triangles, and squares in Fig. 3(a)]. Performance with masker-only modulation (pentagrams) and out-of-phase modulation (hexagrams) was better, with lower thresholds.

Although the rank ordering of thresholds was the same for colocated and spatially separated conditions, the dependence of the thresholds on modulation was slightly stronger when the sources were colocated than when they were spatially separated (large open symbols span a range of nearly 7 dB, while the large filled symbols span a range of about 4 dB), suggesting that spatial separation affects performance differently for different modulation conditions. This SRM [the difference between the open and filled symbols in Fig. 3(a)] is plotted as a function of the modulation condition in Fig. 3(b). This panel shows the across-subject mean (and the within-subject standard deviation[5]) of the difference between the spatially separated and corresponding colocated thresholds from panel A.

One-way repeated-measures ANOVA found a significant effect of modulation on SRM ($F_{4,24}=11.44$, $p<0.0001$). The results of Bonferroni-corrected *post hoc* pairwise t-tests (which account for heterogeneity of variances; e.g., Ury and Wiggins, 1971) as implemented in the CLEAVE package (Herron, 2005) are also shown in Fig. 3(b). The horizontal lines under the large symbols in Fig. 3(b) indicate those pairs of conditions in Experiment 1 that did not differ at the $p<0.01$ significance level (all other pairs were significantly different from one another). The no-modulation, in-phase modulation, and target-only modulation SRMs were not significantly different from one another. Similarly, the masker-only modulation versus out-of-phase modulation SRMs were not significantly different from one another. However, the modulation type had a small but significant effect on the SRM: compared to no-modulation, in-phase modulation, or target-only modulation [circle, triangle, and square in Fig. 3(b)], modulating only the masker (pentagram) or modulating the target and masker stimuli with opposite phases (hexagram) decreased the SRM by roughly 1.5–2 dB ($p<0.01$), resulting in no benefit of spatial separation in the latter modulation conditions.

Finally, as discussed in the Appendix, learning affected SRM: in the first of the five repeats of this experiment, the SRM was essentially the same for all types of modulation (the largest difference was less than 1 dB). However, by the fifth repeat, the difference between the target-only modulation and the out-of-phase modulation grew to more than 4 dB. Thus, the average effect plotted in the data collapsed across the repeats is smaller than might be seen after extensive training.

B. Experiment 2: Masker level roved

To isolate the contribution of the overall level cue to performance, Experiment 2 was performed with the masker level roved between the intervals within a trial, a strategy used extensively in the profile analysis literature (Mason *et al.*, 1984; Kidd *et al.*, 1989). The (*T* 90°, *M* 90°) colocated condition was added to balance the number of colocated and spatially separated conditions; otherwise, Experiment 2 was identical to Experiment 1, except with a random ±5 dB intensity rove added from interval to interval.

1. Overall results

Panels D, E, and F in Fig. 1 present the results of Experiment 2 in a format identical to Experiment 1 (see Sec. IV A). The raw data in Fig. 1(d) followed the same trends as in Experiment 1. The spatially separated thresholds (filled symbols) were almost identical to those found in Experiment 1. The colocated thresholds for the no-modulation (circles) and in-phase modulation (triangles) conditions tended to be worse than in Experiment 1. However, the level rove had little effect on the remaining colocated configurations (a direct comparison is presented below). This result suggests that overall level was the main cue used for detection only in the colocated configurations in which the target and masker had identical temporal envelopes, a conclusion that was confirmed by a comparison of the data in panels E and F to respective panels B and C. (ANOVAs performed on the raw and corrected Experiment 2 data from panels D, E, and F found the same significant main effects and interactions as the respective ANOVAs performed on the Experiment 1 data.)

2. Results collapsed across the masker location

In order to analyze the interaction between modulation and spatial separation, the data were collapsed across the masker locations. To allow a direct comparison of the effect of masker level uncertainty, Fig. 3 shows the results for Experiment 2 (the small symbols slightly offset to the right) plotted alongside the data from Experiment 1 (larger symbols).

The filled symbols in Fig. 3(a) show the spatially separated thresholds. Roving the masker level had essentially no effect on any of the spatially separated thresholds (compare the small and large filled symbols from Experiments 2 and 1, respectively). In contrast, all colocated thresholds were larger in Experiment 2 than in Experiment 1 (the small open symbols fell above the corresponding large open symbols). The largest increase (around 7 dB) was observed when the target and masker had identical temporal envelopes (i.e., in the no-modulation and in-phase comodulation conditions; circles and triangles). In the three remaining modulation conditions, the masker-level rove increased thresholds by approximately 2 dB.

Figure 3(b) shows that, as a consequence of the effects of the level rove on the colocated configurations, the SRM was much larger in Experiment 2 than in Experiment 1 in the conditions in which the target and masker had the same temporal envelope. A one-way repeated-measure ANOVA revealed a significant effect of modulation on the SRM ($F_{4,24} = 35.55$, $p < 0.0001$). Bonferroni-corrected *post hoc* pairwise t-tests found no significant differences between unmodulated and comodulated SRMs, target-only and masker-only modulated SRMs, or the masker-only and out-of-phase modulated SRMs (see the horizontal bars above the pairs of small symbols that were not significantly different; $p > 0.01$). All other pairs of modulation conditions showed statistically significant differences.

The results in Fig. 3 suggest that overall level was used to detect the target when the colocated target and masker had the same envelope. For wideband noise, the smallest detectable intensity change ΔI is proportional to the base line intensity, I, so that $\Delta I/I$ is approximately constant with values between -9 and -6 dB over a large range of I (20–100 dB above the absolute thresholds; Moore, 2003). The results for colocated identically modulated stimuli in Experiment 1 match these data well, with TMR thresholds of approximately -5 dB [large open circles and triangles in Fig. 3(a)]. If overall level was the only available cue in this two-alternative forced-choice task and the external noise of the 10 dB rove dominated performance, then the TMR at detection threshold would be 1.07 dB for an ideal observer (Durlach *et al.*, 1986; Green, 1988), which is remarkably close to the actual thresholds observed for the identically modulated and in-phase modulated conditions, where threshold TMRs were around 2 dB. In most previous studies of the effect of rove on profile analysis, the rove yielded performance that was worse than was predicted for an optimal observer (Spiegel *et al.*, 1981; Mason *et al.*, 1984). Thus, even the fact that thresholds are slightly higher than the ideal-observer prediction is consistent with past work. Moreover, the no-modulation and in-phase comodulation thresholds were very similar to each other, suggesting that the fluctuating envelope in the latter condition did not make it harder to judge the levels in the different intervals.

In conditions for which target and masker were colocated but had different temporal envelopes, performance was much better than would be predicted if the main cue used for target detection was overall intensity, showing that some other nonlevel cue was the main feature used to detect the target. Nevertheless, in such conditions, the rove interfered slightly with performance, a result that suggests that the intensity rove made it more difficult for listeners to extract whatever feature was the main detection cue when target and masker were colocated.

C. Modulation detection

To understand the effects of modulation on performance, two analyses were performed. First, the instantaneous TMR was analyzed. In this analysis, predictions were based on detecting the target by hearing its effect at the best instant in time. A second analysis assumed that the listeners detected the target+masker interval by detecting a modulation depth that was different from the masker-only modulation (in the nontarget intervals).

1. Listening at peaks and dips: Instantaneous TMR analysis

The presence of modulation in the stimuli caused the instantaneous TMR to change over time. Humans appear to utilize these changes and detect the target in moments when the TMR is most favorable, both in monaural (Buus *et al.*, 1996) and binaural (Buss *et al.*, 2003) listening tasks, even though this ability can differ across subjects (e.g., see Buss *et al.*, 2007). Of course, given that the ability to utilize these cues is limited by the temporal resolution of the auditory system, factors like forward masking are likely to influence the ability to listen in dips (Widin *et al.*, 1986; Wojtczak and Viemeister, 2005). While the present analysis does not consider these limitations, it does provide an upper limit on how much the listeners could have benefited from changes in the instantaneous TMR. Specifically, if one assumes that the peak TMR produced after temporal integration over some fixed time window predicts performance, the current analysis gives the limit of performance if temporal resolution is infinitely precise, leading to an effective time window that is infinitely narrow. Conversely, the overall-TMR analysis shown in Fig. 3(a) shows predictions for an infinitely long time window. Any finite-length time window must produce results intermediate between these two extremes.

In the colocated conditions with identical modulation (no modulation and in-phase comodulation; circles and triangles), the TMR was constant over the duration of the stimulus. In the conditions with different target and masker modulations, the difference between the long-term TMR and the peak instantaneous TMR depended on which stimulus was modulated. Because the modulation envelope was sinusoidal in pressure units, the effect of modulation on the instantaneous sound pressure level was not symmetrical in decibel units. For sinusoidal modulation with a modulation

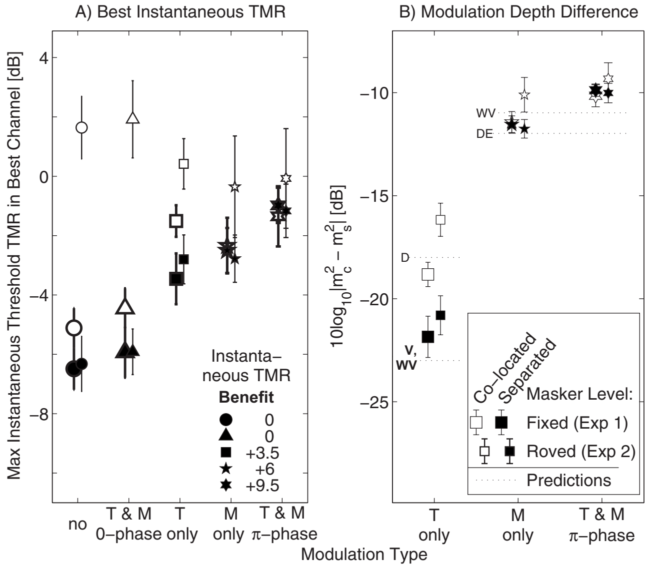

FIG. 4. (A) Peak instantaneous TMRs at threshold in the best 1/3-octave channel [derived from Fig. 3(a) by applying the instantaneous TMR benefit corrections, listed in the inset, to both colocated and separated thresholds of both Experiments 1 and 2] (B) Modulation depth (across-subject means and within-subject 95% confidence interval) at the threshold TMR in the three modulation conditions in which modulation of the target and masker differed. Data are compared to predictions based on the data of Wakefield and Viemeister (1990)—WV, Dau and Ewert (2004)—DE, Viemeister (1979), and Dau (1996)—D. The legend of panel B applies to data in both panels and to all modulation conditions.

depth of 0.5 (used in this study), the instantaneous signal level at the minima of the modulation envelope was 6 dB lower than the level with no modulation, while the level at the peaks of the modulation envelope was 3.5 dB higher than the unmodulated level.

Figure 4(a) plots the best instantaneous TMR in the frequency-corrected best 1/3-octave channel at threshold, determined by adding the instantaneous-TMR-benefit corrections (described above and listed in the inset) to the long-term frequency-corrected TMR thresholds in the best frequency channel [from Fig. 3(a)]. [Note that for each modulation condition, colocated and spatially separated thresholds have the same instantaneous-TMR-benefit correction, so that this correction does not influence SRM, shown in Fig. 3(b).]

As seen in Fig. 4(a), the peak instantaneous TMR at detection threshold falls between −4 and 0 dB for the conditions in which the target and masker envelopes differ [target modulation, masker modulation, and out-of-phase modulation conditions; large open squares, pentagrams, and hexagrams in Fig. 4(a)]. These values are higher than the intensity just noticeable difference (JND) (−9 to −6 dB, as discussed above), suggesting that listeners were unable to make use of the peak instantaneous TMR to detect the target based on changes in overall intensity. Given that the long-term average TMR does not capture the differences in thresholds as a function of modulation type [if it did then the thresholds represented by the large open squares, pentagrams, and hexagrams would be constant in Fig. 3(a)], while the instantaneous TMR predicts performance that is too poor (even though it is approximately constant), it is possible that predictions based on the TMR averaged over an appropriate finite-length time window could account for detection based on changes in intensity. However, if performance were based on the same intensity cue for cases when target and masker had the same envelope and cases when the target and masker envelopes differed, the effect of intensity rove should be similar in all conditions. Instead, intensity rove affected performance in the different conditions very differently, suggesting that some cue other than overall intensity integrated over some finite-duration time window enabled target detection when target and masker envelopes differed.

2. Effect of the target on the masker envelope modulation

One attribute that is affected by the addition of the target to the masker is the shape of the total stimulus envelope (Dau *et al.*, 1997). The salience of any change in the envelope due to the presence of the target depends on the relative levels of the target and masker as well as on the modulation condition. In the target-only-modulated condition, modulation is only present in the target interval and listeners may detect the target by detecting the presence of modulation. In the masker-only-modulated and the target-and-masker-modulated-out-of-phase conditions, the addition of the target decreases modulation depth from the 0.5 depth in the nontarget intervals and listeners may discriminate changes in the modulation depth to detect the target.

Detection and discrimination thresholds for modulation can be expressed as the modulation index $10\log_{10}|m_c^2 - m_s^2|$, where m_s represents the modulation depth of the standard (i.e., in the nontarget interval) and m_c is the modulation depth of the stimulus at discrimination threshold (i.e., the modulation depth of the combined target+masker signal in the target interval). The current target-modulated thresholds can be estimated either from previous modulation detection data (Viemeister, 1979; Dau, 1996) or from discrimination data using a standard with a very low modulation depth (Wakefield and Viemeister, 1990, and Dau and Ewert, 2004; summarized in Fig. 2 of Dau and Ewert, 2004).[6] For modulation detection, the modulation index at threshold is in the range from −23 dB (Viemeister, 1979) to −18 dB at threshold (Dau, 1996). The results from modulation discrimination experiments (Wakefield and Viemeister, 1990) suggest that modulation index thresholds are near −23 dB for standard depths less than −30 dB.

Thresholds from a previous modulation discrimination study (e.g., Dau and Ewert, 2004) can be linearly approximated as $10\log_{10}(m_c^2 - m_s^2) = 10\log_{10} m_s^2 - 4$, from which the predicted threshold for a decrease in modulation from the standard of $m_c = -6$ dB can be estimated as $10\log_{10}(m_c^2 - m_s^2) = -11$ dB [thresholds from Wakefield and Viemeister (1990) are approximately 1 dB larger than the Dau and Ewert (2004) thresholds when analyzed in this way].

In order to compare the current data to these predictions, the relationship between the threshold TMRs and the modulation depth of the combined stimulus was examined for our stimuli. However, combining a SAM noise and an unmodulated noise does not produce a stimulus with sinusoidal amplitude modulation. The relation between the threshold modulation and threshold TMR was estimated by determining the maximum and minimum amplitudes, of the combined stimulus envelope and then finding the modulation depth of a SAM stimulus that would give the same maximum and minimum (although the exact shape of the modulation envelope differs, the difference is relatively small, especially near threshold). The resulting relationships for the three differential modulation conditions in this study (and for the target and/or masker modulation of 0.5) are as follows.

In target-only modulated,

$$m = \frac{\sqrt{1 + 1.5^2\mathrm{TMR}^2} - \sqrt{1 + 0.5^2\mathrm{TMR}^2}}{\sqrt{1 + 1.5^2\mathrm{TMR}^2} + \sqrt{1 + 0.5^2\mathrm{TMR}^2}}.$$

In masker-only modulated,

$$m = \frac{\sqrt{1.5^2 + \mathrm{TMR}^2} - \sqrt{0.5^2 + \mathrm{TMR}^2}}{\sqrt{1.5^2 + \mathrm{TMR}^2} + \sqrt{0.5^2 + \mathrm{TMR}^2}}.$$

In stimuli modulated out of phase,

$$m = \frac{\sqrt{1.5^2 + 0.5^2\mathrm{TMR}^2} - \sqrt{0.5^2 + 1.5^2\mathrm{TMR}^2}}{\sqrt{1.5^2 + 0.5^2\mathrm{TMR}^2} + \sqrt{0.5^2 + 1.5^2\mathrm{TMR}^2}},$$

where TMR is the threshold TMR in the best channel (from Fig. 3) in pressure units and m is the threshold modulation depth of an equivalent SAM noise. These equations can be inverted to estimate the target+masker modulation depth at target detection threshold for the measured results.

Figure 4(b) shows data for the three modulation conditions in which target modulation is different from the masker modulation, expressed as the difference in modulation depth between the target+masker interval and the reference masker-alone interval (the modulation conditions for which the target and masker have the same envelope were not included in this analysis because there is no change in modulation with addition of the target). Also shown are the predictions estimated from results of Viemeister (1979), Wakefield and Viemeister (1990), Dau and Ewert (2004), and Dau (1996; see dashed lines).

The thresholds for the colocated stimuli with fixed masker levels (open large symbols) generally match the previous detection and discrimination data fairly well for all three types of modulation, suggesting that the listeners detected changes in modulation depth in these conditions. The spatially separated thresholds are only lower (detection is easier) than the colocated thresholds in the target-modulation condition, when the listeners do not ever get a good “glimpse” of the target (large filled versus open squares). At first glance, the fact that the spatially separated thresholds fall within the range of the previous modulation detection data (i.e., between the dotted lines marked by D and V, VW) seems to suggest that the listeners did not benefit from spatial cues in this condition. However, given the large difference between the D and the V, VW thresholds, and given that there is a consistent difference between the colocated and spatially separated thresholds in the current study, it is clear that the listeners did use the spatial separation cue, in addition to modulation, here.

Finally, although the effect is small, colocated roved thresholds (open small symbols) consistently fall above the range of thresholds observed in previous studies which did not rove the stimulus presentation level. This shows that overall level rove impaired the listeners’ ability to detect or discriminate modulation in the current study.

V. DISCUSSION

Noise-on-noise threshold TMRs changed over a range of 30 dB [Figs. 1(a) and 1(d)], and were influenced by the spatial configuration of the target and masker, the type of modulation present in the stimuli, and a rove of the masker level. Moreover, as discussed in the Appendix, these differences appear to increase with experience. A large part of the variability in performance across the tested conditions (as much as 20 dB) came from the changes in the target and masker energy levels received at the ears when the target and masker locations changed. Specifically, if one considers the TMR within the best 1/3-octave frequency channel in the acoustically better ear, threshold TMRs ranged only over 5 dB across different spatial configurations. If one then corrects these detection thresholds based on the detection threshold differences across frequency,[3] threshold TMRs were even closer, spanning a range of only about 1 dB across the different spatial configurations for a given modulation condition.

As shown in Fig. 2, the way in which TMR varies with center frequency differs dramatically across the spatial configurations used in this study. Therefore, any contributions of ITD and across-frequency processing to performance are likely to depend on masker location. However, no large differences were observed after applying frequency-dependent corrections to the TMR in the best frequency channel. Thus, for the broadband stimuli used here, both binaural and across-frequency contributions to performance appear to be modest. Frequency-dependent TMR thresholds could also explain the results of a previous related experiment without considering any across-frequency integration or binaural processing (Lane *et al.*, 2004). Together, these results suggest that low-level binaural processing does not contribute very much to spatial unmasking when detecting a broadband target in a broadband masker (although it can contribute significantly when the target is narrowband; e.g., see Kopco and Shinn-Cunningham, 2003).

The benefit of spatial separation found in the current results is similar for all spatial configurations, even though the best frequency channel is sometimes in a low-frequency region where binaural processing is expected to provide a large benefit and sometimes in a high-frequency region where binaural processing typically provides much more modest benefits (Zurek, 1993; Kopco and Shinn-Cunningham, 2003). This suggests that differences in the perceived *spatial attributes* of the stimuli (which depend both on low-frequency ITDs as well as high-frequency interaural level differences and spectral cues) are responsible for the spatial unmasking not explained by changes in the TMR at the better ear, rather than *binaural processing* that operates primarily at low-frequencies (unmasking caused by interaural decorrelation; Colburn, 1977); (see Freyman *et al.*, 1999, for another study contrasting how spatial perception and binaural processing contribute to spatial unmasking).

Both modulation and intensity rove influenced the SRM, defined as the difference between the best-channel threshold TMRs with colocated and spatially separated stimuli. With the masker level fixed, SRM was comparable for no-modulation, target and masker in-phase modulation, and target-only modulation configurations, but SRM was statistically insignificant when only the masker was modulated or target and masker were modulated out of phase [see Fig. 3(b)]. Uncertainty about the masker level increased SRM in all modulation conditions, but the size of this effect depended on the modulation in the stimuli. For the level-roved stimuli, SRM was 7 dB larger when the target and masker have the same temporal envelope, but only 2 dB larger when the stimuli had different modulation. These results can be understood by considering how and when listeners use overall level, modulation, and spatial cues to detect the presence of the target.

A. Overall level

Detection in the colocated, identically modulated conditions [i.e., when neither modulation nor spatial cues were available for target detection; open circles and triangles in Fig. 3(a)] appears to be based on detecting changes in overall intensity. This conclusion is supported by (1) the observed good match between thresholds in these conditions and predictions from previous intensity JND studies (Experiment 1) and (2) the effect of the intensity rove in these conditions (Experiment 2), which increased detection thresholds to just above that expected for an ideal observer using overall level as the detection cue (Green, 1988). (However, note that there were small gating asynchronies and spectral differences between the target and masker signals that could have contributed to the detection of colocated identically modulated targets.)

B. Space cue alone

When stimuli differed in their spatial locations but not in their modulation [filled circles and triangles in Fig. 3(a)], a consistent improvement in performance was observed, showing that spatial separation provided benefits beyond the improvements in the better-ear TMRs. Changes in the spatial attributes of the target+masker versus masker-only stimuli (such as perceived spatial width) likely were used to detect the target at threshold, a conclusion particularly supported by the fact that the threshold was not influenced by the intensity rove [large and small filled circles and triangles are the same in Fig. 3(a)].

C. Modulation cue alone

Differences in the target and masker modulations led to some improvements in detection when the target and masker had the same location, but not in all conditions. Modulation led to lower thresholds when only the masker was modulated and when the target and masker were modulated out of phase, independent of whether the overall level was roved or not [compare open pentagrams and hexagrams to open circles and triangles in Fig. 3(a)]. When the level was roved, modulation also improved detection when only the target was modulated [compare small open square to small open circle and triangle in Fig. 3(a)]. However, when the level was fixed, the target-only modulation did not improve performance compared to when there were no modulation cues to detect the target [compare large open square to large open circle and triangle in Fig. 3(a)].

The intensity rove caused modest degradations in performance when colocated target and masker had different modulation envelopes, hinting that the listeners might have used the overall level cue (selected at the most favorable TMR instances) instead of the modulation cue in these conditions. However, given that the rove effects were much smaller than when target and masker had identical envelopes, and that the thresholds in these cases were better than (i.e., below) those predicted for an ideal observer using intensity increments to detect the target (Green, 1988), it is unlikely that the listeners used overall level to detect the presence of the target in these conditions [small open squares, pentagrams, and hexagrams in Fig. 4(a)]. Instead, it seems that roving overall level made it slightly harder to judge the changes in modulation caused by adding a target to a masker in these tasks. However, in the target-only modulation condition, the long-term TMR threshold is comparable to that for the no-modulation and in-phase modulation conditions when the level is fixed [large open square, triangle, and circle are comparable in Fig. 3(a)]. Moreover, when the level was not roved, the spatial separation improved performance by similar amounts when only the target was modulated and in the cases where the level was clearly the cue for detection (no modulation, in-phase modulation). Thus, for the target-only modulation condition, it is possible that the subjects used an overall level to detect the target when the level was roved and used a modulation to detect the target when the level varied randomly from interval to interval.

Another result hinting that the subjects' behavior might have been more complex than just detecting the modulation depth is that no similar effect of an intensity rove was seen in a previous study that measured modulation discrimination (Stellmack *et al.*, 2006). However, this difference in the effect of an intensity rove in the two studies may be due to the differences in the instructions given to subjects. In the previous study, listeners were instructed to detect changes in the modulation depth of a single stimulus, while in the current study they were presented with examples of the masker and target at the start of each block and instructed to detect the presence of the target. This priming may have enhanced the likelihood that listeners perceptually segregated the target from the masker in the current study, or that they switched cues between the rove and no-rove experiments, rather than detecting the target+masker interval by perceiving a change in masker attributes. However, further experiments are required to explore which of these alternatives is correct.

D. Space and modulation

Spatial separation did not always improve detection beyond performance for colocated sources after accounting for the TMR at the best frequency in the better acoustic ear. Specifically, spatial separation did not improve detection very much, other than by changing TMR, when the masker envelope had dips, providing good glimpses of the target (in the out-of-phase and masker-only modulation conditions). As noted above, in these conditions, listeners appear to have detected the target by detecting changes in the modulation depth between the masker-only and target+masker intervals, and spatial cues did not help in detecting these modulation changes. However, when the intensity rove was added in these conditions, the modulation-based colocated detection performance was impaired, while the spatially separated performance was not. Thus, spatial cues helped, bringing the spatially separated threshold to the no-rove levels, possibly by making it easier to use the modulation cue optimally.

When only the target was modulated, spatial cues provided a significant improvement in performance both when intensity was fixed across intervals in a trial (Experiment 1) and when intensity was roved (Experiment 2). For these stimuli, listeners were never given a good glimpse of the target, because the masker envelope was constant. In addition, the spatially separated thresholds were almost identical to the thresholds in the no-modulation and in-phase modulation conditions, and the size of the spatial benefit in the no-rove experiment was nearly identical to that in the no-modulation cue conditions. There are two possible explanations for the listeners' behavior in the target-only modulation condition when overall level was not roved. One possibility is that when the target and masker were colocated, listeners used an overall level to detect the target, and when target and masker were spatially separated, listeners used a spatial cue to detect the target. If so, then the modulation and spatial cues were subadditive in the target-only modulation case: listeners either used space or modulation. Alternatively, listeners may have used the modulation cue in the colocated target-only modulated condition and a combination of modulation and space cues in the spatially separated condition. If so, then spatial and modulation cues combined additively for this condition, but were combined subadditively in the masker-only and out-of-phase modulation conditions.

E. Final comments

After accounting for the better-ear acoustic benefit of spatial separation, the current study did not find any evidence for superadditive combination of modulation and space cues for detecting a broadband target embedded in a broadband masker. The results are consistent with two interpretations of the behavior when both cues were available and the level was fixed: (1) the subjects always used one of the cues, getting no benefit from the other one, or (2) the combination of modulation and space cues was additive when only the target was modulated, but the space cue contributed nothing to detection in the conditions in which the masker envelope was modulated and provided glimpses of the target. However, when the overall level was roved, spatial cues always helped performance when modulation was the main detection cue.

These results confirm the first of the proposed hypotheses (H1). The combined effect of modulation and spatial separation on detection is asymmetrical in that spatial separation improves detection performance more when the target is modulated and the masker is unmodulated than when the masker is modulated.

The results contradict our second hypothesis (H2). The combined effect of modulation and separation does not depend on the specific location of the target and masker, even though the contribution of binaural and across-frequency processing likely would vary in the different configurations. This result argues that the combined effect of modulation and spatial cues occurs at a stage that is later in the processing stream than the binaural processing occurring in the brainstem.

In contrast to the current stimuli, everyday auditory scenes contain objects that differ along many more dimensions than just their temporal envelopes and locations. It is difficult to extrapolate these findings to predict how modulation and spatial cues may interact for more complex stimuli. Nonetheless, it is likely that the main result, that modulation and space cues tend to contribute to detection subadditively, will also hold true for other stimuli differing in their spatial positions and modulation structure. However, it is also important to consider how our detection results compare to suprathreshold tasks, such as understanding speech embedded in fluctuating maskers. We find it intriguing that there is essentially no evidence for across-frequency integration in our experiments. In contrast, across-frequency integration is the basis of models that predict speech intelligibility in noise (e.g., see Zurek, 1993). We believe that the key difference between these results is that in our simpler detection task, any glimpse of the target (at any frequency) is sufficient for detection. In contrast, understanding speech requires the integration of information from different frequency bands and estimation of the absolute spectrotemporal content of the speech target. Thus, while the current results may be helpful in predicting how listeners detect a complex signal embedded in a competing fluctuating masker, they are only a first step in understanding how we analyze and understand the content of a complex signal in a setting containing multiple sound sources.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Institute on Deafness and Other Communication Disorders (5R01DC005778-03) and a grant from the Slovak Science Grant Agency (VEGA 1/3134/06). Rich Freyman gave a number of extraordinarily helpful suggestions as editor. In addition, two anonymous reviewers provided very helpful feedback on earlier versions of this work. The authors wish to thank H. Steven Colburn, Constantine Trahiotis, Les Bernstein, Bertrand Delgutte, Chris Mason, Eric Thompson, and Gin Best, for their helpful comments, and Jackie Jacobsen for help with data collection.

APPENDIX: LEARNING

Previous studies show that modulation detection performance improves with training over the course of hours (Wakefield and Viemeister, 1990; Dau and Ewert, 2004; Fitzgerald and Wright, 2005). In the current study, subjects did not receive extensive training prior to the experiment; each performed only one practice session in which thresholds for all conditions were measured once each (25 combinations of modulation and spatial configuration in Experiment 1 and 30 combinations in Experiment 2). To evaluate how learning influenced the results, data were analyzed as a function of the experimental session.

A three-way repeated-measure ANOVA was performed for both experiments on the data collapsed across masker locations [as in Fig. 3(a)], with factors of repeat (five levels), modulation type (five levels), and spatial separation (two levels). For Experiment 1, all two-way interactions were significant (repeat × modulation: $F_{16,96}=2.11$, $p=0.0134$; repeat × separation: $F_{4,24}=6.03$, $p=0.0017$; modulation × separation: $F_{4,24}=230$, $p<0.0001$), as were the main effects of modulation and separation ($p<0.0001$). For Experiment 2, the results were very similar (repeat × modulation: $F_{16,96}=1.69$, $p=0.062$; repeat × separation: $F_{4,24}=20.96$, $p<0.0001$; modulation × separation: $F_{4,24}=212$, $p<0.0001$; main effects of modulation and separation: $p<0.0001$). These results show that performance changes over time, and that the change depends on the specific combinations of modulation and of spatial separation.

Post hoc inspection reveals that the largest changes in SRM over time arose when only the target was modulated and when the target and masker were modulated out of phase. Panel A of Fig. 5 shows the thresholds for these conditions (target-only shown as squares; out-of-phase target and masker modulation shown as hexagrams), collapsed across the masker location and plotted as a function of the repeat, for both spatially colocated (open) and separated (filled) conditions. Panel B shows the SRM. The left-hand and right-hand panels show data from Experiments 1 and 2, respectively. Each symbol represents the across-subject mean (and within-subject 95% CI) of the thresholds obtained for one combination of repeat, spatial configuration, and modulation types.

Overall, TMR thresholds generally improved over time, as illustrated by the downward trend in all the graphs in panel A. However, a more detailed inspection shows that the size of this learning effect differed in the different conditions, and that these differences were consistent across the two experiments. When the stimuli were spatially separated, the target-only modulated thresholds (filled squares) improved by 2–3 dB over the five repeats, while the out-of-phase modulated thresholds (filled hexagrams) improved by 1 dB or less. On the other hand, when the stimuli were colocated, there was a roughly 3 dB improvement in the out-of-phase modulated thresholds (open hexagrams), while the improvement was negligible in the target-only modulated thresholds (open squares). As a result, the SRM tended to increase across sessions for target-only modulation stimuli but to decrease when the target and masker were modulated out of phase (panel B). Thus, while the SRMs for these two conditions differed by only about 1 dB in the first repeat, they differed by more than 4 dB by the fifth repeat.

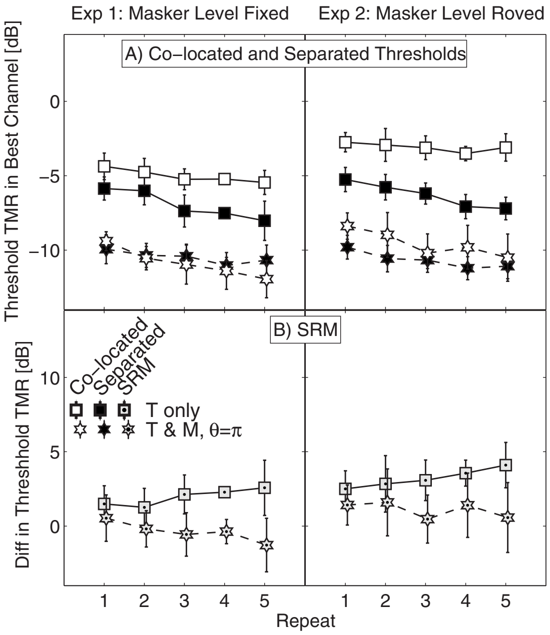

FIG. 5. Threshold TMR in the frequency-corrected best 1/3-octave channel (panel A) and the SRM (panel B) as a function of the measurement repeat in Experiment 1 (left-hand panels) and Experiment 2 (right-hand panels). Panel A: For each repeat, the data represent the across-subject mean (and within-subject 95% confidence interval) of the thresholds collapsed across the corresponding spatially separated or colocated conditions. Panel B: SRM, determined as the difference of the respective thresholds from Panel A.

At first glance, these changes seem difficult to understand. However, as discussed in the main text, spatial cues are generally not helpful for the out-of-phase conditions (hexagrams); in those conditions, performance is based on detecting (nonspatial) changes in modulation. The only effect of spatial cues in the out-of-phase modulation conditions was to make it easier to focus on this change in modulation (e.g., ignoring the distracting effects of intensity rove). Consistent with this, the main effect of learning in the out-of-phase modulation conditions is to improve how well listeners do when there are no spatial cues present and it is most difficult to focus attention on the modulation cue that underlies detection (open hexagrams).

In contrast, in the target-only modulation condition (squares), spatial cues provide a real advantage in target detection and allow detection at lower thresholds than when only monaural modulation and/or level cues are available. In these conditions, listeners improve most in their ability to use this subtle spatial cue (filled squares). However, listeners show little improvement in their ability to detect nonspatial changes in modulation or level with practice (open squares), perhaps because detection of modulation or detection of changes in level increases is a relatively simple task in which near-asymptotic performance is reached much faster (compared to the discrimination of modulation depth or detection of subtle spatial changes). As a result, SRM grows with time for the target-only modulation condition.

[1] The colocated spatial configuration with the target and masker at 90° was not measured in Experiment 1.

[2] In the case of a sinusoidal target, this correction can be computed by considering only the TMR change at the target frequency (Shinn-Cunningham *et al.*, 2005). If the relative contribution of each frequency to task performance is known for a broadband signal, the frequency-dependent TMR function can be used to predict performance (e.g., Zurek, 1993).

[3] This simple linear correction is purely phenomenological, rather than based on theoretical considerations. To the extent that this is the right correction to apply, it may reflect systematic deviations in the degree to which 1/3-octave filters approximate peripheral filtering as a function of frequency, differences in the internal noise of different frequency channels, or other systematic effects of frequency.

[4] Binaural and across-frequency processing may explain some of the dependence of the uncorrected thresholds on the masker locations. Specifically, in the spatially separated configuration of Fig. 2(a), the largest TMRs occur at low frequencies (below 2 kHz, full thick line) and the TMR profile in the right ear is relatively flat as a function of frequency. On the other hand, in the configurations of Figs. 2(b) and 2(c), the best frequency channel is at high frequencies and the TMRs vary significantly with frequency. These differences in the dominant spectral region suggest that binaural and across-frequency processing may contribute more to performance for the conditions of Fig. 2(a) than in the other two configurations, consistent with results in Figs. 1(b) and 1(e) (filled symbols in the *M* 0°, *T* 90° configuration are below the filled symbols for the other two configurations). However, while the binaural and across-frequency processing may explain why threshold TMRs tend to be lower when the masker is at 0° compared to the other configurations [leftmost versus middle and rightmost plots of Fig. 1(b)], they are not analyzed because (1) these factors cannot explain why some spatially separated thresholds are worse than the corresponding colocated thresholds in Figs. 1(b) and 1(e), and 2 the correction based on the frequency-dependent best-channel TMRs accounts for these differences, without considering binaural and across-frequency processing.

[5] Within-subject standard deviations are computed by subtracting out the mean performance (averaged across conditions) for each subject prior to the computation of variability. This method for computing variability is analogous to using subject as a factor in ANOVA analysis. In particular, the remaining variability shows how variable the across-condition results are after removing differences in overall performance across subjects. See the Appendix of Kopco *et al.* (2007) for further descriptions of this analysis.

[6] Comparisons of the current and previous results should be made with caution, as there are important differences in experimental procedures: for instance, none of the previous studies (Viemeister, 1979; Wakefield and Viemeister, 1990; Dau and Ewert, 2004; Dau, 1996) used the 40 Hz modulation frequency adopted in the present study. In addition, the current stimuli differ from the stimuli in the previous studies in their spectral content as they are filtered by the HRIRs.

ANSI (**1986**). *Specifications for Octave-Band and Fractional-Octave-Band Analog and Digital Filters, S1.1 (ASA 65-1986)* (American National Standards Institute, Inc., New York).

Asemi, N., Sugita, Y., and Suzuki, Y. (**2003**). “Auditory search asymmetry between pure tone and temporal fluctuating sounds distributed on the frontal-horizontal plane,” Acta. Acust. Acust. **89**, 346–354.

Bernstein, L. R., and Trahiotis, C. (**1994**). “Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise,” J. Acoust. Soc. Am. **95**, 3561–3567.

Bernstein, L. R., and Trahiotis, C. (**2002**). “Enhancing sensitivity to interaural delays at high frequencies by using “transposed stimuli”,” J. Acoust. Soc. Am. **112**, 1026–1036.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2003**). “The masking level difference for signals placed in masker envelope minima and maxima,” J. Acoust. Soc. Am. **114**, 1557–1564.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2007**). “Individual differences in the masking level difference with a narrowband masker at 500 or 2000 Hz,” J. Acoust. Soc. Am. **121**, 411–419.

Buus, S., Zhang, L., and Florentine, M. (**1996**). “Stimulus-driven, time-varying weights for comodulation masking release,” J. Acoust. Soc. Am. **99**, 2288–2297.

Colburn, H. S. (**1977a**). “Theory of binaural interaction based on auditory-nerve data. II: Detection of tones in noise,” J. Acoust. Soc. Am. **61**, 525–533; Colburn, H. S. (**1977b**). “Theory of binaural interaction based on auditory-nerve data. II: Detection of tones in noise. Supplementary material,” J. Acoust. Soc. Am. AIP document no. PAPS JASMA-61–525–98.

Dau, T. (**1996**). “Modeling auditory processing of amplitude modulation,” (Universität Oldenburg, Germany).

Dau, T., and Ewert, S. D. (**2004**). “External and internal limitations in amplitude-modulation processing,” J. Acoust. Soc. Am. **116**, 478–490.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). “Modeling auditory processing of amplitude modulation I. Detection and masking with narrow-band carriers,” J. Acoust. Soc. Am. **102**, 2892–2905.

Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). “Towards a model for the discrimination of broadband stimuli,” J. Acoust. Soc. Am. **80**, 63–72.

Fitzgerald, M. B., and Wright, B. A. (**2005**). “A perceptual learning investigation of the pitch elicited by amplitude-modulated noise,” J. Acoust. Soc. Am. **118**, 3794–3803.

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.

Gilkey, R. H., and Good, M. D. (**1995**). “Effects of frequency on free-field masking,” Hum. Factors **37**, 835–843.

Good, M. D., Gilkey, R. H., and Ball, J. M. (**1997**). “The relation between detection in noise and localization in noise in the free field,” in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. Gilkey and T. Anderson (Erlbaum, New York), pp. 349–376.

Green, D. M. (**1988**). *Profile Analysis. Auditory Intensity Discrimination* (Oxford University Press, New York).

Hall, J. W., III, Buss, E., and Grose, J. H. (**2006**). “Binaural comodulation masking release: Effects of masker interaural correlation,” J. Acoust. Soc. Am. **120**, 3878–3888.

Hall, J. W., III, Haggard, M. P., and Fernandes, M. A. (**1984**). “Detection in noise by spectrotemporal pattern analysis,” J. Acoust. Soc. Am. **76**, 50–56.

Herron, T. (**2005**). “C Language Exploratory Analysis of Variance with Enhancements,” (Janunary 30, 2005 version. URL: http://www.ebire.org/hcnlab/software/cleave.html, date last viewed: May 19, 2008).

Kidd, G., Jr., Mason, C. R., Brantley, M. A., and Owen, G. A. (**1989**). “Roving-level tone-in-noise detection,” J. Acoust. Soc. Am. **86**, 1310–1317.

Kopco, N. (**2005**). “Across-frequency integration in spatial release from masking,” *Forum Acusticum* (OPAKFI, Budapest, Hungary), pp. 1607–1612.

Kopco, N., Best, V., and Shinn-Cunningham, B. G. (**2007**). “Sound localization with a preceding distractor,” J. Acoust. Soc. Am. **121**, 420–432.

Kopco, N., and Shinn-Cunningham, B. G. (**2003**). “Spatial unmasking of nearby pure-tone targets in a simulated anechoic environment,” J. Acoust. Soc. Am. **114**, 2856–2870.

Lane, C., Kopco, N., Delgutte, B., Shinn-Cunningham, B., and Colburn, H. (**2004**). “A cat’s cocktail party: Psychophysical, neurophysiological, and computational studies of spatial release from masking,” in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. d. Cheveigne, S. McAdams, and L. Collet (Springer, Dourdan, France), pp. 405–413.

Lane, C. C., and Delgutte, B. (**2005**). “Neural correlates and mechanisms of spatial release from masking: Single-unit and population responses in the inferior colliculus,” J. Neurophysiol. **94**, 1180–1198.

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

Mason, C. R., Kidd, G., Hanna, T. E., and Green, D. M. (**1984**). “Profile analysis and level variation,” Hear. Res. **13**, 269–275.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego, CA).

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (**1991**). “Free-field release from masking,” J. Acoust. Soc. Am. **90**, 1355–1370.

Sheft, S., and Yost, W. A. (**1997**). “Binaural modulation detection interference,” J. Acoust. Soc. Am. **102**, 1791–1798.

Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). “Bottom-up and top-down influences on spatial unmasking,” Acta. Acust. Acust. **91**, 967–979.

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

Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). “Signal and masker uncertainty in intensity discrimination," J. Acoust. Soc. Am. **70**, 1015–1019.

Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2005**). "Monaural and interaural temporal modulation transfer functions measured with 5-kHz carriers," J. Acoust. Soc. Am. **118**, 2507–2518.

Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2006**). "Discrimantion of depth of sinusoidal amplitude modulation with and without roved carrier levels (L)," J. Acoust. Soc. Am. **119**, 37–40.

Sterbing, S. J., D'Angelo, W. R., Ostapoff, E.-M., and Kuwada, S. (**2003**). "Effects of amplitude modulation on the coding of interaural time differences of low-frequency sounds in the inferior colliculus. I. Response properties," J. Neurophysiol. **90**, 2818–2826.

Ury, H. K., and Wiggins, A. D. (**1971**). "Large sample and other multiple comparisons among means," Br. J. Math. Stat. Psychol. **24**, 174–194.

van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.

van de Par, S., and Kohlrausch, A. (**1998**). "Comparison of monaural (CMR) and binaural (BMLD) masking release," J. Acoust. Soc. Am. **103**, 1573–1579.

van de Par, S., Kohlrausch, A., Breebaart, J., and McKinney, M. (**2004**). "Discrimination of different temporal envelope structures of diotic and dichotic target signals within diotic wide-band noise," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 398–404.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Wakefield, G. H., and Viemeister, N. F. (**1990**). "Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise," J. Acoust. Soc. Am. **88**, 1367–1373.

Widin, G. P., Viemeister, N. F., and Bacon, S. P. (**1986**). "Effects of forward and simultaneous masking on intensity discrimination," J. Acoust. Soc. Am. **80**, 108–111.

Winter, I. M., Neuert, V., and Verhey, J. L. (**2004**). "Comodulation masking release and the role of wideband inhibition in the cochlear nucleus," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 321–327.

Wojtczak, M., and Viemeister, N. F. (**2005**). "Forward masking of amplitude modulation: Basic characteristics," J. Acoust. Soc. Am. **118**, 3198–3210.

Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. Studebaker and I. Hochberg (College-Hill Press, Boston, MA).

Zurek, P. M., Freyman, R. L., and Balakrishnan, U. (**2004**). "Auditory target detection in reverberation," J. Acoust. Soc. Am. **115**, 1609–1620.