Spatial unmasking of nearby pure-tone targets in a simulated anechoic environment

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

(Received 31 December 2002; revised 12 August 2003; accepted 15 August 2003)

Detection thresholds were measured for different spatial configurations of 500- and 1000-Hz pure-tone targets and broadband maskers. Sources were simulated using individually measured head-related transfer functions (HRTFs) for source positions varying in both azimuth and distance. For the spatial configurations tested, thresholds ranged over 50 dB, primarily as a result of large changes in the target-to-masker ratio (TMR) with changes in target and masker locations. Intersubject differences in both HRTFs and in binaural sensitivity were large; however, the overall pattern of results was similar across subjects. As expected, detection thresholds were generally smaller when the target and masker were separated in azimuth than when they were at the same location. However, in some cases, azimuthal separation of target and masker yielded little change or even a small increase in detection threshold. Significant intersubject differences occurred as a result both of differences in monaural and binaural acoustic cues in the individualized HRTFs and of different binaural contributions to performance. Model predictions captured general trends in the pattern of spatial unmasking. However, subject-specific model predictions did not account for the observed individual differences in performance, even after taking into account individual differences in HRTF measurements and overall binaural sensitivity. These results suggest that individuals differ not only in their overall sensitivity to binaural cues, but also in how their binaural sensitivity varies with the spatial position of (and interaural differences in) the masker. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1616577]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Qp [LRB] Pages: 2856–2870

I. INTRODUCTION AND BACKGROUND

When listening for a target sound in the presence of a masking sound, a listener's ability to detect the target is influenced by the locations of both target and masker. When target and masker are at the same distance, it is generally easier to detect or recognize the target when it is spatially separated from the masker compared to when the target and masker are at the same position. This ''spatial unmasking'' effect has been studied for many types of stimuli, including speech (e.g., see Freyman *et al.*, 1999; Shinn-Cunningham *et al.*, 2001), click-trains (e.g., see Saberi *et al.*, 1991; Good *et al.*, 1997), and tone complexes (e.g., see Kidd *et al.*, 1998).

For broadband noise maskers, spatial unmasking arises primarily from acoustic ''better-ear'' effects (moving a sound source in space changes the levels of the signal reaching the ears of the listener) and ''binaural'' effects. ''Better-ear'' effects lead to unmasking because the target-to-masker ratio (TMR) generally increases at one ear when target and masker are in different directions compared to when they are in the same direction. Binaural unmasking can occur when the interaural time and intensity differences in the target and masker differ.

There have been many studies of how binaural differences affect tone detectability in noise [see Durlach and Colburn (1978) for a review of this classic literature]. However, most of these studies were performed under headphones using interaural differences that do not occur naturally. There are only a few studies that have measured how tone detection is affected by the spatial locations of target and masker (examples include Ebata *et al.*, 1968; Gatehouse, 1987; Santon, 1987; Doll and Hanna, 1995). Moreover, results of these studies are inconsistent, finding spatial unmasking ranging from as little as 7 or 8 dB [Santon (1987) and Doll and Hanna (1995), respectively] to as much as 24 dB (Gatehouse, 1987). These apparent discrepancies may be caused by differences in the spatial configurations tested. However, none of these studies analyzed how the TMR at the ears changed with spatial configuration and did not factor out how better-ear (versus binaural) factors may have contributed to the observed spatial unmasking.

Previous studies of spatial unmasking for pure-tone targets considered sources relatively far from the listener and looked only at unmasking resulting from changes in source direction, ignoring any effects of source distance. For sources more than about a meter from the listener, the only significant effect of changing source distance is a change in signal level that is equal at the two ears. However, changes in source distance for sources within reach of the listener produce changes in signal level that differ at the two ears, resulting in exceptionally large interaural level differences (ILDs; see Brungart and Rabinowitz, 1999; Shinn-Cunningham *et al.*, 2000), even at low frequencies for which ILDs are essentially zero for relatively distant sources. In addition, for near sources, relatively small positional changes

[a)]Author to whom correspondence should be addressed: Department of Cognitive and Neural Systems, Boston University, 677 Beacon St., Room 311, Boston, MA 02215. Electronic mail: shinn@cns.bu.edu

can lead to large changes in the energy of the target and masker reaching the two ears. A few previous studies hint that, in some conditions, binaural performance can be worse than monaural performance using the better ear, particularly when there are large ILDs in the stimuli (e.g., see Bronkhorst and Plomp, 1988; Shinn-Cunningham *et al.*, 2001). Given that large ILDs can arise when sources are within reach of the listener, studies of binaural unmasking for nearby sound sources may shed light on these reports.

The current study examined spatial unmasking of pure tone sources within reach of a listener in a simulated anechoic environment. Individually measured head-related transfer functions (HRTFs) were used to simulate sources. This approach allowed realistic spatial acoustic cues to be presented to the subjects while still allowing detailed analyses of the stimuli reaching the subjects during the experiment. The main goals of the study were to (1) measure how target threshold depends on target and masker azimuth and distance for nearby sources, (2) characterize better-ear effects by analyzing how the TMR varies with the spatial configurations tested, (3) evaluate the binaural contribution to spatial unmasking, particularly for spatial configurations in which large ILDs arise, and (4) investigate the degree to which results can be accounted for by a model of binaural interaction.

II. SPATIAL UNMASKING OF NEARBY PURE TONE TARGETS

A. Methods

1. Subjects

Four graduate students with prior experience in psychoacoustic experiments (including author NK) participated in the study. One subject was female and three were male. Subject ages ranged from 25 to 28 years. All subjects had normal hearing as confirmed by an audiometric screening.

2. HRTF measurement

Individualized HRTF measurements were made with subjects seated in the center of a quiet classroom (rough dimensions of $5\times9\times3.5$ m; broadband T_{60} of approximately 700 ms). Subjects were seated with their heads in a headrest so that their ears were approximately 1.5-m above the floor. Measurements were taken for sources in the right front horizontal plane (at ear height) for all six combinations of azimuths (0°, 45°, 90°) and distances (0.15 m, 1 m) relative to the center of the head (defined as the intersection of the interaural axis and the median plane) as shown in Fig. 1.

The Maximum-Length-Sequence (MLS) technique (e.g., see Vanderkooy, 1994) was used to measure HRTFs. Two identical 32 767-long maximum length sequences were concatenated and presented through a small loudspeaker using a 44.1-kHz sampling rate (details regarding the equipment are described below). The response to the second sequence was recorded.[1] This measurement was repeated ten times and the raw measurements averaged in the time domain. This average response was then used to estimate a 743-ms-long head-related impulse response.

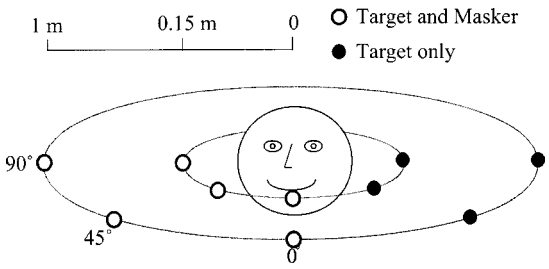

FIG. 1. Spatial positions used in the study. HRTFs were measured at the positions denoted by open symbols. Target detection thresholds were measured for all spatial combination of six masker positions (open symbols) and ten target positions (filled and open symbols; targets simulated at the filled symbols used the corresponding HRTFs from the contralateral hemifield with left- and right-ear signals reversed).

HRTFs were measured using a Tucker-Davis Technologies (TDT) signal processing system under computer control. For each measurement, the concatenated MLS sequence was read from a PC hard-drive and sent to a TDT D/A converter (TDT PD1), which drove a Crown amplifier connected to a BOSE mini-cube loudspeaker. At the start of the measurement session, the subject was positioned so that the center of his/her head was at a location marked on the floor of the room. The subject's head position was read from a Polhemus FastTrak electromagnetic tracker worn on the head to ensure that the center of the head was within 1-cm of the correct location in the room, marked on the floor. The experimenter used other angular and distance markings on the floor to hand-position the loudspeaker to the appropriate azimuth and distance prior to each measurement. Miniature microphones (Knowles FG-3329c) mounted in earplugs and inserted into the entrance of the subjects' ear canals (to produce blocked-meatus HRTF recordings) measured the raw acoustic responses to the MLS sequence. Microphone outputs drove a custom-built microphone amplifier that was connected to a TDT A/D converter (TDT PD1). These raw results were stored in digital form on the computer hard-drive for off-line processing to produce the estimated HRTFs.

No correction for the measurement system transfer function was performed, but the amplitude spectrum of the transfer-function of this measurement system was examined and found to vary by less than 2 dB and to cause no significant interaural distortion for frequencies between 400 and 1500 Hz (the frequency region important for the current study). The useful dynamic range of the measurements (taking into account the ambient acoustic and electrical noise) was at least 50 dB for all frequencies greater than 300 Hz.

HRTFs measured as described above include room echoes and reverberation. To eliminate room effects, time-domain impulse responses were multiplied by a 6-ms-long cos-squared time window (rise/fall time of 1 ms) to exclude all of the reverberant energy while retaining all of the direct-sound energy. The resulting ''pseudo-anechoic'' HRTFs were used to simulate sources (and in all subsequent analyses).

HRTFs were measured only for sources in the right hemifield. To simulate sources in the left hemifield, HRTFs from the corresponding right-hemifield position were used, exchanging the left and right channels (i.e., left/right symmetry was assumed; given that only pure tone targets were

simulated in the left hemifield, this approximation should introduce no significant perceptual artifacts in the simulated stimuli).

The measured HRTFs reflect the radiation characteristics of the loudspeaker used, which is not a uniformly radiating point source. For sources relatively far from the head, any differences in the measurement caused by the directivity of the source should be minor. For sources 15-cm from the center of the head, the effect of the source directivity may be significant. Therefore, the current study focuses on how distance influenced the signals reaching the ears for the particular source used (the Bose loudspeaker in question). The issue of how well the current results may generalize to other nearby sources is considered further in Sec. III, where empirical HRTF measurements are compared with theoretical predictions from a spherical head model that assumes a perfect point source.

In a similar vein, HRTFs measured for sources close to the head are much more sensitive to small displacements in the source (re: the intended source location) than more distant sources. However, given that all acoustic analyses and predictions of performance were made using the same measured HRTFs used to simulate the headphone-presented stimuli, any conclusions regarding which acoustic factors influence performance are justified, even if other measurement techniques might yield slightly different estimates of near-source HRTFs for the positions reported here.

3. Stimulus generation

Target stimuli consisted of 165-ms-long pure tones of 500 or 1000 Hz gated on and off by 30-ms cos-squared ramps. The 500-Hz target frequency was chosen so that results could be compared with previous studies of binaural masking level differences (BMLDs) and spatial unmasking of tones, most of which include a 500-Hz target condition. The 1000-Hz target was included in order to examine what happens for a higher target frequency where target and masker ITDs are still likely to have a large impact on detection but ILDs are larger than at 500 Hz. The target was temporally centered within a broadband, 250-ms-long masker. On each trial, the masker token was randomly chosen from a set of 100 pregenerated samples of broadband noise that were digitally low-pass filtered with a 5000 Hz cutoff frequency (ninth-order Butterworth filter, as implemented in the signal-processing toolbox in Matlab, the Mathworks, Natick, MA).

In most cases, target and masker were simulated as arising from different locations in anechoic space by convolving the stimuli with appropriate individualized head-related impulse responses (time-domain representation of the HRTFs). The simulated spatial configurations included all combinations of target at azimuths ($-90°$, $-45°$, $0°$, $45°$, $90°$) and distances (0.15 m, 1 m) and masker at azimuths ($0°$, $45°$, $90°$) and distances (0.15 m, 1 m). A total of 60 spatial configurations was tested (10 target locations $\times$ 6 masker locations; see Fig. 1). In a subset of trials, traditional BMLDs were measured using the same stimuli without HRTF processing.

For nearby sources, keeping the masker presentation level constant would result in the received level (at the subject's ears) varying widely with masker position. In order to keep the received level of masker relatively constant, the levels of the HRTF-processed masker stimuli were normalized to keep constant the rms energy falling within the equivalent rectangular band (ERB; Moore, 1997) centered on the target frequency at the ear receiving the more intense masker signal (the right ear for all of the tested configurations). In other words, the virtual stimuli actually simulated a masker whose distal energy level was adjusted up or down (depending on the masker spatial location) until the proximal stimulus level was constant at the more intense ear. In our analysis, the amounts by which the distal masker was adjusted were added back to the raw thresholds to predict the amount of spatial unmasking that would have occurred if the distal masker level had been constant.[2]

For the 500-Hz center frequency, the rms levels were adjusted using a 100-Hz-wide ERB. For the 1000-Hz target, the ERB width was set to 136 Hz. The masker signals were preprocessed in Matlab so that the right- (more-intense-) ear rms masker level in the ERB would be 64 dB SPL when played via headphones. BMLDs were measured with the low-pass-filtered noise spectral level fixed at 64 dB SPL.

Stimulus files, generated at a sampling rate of 44.1 kHz, were stored on the hard disk of the control computer (IBM PC compatible). On each trial, appropriate target and masker signals were presented through TDT hardware. Left- and right-ear target and masker signals were processed through four separate D/A converters (TDT PD1). Target signals were scaled to the appropriate presentation level by a programmable attenuator (TDT PA4), summed with the fixed-level masker signals (TDT SM3), and amplified through a headphone buffer (TDT HB6). The resulting binaural stimuli were presented via Etymotic Research ER-1 insert earphones. No filtering was done to compensate for the transfer characteristics of the playback system. A handheld RS 232 terminal (QTERM) was used to gather subject responses and provide feedback.

4. Experimental procedure

Behavioral experiments were performed in a single-walled sound-treated booth.

Each trial consisted of three intervals, each of which contained a noise burst. Either the second or third interval (randomly chosen, with equal probability, on each trial) also contained the tone-burst target. Subjects performed a two-alternative, forced-choice task in which they were asked to identify which interval, the second or third, contained the target tone. 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 the target 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.

TABLE I. Binaural masking level differences for individual subjects. Note that subjects S1 and S3 performed detection experiments for both 500- and 1000-Hz targets; S2 and S4 only performed the experiments for one target frequency (500 and 1000 Hz, respectively). Symbols give the convention used in the figures when plotting individual subject results.

Target frequency	Individual subject results				Across-subject average
	S1 ○	S2 ▽	S3 □	S4 △	
500 Hz	15.6	11.0	14.5	NA	13.7
1000 Hz	13.1	NA	7.5	8.7	9.8

At least three separate runs were performed for each subject in each condition. Final threshold estimates were computed by taking the average threshold across the repeated adaptive threshold estimates. Additional adaptive runs were performed as needed for every subject and condition to ensure that the standard error in this final threshold estimate was less than or equal to 1 dB for each condition and spatial configuration tested.

The study was divided into two parts, one measuring thresholds for the 500-Hz target and one for the 1000-Hz target. Three subjects performed each part (two of the four subjects performed both). For each target, subjects performed multiple sessions consisting of ten runs. Subjects were allowed to take short breaks between runs within one session, with a minimum 4-h break required between sessions. Each subject performed one initial practice session consisting of four practice runs and six runs measuring detection thresholds for NoSo and NoSπ conditions (where NoSo represents a sinusoidal diotic signal, i.e., with zero interaural phase difference, in the presence of a diotic noise; NoSπ represents a sinusoidal signal with interaural phase difference equal to π in the presence of a diotic noise). Subjects then performed 18 additional sessions (180 runs; 3 runs each of every combination for 6 target positions and 10 masker positions). In each of these sessions, a full set of thresholds was determined for one masker position (the order of the ten target positions was randomized within each session). These sessions were grouped into three blocks of six with each block containing a full set of thresholds. The order of masker positions was separately randomized for each block and subject. Any additional runs were performed after completion of the initial 19 sessions. Each subject performed approximately 20 h of testing per target frequency.

B. Results

1. Binaural masking level difference

Table I shows the BMLD (see Durlach and Colburn, 1978), defined as the difference in target detection threshold in the NoSo and NoSπ conditions. Results are consistent with those from previous, similar experiments. BMLDs are larger for the 500-Hz target (where BMLDs ranged from 11 to 16 dB) than the 1000-Hz target (where BMLDs ranged from 7 to 14 dB).

2. Spatial unmasking

The amount of ''spatial unmasking'' is defined as the change in the energy a target emits at threshold for a particular target location compared to when the target is at the same position as the masker. In order to estimate the target detection threshold when the emitted level of the masker is held constant, the amount by which the masker was normalized (to equate the masker level at the more intense ear) was first added back to the raw target detection thresholds. To estimate spatial unmasking (i.e., the amount by which detection thresholds improve with spatial separation of target and masker), the average of all thresholds when target and masker were at the same location was computed and this value was subtracted from all the renormalized thresholds.

Figures 2 and 3 plot the amount of spatial unmasking for 500- and 1000-Hz targets, respectively. Each panel shows the amount of spatial unmasking (improvement in target threshold relative to when target and masker are at the same location) for one masker location (shown graphically in the inset legend in each panel). The abscissa shows the target azimuth. Thick lines and filled symbols show results for the near target; thin lines and open symbols show results for the far target. Symbols show individual subject results and solid lines give the across-subject mean. Dashed lines represent the estimates of the better-ear contribution to spatial unmasking (averaged across subjects), discussed in detail in Sec. IV.

For the spatial configurations tested, the amount of spatial unmasking spans a range of over 50 dB [e.g., compare the thresholds for a 500-Hz target at (0°, 1 m), the center of the thin line in Fig. 2(d), to the thresholds for the 500-Hz target at (90°, 15 cm), the rightmost point of the thick line in Fig. 2(a)]. While subjects generally show similar patterns of results, intersubject differences are large. For instance, in Fig. 2(a) when the masker is at (0°, 1 m) and the 500-Hz target is at 15-cm, subject S1 (filled circles) consistently shows as much as 10 dB more unmasking than the other subjects (other filled symbols). However, this same subject consistently shows the least unmasking in other cases [e.g., in Fig. 2(f) when the masker is at (90°, 15 cm) and the target is at 1-m; compare open circles to the other open symbols].

Despite the large intersubject differences, overall trends are similar across subjects and for both 500- and 1000-Hz targets, and are summarized below.

To a first-order approximation, changing either target or source distance influences spatial unmasking in a straightforward way predicted by a simple change in the stimulus levels at the ears. For instance, looking within any single panel in Fig. 2 or 3 shows that positioning the target near the subject (thick lines) improves target detectability compared to when the target is far from the subject (thin lines; i.e., within any single panel thick lines are grossly similar to thin line results shifted upward by 10–20 dB). Similarly, comparison of the upper panels (a, b, and c) to the lower panels (d, e, and f) shows that positioning the masker near the subject (lower panels) degrades target detectability compared to when the masker is farther from the subject (upper panels; i.e., results in the upper panels are grossly similar to results in the lower panels shifted upward by 10–15 dB). However, closer inspection shows that the detailed pattern of spatial unmasking varies in a more complex way with both target and masker distance than a simple shift in threshold.

Spatial unmasking resulting from a fixed angular separation of target and masker is larger for nearby targets than

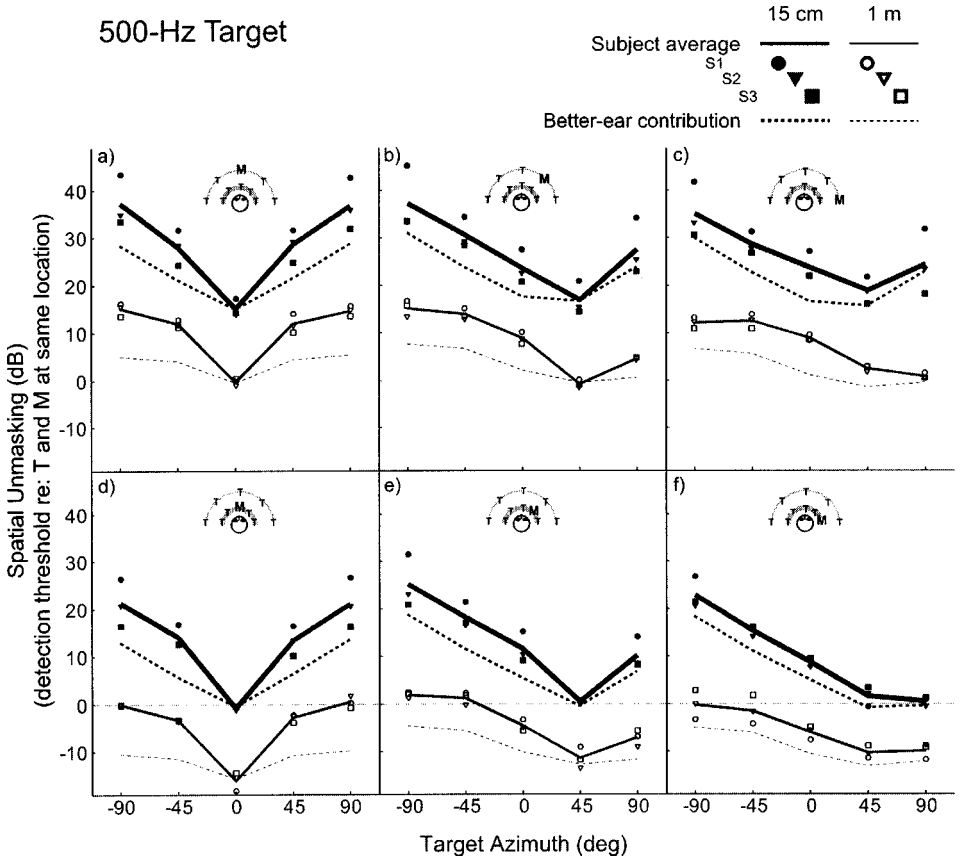

FIG. 2. Spatial unmasking for the 500-Hz target. Each panel plots spatial unmasking (the difference between target detection threshold when target and masker are at the same spatial location and when target and masker are in the spatial configuration denoted in the plot) as a function of target azimuth for a fixed masker location. Across-subject averages are plotted for target distances of 15-cm (thick solid lines) and 1-m (thin solid lines). Individual subject results are plotted as symbols. Dashed lines show the estimated better-ear contribution to spatial unmasking. The spatial configurations of target and masker represented in each panel are denoted in the panel legend.

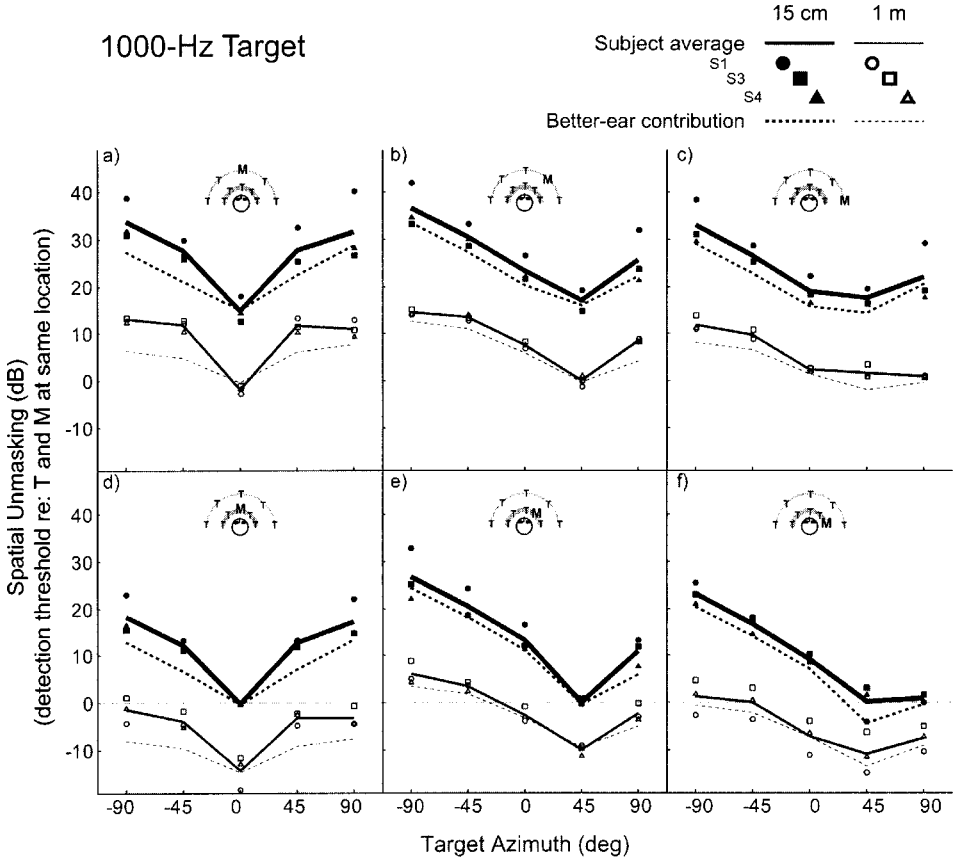

FIG. 3. Spatial unmasking for the 1000-Hz target. See caption for Fig. 2.

for distant targets. For example, in Fig. 3e, the difference between thresholds for the −90° and 45° targets is more than 25 dB for nearby targets (thick line) but less than 20 dB for distant targets (thin line).

Similarly, spatial unmasking resulting from a fixed angular separation of target and masker is larger for nearby maskers than for distant maskers. For example, as discussed above, for a 1000-Hz target when the masker is at (45°, 15 cm) [Fig. 3(e)], spatial unmasking for a 15-cm target (thick line) decreases by more than 25 dB when the target azimuth changes from −90° to +45°. However, when the masker is at (45°, 1 m) [Fig. 3(b)], this same angular displacement of the 15-cm target (thick line) produces a change in spatial unmasking of roughly 20 dB (compare the leftmost point and the point producing the least spatial unmasking, where the target is at 45°).

Angular separation of target and masker can actually make performance worse when target distance differs from masker distance. Usually, separating target and masker in azimuth improves target detectability compared to when the target and masker are in the same direction, but not in every case. When the masker is at 0° (panels a and d in both Figs. 2 and 3) the least amount of spatial unmasking occurs (thresholds are highest) when the target is at 0° (the same direction as the masker); when the masker is at 45° (panels b and e in Figs. 2 and 3) the least unmasking arises when the target is in the 45° masker direction. However, when the masker is at 90° (panels c and f in Figs. 2 and 3), angular separation of target and masker does not always increase the amount of unmasking. Specifically, for a masker at (90°, 1 m) [Figs. 2(c) and 3(c)] there is less spatial unmasking when the 15-cm target (thick line) is at 45° than when it is at 90°. Similarly, for a masker at (90°, 15 cm) [Figs. 2(f) and 3(f)] the amount of spatial unmasking for a 1-m target (thin line) is either equal [500-Hz target; Fig. 2(f)] or greater [1000-Hz target; Fig. 3(f)] when the target is at 90° compared to 45°.

Finally, independent of target or masker distance, the same angular separation of target and masker tends to produce less spatial unmasking as the masker laterality increases. For example, in Fig. 2(d) when the masker is at (0°, 15 cm) and the 500-Hz target is at a distance of 15-cm (thick line), a 90° angular separation of target and masker yields nearly 20 dB of unmasking. However, in Fig. 2(f), when the masker is at (90°, 15 cm) and the target is at 15-cm (thick line), the same angular separation of target and masker produces only 10 dB of unmasking.

C. Discussion

Intersubject differences in spatial unmasking may be partially explained by intersubject differences in the size of the BMLD. For instance, subject S1 has the largest BMLDs and exhibits the most spatial unmasking. However, intersubject differences in spatial unmasking could also be caused by differences in the acoustic parameters in the individually measured HRTFs. Analysis of acoustic differences in the measurements and the binaural contribution to spatial unmasking, which are considered further in Sec. IV, suggest that intersubject differences in spatial unmasking are affected both by subject-specific differences in acoustic cues and in different sensitivities to binaural cues.

Many of the current results follow easily predicted patterns. Moving the target closer to the subject improves detection performance (as expected on the basis of an increase in the level of the target reaching the listener); conversely, moving the masker closer degrades detection performance (as expected when the level of the masker at the ears increases). Separating target and masker in angle improves detection performance for most spatial configurations. However, there are other effects that are less intuitive. Unmasking varies more with target azimuth for a 15-cm masker than for a 1-m masker and for a 15-cm target than for a 1-m target. The masker laterality influences the effectiveness of a given angular separation of target and masker, decreasing with masker laterality. Finally, when target and masker are at different distances and the masker is at 90°, the amount of unmasking can actually decrease when the target is at 45° compared to when the target is in the same direction as the masker (this is essentially a case where there is ''spatial masking,'' i.e., where performance is actually worse when the sources are spatially separated compared to when they are at the same location).

Apparent discrepancies in the amount of spatial unmasking observed in previous studies are actually consistent with the current results. For example, the current study found more spatial unmasking for 1-m sources when the masker is at 0° compared to when the masker is at 90°. Thus, the relatively large amount of spatial unmasking observed by Gatehouse (1987) compared to that found by Santon (1987) and Doll and Hanna (1995) may be caused by the fact that Gatehouse fixed the masker in front of the listener and varied target azimuth, whereas Santon and Doll and Hanna fixed the target in front of the listener and varied masker azimuth.

III. HRTF MEASUREMENTS

The acoustic factors that influence spatial unmasking can be characterized by analysis of the HRTFs used in the simulations. Three acoustic characteristics of the HRTFs influence the performance in a spatial unmasking task: the magnitude spectra of, the interaural level differences (ILDs) in, and the interaural time differences (ITDs) in the signals reaching the two ears. The magnitude spectra of the HRTFs determine the intensity of the sound at the ears and thus the amount of spatial unmasking resulting from better-ear effects. ITDs and ILDs determine the amount of binaural unmasking. In this section, these parameters are analyzed for the individually measured HRTFs.

Individual HRTFs for the four human subjects are compared both to values measured for a KEMAR acoustic manikin (using the same measurement techniques used for the individual subjects) and those predicted from a spherical model of the head assuming a perfect point source. While the literature contains descriptions of both KEMAR (Brungart and Rabinowitz, 1999) and spherical-head model (Duda and Martens, 1998; Shinn-Cunningham *et al.*, 2000) HRTFs for sources near the listener, the current analysis compares these ''generic'' models to human measurements to determine

whether the models capture the acoustic effects that are important for predicting the amount of spatial unmasking as a function of nearby target and masker locations. As noted in Sec. II, the current measurements do not try to compensate for the radiation characteristics of the loudspeaker used; as such, any consistent discrepancies between predictions from a spherical-head model and measured results (from KEMAR and the human subjects) may reflect influences of the radiation characteristics of the loudspeaker used (which is not a point source) or other differences between the assumptions of the spherical-head model and properties of the physical sources and heads measured.

A. Methods

KEMAR HRTFs were measured using a procedure identical to that used for the human listeners (see description in Sec. II). HRTF predictions for a spherical head model (Brungart and Rabinowitz, 1999; Shinn-Cunningham *et al.*, 2000) were computed using a head with radius of 9-cm and diametrically opposed ears. These results are compared to the HRTFs measured for the four subjects who participated in the spatial unmasking experiment.

For all of the HRTFs, the magnitude spectra, ILD, and ITD were determined for the equivalent rectangular band (ERB) centered at a given frequency. Magnitude spectra were calculated as the rms energy in the HRTF falling within each ERB filter (100-Hz width centered at 500 Hz and 136-Hz width centered at 1000 Hz). ILDs were computed as the difference in the magnitude spectra for the left and right ears. ITD was first estimated as a function of frequency by taking the difference between the right- and left-ear HRTF phase angles at each frequency f and dividing by $2\pi f$. The ITD in each ERB filter was then estimated as the average of the ITD values for the frequencies falling within each ERB filter.

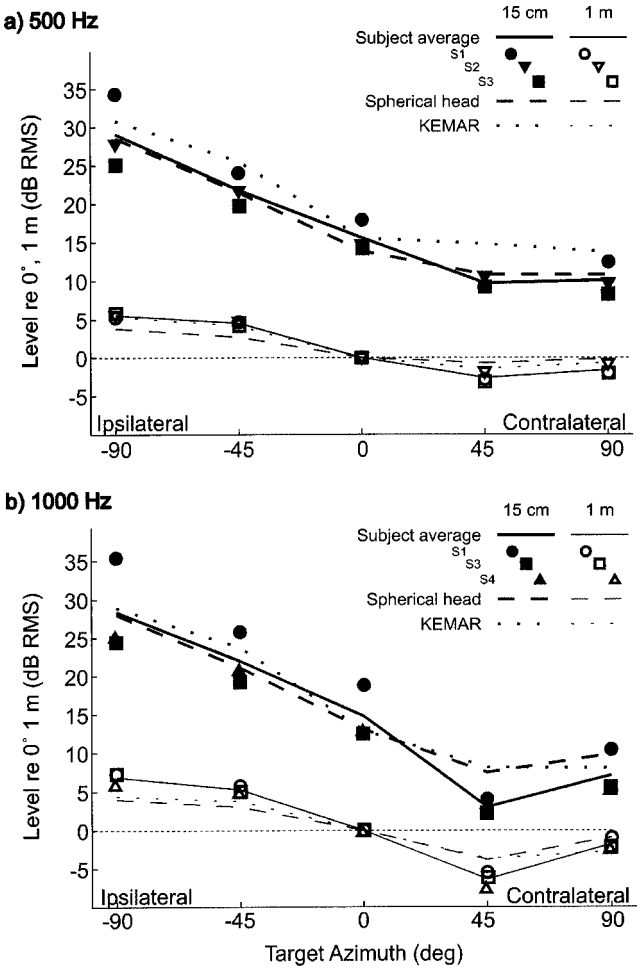

FIG. 4. Left-ear HRTF spectrum levels in ERB filters, relative to the left-ear HRTF for a source at (0°, 1 m). Results are shown for individual listeners, KEMAR, and the spherical head model as a function of source position. (a) 500 Hz. (b) 1000 Hz.

B. Results

1. Intensity effects

Figure 4 shows the magnitude of the ERB-filtered HRTFs at 500 [Fig. 4(a)] and 1000 Hz [Fig. 4(b)] for the left ear relative to a source at (0°, 1 m). (Recall that HRTFs were measured only for sources to the right of the listener and that this analysis assumes left-right symmetry.) Results are shown as a function of the target azimuth for individual human subjects (symbols), the across-human-subject average (solid line), KEMAR (dotted line), and a spherical head model (dashed line). Distant sources are represented by open symbols and thin lines; near sources are shown by filled symbols and thick lines.

Not surprisingly, for both frequencies the spectral gain is larger for near sources (thick lines) than far sources (thin lines). However, in addition to an overall shift in level, the dependence of the HRTF level on source azimuth differs for the two distances. Specifically, for the 15-cm distance (thick lines), the gain to the ipsilateral ear (i.e., the gain for sources at negative azimuths) grows rapidly with source eccentricity compared to the 1-m distance, while the gain to the contralateral ear (positive azimuths) changes similarly with source angle for both distances (compare thick and thin lines).

Overall, intersubject differences are modest for the more distant source (consider the open symbols in each panel). However, there are larger intersubject differences for the 15-cm source positions (filled symbols). For instance, at both frequencies [Figs. 4(a) and (b)], the 15-cm HRTF gain for subject S1 (filled circles) is generally 5–10 dB larger than for the other subjects, except at 45° where all HRTFs are similar.

For a 15-cm source at both 500 Hz [Fig. 4(a)] and 1000 Hz [Fig. 4(b)], KEMAR (thick dotted lines) and spherical-head gains (thick dashed lines) generally fall within the range of values observed for the four human subjects (filled symbols) measured in this study. However, in Fig. 4(b) for a 1-m source, KEMAR measurements (thin dotted lines) and model predictions (thin dashed lines) slightly underestimate the 1000-Hz gain to the ipsilateral ear compared to the individual subject results (lines fall below symbols for azimuths of −45° and −90°). At 500-Hz [Fig. 4(a)], the 1-m KEMAR measurements (thin dotted lines) fall within the range of results obtained from the human subjects (open symbols); however, the spherical head model results (thin dashed lines) fall below the subject measurements (open symbols) for ipsilateral sources (sources at −45° and −90°).

While, intuitively, we expect the level of the signal

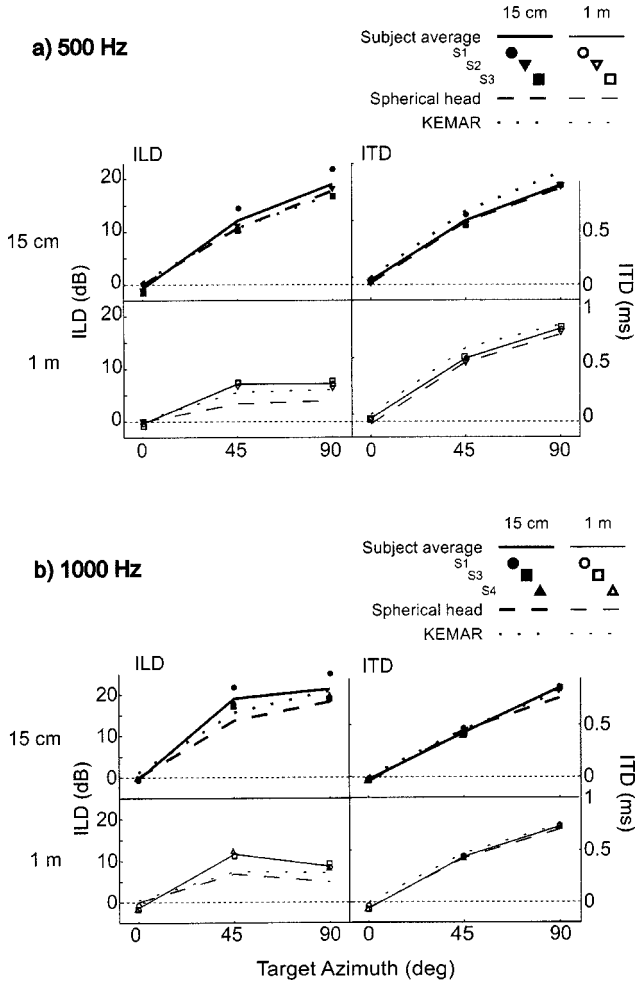

FIG. 5. ILDs and ITDs in HRTFs for individual subjects, KEMAR manikin, and the spherical head model. (a) 500 Hz. (b) 1000 Hz.

reaching the ears to vary monotonically with lateral angle of the source, human HRTF measurements show that this is not strictly true. In particular, the 1000-Hz human measurements [symbols and solid lines in Fig. 4(b)] show that less energy reaches the contralateral ear when a source is at 45° than when it is at 90° for both source distances (thick and thin lines are nonmonotonic with azimuth) Similarly, at 500 Hz [Fig. 4(a)] the gain to the contralateral ear is comparable for 45° and 90° sources rather than decreasing for the 90° source (thick and thin lines). This nonmonotonicity [which may in part be a consequence of the acoustic ''bright spot;'' e.g., see Brungart and Rabinowitz (1999)] is underestimated in both the spherical-head model (dashed lines) and KEMAR (dotted lines) HRTFs, especially at 1000 Hz [compare lines to human subject results for sources at 45°, especially in Fig. 4(b)].

2. Interaural differences

Figure 5 shows the ILDs and ITDs in the measured HRTFs at 500 and 1000 Hz [Figs. 5(a) and (b), respectively] for the spatial positions used in the study. As in Fig. 4, results for individual subjects (symbols), the across-human-subject average (full lines), KEMAR (dotted lines), and a spherical head model (dashed lines) are shown as a function of target azimuth. Results for near sources are shown in the top of each subplot with heavy lines and filled symbols. Thin lines and open symbols plot results for far sources (bottom row of each half of the figure). The left column shows ILD results and the right column shows ITD results.

ILDs were calculated directly from the measurements plotted in Fig. 4. As a result, there are large intersubject differences in the ILDs (left panels in Fig. 5) that are directly related to the intersubject differences in the monaural spectral gains. For instance, subject S1 has much larger ILDs at both 500 and 1000 Hz for the 15-cm source [filled circles in the left columns of Figs. 5(a) and (b)] than any of the other subjects (other filled symbols).

As expected, for both frequencies [Figs. 5(a) and (b)] ILDs are much larger for sources at 15-cm (thick lines in top left panels) compared to 1-m (thin lines in the bottom left panels) with ILDs at 500 and 1000 Hz approaching 20 dB for the nearby sources at 90° (rightmost point in the top left panels). The spherical-head (dashed lines) and KEMAR (dotted lines) results tend to underestimate ILDs for lateral sources, although for the 500-Hz, 15-cm sources [Fig. 5(a), top left panel], both spherical-head and KEMAR results are within the range of human observations. Discrepancies between human and model results are most pronounced for a 1000-Hz source at a distance of 1-m [Fig. 5(b), bottom left panel] and are greater for the spherical-head predictions (dashed lines) than KEMAR measurements (dotted lines).

ITDs [the right panels in Figs. 5(a) and (b)] vary primarily with source angle and change only slightly with distance and frequency. For most of the measured locations, both spherical-head and KEMAR results are in close agreement with human measurements.

C. Discussion

Both spherical-head and KEMAR HRTFs provide reasonable approximations to how acoustic parameters in human HRTFs vary with source location. In general, both KEMAR and the spherical head measurements fall within the range spanned by the individual subject measurements. However, both spherical-head predictions and KEMAR measurements slightly overestimate the gain at the contralateral ear when a source is at 45° (especially at 1000 Hz) and tend to modestly underestimate the ILD for sources off midline, particularly at the 1-m distance. These small differences cannot be attributed to loudspeaker characteristics, given that (1) the discrepancies are similar for both KEMAR measurements (using the same loudspeaker) and spherical-head predictions (assuming a perfect point source) and (2) the differences are, if anything, larger for the more distant, 1-m source (where the loudspeaker directivity is less influential) than the nearby source. Thus, we conclude that generic HRTF models capture the important features of the HRTFs measured in human subjects and that the effects of the source transmission characteristics do not strongly influence the signals reaching the ears even for nearby sources, at least for the frequencies considered in the current study.

Intersubject differences in the HRTFs are large, especially for nearby sources. Of the four subjects, one subject showed consistently larger spectral gains and consistently

larger ILDs than the other subjects when the source was at 15-cm. While it is possible that some of the intersubject differences arise from inaccuracies in HRTF measurement (e.g., from hand-positioning the loudspeaker), the fact that one subject has consistently larger gains and ILDs for all nearby source locations suggests that real anatomical differences rather than measurement errors are responsible for the observed effects. It is also interesting to note that the observed intersubject differences are much smaller for the 1-m source, suggesting that intersubject differences in HRTFs are especially important when considering sources very close to the listener.

IV. BETTER-EAR AND BINAURAL CONTRIBUTIONS TO SPATIAL UNMASKING

A. Analysis

For each subject, estimates of the better-ear and binaural contributions to spatial unmasking were derived from the acoustic parameters of the HRTFs and the behavioral thresholds.

The better-ear contribution to spatial unmasking was estimated by calculating the TMR in the ERB filter centered on the target frequency at the better ear for each spatial configuration when target and masker emit the same level (and thus would yield a TMR of zero when at the same location). The resulting TMR predicts the amount by which target thresholds decrease or increase simply because of acoustic effects at the better ear (i.e., if the calculated TMR is +2 dB, it implies that at detection threshold, the intensity of the target at the better ear was 2 dB more for the given spatial configuration than if the target and masker were at the same spatial location; thus, the better-ear contribution for such a configuration is +2 dB). The subject-specific binaural contribution to spatial unmasking was estimated by subtracting the estimated better-ear contribution to spatial unmasking (derived from individually-measured HRTFs) from the individual behavioral estimates of spatial unmasking.

B. Results

1. Better-ear contributions to spatial unmasking

While intersubject differences in the better-ear contribution to spatial unmasking are large, the trends in the across-subject average data capture the important features of the individual data. For brevity, only the across-subject averages are presented in Figs. 2 and 3 for the 500- and 1000-Hz target, respectively, as dashed lines. For all spatial configurations tested, the behaviorally observed amount of spatial unmasking either equals or is larger than the predicted spatial unmasking from better-ear effects (dashed lines fall below or at measured values in all graphs). Thus, even when there are large ILDs in the signals reaching the listener, binaural performance is always better than or equal to predicted performance when listening monaurally with the acoustically better ear.

Better-ear effects account for a large portion of the observed spatial unmasking when target and masker are in the same direction and for the large influence of target and/or masker distance on spatial unmasking. Specifically, the predicted results (dashed lines) are in good agreement with the measured results when the target is at 0° in the left column, at 45° in the middle column, and at 90° in the right column. Generally, angular separation of target and masker increases the better-ear contribution to unmasking (dashed-line predictions generally increase as the target azimuth moves away from the masker azimuth). However, when the masker is at 90° (the right columns in Figs. 2 and 3), better-ear effects either decrease or are roughly the same when the target is at 45° compared to 90° (dashed-line predictions are either constant or decrease as the target azimuth moves from 90° to 45°). Better-ear contributions to unmasking change more with target azimuth when the target is at 15-cm (thick dashed lines) than at 1-m (thin dashed lines), primarily because, for nearby sources, small positional changes cause large changes in the relative distance from source to the better ear.

Finally, differences between mean subject results (solid lines) and predicted better-ear effects (dashed lines) are generally larger for the 500-Hz target (Fig. 2) than the 1000-Hz target (Fig. 3), suggesting that the better-ear contributions to unmasking are relatively more important (i.e., account for a greater portion of the observed amount of spatial unmasking) for the 1000-Hz target than the 500-Hz target. This is true both because the better-ear effects are larger in absolute terms and because the additional spatial unmasking for which better-ear effects cannot account is smaller at 1000 Hz than at 500 Hz.

2. Binaural contributions to spatial unmasking

Figures 6 and 7 show the estimated binaural contribution to spatial unmasking for the 500- and 1000-Hz target, respectively. The binaural contribution was calculated for each individual subject by subtracting the estimated better-ear contribution (the across-subject average of which is shown by dashed lines in Figs. 2 and 3) from the total amount of spatial unmasking (symbols in Figs. 2 and 3). Both Figs. 6 and 7 show results for each subject who performed that condition in a separate subplot. Each subplot is divided into six panels corresponding to the six masker locations (laid out as indicated in the legend). In each panel, symbols plot the mean binaural contribution to spatial unmasking (averaged across the repeated adaptive runs). The error bars show the *range* of thresholds obtained across the repeated adaptive runs for each condition. Results are shown for both the far target (gray) and the near target (black) as a function of target azimuth. Figures 6 and 7 also show model predictions (lines), which are derived and discussed in Sec. V.

Even though intersubject differences are large, there are a number of trends that are consistent across subjects. Not surprisingly, for both target frequencies (Figs. 6 and 7) there is no unmasking beyond the better-ear contribution when target and masker are at the same spatial location (the binaural gain is near zero when the target is at 0° in the left columns, at 45° in the middle columns, and at 90° in the right columns of Figs. 6 and 7). In fact, only the 500-Hz results for subject S1 [Fig. 6(a)] show any binaural unmasking when target and masker are at the same off-median-plane direction but at different distances. For example, looking at the top right panel of Fig. 6(a) [masker at (90°, 1 m)], the binaural gain is posi-

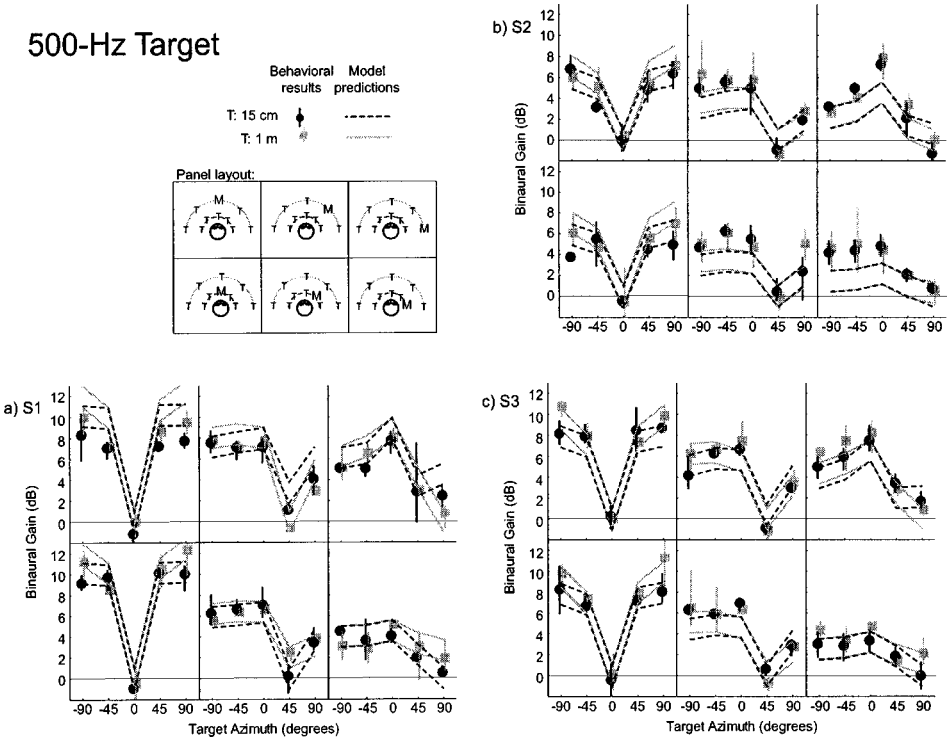

FIG. 6. Estimated binaural contribution to spatial unmasking for the 500-Hz target. Each panel plots the amount of binaural unmasking for one masker position for both the 15-cm and 1-m target. Symbols show estimates for individual subjects with error bars showing the range of results across multiple adaptive runs. Lines trace a 2-dB range around the predicted amount of binaural unmasking from the Colburn (1977a) model for the 15-cm (dashed black lines) and 1-m (solid gray lines) target. The layout of the spatial configurations of target and masker represented in each panel are shown in the legend. (a) Subject S1. (b) Subject S2. (c) Subject S3.

tive when the target is at (90°, 15 cm) (black circle); in the bottom right panel of Fig. 6(a) [masker at (90°, 15 cm)], the binaural gain is positive when the target is at (90°, 1 m) (gray square).

Overall, target distance has relatively little impact on the binaural component of the spatial release from masking (black and gray symbols are generally comparable within each panel). However, masker distance influences results for all subjects, particularly for the 500-Hz results (Fig. 6) when the masker is located at 90° (right panels). In these configurations, binaural unmasking is smaller when the masker is at 15-cm (lower right panel) than when it is at 1-m (upper right panel).

In general, the binaural contribution to spatial unmasking is larger for the 500-Hz target (Fig. 6) than the 1000-Hz target (Fig. 7). For both target frequencies, the amount of binaural unmasking tends to be largest when the masker is at 0° (left panels in each subplot) and decrease as the masker is displaced laterally (center and right panels in each subplot). Similarly, the change in binaural unmasking with target

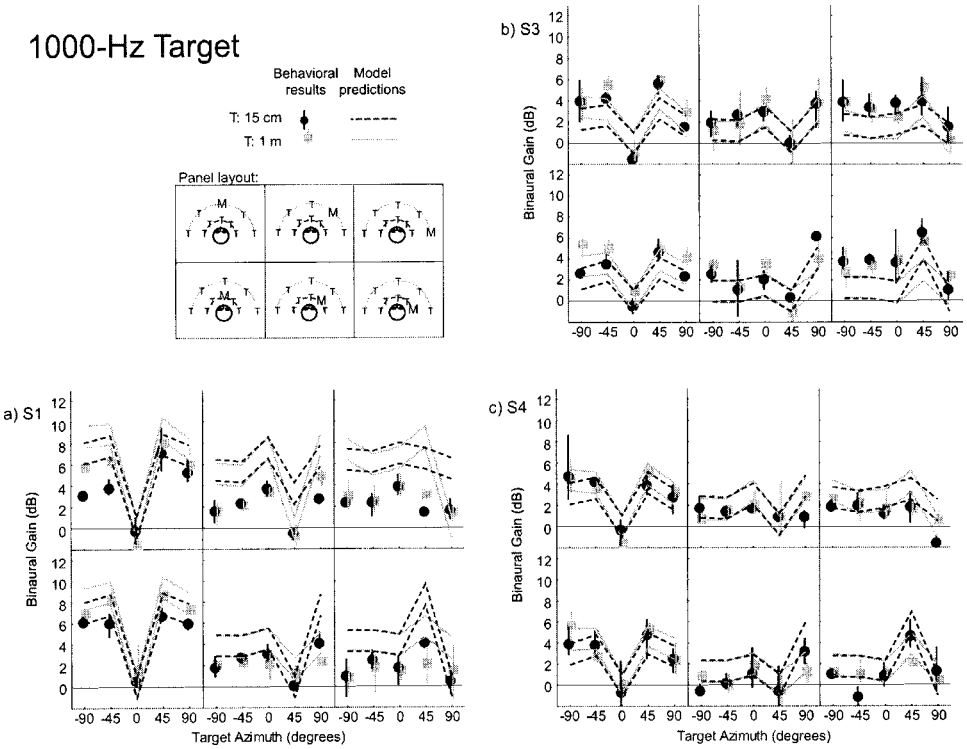

FIG. 7. Estimated binaural contribution to spatial unmasking for the 500-Hz target. See caption for Fig. 6. (a) Subject S1. (b) Subject S3. (c) Subject S4.

angle (i.e., the modulation of binaural gain with target azimuth) is smaller when the masker is laterally displaced (right panels) than when the masker is at 0° (left panels), particularly for the 1000-Hz target (Fig. 7). For instance, looking at the bottom left panel of Fig. 7(a), when the masker is at (0°, 15 cm) the binaural contributions to spatial unmasking for the 1000-Hz target for subject S1 range from 0 to 8 dB depending on the target azimuth. However, when the masker is at (90°, 15 cm) [bottom right panel in Fig. 7(a)], binaural unmasking is roughly constant, independent of target angle (roughly 0–2 dB).

The angular separation of target and masker that leads to the greatest amount of binaural unmasking depends on target frequency. For the 500-Hz target (Fig. 6), binaural unmasking tends to be greatest when target and masker angles differ by about 90° (for example, in the right columns of Fig. 6 where the masker is at 90°, the unmasking is generally greatest when the target is at 0°). However, for the 1000-Hz target (Fig. 7), binaural unmasking tends to be greatest when target and masker angles differ by roughly 45° (in the right columns of Fig. 7 where the masker is at 90°, the amount of unmasking tends to be greatest when the target is at 45°).

C. Discussion

Better-ear factors contribute significantly to spatial unmasking for all of the spatial configurations tested. Better-ear effects are larger at 1000 Hz than 500 Hz and are larger when the target is at 15-cm compared to when the target is at 1-m. The better-ear contribution to spatial unmasking does not always increase monotonically with angular separation of target and masker. In particular, when the masker is at 90°, displacing the target toward the median plane can lead to decreases in the TMR at the better ear, especially if the target and masker are at different distances. This result helps explain why angular separation of target and masker does not always improve detection performance.

Subjects show large differences in their ability to use binaural cues in detection tasks. For subject S1, binaural differences can decrease detection thresholds by as much as 12 dB at 500 Hz [see Fig. 6(a)]; for subject S2 binaural differences provide at most 7 dB of unmasking [Fig. 6(b)]. These intersubject differences in the binaural component of spatial unmasking roughly correlate with differences in BMLDs (Table I); however, intersubject differences in binaural sensitivity for one masker location do not predict results in other spatial configurations. For example, in the 500-Hz conditions when the masker is at 0°, subjects S1 and S3 [left columns in Figs. 6(a) and (c)] have larger binaural components of spatial unmasking than subject S2 [left column in Fig. 6(b)]. However, when the masker is at 90° [right columns of Figs. 6(a)–(c)], all three subjects exhibit essentially the same amount of binaural unmasking. This result suggests that intersubject differences in binaural sensitivity cannot be fully captured with a single ''binaural sensitivity'' parameter at each frequency [the degree to which intersubject differences can be predicted by Colburn's (1977b) model is considered further in Sec. V].

The magnitude of interaural level differences in the masker appears to have a large effect on the amount of binaural masking. For both target frequencies (Figs. 6 and 7), binaural unmasking is greatest when the masker is at 0° (and ITDs and ILDs in the masker are near zero; left columns in each subplot); when the masker is at 45° and 90° (center and right columns in each subplot), the amount of binaural unmasking decreases for the same angular separation of target and masker (i.e., even for roughly the same difference in target and masker ITD). When the masker is off to the side (right columns in the subplots of Figs. 6 and 7), the binaural contribution to spatial unmasking is also smaller when the masker is at 15-cm (when ILDs are very large; bottom right panels) compared to 1-m (when ILDs are smaller; top right panels). These effects are consistent with past reports showing that the BMLD decreases with masker ILD (e.g., see Durlach and Colburn, 1978, p. 433).

In general, the maximum difference in interaural phase difference (IPD) cues for target and masker arises when the ITDs for target and masker differ by one-half the period of the target frequency. For a 500-Hz target, the ITDs in target and masker need to differ by roughly 1 ms to maximize binaural unmasking. For a 1000-Hz target, the ITDs in target and masker need to differ by roughly 500 μs. This explains the dependence of maximal binaural unmasking on target and masker separation and frequency: results in Fig. 5 show that an angular separation of about 90° causes target and masker ITDs to differ by roughly 1 ms (maximizing IPD differences in target and masker for a 500-Hz target) whereas an angular separation of about 45° causes target and masker ITDs to differ by roughly 500 μs.

V. BINAURAL MODEL PREDICTIONS

A. Analysis

Subject-specific predictions of binaural unmasking were calculated using a modified version of the Colburn (1977a, 1977b) model (a description of the current implementation of the model is provided in the Appendix). Predictions depend on six parameters, evaluated at the target frequency: the ITDs and ILDs in both target and masker; the binaural sensitivity of the listener; and the spectrum level of the masker at the more intense ear relative to the absolute, monaural detection threshold in quiet.

The ITDs and ILDs used in the predictions were taken from the analysis of the cues present in the HRTFs. The ITD and ILD in masker were calculated from the values averaged over the ERB filter centered on the target frequency (see Fig. 5). The ITD and ILD in the target were taken directly from the HRTF values at the target frequency (not averaged over the ERB). Binaural sensitivity at each frequency was set to the measured BMLD for each subject and target frequency (Table I). For both the 500- and 1000-Hz targets, the monaural detection threshold (parameter K in the model) was set to 44 dB/Hz.

B. Results

Model predictions are plotted alongside behavioral estimates of the binaural contribution to spatial unmasking in Figs. 6 and 7 (for the 500- and 1000-Hz targets, respectively). In order to be somewhat conservative in identifying

conditions where the model fails to account for behavioral data, parallel lines plot a range of ±1 dB around the actual model predictions. Predictions for the nearby target are shown as dashed black lines; predictions for the far target are shown as solid gray lines.

Model predictions of binaural unmasking are nonnegative for all spatial configurations. Predictions are exactly zero whenever the target and masker are at the same spatial location and positive whenever the target and masker have differences in either their IPDs or ILDs at the target frequency. Thus, in theory, predictions of binaural unmasking are positive whenever the target and masker are at different distances but in the same direction off the median plane because of differences in ILDs in target and masker. However, in practice, predictions are near zero for all configurations when the target and masker are in the same direction for subjects S2, S3, and S4 [Figs. 6(b), 6(c), 7(b), and 7(c)]. Predictions for subject S1 [who has the largest ILDs for 15-cm sources and the largest BMLDs at both frequencies; Figs. 6(a) and 7(a)] are greater than zero for both target frequencies when the target and masker are at different distances but the same (off-median-plane) direction. For instance, in the top center and top right panels of Figs. 6(a) and 7(a) [masker at (45°, 1 m) and (90°, 1 m)], the black dotted lines (predictions for the target at 15 cm) are above zero for all target azimuths, including the target at 90°; in the bottom center and right panels of Figs. 6(a) and 7(a) [masker at (45°, 15 cm) and (90°, 15 cm)], the gray solid lines (predictions for the target at 1 m) are positive for all azimuths.

Binaural unmasking predictions are generally larger at 500 Hz (Fig. 6) than 1000 Hz (Fig. 7). At both frequencies, binaural unmasking varies with angular separation of target and masker; however, the angular separation that maximizes the predicted spatial unmasking depends on frequency. As in the behavioral results, predicted binaural unmasking is greatest when the target and masker are separated in azimuth by 90° for the 500-Hz target (Fig. 6) and 45° for the 1000-Hz target (Fig. 7), corresponding to separations that maximize the differences in target and masker IPD at the target frequency (e.g., in the left column of Fig. 6, when the 500-Hz masker is at 0°, the maximum predicted unmasking, shown by the lines, occurs for targets at +90° and −90°; however, in the left column of Fig. 7, when the 1000-Hz masker is at 0°, the maximum predicted unmasking generally occurs for targets at +45° and −45°).

Also consistent with behavioral results, the maximum predicted amount of binaural unmasking decreases with masker ILD. As a result, the predicted amount of binaural unmasking varies with masker location, systematically decreasing with increasing masker angle and decreasing when the masker is at 15-cm compared to 1-m. For instance, predicted levels of unmasking are generally largest when the masker is at 0° (left columns of Figs. 6 and 7) and decrease as the masker is laterally displaced (center and right columns). Similarly, the amount of unmasking tends to be larger for the top rows of data in Figs. 6 and 7, when the masker is at 1-m, than in the bottom rows of data, when the masker is at 15-cm.

Model predictions capture much of the variation in binaural unmasking; however, there are systematic prediction errors that are large compared to the intrasubject variability. (Note that the standard error in the mean behavioral results is less than or equal to 1 dB as a direct result of the experimental procedure. The error bars in the figure are even more conservative, showing the *range* of thresholds obtained over multiple runs.)

Predictions are first compared to behavioral results for the 500-Hz target (Fig. 6). Predictions for subject S1 agree well with behavioral results when the masker is at (0°, 15 cm) [bottom left panel of Fig. 6(a)] and reasonably well for three other masker locations [(45°, 15 cm), (90°, 15 cm), and (90°, 1 m); bottom center, bottom right, and top right panels of Fig. 6(a), respectively]. However, S1 predictions tend to overestimate binaural unmasking for two masker locations [(0°, 1 m) and (45°, 1 m); top left and top center panels of Fig. 6(a)]. For subject S2, predictions match behavioral results reasonably well when the masker is at 0° [see the top left and bottom left panels of Fig. 6(b)], independent of masker distance (although there are isolated data points for which the model overestimates binaural unmasking), but systematically underestimate binaural unmasking when the masker is at 45° and 90° for both masker distances [see center and right panels of Fig. 6(b), where symbols fall above lines]. Results for subject S3 are similar to those of subject S2: predictions are in good agreement with measurements when the masker is in the median plane [left panels of Fig. 6(c)] but underestimate binaural unmasking when the masker is laterally displaced [center and right panels of Fig. 6(c)].

Focusing on the 1000-Hz results (Fig. 7), subject S1 predictions generally overestimate binaural unmasking (in all panels in Fig. 7(a), symbols fall below lines). For subject S3, predictions generally underestimate binaural unmasking, except when the masker is at (45°, 1 m), where predictions and measurements are reasonably close [agreement between the measured data points and the prediction lines is good only for the top center panel of Fig. 7(b); for all other panels, symbols fall above lines]. Finally, predictions for subject S4 either fit reasonably well or underestimate binaural unmasking when the masker is at 0° [left panels of Fig. 7(c)] but overestimate binaural unmasking when the masker is at 45° or 90°, independent of masker distance [see center and right panels of Fig. 7(c), where symbols fall below lines].

Overall, predictions and behavioral results are in better agreement when the masker is in the median plane than when the masker is at 45° or 90° and for the 500-Hz data compared to the 1000-Hz data.

C. Discussion

The Colburn model assumes that a single value representing binaural sensitivity at a particular frequency can account for intersubject differences in binaural unmasking. This binaural sensitivity parameter was set from BMLD measures taken with a diotic masker and target that was either diotic (NoSo) or inverted at one ear to produce an interaural phase difference of π (NoSπ). These conditions are most analogous to the spatial configurations in which the masker is directly in front of the listener (and the masker is

essentially diotic). For most of the configurations with the masker at 0°, model predictions agree well with observed results. In contrast, larger discrepancies between the modeled and measured results arise when the masker is at 45° and 90° (conditions in which there are significant ILDs in the masker).

While there are some conditions in which the model predictions consistently over- or underestimate binaural unmasking [e.g., results for subject S1 at 1000 Hz in Fig. 7(a) or for subject S3 at 1000 Hz in Fig. 7(b)], there are other conditions for which changing the single subject-specific ''binaural sensitivity'' of the model cannot account for discrepancies between the model predictions and the measurements [e.g., results for subject S2 at 500 Hz in Fig. 6(b) or for subject S4 at 1000 Hz in Fig. 7(c)].

The current results suggest that subjects differ not only in their overall sensitivity to binaural differences, but also in the dependence of binaural sensitivity on the interaural parameters in masker and/or target. In particular, binaural sensitivity appears to depend on the interaural level difference in the masker differently for different subjects. As a result, individualized model prediction errors are generally larger when there are large ILDs in the masker than when the masker has near-zero ILD. While the Colburn model has been tested (and shown to predict results relatively well) in many studies in which target and masker vary in their interaural phase parameters, there are few studies that manipulate the target and masker ILD. These results suggest the need for additional behavioral and theoretical studies of the effects of ILD in binaural detection tasks.

Even though there are specific conditions for which predictions fail to account for the results for a particular subject, the model captures many of the general patterns in results, including the tendency for binaural unmasking to decrease as the ILD in the masker increases and how the amount of binaural unmasking depends on the angular separation of target and masker and the frequency of the target.

VI. GENERAL DISCUSSION

The current study is unique in measuring how tone detection thresholds are affected by target and masker location when sources are very close to the listener. Results show that for sources very close to the listener, small changes in source location can lead to large changes in detection threshold. These large changes arise from changes in both the TMR (affecting the better-ear contribution to spatial unmasking) and ILDs (affecting the binaural contribution to spatial unmasking).

The current results demonstrate how the relative importance of better-ear and binaural contributions to spatial unmasking change with target and masker location, including source distance (in contrast to previous studies that considered only angular separation of relatively distant sources). The relative importance of better-ear contributions to spatial unmasking increases as masker distance decreases, probably because of increases in the ILD in the masker, which reduce the amount of binaural unmasking. The better-ear contribution also increases as target distance decreases, primarily because the TMR changes more rapidly with target angle when the target is near the listener. The relative importance of the better-ear contribution to spatial unmasking increases with target frequency, both because the absolute magnitude of better-ear factors increases and because the binaural contribution to unmasking decreases. For a 500-Hz target, binaural and better-ear factors are roughly equally important when the masker is in the median plane. However, better-ear factors become relatively more important as the masker is displaced laterally, in part because the amount of binaural spatial unmasking decreases with masker ILD. This trend, which is predicted by the Colburn model, helps to explain large differences in the amount of spatial unmasking observed in previous studies (e.g., Ebata *et al.*, 1968; Gatehouse, 1987; Santon, 1987). Specifically, more spatial unmasking arises when the masker is positioned in front of the listener and the target location is varied (leading to near-zero ILDs in the masker) than when the target is fixed in location and the angle of masker is varied (leading to progressively larger ILDs in the masker with spatial separation of target and masker).

Binaural processing contributes up to 10 dB to spatial unmasking for the spatial configurations tested. In theory, differences in target and masker distance cause differences in target and masker ILD when the sources are off the median plane, leading to binaural unmasking. However, in the current study evidence of binaural unmasking resulting from differences in target and masker distance was observed only for Subject S1, who had both the largest BMLDs and the largest ILDs of the four subjects in the study.

Although monaural detection thresholds were not directly measured in the current study, binaural performance is always better than or equal to the performance predicted by analysis of the TMR at the better ear. Thus, the current study does not help to explain results suggesting that binaural performance sometimes falls below monaural performance using the better ear alone, particularly for configurations with large ILDs (Bronkhorst and Plomp, 1988; Shinn-Cunningham *et al.*, 2001). One important distinction between the current study and these previous reports is that the current study measured tone detection for relatively low-frequency tones, whereas both of the previously cited studies measured speech intelligibility, a suprathreshold task that emphasizes information at higher frequencies. Further studies are necessary to help determine when binaural stimulation may actually degrade performance compared to monaural, better-ear performance.

Intersubject differences in the amount of spatial unmasking are large and arise from individual differences in (1) HRTFs, (2) overall binaural sensitivity, and (3) the way in which binaural sensitivity varies with spatial configuration of target and masker. The Colburn (1977b) model of binaural processing predicts overall trends in behavioral measures of binaural unmasking, but fails to capture subject-specific variations in performance. The spatial configurations for which model predictions are least accurate are the positions for which large ILDs arise in masker and/or target, conditions that have not been extensively tested in previous studies. The current results suggest that the Colburn model must be modified so that subject differences in binaural sensitivity

vary not only in overall magnitude but as a function of the interaural differences in the masker.

While predictions from the Colburn model (taking into account differences in the stimuli presented to the individual subjects as well as individual differences in binaural sensitivity) cannot account for some small but significant intersubject differences in spatial unmasking, rough predictions of the amount of spatial unmasking capture most of the observed changes in detection threshold with spatial configuration. For instance, generic acoustic models of HRTFs (e.g., KEMAR measurements or spherical-head model predictions) combined with predictions of binaural unmasking using ''average'' model parameters should produce predictions that fall within the range of behavior observed across a population of subjects.

VII. CONCLUSIONS

(1) Acoustic cues (particularly TMR and ILD) vary dramatically with source distance and direction for nearby sources. Therefore, when source distance varies, the effect of source location on both the better-ear and binaural contributions to spatial unmasking is complex.
(2) For nearby sources, the better-ear contribution to pure-tone spatial unmasking can be very large (as much as 25 dB) compared to conditions where sources are relatively far from the listener.
(3) The binaural contribution to spatial unmasking decreases with increasing masker ILD. As a result, the binaural contribution to spatial unmasking is smaller for lateral sources very near the head than for more distant sources at the same lateral angle relative to the listener.
(4) Intersubject differences in spatial unmasking are larger for nearby sources than for far sources, in part because there are larger acoustic differences in HRTFs for nearby sources compared to more distant sources. However, there also are subject-specific differences both in binaural sensitivity and on how ILDs influence binaural sensitivity.
(5) Predictions based on Colburn's analysis (1977b) show the correct general trends in binaural detection for both near and far sources, but cannot account for small, but consistent, subject-specific differences in performance, particularly when large ILDs are present in the masker.

ACKNOWLEDGMENTS

This work was supported in part by AFOSR Grant No. F49620-98-1-0108 and the Alfred P. Sloan Foundation. Portions of this work were presented at the Spring 2000 meeting of the Acoustical Society of America. H. Steven Colburn provided valuable input throughout this work, including help in putting the results in appropriate context. Les Bernstein, Adelbert Bronkhorst, and an anonymous reviewer provided valuable criticism and comments on a previous draft of this paper.

APPENDIX: BINAURAL MODELING

A modified version of the model presented in Colburn (1977b) was used to predict the amount of binaural unmasking, defined as the difference in detection thresholds when target and masker are at the same spatial location and when they are in different locations. The predicted amount of binaural unmasking for a target at frequency f_0 is computed as

$$s(f_0,\alpha_T,\phi_T,\alpha_M,\phi_M,\mathrm{BMLD},K) = \sqrt{\max\left(1,\frac{\alpha_T^4}{\alpha_M^4}\right)+(2\cdot 10^{\mathrm{BMLD}/10}-1)R(\alpha_M,10^{K/10})\frac{F^2(\phi_M,f_0)}{16}\left(1+\frac{\alpha_T^2}{\alpha_M^2}-2\frac{\alpha_T}{\alpha_M}\cos(\phi_M-\phi_T)\right)^2}, \quad \text{(A1)}$$

where $\alpha_T = 10^{\mathrm{ILD}-T/20}$; $\alpha_M = 10^{\mathrm{ILD}-M/20}$; ILD$-T$ and ILD$-M$ are the interaural level differences in target and masker (respectively) in dB; ϕ_T and ϕ_M are the IPDs of target and masker (respectively) in radians; BMLD is the (subject-specific) binaural masking level difference in dB; K is the level of masker relative to absolute detection threshold in quiet, in dB; and the functions F^2 and R are defined below (all evaluated at the target frequency).

Function F^2 represents the extent to which phase shifts in masker cannot be compensated by internal time delays. This function is given by

$$F^2(\phi_M,f_0) = \frac{\Sigma_{k=-1000}^{1000} p(\phi_M/2\pi f_0 + k/f_0, f_0)\exp\{-G^2(f_0)[1-\gamma(\phi_M/2\pi f_0 + k/f_0)]\}}{\Sigma_{k=-1000}^{1000} p(k/f_0, f_0)\exp\{-G^2(f_0)[1-\gamma(k/f_0)]\}}, \quad \text{(A2)}$$

where $p(\tau,f)$ represents the relative number of interaural coincidence detectors (i.e., neurons in the medial superior olive) tuned to ITD τ and frequency f; $G(f)$ represents the synchrony of firings of the auditory nerve at frequency f (squared to account for the sharpening of synchrony in the cochlear nucleus); and $\gamma(\tau)$ is the envelope of the autocorrelation function of the auditory nerve fiber impulse response at autocorrelation delay τ. In the current realization of the model, function $p(\tau,f)$ was modified to allow for a frequency dependence in the distribution of interaural coincidence detectors (as suggested by Stern and Shear, 1996), using

$$p(\tau,f_0)=\begin{cases} C(e^{-2\pi k_l|0.2|}-e^{-2\pi k_h|0.2|})/0.2, & |\tau|<0.2 \text{ ms},\\ C(e^{-2\pi k_l|0.2|}-e^{-2\pi k_h|0.2|})/|\tau|, & |\tau|\geqslant 0.2 \text{ ms},\end{cases}$$

$$k_h=3\times 10^6,$$

$$k_l\begin{cases} 0.1(f_0 10^{-3})^{1.1}, & f_0\leqslant 1200 \text{ Hz},\\ 0.1(1200\times 10^{-3})^{1.1}, & f_0>1200 \text{ Hz},\end{cases} \tag{A3}$$

$$C=\begin{cases} 0.1534, & f_0=500 \text{ Hz},\\ 0.2000, & f_0=1000 \text{ Hz}.\end{cases}$$

$G(f)$ is given by

$$G(f_0)=\begin{cases} \sqrt{10}, & f_0\leqslant 800 \text{ Hz},\\ \sqrt{10}\dfrac{800}{f_0}, & f_0>800 \text{ Hz}.\end{cases} \tag{A4}$$

$\gamma(\tau)$ is given by

$$\gamma(\tau)=\begin{cases} 2.359\times 10^{-4}+1.5207\times 10^8\tau^4-1.764\times 10^4\tau^2 & \\ \qquad +0.993, & |\tau|\leqslant 0.006,\\ -97.3236|\tau|+1.139, & |\tau|>0.006,\end{cases} \tag{A5}$$

where τ is in milliseconds.

Finally, function $R(\alpha,K)$ characterizes the decrease in the number of activated auditory nerve fibers in the ear receiving the less intense signal as a function of masker ILD. The current implementation uses a modified version of Eq. (35) from Colburn (1977b):

$$R(\alpha_n)=\begin{cases} \left(\dfrac{10\log_{10}\alpha_n^{-2}K}{40}\right)^2, & \alpha_n^{-2}K\leqslant 10^4,\\ 1, & \alpha_n^{-2}K>10^4,\end{cases} \tag{A6}$$

where K is the ratio of the spectrum level at the more intense ear to the detection threshold in quiet. This implementation of the model assumes that the auditory nerve fibers at each target frequency have thresholds uniformly distributed (on a dB scale) over a 40-dB range above the absolute detection threshold for that frequency.

[1]System identification using a MLS depends on circular convolution techniques. Theoretically, the approach requires the MLS to be concatenated with itself and presented an infinite number of times to ensure that the system is in its steady-state response prior to measuring the response (see Vanderkooy, 1994). The resulting estimated system response is a time-aliased version of the true system response. In the current measures, the MLS was presented twice and the response to the second repetition was recorded. Given the length of the MLS used, the room characteristics of and ambient noise in the environment in which we were measuring, and the noise in our measurement system, the steady-state response can be approximated with only two repetitions of the MLS and no significant time aliasing is present in our measurements.

[2]Note that this analysis assumes that detection performance depends only on the target-to-masker ratio or TMR and is independent of the overall masker level, an assumption that is not valid if the masker is near absolute threshold or at very high presentation levels. For instance, imagine two masker locations so distant from the listener that the masker is inaudible. These masker locations would produce identical signal detection thresholds if the experiment were performed with the distal stimulus intensity fixed; however, our technique might adjust the masker by different amounts for these two masker locations in order to achieve a fixed proximal stimulus level at the ear of the listener, producing two different estimates of spatial unmasking. While holding the distal masker intensity fixed may seem more natural and intuitive than holding the proximal stimulus level constant, the overall presentation level of the masker would span an extraordinarily large range in the current experiments because the masker distance varied between 15 cm and 1 m in addition to varying in direction. Therefore, we elected to fix the proximal masker intensity.

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

Brungart, D. S., and Rabinowitz, W. M. (**1999**). ''Auditory localization of nearby sources. I. Head-related transfer functions,'' J. Acoust. Soc. Am. **106**, 1465–1479.

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.

Doll, T. J., and Hanna, T. E. (**1995**). ''Spatial and spectral release from masking in three-dimensional auditory displays,'' Hum. Factors **37**, 341–355.

Duda, R. O., and Martens, W. L. (**1998**). ''Range dependence of the response of a spherical head model,'' J. Acoust. Soc. Am. **104**, 3048–3058.

Durlach, N. I., and Colburn, H. S. (**1978**). ''Binaural phenomena,'' in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), pp. 365–466.

Ebata, M., Sone, T., and Nimura, T. (**1968**). ''Improvement of hearing ability by directional information,'' J. Acoust. Soc. Am. **43**, 289–297.

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.

Gatehouse, R. W. (**1987**). ''Further research on free-field masking,'' J. Acoust. Soc. Am. Suppl. 1 **82**, S108.

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.

Kidd, Jr., G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). ''Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**, 422–431.

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

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego).

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.

Santon, F. (**1987**). ''Detection d'un son pur dans un bruit masquant suivant l'angle d'incidence du bruit. Relation avec le seuil de reception de la parole,'' Acustica **63**, 222–230.

Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. (**2000**). ''Tori of confusion: Binaural localization cues for sources within reach of a listener,'' J. Acoust. Soc. Am. **107**, 1627–1636.

Shinn-Cunningham, B. G., Schickler, J., Kopco, N., and Litovsky, R. Y. (**2001**). ''Spatial unmasking of nearby speech sources in a simulated anechoic environment,'' J. Acoust. Soc. Am. **110**, 1118–1129.

Stern, R. M., and Shear, G. D. (**1996**). ''Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay,'' J. Acoust. Soc. Am. **100**, 2278–2288.

Vanderkooy, J. (**1994**). ''Aspects of MLS measuring systems,'' J. Audio Eng. Soc. **42**, 219–231.