Spatial unmasking of nearby speech sources in a simulated anechoic environment

Barbara G. Shinn-Cunningham[a)]
Boston University Hearing Research Center, Departments of Cognitive and Neural Systems and Biomedical Engineering, Boston University, 677 Beacon St., Room 311, Boston, Massachusetts 02215

Jason Schickler
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Norbert Kopčo
Boston University Hearing Research Center, Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215

Ruth Litovsky
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Spatial unmasking of speech has traditionally been studied with target and masker at the same, relatively large distance. The present study investigated spatial unmasking for configurations in which the simulated sources varied in azimuth and could be either near or far from the head. Target sentences and speech-shaped noise maskers were simulated over headphones using head-related transfer functions derived from a spherical-head model. Speech reception thresholds were measured adaptively, varying target level while keeping the masker level constant at the "better" ear. Results demonstrate that small positional changes can result in very large changes in speech intelligibility when sources are near the listener as a result of large changes in the overall level of the stimuli reaching the ears. In addition, the difference in the target-to-masker ratios at the two ears can be substantially larger for nearby sources than for relatively distant sources. Predictions from an existing model of binaural speech intelligibility are in good agreement with results from all conditions comparable to those that have been tested previously. However, small but important deviations between the measured and predicted results are observed for other spatial configurations, suggesting that current theories do not accurately account for speech intelligibility for some of the novel spatial configurations tested. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1386633]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.71.An, 43.66.Rq [LRB]

I. INTRODUCTION

When a target of interest (T) is heard concurrently with an interfering sound (a "masker," M), the locations of both target and masker have a large effect on the ability to detect and perceive the target. Previous studies have examined how T and M locations affect performance in both detection (e.g., see the review in Durlach and Colburn, 1978 or, for example, recent work such as Good, Gilkey, and Ball, 1997) and speech intelligibility tasks (e.g., see the recent review by Bronkhorst, 2000). Generally speaking, when the T and M are located at the same position, the ability to detect or understand T is greatly affected by the presence of M; when either T or M is displaced, performance improves.

While there are many studies of spatial unmasking for speech (e.g., see Hirsh, 1950; Dirks and Wilson, 1969; MacKeith and Coles, 1971; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988; Bronkhorst and Plomp, 1990; Peissig and Kollmeier, 1997; Hawley, Litovsky, and Colburn,

1999), all of the previous studies examined targets and maskers that were located far from the listener. These studies examined spatial unmasking as a function of angular separation of T and M without considering the effect of distance. One goal of the current study was to measure spatial unmasking for a speech reception task when a speech target and a speech-shaped noise masker are within 1 meter of the listener. In this situation, changes in source location can give rise to substantial changes in both the overall level and the binaural cues in the stimuli reaching the ears (e.g., see Duda and Martens, 1997; Brungart and Rabinowitz, 1999; Shinn-Cunningham, Santarelli, and Kopčo, 2000). Because the acoustics for nearby sources can differ dramatically from those of more distant sources, insights gleaned from previous studies may not apply in these situations. In addition, previous models (which do a reasonably good job of predicting performance on similar tasks; e.g., see Zurek, 1993) may not be able to predict what occurs when sources are close to the listener precisely because the acoustic cues at the ears are so different than those that arise for relatively distant sources.

For noise maskers that are statistically stationary (such

[a)]Electronic mail: shinn@cns.bu.edu

as steady-state broadband noise in anechoic settings, but not, for instance, amplitude-modulated noise or speech maskers), spatial unmasking can be predicted from simple changes in the acoustic signals reaching the ears (e.g., see Bronkhorst and Plomp, 1988; Zurek, 1993). For T fixed directly in front of a listener, lateral displacement of M causes changes in (1) the relative level of the T and M at the ears (i.e., the target to masker level ratio, or TMR), which will differ at the two ears (a monaural effect) and (2) the interaural differences in T compared to M (a binaural effect, e.g., see Zurek, 1993). For relatively distant sources, the first effect arises because the level of the masker reaching the farther ear decreases (particularly at moderate and high frequencies) as the masker is displaced laterally (giving rise to the acoustic "head shadow"). Thus, as M is displaced from T, one of the two ears will receive less energy from M, resulting in a "better-ear advantage." Also, for relatively distant sources the most important binaural contribution to unmasking occurs when T and M give rise to different interaural time differences (ITDs), resulting in differences in interaural phase differences (IPDs) in T and M, at least at some frequencies (e.g., see Zurek, 1993). The overall size of the release from masking that can be obtained when T is located in front of the listener and a steady-state M is laterally displaced (and both are relatively distant from the listener) is on the order of 10 dB (e.g., see Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988; Peissig and Kollmeier, 1997; Bronkhorst, 2000). Of this 10 dB, roughly 2–3 dB can be attributed to binaural processing of IPDs, with the remainder resulting from head shadow effects (e.g., see Bronkhorst, 2000).

If one restricts the target and masker to be at least 1 meter from the listener, the only robust effect of distance on the stimuli at the ears is a change in overall level (e.g., see Brungart and Rabinowitz, 1999). Thus, for relatively distant sources, the effect of distance can be predicted simply from considering the dependence of overall target and masker level on distance; there are no changes in binaural cues, the better-ear-advantage, or the difference in the TMR at the better and worse ears.

There are important differences between how the acoustic stimuli reaching the ears change when a sound source is within a meter of and when a source is more than a meter from the listener (e.g, see Duda and Martens, 1997; Brungart and Rabinowitz, 1999; Shinn-Cunningham et al., 2000). For instance, a small displacement of the source towards the listener can cause relatively large increases in the levels of the stimuli at the ears. In addition, for nearby sources, the interaural level difference (ILD) varies not only with frequency and laterality but also with source distance. Even at relatively low frequencies, for which naturally occurring ILDs are often assumed to be zero (i.e., for sources more than about a meter from the head), ILDs can be extremely large. In fact, these ILDs can be broken down into the traditional "head shadow" component, which varies with direction and frequency, and an additional component that is frequency independent and varies with source laterality and distance (Shinn-Cunningham et al., 2000).

In the "distant" source configurations previously studied, the better ear is only affected by the relative laterality of

T versus M; the only spatial unmasking that can arise for T and M in the same direction is a result of equal overall level changes in the stimuli at the two ears. Moving T closer than M will improve the SRT while moving T farther away will decrease performance, simply because the level of the target at both ears varies with distance (equivalently). In contrast, when a source is within a meter of the head, the relative level of the source at the two ears depends on distance. Changing the distance of T or M can lead not only to changes in overall energy, but changes in the amount of unmasking that can be attributed to binaural factors, the difference in the TMR at the two ears (as a function of frequency), and even which is the better ear. In addition, overall changes in the level at the ears can be very large, even for small absolute changes in distance. Although the distances for which these effects arise are small, in a real "cocktail party" it is not unusual for a listener to be within 1 meter of a target of interest (i.e., in the range for which these effects are evident).

We are aware of only one previous study of spatial unmasking for speech intelligibility in which large ILDs were present in both T and M (Bronkhorst and Plomp, 1988). In this study, the total signal to one ear was attenuated in order to simulate monaural hearing impairment. Unlike the Bronkhorst and Plomp study, the current study focuses on the spatial unmasking effects that occur when realistic combinations of IPD and ILD, consistent with sources within 1 m of the listener, are simulated for different T and M geometries.

II. EXPERIMENTAL APPROACH

A common measure used to assess spatial unmasking effects on speech tasks is the speech reception threshold (SRT), or the level at which the target must be presented in order for speech intelligibility to reach some predetermined threshold level. The amount of spatial unmasking can be summarized as the difference (in dB) between the SRT for the target/masker configuration of interest and the SRT when T and M are located at the same position.

In these experiments, SRT was measured for both "nearby" sources (15 cm from the center of the listener's head) and "distant" sources (1 m from the listener). Tested conditions included those in which (1) the speech target was in front of the listener and M was displaced in angle and distance; (2) M was in front of the listener and T displaced in angle and distance; and (3) T and M were both located on the side, but T and M distances were manipulated.

The goals of this study were to (1) measure how changes in spatial configuration of T and M affect SRT for sources near the listener; (2) explore how the interaural level differences that arise for nearby sources affect spatial unmasking; and (3) quantify the changes in the acoustic cues reaching the two ears when T and/or M are near the listener.

A. Subjects

Four healthy undergraduate students (ages ranging from 19–23 years) performed the tests. All subjects had normal hearing thresholds (within 15 dB HL) between 250 and 8000 Hz as verified by an audiometric screening. All subjects were native English speakers. One of the subjects was author JS
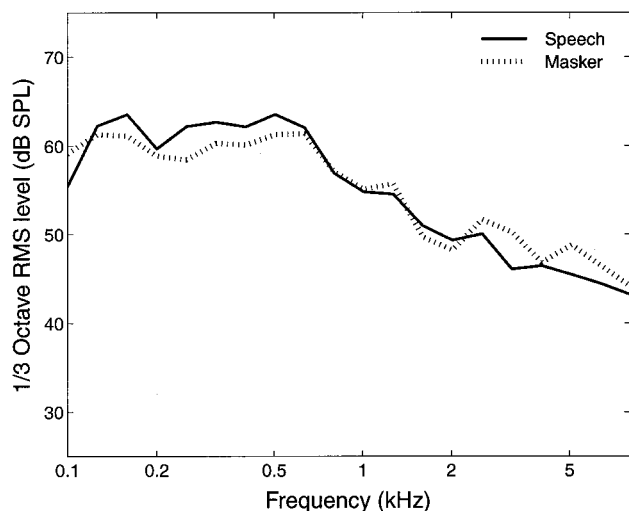
FIG. 1. Average spectral shape of speech-shaped noise masker and speech targets, prior to HRTF processing.

with relatively little experience in psychoacoustic experiments; the other three subjects were naive listeners with no prior experience.

B. Stimuli

1. Source characteristics

In the experiments, the target (T) consisted of a high-context sentence selected from the IEEE corpus (IEEE, 1969). Sentences were chosen from 720 recordings made by two different male speakers. These materials have been employed previously in similar speech intelligibility experiments (Hawley *et al.*, 1999). The recordings, ranging from 2.41–3.52 s in duration, were scaled to have the same rms pressure value in their "raw" (nonspatialized) forms. An example sentence is "The DESK and BOTH CHAIRS were PAINTED TAN," with capitalized words representing "key words" that are scored in the experiment (see Sec. C).

The masker (M) was speech-shaped noise generated to have the same spectral shape as the average of the speech tokens used in the study. For each masker presentation, a random 3.57-s sample was taken from a long (24-s) sample of speech-shaped noise (this length guaranteed that all words in all sentences were masked by the noise). Figure 1 shows the rms pressure level in 1/3-octave bands (dB SPL) of the 24-s-long masking noise and the average of the spectra of the speech samples used in the study.

2. Stimulus generation

Raw digital stimuli (i.e., IEEE sentences and speech-shaped noise sampled at 20 kHz) were convolved with spherical-head head-related transfer functions (HRTFs) off-line (see below). T and M were then scaled (in software) to the appropriate level for the current configuration and trial. The resulting binaural T and M were then summed in software and sent to Tucker-Davis Technologies (TDT) hardware to be converted into acoustic stimuli (using the same equipment setup described in Hawley *et al.*, 1999). Digital signals were processed through left- and right-channel D/A converters (TDT DD3-8), low-pass filters (10-kHz cutoff; TDT

FT5), and attenuators (TDT PA4). The resulting binaural analog signals were passed through a Tascam power amplifier (PA-20 MKII) connected to Sennheiser headphones (HD 520 II). No compensation for the headphone transfer function was performed. A personal computer (Gateway 2000 486DX) controlled all equipment and recorded results.

3. Spatial cues

In order to simulate sources at different positions around the listener, spherical-head HRTFs were generated for all the positions from which sources were to be simulated. These HRTFs were generated using a mathematical model of a spherical (9-cm-radius) head with diametrically opposed point receivers (ears; for more details about the model or traits of the resulting HRTFs see Rabinowitz *et al.*, 1993; Brungart and Rabinowitz, 1999; Shinn-Cunningham *et al.*, 2000). Source stimuli (T and M) were convolved to generate binaural signals similar to those that a listener would experience if the T and M were played from specific positions in anechoic space.

It should be noted that the spherical-head HRTFs are not particularly realistic. They contain no pinnae cues (i.e., contain no elevation information), are more symmetrical than true HRTFs, and are not tailored to the individual listener. As a result, sources simulated from these HRTFs are distinguishably different from sounds that would be heard in a real-world anechoic space. As a result, the sources simulated with these HRTFs may not have been particularly "externalized," although they were generally localized at the simulated direction. There was no attempt to evaluate the realism, externalization, or localizability of the simulated sources using the spherical-head HRTFs. Nonetheless, the spherical-head HRTFs contain all the acoustic cues that are unique to sources within 1 m of the listener (i.e., large ILDs that depend on distance, direction, and frequency; changes in IPD with changes in distance), a result confirmed by comparisons with measurements of human subject and KEMAR HRTFs for sources within 1 m (see, for example, Brown, 2000; Shinn-Cunningham, 2000). Further, because the unique acoustic attributes that arise for free-field near sources are captured in these HRTFs, we believe that any unique behavioral consequences of listening to targets and maskers that are near the listener will be observed in these experiments.

4. Spatial configurations

In different conditions, the target and masker were simulated from any of six locations in the horizontal plane containing the ears; that is, at three azimuths (0°, 45°, and 90° to the right of midline) and two distances from the center of the head (15 cm and 1 m). The 15 spatial configurations investigated in this study are illustrated in Fig. 2. The three panels depict three different conditions: target location fixed at (0°, 1 m) [Fig. 2(a)], masker fixed at (0°, 1 m) [Fig. 2(b)] and target and masker both at 90° [Fig. 2(c)]. All subsequent graphs are arranged similarly. Note that the configuration in which T and M are both located at (0°, 1 m) appears in both panels (a) and (b) of Fig. 2; this spatial configuration was the (diotic) reference used in computing spatial masking effects.
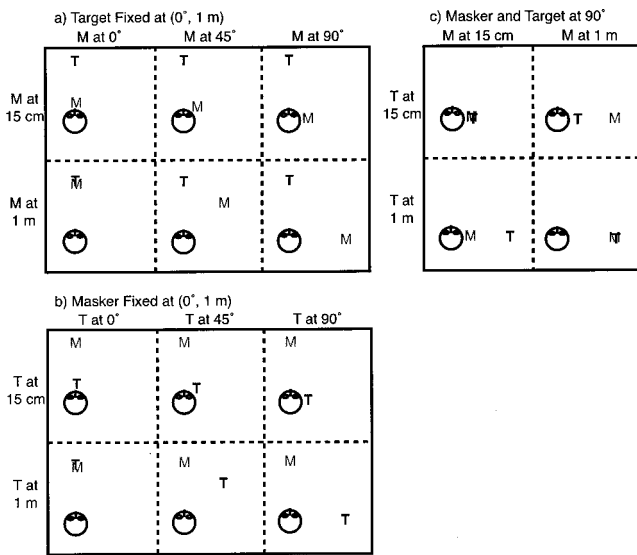
FIG. 2. Spatial configurations of target (T) and masker (M). Conditions: (a) T fixed (0°, 1 m); (b) M fixed (0°, 1 m); and (c) T and M at 90°.

5. Presentation level

If we had simulated a masking source emitting the same energy from different distances and directions, the level of the masker reaching the better ear would vary dramatically with the simulated position of M. In addition, depending on the location of M, the better ear can be either the ear nearer or farther from T. For instance, if T is located at (90°, 1 m) and M is located at (90°, 15 cm) [see Fig. 2(c), bottom left panel], T is nearer to the right ear, but the left ear will be the "better ear."

In order to roughly equate the masker energy *reaching the better ear* (as opposed to keeping constant the distal energy of the simulated masker), masker level was normalized so that the root-mean-square (rms) pressure of M at the better ear was always 72 dB SPL. With this choice, the masker was always clearly audible at the worse ear (even when the masker level was lower at the worse ear) and at a comfortable listening level at the worse ear (even when the masker level was higher at the worse ear). Of course, the worse-ear masker level varied with spatial configuration, and could either be greater or less than 72 dB SPL depending on the locations of T and M.

C. Experimental procedure

All experiments were performed in a double-walled sound-treated booth in the Binaural Hearing Laboratory of the Boston University Hearing Research Center.

An adaptive procedure was used to estimate the SRT for each spatial configuration of T and M. In each adaptive run, the T level was adaptively varied to estimate the SRT, which was defined as the level at which subjects correctly identified 50% of the T sentence key words.

For each configuration, at least three independent, adaptive-run threshold estimates were averaged to form the final threshold estimate. If the standard error in the repeated measures was greater than 1 dB, additional adaptive runs

were performed until the standard error in this final average was equal to or less than 1 dB.

The T and M locations were not known *a priori* by the subject, but were held constant through a run, which consisted of ten trials. Runs were ordered randomly and broken into sessions consisting of approximately seven runs each.

Within a run, the first sentence of each block was repeated multiple times in order to set the T level for subsequent trials. The first sentence in each run was first played at 44 dB SPL in the better ear. The sentence was played repeatedly, with its intensity increased by 4 dB with each repetition, until the subject indicated (by subjective report) that he could hear the sentence. The level at which the listener reported understanding the initial sentence set the T level for the second trial in the run. On each subsequent trial, a new sentence was presented to the subject. The subject typed in the perceived sentence on a computer keyboard. The actual sentence was then displayed (along with the subject's typed response) on a computer monitor (visible to the subject) with five "key words" capitalized. The subject then counted up and entered into the computer the number of correct key words perceived. Scoring was strict, with incorrect suffixes scored as "incorrect;" however, homophones and misspellings were not penalized. Listeners heard only one presentation of each T sentence.

If the subject identified at least three of the five key words correctly, the level of the T was decreased by 2 dB on the subsequent trial. Otherwise (i.e., if the subject identified two or fewer key words), the level of the T was increased by 2 dB. Thus, if the subject performed at or above 60% correct, the task was made more difficult; if the subject performed at or below 40% correct, the task was made easier. This procedure (which, in the limit, will converge to the presentation level at which the subject will achieve 50% correct) was repeated until ten trials were scored. SRT was estimated as the average of the presentation levels of the T on the last eight (of ten) trials.

III. RESULTS

A. Target-to-masker levels at speech reception threshold

In order to visualize the changes in relative spectral levels of T and M with spatial configuration, the average TMR in third-octave spectral bands was computed as a function of center frequency at 50%-correct SRT and plotted in Fig. 3.

By construction (because T and M have the same spectral shape), the TMR is equal in both ears and independent of frequency for configurations in which T and M are located at the same position (i.e., for two diotic configurations and two configurations with T and M at 90°). However, in general, the overall spectral shape of both T and M depends on spatial configuration and the TMR varies with frequency.

In the diotic reference configuration, the TMR is −7.6 dB [e.g., see Fig. 3(a), bottom left panel]. In other words, when the diotic sentence is presented at a level 7.6 dB below the diotic speech-shaped noise, subjects achieve threshold performance in the reference configuration. This diotic reference TMR is plotted as a dashed horizontal line in all panels
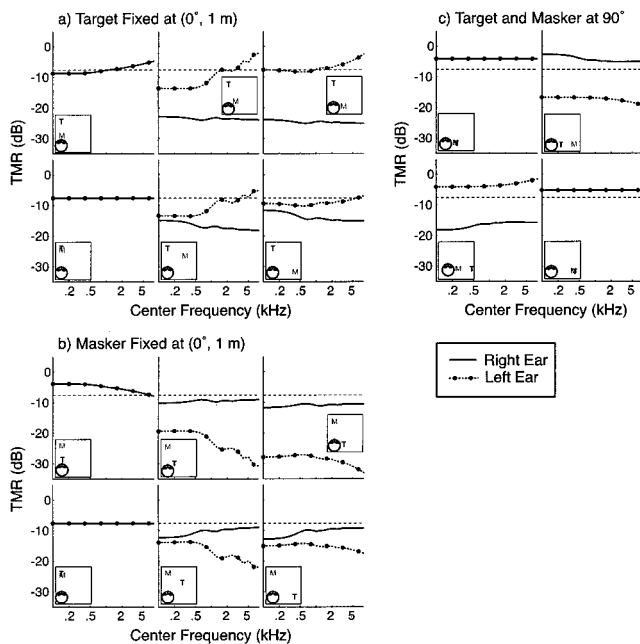
FIG. 3. Target-to-masker level ratio (TMR) in 1/3-octave frequency bands for left (dotted lines with symbols) and right (solid lines) ears as a function of center frequency at speech reception threshold. Conditions: (a) T fixed (0°, 1 m); (b) M fixed (0°, 1 m); and (c) T and M at 90°.

in order to make clear how the TMR varies with spatial configuration. When threshold TMR at the better ear is lower than the diotic reference TMR, the results indicate the presence of spatial masking effects that cannot be explained by overall level changes. In such cases, other factors, such as differences in binaural cues in T and M, are likely to be responsible for the improvements in SRT.

Figure 3(a) shows the results when T is fixed at (0°, 1 m). For these spatial configurations, the TMR at the better (left) ear (dotted line with symbols) is generally equal to or smaller than the reference TMR. TMR is lowest when M is located at (45°, 1 m) (bottom center panel); in this case, the TMR at low frequencies is as much as 14 dB below the diotic reference TMR (the TMR at higher frequencies is approximately equal to the diotic reference TMR). The worse-ear TMR (right ear; solid line) is often much smaller than that of the better ear, particularly when M is at 15 cm.

When the masker is fixed at the reference position (0°, 1 m) [Fig. 3(b)], the TMR at the better (right) ear (solid line) is below the reference TMR at all frequencies for all four cases in which T is laterally displaced. The magnitude of this improvement is roughly the same (2–3 dB) whether T is near or far, at 45° or 90°. In the diotic case for which T is at (0°, 1 m) and M is at (0°, 15 cm) [top-left panel in Fig. 3(b)], the TMR is roughly 4 dB larger than in the diotic reference configuration. This result indicates a small spatial *disadvantage* in this diotic configuration compared to the "typical" diotic reference configuration when T and M are both distant after taking into account the overall level of M.

In all four configurations for which both T and M are located laterally [Fig. 3(c)], the TMR at the better ear is roughly 3–4 dB larger at all frequencies than the diotic reference TMR. In other words, listeners need a laterally lo-

cated speech source to be presented at a relatively high level when it competes with a masker located in the same lateral direction. This is even true when M is at 1 m and T is at 15 cm [top right panel of Fig. 3(c)], despite the fact that the better- (right-) ear stimulus is at a substantially higher overall level than the worse- (left-) ear stimulus in this configuration.

B. Mean difference in monaural TMRs

The results in Fig. 3 show that the difference in the TMRs at the two ears can be very large when either T or M is near the listener (a direct consequence of the very large ILDs that arise for these sources). This difference is important for understanding and quantifying the advantage of having two ears, independent of any binaural processing advantage. For instance, if a monaurally impaired listener's intact ear is the acoustically worse ear, the impaired listener will be at a larger disadvantage for many of the tested configurations than when both T and M are distant. In order to quantify the magnitude of these acoustic effects, the absolute value of the mean of the difference in left- and right-ear TMR was calculated, averaged across frequencies up to 8000 Hz.

The leftmost data column in Table I gives the mean of $|\mathrm{TMR_{right}} - \mathrm{TMR_{left}}|$ at SRT, averaged across frequency. Because the TMRs change with frequency, this estimate cannot predict SRT directly; for instance, moderate frequencies (e.g., 2000–5000 Hz) convey substantially more speech information than lower frequencies. Nonetheless, these calculations give an objective, acoustic measure, weighting all frequencies equally, of differences in the better and worse ear signals.

From symmetry and because T and M have the same spectral shape, the difference in better- and worse-ear TMR is the same if M is held at (0°, 1 m) and T is moved or T is fixed and M is moved (see Table I, comparing top and center sections).

For configurations in which both T and M are far from the head, the acoustic difference in the TMRs at the two ears ranges from 5–10 dB, depending on the angular separation of T and M. If T remains fixed and a laterally located M is moved from 1 m to 15 cm (or vice versa), the difference between the better and worse ear TMR increases substantially. For instance, with T fixed at (0°, 1 m) and M at (90°, 15 cm), the difference in TMR is nearly 20 dB (third line in Table I). For spatial configurations in which one source is near the head but not in the median plane, part of this difference in better- and worse-ear TMR arises from "normal" head-shadow effects and part arises due to differences in the relative distance from the source to the two ears (Shinn-Cunningham *et al.*, 2000).

In the configurations for which both T and M are located at 90°, there is no difference in the TMR at the ears when T and M are at the same distance. When one source is near and one is far, the TMR at the ears differs by roughly 13 dB.

It should be noted that there are even more extreme spatial configurations than those tested here. For instance, with T at (−90°, 15 cm) and M at (+90°, 15 cm) the acoustic difference in the TMRs at the two ears would be on the order of 40 dB (i.e., twice the difference obtained when one

Shinn-Cunningham *et al.*: Spatial unmasking of nearby speech sources

TABLE I. Spatial effects for different spatial configurations tested. Leftmost data column shows the mean of the absolute difference $|\text{TMR}_{\text{right}} - \text{TMR}_{\text{left}}|$ at SRT, averaged across frequencies up to 8000 Hz. The second data column gives the predicted magnitude of the difference in the monaural left- and right-ear SRTs from the Zurek model calculations. The third data column gives the binaural advantage calculated from Zurek model calculations (the difference in predicted SRT for binaural and monaural better-ear listening conditions).

			Left/right asymmetry (acoustic analysis) (dB)	Left/right asymmetry (Zurek predictions) (dB)	Binaural advantage (Zurek predictions) (dB)
T (0°, 1 m)	M (15 cm)	M (0°)	0	0	0
		M (45°)	17.5	14.6	2.0
		M (90°)	19.6	17.9	1.5
	M (1 m)	M (0°)	0	0	0
		M (45°)	9.8	7.5	2.4
		M (90°)	6.4	5.2	2.2
M (0°, 1 m)	T (15 cm)	T (0°)	0	0	0
		T (45°)	17.5	14.5	1.5
		T (90°)	19.6	17.2	1.5
	T (1 m)	T (0°)	0	0	0
		T (45°)	9.8	7.5	1.9
		T (90°)	6.4	5.2	2.2
T & M (90°)	T (15 cm)	M (15 cm)	0	0	0
		M (1 m)	13.2	12.6	0.8
	T (1 m)	M (15 cm)	13.2	12.6	0.9
		M (1 m)	0	0	0

source is diotic and one source is at 90°, 15 cm). This analysis demonstrates that one novel outcome of T and M being very close to the head is that the difference in the TMRs at the two ears can be dramatically larger than in previously tested configurations.

C. Spatial unmasking

Figure 4 plots the amount of spatial unmasking for each spatial configuration.[1] In the figure, the amount of "spatial unmasking" equals the decrease in the distal energy the target source must emit for subjects to correctly identify 50% of the target key words if the distal energy emitted by the masking source were held constant. This analysis includes changes in the overall level of T and M reaching the ears with changes in source position (and assumes that SRT depends only on TMR and is independent of the absolute level of the masker for the range of levels considered).

When T is fixed at (0°, 1 m) [Fig. 4(a)], the release from masking is largest when the 1-m M is at 45° and decreases slightly when M is at 90°. The dependence of the unmasking on M distance is roughly the same for all M directions: moving M from 1 m to 15 cm increases the required T level by roughly 13 dB for M in all tested directions (0°, 45°, and 90°).

When M is fixed ahead [Fig. 4(b)], moving the 1-m-distant T to either 45° or 90° results in the same unmasking. Moving the T close to the head (15 cm) results in a large amount of spatial unmasking, primarily due to increases in the level of T reaching the ears. For a given T direction, the effect of decreasing the distance of T increases with its lateral angle.

Figure 4(c) shows the spatial unmasking that arises when T and M are both located at 90°. When T and M are at the same distance [either at 15 cm, circles at left of Fig. 4(c); or at 1 m, squares at right of Fig. 4(c)], there is a 3-dB increase in the level the target source must emit compared to the reference configuration. When T and M are at different distances, spatial unmasking results are dominated by differences in the relative distances to the head.
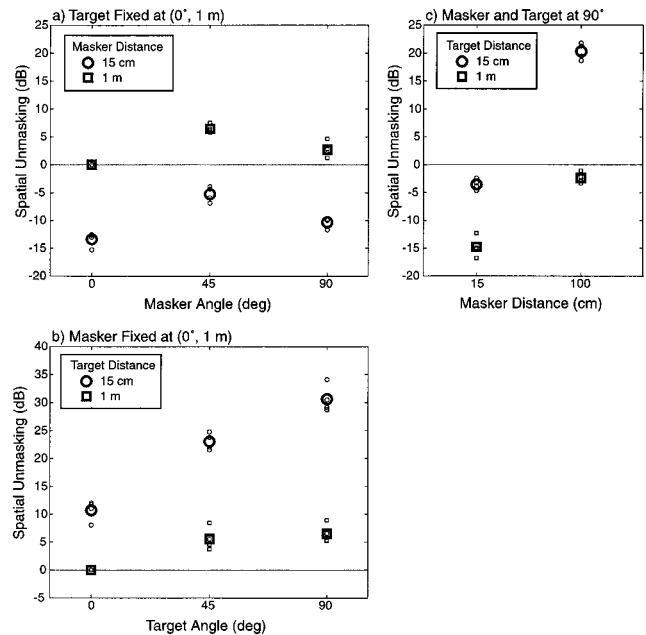

FIG. 4. Spatial advantage (energy a target emits at threshold for a constant-energy masker) relative to the diotic configuration. Positive values are decreases in emitted target energy. Large symbols give the across-subject mean; small symbols show individual subject results. Conditions: (a) T fixed (0°, 1 m); (b) M fixed (0°, 1 m); and (c) T and M at 90°.

D. Discussion

Our findings are generally consistent with previous results that show that speech intelligibility improves when T and M give rise to different IPDs, and that spatially separating a masker and target tends to reduce threshold TMR.

However, in some of the spatial configurations tested, the threshold TMR at the better ear is greater than the TMR in the diotic reference configuration. For instance, in all four spatial configurations with T and M at 90° [Fig. 3(c)], the better-ear TMR is roughly the same (independent of the relative levels of the better and worse ears) and elevated compared to the TMR in the diotic reference configuration. These results are inconsistent with predictions from previous models, which generally assume that binaural performance is always at least as good as would be observed if listeners were presented with the better-ear stimulus monaurally. Discrepancies between the current findings and predictions from an existing model (Zurek, 1993) are considered in detail in the next section.

For distant sources, changing the distance of T or M may change the overall level at the better ear, but it causes an essentially identical change at the worse ear. Thus, the difference between listening with the worse and the better ears is independent of T and M distance when T and M are at least 1 m from the listener. One of the novel effects that arises when either T or M is within 1 meter of the head is that the difference between the TMR at the better and worse ears can be dramatically larger than if both T and M are distant (see Table I). For the configurations tested, the difference in the TMRs at the two ears can be nearly double the difference that occurs when both T and M are at least a meter from the listener [e.g., 19.6 dB for a diotic T and M at (90°, 15 cm) versus 9.8 dB for diotic T and M at (90°, 1 m)].

Analysis of the spatial unmasking (Fig. 4) emphasizes the large changes in overall level that can arise with small displacements of a source near the listener. For the configurations tested, the change in the level that the target must emit to be intelligible against a constant level masker ranges from −31 to +15 dB (relative to the diotic reference configuration).

IV. MODEL PREDICTIONS

A. Zurek model of spatial unmasking of speech

Zurek (1993) developed a model based on the Articulation Index (AI,[2] Fletcher and Galt, 1950; ANSI, 1969; Pavlovic, 1987) to predict speech intelligibility as a function of target and masker location. AI is typically computed for a single-channel system as a weighted sum of target-to-masker ratios (TMRs) across third-octave frequency bands. In Zurek's model, the TMRs at both ears are considered, along with interaural differences in the T and M.

To compute the predicted intelligibility, Zurek's model first computes the actual TMR at each ear in each of 15 third-octave frequency bands (spaced logarithmically between 200 to 5000 Hz). The "effective TMR" (R_i) in each frequency band i is the sum of (1) the larger of the two true TMRs at the left and right ears and (2) an estimate of the "binaural advantage" in band i. The binaural advantage in
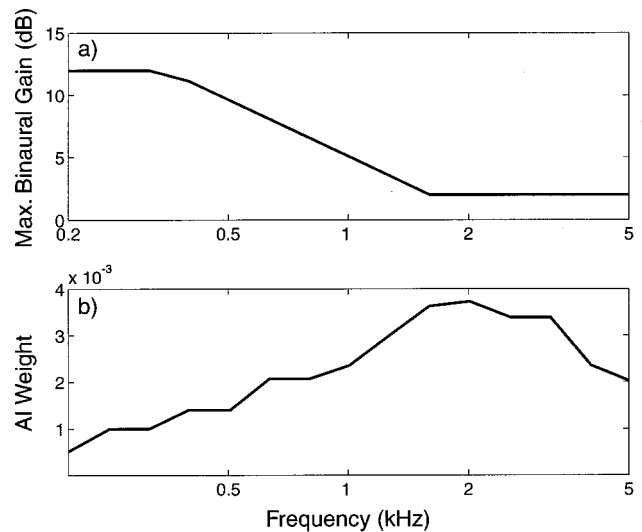

FIG. 5. Binaural AI model assumptions (Zurek, 1993). Panel (a) shows maximal binaural advantage (improvement in effective target-to-masker level ratio or TMR) as a function of frequency, which only arises when IPD of T and M differ by 180°. Panel (b) shows weighting of information at each frequency for speech intelligibility.

each band, derived from a simplified version of Colburn's model of binaural interaction (Colburn, 1977a, b), depends jointly on center frequency and the relative IPD of target and masker at the center frequency of the band. The advantage in a particular frequency band equals the estimated binaural masking level difference (BMLD) for a "comparable" tone-in-noise detection task. Specifically, if the difference in the IPD of T and M at the center frequency of band i is equal to x rad, the binaural advantage in band i is estimated as the expected BMLD when detecting a tone at the band center frequency in the presence of a diotic masker when the tone has an IPD of x rad. The maximum binaural advantage in a band [taken directly from Zurek, 1993, Fig. 15.2, and shown in Fig. 5(a) as a function of frequency] occurs when, at the band center frequency, the IPD of T and M differ by π rad. When the difference in the T and M IPD at the band center frequency is less than π rad, the binaural advantage in the band is lower (in accord with the Colburn model). The amount of information (γ_i) in each band (the "band efficiency") is computed as

$$\gamma_i = \begin{cases} 0, & R_i < -12 \text{ dB} \\ R_i + 12, & -12 \text{ dB} < R_i < 18 \text{ dB} \\ 30, & R_i > 18 \text{ dB} \end{cases} \quad (1)$$

This operation assumes that there is no incremental improvement in target audibility with increases in TMR above some asymptote (i.e., 18 dB) and no decrease in target audibility with additional decrements in TMR once the target is below masked threshold (i.e., −12 dB). The analysis implicitly assumes that the target is well above absolute threshold. Finally, the values of γ_i are multiplied by the frequency-dependent weights shown in Fig. 5(b) (which represent the relative importance of each frequency band for understanding speech) and summed to estimate the effective AI. The effective AI can take on values between 0.0 (if all R_i are less than or equal to 12 dB) and 1.0 (if all R_i are greater than or
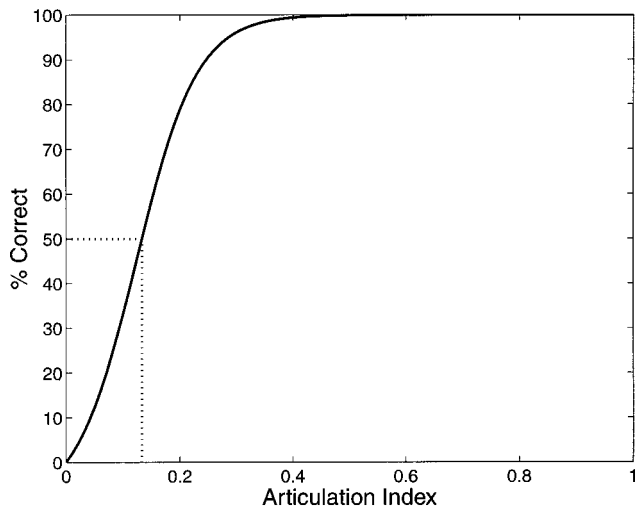
FIG. 6. Assumed relationship between AI and percent words correct assumed for high-context speech (as described in Hawley, 2000). Dashed lines show threshold level for the experiments reported herein.
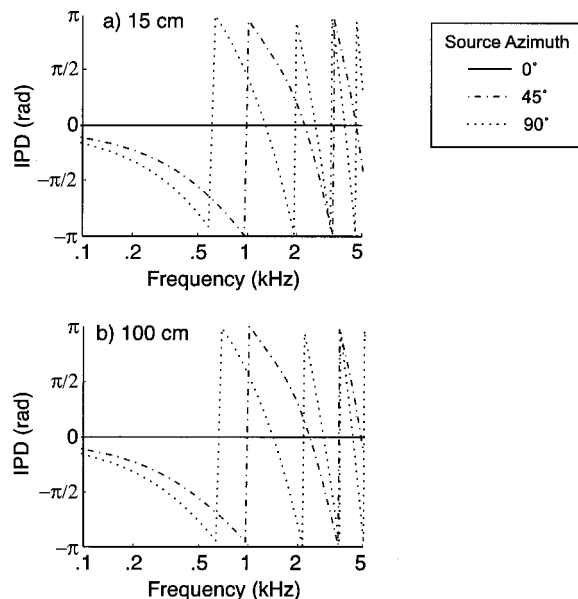

FIG. 7. Interaural phase differences as a function of frequency for the spherical-head HRTFs. (a) Near distance (15 cm) in top panel. (b) Far distance (1 m).

equal to 18 dB). For a given speech intelligibility task and a given set of speech materials, percent correct is a monotonic function of AI (e.g., see Kryter, 1962); for the high-context speech materials used in the present study, this correspondence, as derived by Hawley (2000), is shown in Fig. 6.

Using this model, Zurek (1993) was able to predict the spatial unmasking effects observed in a number of studies that used steady-state maskers (such as broadband noise) and positioned both T and M at a distance of at least 1 m from the subject (e.g., Dirks and Wilson, 1969; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988, among others). In this paper, we apply this model to cases when the target and/or masker are close to the subject (i.e., 15 cm).

B. Predicted speech intelligibility at speech reception threshold

In order to calculate model predictions of the current results, the IPDs in the spherical-head HRTFs were analyzed. Figure 7, which plots the IPD in the HRTFs (as a function of frequency) for the positions used in the study, shows that IPD varies dramatically with source laterality and only slightly with distance (e.g., see Brungart and Rabinowitz, 1999; Shinn-Cunningham *et al.*, 2000). Using the left- and right-ear TMRs at the measured SRT (Fig. 3), the difference in T and M IPD was used to compute the effective TMR (the TMR at the better ear, adjusted for binaural gain) and the "band efficiency" in each frequency band. From these values, the AI was calculated and used to predict percentage correct key words using the mapping shown in Fig. 6.

We applied a similar analysis to the left and right ear stimuli in isolation (i.e., for a comparable configuration but with one of the ears "turned off"). To generate these monaural predictions, the appropriate monaural TMR (Fig. 3) was used to compute the AI directly (excluding any binaural contributions). In this way, we predicted not only the percentage-correct words for binaural stimuli but also left- and right-ear monaural stimuli.

Figure 8 shows the predicted percentage correct on our

high-context speech task when the T and M levels equaled those presented at SRT. Predictions are shown for binaural listeners (x's) as well as monaural-left and monaural-right listeners (triangles and circles, respectively). The relative levels of T and M used in the predictions are those at which subjects correctly identified approximately 50% of the sentence key words. Thus, the model correctly predicts an observed result when the prediction is close to 50%. For our purposes, predictions falling within the gray area in each panel (within 10% of the defined 50%-correct threshold) are considered to match measured performance.[3] Note that in the
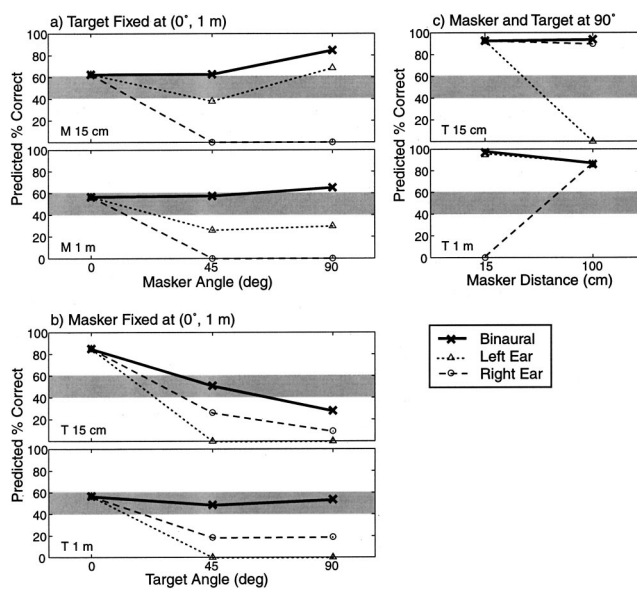

FIG. 8. Predicted percent-correct word scores from model using TMRs and binaural cues present at threshold (actual performance indicated by gray region). Bold exes show binaural model predictions; triangles and circles give monaural, left- and right-ear predictions, respectively. Conditions: (a) T fixed (0°, 1 m) and M at each of 6 locations; (b) M fixed (0°, 1 m) and T at each of 6 locations; and (c) T and M at 90° and 15 cm or 1 m.

model, predicted monaural performance (triangles or circles) is always less than or equal to binaural performance (exes), because any binaural processing will only increase the AI calculated from the better ear (and hence the predicted level of performance).

The one constant feature in Fig. 8 concerns the worse-ear monaural predictions. In every configuration for which the TMR differs in the two ears [four in Fig. 8(a) (circles), four in Fig. 8(b) (triangles), and two in Fig. 8(c) (rightmost triangle in top panel, leftmost circle in bottom panel)] the worse-ear, predicted percent correct is 0%.

Figure 8(a) shows predictions for T fixed ahead. For the diotic configurations [left side of Fig. 8(a)] both ears receive the same stimulus, left- and right-ear monaural predictions are identical, and there is no predicted benefit from listening binaurally. For all configurations in which M is at 1 m [lower panel, Fig. 8(a)], binaural predictions fall within or slightly above the expected range. Predictions for the better (left) ear are near 30% correct when the 1-m M is positioned laterally. When M is at 15 cm [upper panel in Fig. 8(a)], the binaural model predictions are generally higher than observed performance, but the error is only significant when M is at (90°, 15 cm) (binaural prediction near 90% correct). The monaural better-ear prediction is slightly below measured performance when M is at (45°, 15 cm) and substantially above measured performance when M is at (90°, 15 cm).

Figure 8(b) shows the predictions when M is fixed at (0°, 1 m). For this condition, the binaural predictions fit the data well for all configurations in which T is at the farther (1 m) distance [lower panel in Fig. 8(b)]. For the distant, laterally displaced T, better-ear predictions fall well below true binaural performance (19% correct for T at 45° and 90°). When T is at 15 cm, the binaural model predictions are less accurate, overestimating performance for T at 0° and underestimating performance for T at 90°.

In all four configurations in which T and M are positioned at 90° [Fig. 8(c)], the model predicts that both binaural performance and monaural better-ear performance should be much better than what was actually observed, with the predictions ranging from 86% to 95% correct.

C. Predicted spatial unmasking

The Zurek model (1993) was also used to predict the magnitude of the spatial unmasking in the various spatial configurations. To make these predictions, the mapping in Fig. 6 was used to predict the AI at which 50% of the key words are identified (see the dashed lines in Fig. 6). We then computed the level that T would have to emit in order to yield this threshold AI for each spatial configuration (assuming that the level emitted by M is fixed) and subtracted the level T would have to emit in the diotic reference configuration. Similar analysis was performed for left- and right-ear monaural signals in order to predict the impact of having only one functional ear.

Results of these predictions are shown in Fig. 9. In the figure, the large symbols show the mean unmasking found in the binaural experiments (presented previously in Fig. 4), while the lines with small symbols show the corresponding binaural (solid lines), left-ear (dashed lines), and right-ear
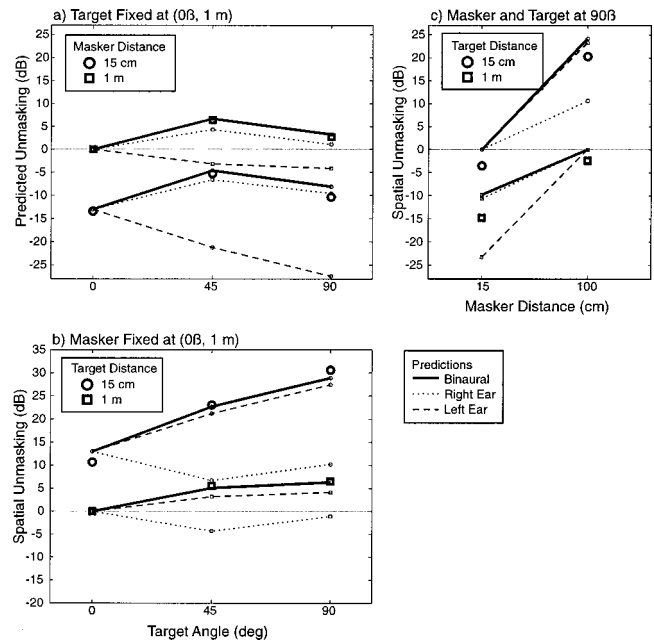

FIG. 9. Spatial advantage (energy a target emits at threshold for a constant-energy masker) and model predictions, relative to diotic reference. Symbols show across-subject means of measured spatial advantage, repeated from Fig. 4. Lines give model predictions: solid line for binaural model; dotted and dashed lines for left and right ears (without binaural processing), respectively. In any one configuration, the difference between the solid line and the better of the dotted or dashed lines gives the predicted binaural contribution to unmasking; the difference between the dotted and dashed lines yields the predicted better-ear advantage.

(dotted lines) predictions. To the extent that the model is accurate, the difference in binaural and better-ear predictions at each spatial configuration gives an estimate of the binaural contribution to spatial unmasking; the difference between the binaural and worse-ear predictions predicts how large the impact of listening with only one ear can be (i.e., if the acoustically better ear is nonfunctional).

The binaural predictions capture the main trends in the data, accounting for 99.05% of the variance in the measurements. The only binaural predictions that are not within the approximate 1-dB standard error in the measurements correspond to the same configurations for which the predicted percent-correct scores fail.

D. Difference between better- and worse-ear thresholds

The spatial unmasking analysis presented in Fig. 9 separately estimates binaural, monaural better-ear, and monaural worse-ear thresholds (in dB). From these values, we can predict the binaural advantage (i.e., the difference between the binaural and the better-ear threshold) and the difference between the better- and worse-ear thresholds (at least to the extend that the Zurek, 1993 model is accurate). These values are presented in Table I. The difference between the better- and worse-ear thresholds (second data column) is calculated as the absolute value of the difference (in dB) of the threshold T levels for left- and right-ear monaural predictions. This difference ranges from 5–18 dB for configurations in which T and M are not in the same location. Comparing these estimates (which weigh the TMR at each frequency according

to the AI calculation) to estimates made from the strict acoustic analysis (which weigh all frequencies up to 8000 Hz equally; first data column) shows (not unexpectedly) that the two methods yield very similar results. The predicted binaural advantage (third data column in Table I), defined as the difference between binaural and monaural better-ear model predictions for each configuration, is uniformly small, ranging from 0–2 dB.

E. Discussion

The Zurek model (1993) does a very good job of predicting the results for all spatial configurations similar to those that have been tested previously. In fact, the model fails only when T and/or M are near the head or when both T and M are located laterally.

Of the 15 independent spatial configurations tested, predicted performance is better than observed for six configurations, worse than observed for one configuration, and in agreement with the measurements in the remaining eight configurations. In six of the seven configurations for which the model prediction differs substantially from observed performance, T and/or M have ILDs that are larger than in previously tested configurations.

The Zurek model uses a simplified version of Colburn's model (1977a, b) of binaural unmasking to predict the binaural gain in each frequency channel, given the interaural differences in T and M. Colburn's original model accounts for the fact that binaural unmasking decreases with the magnitude of the ILD in M because the number of neurons contributing binaural information decreases with increasing ILD. The simplified version of the Colburn model used in Zurek's formulation does not take into account how the noise ILD affects binaural unmasking. If one were to use a more complex version of the Colburn binaural unmasking model, the predicted binaural gain would be smaller for spatial configurations in which there is a large ILD in the masker. Binaural predictions from such a corrected model would fall somewhere between the current binaural and better-ear predictions.

Unfortunately, such a correction will not improve the predictions. In particular, of the seven predictions that differ substantially from the measurements, there is only one case in which decreasing the binaural gain in the model prediction could substantially improve the model fit [T at (0°, 1 m) and M at (90°, 15 cm); see Fig. 9(a), circle at right side of panel]. In five of the remaining configurations in which the predictions fail [circle symbol at left of Fig. 9(b) and all four observations in Fig. 9(c)], even the better-ear model analysis predicts more spatial unmasking than is observed, and in the final configuration [e.g., circle symbol at right of Fig. 9(b)] both the binaural and better-ear analysis predict less unmasking than was observed. In fact, for this configuration, any decrement in the binaural contribution of the model will degrade rather than improve the binaural prediction fit.

The model assumes that binaural processing can only improve performance above what would be achieved if listening with the better ear alone. Current results suggest that this may not always be the case; we found that measured binaural performance is sometimes worse than the predicted performance using the better ear alone. We know of only one study that found a binaural *dis*-advantage for speech unmasking. Bronkhorst and Plomp (1988) manipulated the overall interaural level differences of the signals presented to the subjects in order to simulate monaural hearing loss. Subjects were tested with binaural, better-ear monaural, and worse-ear monaural stimuli as well as conditions in which the total signal to one of the ears was attenuated by 20 dB. In some cases, monaural performance using only the better-ear stimulus was near binaural performance; in these cases, attenuating the worse ear stimulus by 20 dB had a negligible impact on performance. If both ears had roughly the same TMR but the IPDs in T and M differed, binaural performance was best, performance for left- and right-ear monaural conditions was equal (and worse than binaural performance), and attenuating either ear's total stimulus caused a small (1–2 dB) degradation in SRT. Of most interest, in conditions for which there was a clear "better ear" (i.e., when the TMR was much larger in one ear than the other), performance with the better ear attenuated by 20 dB was worse than monaural performance for the better-ear stimulus, even though the better-ear stimulus was always audible. The researchers noted that this degradation in performance appears to be "due to a "disturbing" effect of the relatively loud noise presented in the other ear" (Bronkhorst and Plomp, 1988, p. 1514), because the better-ear stimulus played alone yielded better performance than the binaural stimulus. In the current experiment, some of the configurations for which the binaural predictions exceeded observed performance had a worse-ear signal that was substantially louder than the better-ear signal. However, when T was at (90°, 15 cm) and M was at (90°, 1 m), binaural performance was worse than predicted better-ear performance, even though the worse-ear signal was quieter than the better-ear signal. One possible explanation for these results is that large ILDs in the stimuli can sometimes degrade binaural performance below better-ear monaural performance, even if the worse-ear stimulus is quieter than the better-ear stimulus.

Finally, it should be pointed out that while the overall rms level of the stimuli was held constant at the better ear, the spectral content in T and M changed with spatial position as a result of the HRTF processing. It may be that some of the prediction errors arise from problems with the monaural, not binaural, processing in the model. Further experiments are needed to directly test whether binaural performance is worse than monaural better-ear performance in spatial configurations like those tested.

V. CONCLUSIONS

The results of these experiments demonstrate that the amount of spatial unmasking that can arise when T and/or M are within 1 m of a listener is dramatic. For a masker emitting a fixed-level noise, the level at which a speech target must be played to reach the same intelligibility varies over approximately 45 dB for the spatial configurations considered. Much of this effect is the result of simple changes in stimulus level with changes in source distance; however, other phenomena also influence these results.

It is well known that, on spatial unmasking tasks, monaural listeners are at a disadvantage compared to binaural listeners. In roughly half of the possible spatial configurations, the better-ear advantage is lost and any binaural processing gains are ineffective for these listeners (e.g., see Zurek, 1993). However, the current results suggest that when either T or M are close to the listener, monaural listeners can suffer from disadvantages (compared to normal-hearing listeners) that are as much as 13 dB greater observed for configurations in which T and M are at least 1 meter from the listener [i.e., from Table I, when T is at (0°, 1 m), the estimated left/right asymmetry is 19.6 dB for M at (90°, 15 cm) and only 6.4 for M at (90°, 1 m)]. Specifically, for the configurations tested, the worse-ear TMR can be nearly 20 dB lower than the better-ear TMR. While the current experiments did not measure performance of monaural listeners directly, this analysis supports the view that having two ears provides an enormous advantage to listeners in noisy environments, especially when the sources of interest are close to the listener. However, much of the benefit obtained from listening with two ears appears to derive from having two independent ''mixes'' of T and M, one of which often has a better TMR than the other. The specifically binaural processing advantages expected in the tested configurations are comparable to those observed in previous studies, on the order of 2 dB. Of course, even 2 dB of improvement in TMR can lead to vast improvements in speech intelligibility near SRT, leading to improvements in percent-correct word identification of over 20%.

The current experiments included a number of novel spatial configurations that have not previously been investigated. For many of these configurations, the Zurek model of spatial unmasking of speech fails to predict observed performance. The reasons underlying these failures (which all simulate either T or M very near the listener or have both T and M located at 90°) must be investigated further. One of the failed predictions may be partially corrected by considering a binaural unmasking model that takes into account the ILD in the masker [i.e., when M is at (90°, 15 cm) and T is at (0°, 1 m)]. However, such a correction will not improve the model predictions for any of the remaining configurations for which the model fails.

Analysis suggests that binaural processing of interaural phase decreases SRT by 1–2 dB for the configurations considered in the current study, similar to the gain observed for configurations in which T and M are both at least 1 meter from the listener (e.g., see Bronkhorst, 2000). However, for the configurations in which better-ear monaural predictions of SRT are lower than the SRTs observed with binaural presentations, there may actually be a disadvantage to listening with two ears (compared to listening with the better ear alone). Additional experiments using monaural control conditions must be performed in order to fully explore whether large ILDs degrade speech intelligibility or whether monaural better-ear performance is worse than predicted in these configurations.

[1]Intersubject differences were relatively modest in these experiments, with an average sample standard deviation across the four subjects of 1.7 dB. These subject differences are shown in Fig. 4, but are left off of Figs. 3 and 4 for clarity.

[2]The AI has since been extended and renamed as the Speech Intelligibility Index or SII; see ANSI, 1997.

[3]On average, the standard error in the SRTs across the four subjects is 0.85 dB. From this, we can estimate the corresponding standard error in percent-correct estimates as follows. Near the 50%-correct point, the AI function is roughly linear. Assuming all frequency bands have TMRs between -12 and 18 dB, the AI is linear with TMR. Thus, under these assumptions (i.e., near the 50%-correct point with all frequency bands contributing to the AI and not saturated), percent correct is a linear function of TMR with a slope of roughly 12%/dB. Multiplying the standard error in the estimate of SRT times this slope yields a rough approximation of the standard error in the percentage correct of 0.85*12%=10.2%. Note that if some frequency bands are inaudible or saturated, the estimated error in percent correct will actually be less than 10.2%.

ANSI (**1969**). ANSI 53.5-1969, "American National Standard Methods for the Calculation of the Articulation Index" (American National Standards Institute, New York).

ANSI (**1997**). ANSI 53.5-1997, "Methods for Calculation of the Speech Intelligibility Index" (American National Standards Institute, New York).

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

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

Bronkhorst, A. W., and Plomp, R. (**1990**). "A clinical test for the assessment of binaural speech perception in noise," Audiology **29**, 275–285.

Brown, T. J. (**2000**). "Characterization of Acoustic Head-Related Transfer Functions for Nearby Sources," M.Eng. thesis, Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

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. **64**, 525–533.

Colburn, H. S. (**1977b**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533. See Aip Document No. E-PAPS JASMA-6-525-98.

Dirks, D. D., and Wilson, R. H. (**1969**). "The effect of spatially separated sound sources on speech intelligibility," J. Speech Hear. Res. **12**, 5–38.

Duda, R. O., and Martens, W. L. (**1997**). "Range-dependence of the HRTF for a spherical head," IEEE ASSP Workshop on Applications of DSP to Audio and Acoustics.

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), Vol. 4, pp. 365–466.

Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.

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.

Hawley, M. L. (**2000**). "Speech Intelligibility, Localization and Binaural Hearing in Listeners with Normal and Impaired Hearing," Ph.D. dissertation, Biomedical Engineering, Boston University.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Hirsh, I. J. (**1950**). "The relation between localization and intelligibility," J. Acoust. Soc. Am. **22**, 196–200.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**(3), 225–246.

Kryter, K. D. (**1962**). "Methods for the calculations and use of the Articulation Index," J. Acoust. Soc. Am. **34**, 1689–1697.

MacKeith, N. W., and Coles, R. R. A. (**1971**). "Binaural advantages in hearing of speech," J. Laryngol. Otol. **85**, 213–232.

Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.

Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.

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

Rabinowitz, W. R., Maxwell, J. Shao, Y., and Wei, M. (**1993**). "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," Presence **2**(2), 125–129.

Shinn-Cunningham, B. G. (**2000**). "Distance cues for virtual auditory space," in Proceedings of the IEEE-PCM 2000, pp. 227–230, Sydney, Australia, 13–15 December.

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

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, Boston).