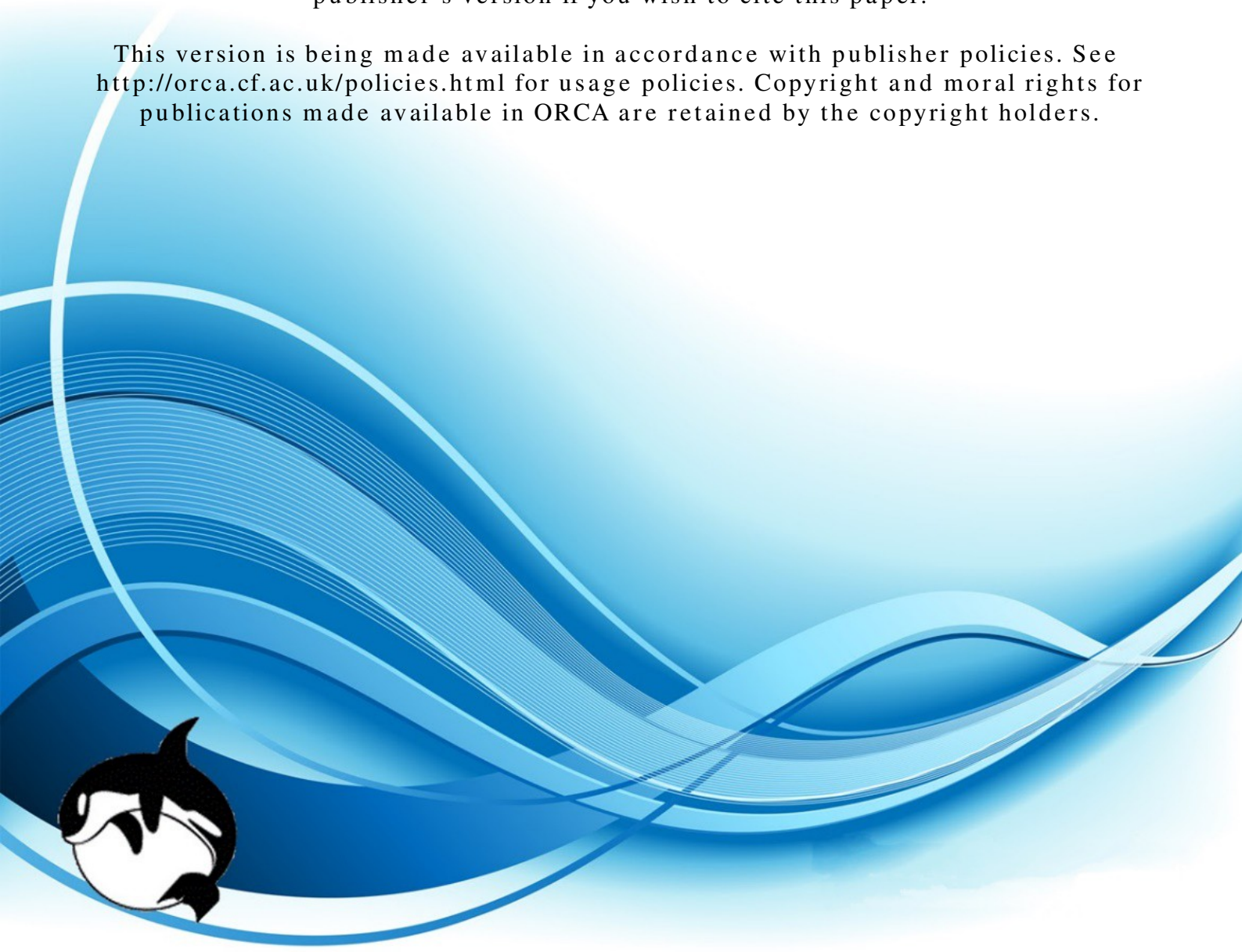

The spatial unmasking of speech: evidence for within-channel processing of interaural time delay

Barrie A. Edmonds and John F. Culling
School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, CF10 3AT, United Kingdom

(Received 1 September 2004; revised 3 February 2005; accepted 4 February 2005)

Across-frequency processing by common interaural time delay (ITD) in spatial unmasking was investigated by measuring speech reception thresholds (SRTs) for high- and low-frequency bands of target speech presented against concurrent speech or a noise masker. Experiment 1 indicated that presenting one of these target bands with an ITD of $+500$ μs and the other with zero ITD (like the masker) provided some release from masking, but full binaural advantage was only measured when both target bands were given an ITD of $+500$ μs. Experiment 2 showed that full binaural advantage could also be achieved when the high- and low-frequency bands were presented with ITDs of equal but opposite magnitude (± 500 μs). In experiment 3, the masker was also split into high- and low-frequency bands with ITDs of equal but opposite magnitude (± 500 μs). The ITD of the low-frequency target band matched that of the high-frequency masking band and vice versa. SRTs indicated that, as long as the target and masker differed in ITD within each frequency band, full binaural advantage could be achieved. These results suggest that the mechanism underlying spatial unmasking exploits differences in ITD independently within each frequency channel. [DOI: 10.1121/1.1880752]

PACS numbers: 43.66.Pn, 43.66.Dc [AK] Pages: 3069–3078

I. INTRODUCTION

The masked threshold of speech is lower when it is spatially separated from its masker than when the two sounds share a common direction. This effect is called the binaural intelligibility level difference (BILD). The BILD has been described as being dependent on improvements in the audibility of the target speech arising from differences in interaural level difference (ILD) and interaural time delay (ITD) between the two sounds (Bronkhorst and Plomp, 1988; Zurek, 1992). This paper focuses on the binaural gain in intelligibility associated with ITD (e.g., Schubert, 1956; Levitt and Rabiner, 1967a) and how ITD is exploited by the auditory system to bring about release from masking. Three experiments are reported in which we tested for the importance of providing a common ITD across different frequency regions to the BILD.

The effect of spatial separation on the segregation of sounds has also been described in terms of selective attention (e.g., Hirsh, 1950; Broadbent, 1954; Darwin and Hukin, 1999; Freyman *et al.*, 1999; Darwin and Hukin, 2000; Freyman *et al.*, 2001, 2004). That is, it is thought that focusing one's attention on the perceived location of the desired speech might aid the formation and perceptual segregation of the target as an auditory event from that of a masking sound. The relationship between lateralization and binaural detection of sounds has been an open question for many years (e.g., Hirsh, 1948; Licklider, 1948; Hafter *et al.*, 1969); a number of investigations have considered the relative importance of spatial location in the segregation of sounds compared to other cues (Bregman, 1990; Kubovy and Van Valkenburg, 2001; Neuhoff, 2003). Given that ITD contributes to both the perceived lateral position of a sound source (Rayleigh, 1876, 1907) and to binaural unmasking, it is tempting to suggest that the latter is dependent on the former. However, two lines of evidence suggest that this is not the case.

First, the perceived location of a sound can be disrupted without any significant effect on binaural release from masking (Licklider, 1948; Carhart *et al.*, 1967, 1968; 1969; Edmonds and Culling, in press). For example, the masked threshold of speech heard against a masker with zero ITD (and therefore perceived centrally) is lower for target speech presented out of phase at the two ears (perceived to be diffusely located) than for target speech that has a fixed ITD and is heard to be clearly lateralized. In addition, theories of speech intelligibility for spatially separated sounds (e.g., Levitt and Rabiner, 1967b; Zurek, 1992) predict improvements in the masked threshold of target speech as a function of binaural unmasking rather than perceived location.

Second, ITD has been demonstrated to be a relatively weak cue for the segregation of competing sounds. For instance, Hukin and Darwin (1995) showed that a single harmonic could be segregated from other harmonics in a vowel sound if its onset time was altered but not if it was given a different ITD. That is, despite the harmonic having a different ITD from the rest of the vowel sound, listeners group the lone harmonic with the other components of the vowel. In addition, listeners do not appear to exploit ITD when grouping sounds across frequency (Culling and Summerfield, 1995) unless they are given considerable amounts of training (Drennan *et al.*, 2003). Culling and Summerfield (1995) presented listeners with four formant-like noise bands (i.e., their frequencies approximated the first and second formants of speech) which could give rise to the perception of two whispered vowel sounds. They found that listeners were unable to correctly identify (with above-chance performance) the two vowels if presented with different ITDs, but could do so

when the two vowels were presented to different ears. Consequently, it has been argued that the auditory system ignores spatial correspondences between different frequency channels, preferring to exploit within-channel interaural differences between concurrent sounds (Culling and Summerfield, 1995; Akeroyd, 2004).

There are a number of models that describe how ITD might be exploited for binaural unmasking (for an overview see Colburn and Durlach, 1978; Blauert, 1983); however, the two most well known are vector theory (Jeffress, 1972) and the equalization-cancellation (E-C) model (Durlach, 1960; 1963; 1972; Breebaart *et al.*, 2001). The Jeffress model assumes that ITD is exploited by a binaural processor consisting of a series of frequency-dependent coincidence detectors connected by delay lines. The auditory system is thought to be able to compare the activity of this binaural processor over a range of interaural delays in order to perform a cross correlation of the input at the two ears. Durlach's model assumes that, if the target sound and its masker are spatially separated, then it should be possible to apply a set of transformations to the signal such that the noise can be eliminated. For instance, when the target has a different ITD from that of the masker, equalization can be achieved by applying an internal delay in order to compensate for the interaural configuration of the noise.[1] The noise can then be canceled from the binaural signal by subtracting the now-equalized target and masker waveforms from one another in order to deliver an improved signal-to-noise ratio. Consequently, the model accurately predicts that the optimal case for binaural unmasking in a given critical band (e.g., the detection of a tone in noise) is when the tone is presented out of phase at the two ears and the noise is presented in phase at the two ears.

Culling and Summerfield (1995) proposed an elaboration of Durlach's model, the modified equalization-cancellation (mE-C), in order to account for the apparent indifference of the auditory system to ITD across frequency for the grouping of sounds. They suggested that, as the grouping of sounds across frequency does not appear to be constrained by spatial correspondences between different frequency channels, then the equalization step of spatial unmasking must be free to use the best ITD within each frequency channel. Subsequently, the mE-C model has been used to explain the results of a number of binaural phenomena (Culling and Summerfield, 1995; Culling, 1998; Culling *et al.* 1998). More recently, Akeroyd (2004) looked for evidence of this within-channel mechanism in the binaural unmasking of complex tones against a broadband masker. Akeroyd found that, even when each component of a harmonic complex was presented with a different ITD, detection of the complex was undiminished. These results suggest that the decision mechanism responsible for choosing the best delay in the equalization process is free to do so independently within each frequency channel.

This paper investigates whether a channel-independent mechanism for exploiting ITD (such as that assumed in the mE-C model) can account for the binaural gains in the intelligibility of speech in noise associated with spatial separation. In particular, the importance of a common ITD to the BILD was tested by presenting listeners with target stimuli that had different ITDs at different frequencies. Three experiments were conducted to explore various strategies for selecting and canceling competing sounds (i.e., target speech heard against either competing speech or a broadband-noise masker) across frequency using ITD; the BILDs measured suggest that the auditory system is able to exploit ITD independently within each frequency channel.

II. GENERAL METHODS

A. Participants

Cardiff University psychology undergraduate students were recruited and awarded course credit in return for their participation. All participants reported normal hearing and spoke English as their first language. Each participant was a naive listener (i.e., they had little or no previous experience in tests of auditory perception) and contributed data to only one experiment in a single session lasting approximately 45 min.

B. Stimuli

Stimuli were presented to the listener using a TDT AP2 array processor via a TDT psychoacoustics rig (DD1, FT6, PA4, HB6) through Sennheiser HD 590 headphones in a single-walled IAC sound-attenuating booth. Sentences from the MIT recordings of the speaker CW reading the Harvard Sentence Lists (IEEE, 1969) were used as target items. The masker was either a sentence from the speaker DA (again from MIT recordings of the Harvard sentence lists) or Brown noise (i.e., a broadband noise with a 6-dB/octave spectral roll-off). Brown noise produces greater energetic masking for low frequencies than for higher frequencies, and roughly approximates the low-frequency emphasis of speech.

C. High- and low-pass filters

In order to test for the importance of a common ITD across frequency, stimuli were spectrally divided into high- and low-pass filtered frequency bands. This manipulation allowed the high- and low-frequency regions of the signal to be configured independently of each other (i.e., given different ITDs). By doing this, the effect of spatial separation on the intelligibility of speech in different frequency regions could be tested.

In experiments 1, 2, and 3, the stimuli were presented as a pair of high- and low-pass filtered frequency bands using 512-point FIR filters with linear phase and >1000 dB/octave cutoffs. The high- and low-frequency bands were separated by a 1-ERB (equivalent rectangular bandwidth) (Moore and Glasberg, 1983) gap centered at splitting frequencies of 750 and 1500 Hz in experiment 1 and 750, 1500, and 3000 Hz in experiments 2 and 3 (see Table I for a summary of the exact filter cutoffs). This gap prevented energy in frequency channels close to the splitting frequency from creating a confounding interaural interaction.

TABLE I. Summary of the upper and lower cutoff frequencies used to spectrally divide the stimuli about a given splitting frequency. The low-frequency band was created by low-pass filtering the stimuli at a cutoff frequency of $\frac{1}{2}$ of the equivalent rectangular bandwidth below the splitting frequency. The high-frequency band was created by high-pass filtering the stimuli at a cutoff frequency of $\frac{1}{2}$ of the equivalent rectangular bandwidth above the splitting frequency.

Splitting frequency (Hz)	Low-pass cutoff (Hz)	High-pass cutoff (Hz)
3000 (experiments 2 and 3)	2821	3186
1500 (all experiments)	1409	1592
750 (all experiments)	700	802

D. Procedure

Speech reception thresholds (SRTs) were measured for each participant in all conditions. The SRT is the masked level in dB of the target speech for a criterion level of understanding. In this case, it was measured for the report of keywords from the target sentence with an accuracy of 50%. The SRT measurement was implemented using the 1-up/1-down adaptive threshold method described by Plomp and Mimpen (1979). Participants were presented with ten trials for each experimental condition; in order to eliminate the effects of order of presentation and of variations in the difficulty of the target materials the conditions were rotated around the different speech materials for successive participants. That is, each participant heard all the target/masker speech materials in the same order; only the order of the conditions was changed. SRTs were also measured for two practice conditions consisting of only monaural stimuli so that listeners could familiarize themselves with the experimental procedure; thresholds for these practice stimuli are not reported.

For the first trial in each condition, the target speech was presented at a very low level (−28 dB) compared to that of the masking sound. A message presented via a computer terminal, viewed through the booth window, prompted the listener to either enter a transcript (using a computer keyboard located inside the booth) or to replay the stimulus. If the participant replayed the stimulus the level of the target speech was increased by 4 dB. The first trial could be replayed in this way until it was loud enough to be judged partially intelligible by the listener (i.e., they felt they could hear approximately half the sentence). At this point, the participant entered a transcript of the words that they thought they had heard. Next, the correct transcript for the current target sentence was displayed on the computer terminal just below the participant's response. This reference transcript contained five keywords (presented in upper case—nonkeywords were presented in lower case). The participant was then prompted to enter the number of keywords that he/she had correctly identified (scoring 0–5). The procedure then entered a second phase in which the stimulus was played only once before the participant was required to transcribe the target sentence.

In the second phase, a fresh target sentence was presented on each of the remaining trials (i.e., trials 2–10) and the level of the target speech for each of these trials was dependent on the listener's reported accuracy in the previous trial. If the participant reported transcribing two or fewer keywords correctly on one trial, the level of the target on the next trial was increased by 2 dB; otherwise, the level of the target was decreased by 2 dB. After all ten trials had been presented, the SRT was determined to be the mean presentation level used for the last seven trials (i.e., trials 3–10) and what would have been the 11th trial.

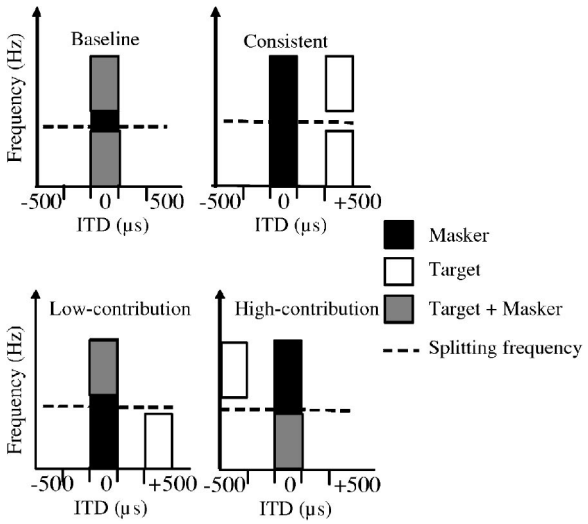

FIG. 1. A schematic illustration of the ITD configurations of experiment 1: Stimuli are represented as high- and low-pass filtered frequency bands presented at distinct ITDs. Target speech bands are depicted in white with black outline, masker bands are depicted in black, and regions that have both target and masker sharing a common ITD are shown in gray. The splitting frequency used to divide the high- and low-pass bands (750 or 1500 Hz) is shown as a dashed line.

III. EXPERIMENT 1

Experiment 1 was a preliminary experiment to establish the importance of both high and low frequencies to speech intelligibility in our experimental paradigm. Its purpose was to ascertain the binaural gain in intelligibility for different frequency regions of target speech. In order to do this we employed a method similar to that of Levitt and Rabiner (1967a). Levitt and Rabiner tested for the importance of different frequency regions of single words heard against a broadband Gaussian noise in binaural release from masking using interaural phase opposition. Here, we measured the binaural advantage due to ITD for high- and low-frequency regions of sentences heard against either a Brown-noise or competing-speech masker.

A. Design

SRTs for target speech presented against a concurrent masker with zero ITD were measured in eight conditions: 2 splitting frequencies (750 and 1500 Hz)×4 ITD configurations (see Fig. 1): *baseline* (both high and low frequencies at zero ITD); *consistent* (both high and low frequencies with +500-μs ITD); *high-contribution* (high frequencies were presented with 500-μs ITD while low frequencies were presented with no ITD); and *low-contribution* (low frequencies were presented with 500-μs ITD while high frequencies were presented with no ITD). Experiment 1 was completed

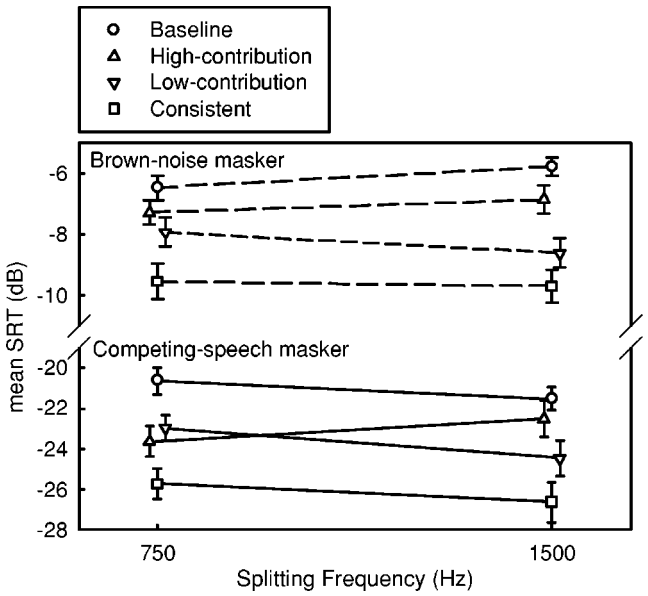

FIG. 2. Mean SRTs of the baseline (circles), high-contribution (upward triangles), low-contribution (downward triangles), and consistent (squares) ITD configurations of experiment 1 for two groups of listeners (Brown-noise masker, dashed lines; competing-speech masker, solid lines). Error bars show standard error. Plots for the high-contribution and low-contribution condition SRTs are offset along the x axis in order to improve visibility of the error bars.

by two groups of participants. SRTs were measured for target speech presented against a Brown-noise masker in experiment 1a (16 participants) and against competing speech in experiment 1b (24 participants).

B. Results and discussion

Figure 2 shows the pattern of SRTs for each condition against Brown-noise (dashed lines) and competing-speech (solid lines) maskers. The baseline condition has the highest SRTs in both groups and the consistent condition the lowest; the high-contribution and low-contribution condition SRTs were intermediate. This result suggests that both the high- and low-frequency regions of the target speech were required in order to achieve full binaural advantage (as measured in the consistent condition). Although the pattern of thresholds measured against both types of masker were very similar, the SRTs measured against the competing-speech masker were approximately 12 dB lower (i.e., speech intelligibility was better against competing speech than against the Brown noise).

A two-way repeated-measures analysis of variance (ANOVA) was performed on the SRTs of experiment 1a, and no effect of splitting frequency or interaction between ITD configuration (baseline, high-contribution, low-contribution, and consistent) and splitting frequency (750 and 1500 Hz) was found. However, there was a significant main effect of ITD configuration [$F(3,15)=34.90$, $p<0.001$]. Tukey pairwise tests showed that the comparison of baseline vs high contribution was not significantly different. However, significant differences were found for other comparisons: baseline vs consistent ($q=13.65$, $p<0.001$), baseline vs low contribution ($q=8.37$, $p<0.001$), high contribution vs consistent ($q=9.95$, $p<0.001$), high contribution vs low contribution ($q=4.67$, $p<0.05$), and low contribution vs consistent ($q=5.28$, $p<0.05$).

For experiment 1b, a two-way repeated-measures ANOVA revealed that there was no main effect of splitting frequency, nor was there a significant interaction with ITD configuration, but there was a significant main effect of ITD configuration [$F(3,23)=17.71$, $p<0.001$]. Tukey HSD tests for the pairwise comparisons of the ITD configurations showed that the comparison of high contribution vs low contribution was not significantly different. However, significant differences were found for all other comparisons: baseline vs consistent ($q=10.21$, $p<0.001$), baseline vs low contribution ($q=5.27$, $p<0.05$), baseline vs high contribution ($q=4.00$, $p<0.05$), high contribution vs consistent ($q=6.21$, $p<0.001$), and low contribution vs consistent ($q=4.94$, $p<0.05$).

A number of researchers have explored the importance of different frequency regions on the intelligibility of speech (e.g., Schubert and Schultz, 1962; Levitt and Rabiner, 1967a) and have typically found that binaural unmasking for detection is largely dependent upon interaural phase differences in the low-frequency (e.g., <1000 Hz) region. Experiment 1 tested for the importance of high- and low-frequency bands of target speech to the BILD at two splitting frequencies, and found that neither band alone (i.e., when presented with a different ITD to that of the masker) was sufficient to produce full binaural advantage. SRTs measured in the consistent ITD configuration were lower than those measured for the high-contribution and low-contribution conditions. However, the low-contribution configuration tended to produce lower thresholds than the high-contribution configuration, especially when combined with a splitting frequency of 1500 Hz.

As noted above, thresholds measured against the competing-speech masker were substantially lower than those measured against the Brown-noise masker. Indeed, these thresholds are much lower than those reported in previous studies that have investigated the effects of spatial separation on speech intelligibility which reported SRTs in the region of −20 dB for stimuli with similar spatial configurations (e.g., Hawley *et al.*, 2004). However, it should be noted that in the current study the competing voice was that of a second male talker and not, as in many other studies, the same talker as the target voice. This is likely to have provided the listener with any number of other cues, arising from differences between the two voices, upon which segregation could be based.

IV. EXPERIMENT 2

Experiment 2 was designed to investigate the importance of common ITD for binaural unmasking. Specifically, we investigated the effect of across-frequency consistency in ITD on the intelligibility of target speech. In order to do this we presented listeners with stimuli that had been manipulated so that different frequency regions of the target speech had either the same or opposing ITDs. If the auditory system is able to exploit ITD independently within each frequency channel, then presenting high- and low-frequency bands of the target speech with different ITDs should have no effect

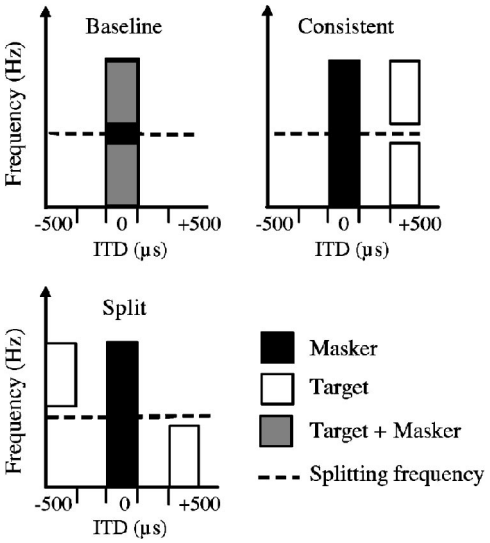

FIG. 3. A schematic illustration of the ITD configurations of experiment 2. Stimuli are represented in high- and low-pass bands presented at distinct ITDs. Target speech bands are depicted in white with black outline, masker bands are depicted in black, and regions that have both target and masker sharing a common ITD are shown in gray. The splitting frequency used to divide the high- and low-pass bands (750, 1500, or 3000 Hz) is shown as a dotted line.

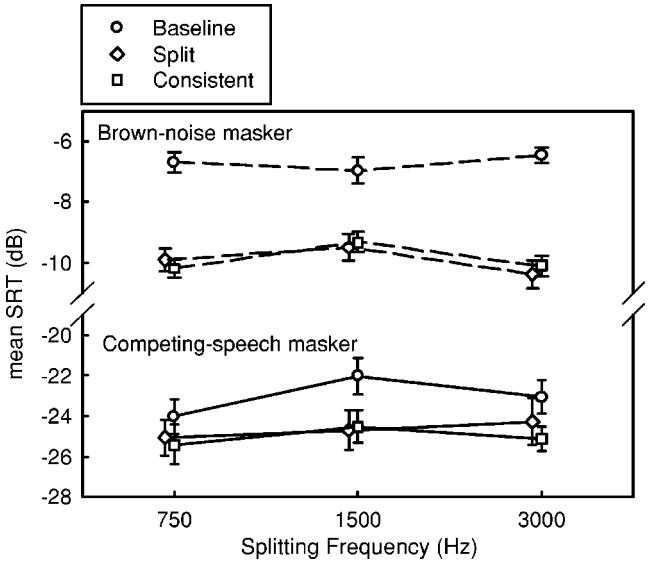

FIG. 4. Mean SRTs of the baseline (circles), split (diamonds), and consistent (squares) ITD configurations of experiment 2 for two groups of listeners (Brown-noise masker, dashed lines; competing-speech masker, solid lines). Error bars show standard error. Plots for the split condition SRTs are offset along the x axis in order to improve visibility of the error bars.

on speech intelligibility. Alternatively, if the BILD is dependent on a strategy involving the selection of information at a common ITD across frequency, then one might predict that speech intelligibility in such a condition would be disrupted, as listeners would be constrained to selecting only one of the two possible target speech bands.

A. Design

SRTs were measured for target speech split into a pair of high- and low-pass filtered frequency bands against a concurrent masker over nine conditions: 3 splitting frequencies (3000, 1500, and 750 Hz)×3 ITD configurations (see Fig. 3). The baseline and consistent conditions from experiment 1 were reused and joined by a third condition: *split* (high frequencies were presented with a +500-μs ITD and low-frequencies were presented with a −500-μs ITD). Experiment 2 was completed by two new groups of participants. SRTs were measured for target speech presented against a Brown-noise masker in experiment 2a (18 participants) and against competing speech in experiment 2b (18 participants).

B. Results and discussion

Figure 4 shows that SRTs were poorest (highest) in the baseline condition, but improved in the consistent and split conditions giving a BILD of approximately 3–4 dB in experiments 2a and 2b. Again, the SRTs measured against the competing-speech masker (solid lines) were approximately 12 dB lower than those obtained against the Brown-noise masker (dashed lines), but the pattern of results for both groups was similar.

A two-way repeated-measures ANOVA was performed on the SRTs, with two within-subject factors (ITD configuration, three levels; splitting frequency, three levels). For experiment 2a, there was no main effect of splitting frequency and no statistically significant interaction between ITD configuration and splitting frequency, but there was a significant main effect of ITD configuration [$F(2,17)=109.91$, $p<0.001$]. Tukey HSD pairwise tests showed that the comparison of consistent vs split was not significantly different. However, significant differences were found for the baseline vs split ($q=18.31$, $p<0.001$) and baseline vs consistent ($q=18.00$, $p<0.001$) comparisons.

For experiment 2b, a two-way ANOVA with repeated measures found no main effect of splitting frequency and no statistically significant interaction with ITD, but there was a main effect of ITD configuration [$F(2,17)=5.23$, $p<0.05$]. Tukey HSD pairwise tests showed that the comparison of consistent vs split was not significantly different. However, significant differences were found for comparisons between baseline vs split ($q=3.53$, $p<0.05$) and baseline vs consistent ($q=4.29$, $p<0.05$).

The results of experiment 2 indicate that the intelligibility of masked speech does not require the target speech to be presented with an ITD consistent with a particular direction across different frequency regions in order for full binaural advantage to be achieved. ITD can be exploited to recover target speech at high and low frequencies even when the ITDs of these frequency bands indicate sources in different hemifields. Consequently, it is argued that listeners do not group information across frequency at a common ITD. Rather, the contribution of the target speech bands presented with opposing ITDs to the BILD suggests that listeners were able to exploit ITD within each frequency band independently. However, there are two alternative explanations that might also account for the BILDs observed in this experiment.

First, one might argue that the SRTs measured in the split condition reflect the contribution of both high and low frequencies, but not their simultaneous contributions. One

could imagine, for example, an attention-switching mechanism which allows the auditory system to select information from different locations over time. Second, one might suggest that, rather than selecting sounds with a fixed ITD across frequency, the auditory system simply cancels interfering sounds at a fixed ITD. Consequently, presenting the high- and low-frequency regions of the target speech with opposing ITDs would have little effect on the unmasking process. These issues were addressed in experiment 3.

V. EXPERIMENT 3

The results of experiment 1 demonstrated that recovery of both the high- and low-frequency target bands is required in order to obtain full binaural advantage. Furthermore, experiment 2 showed that listeners could exploit differences in ITD between target speech and a concurrent masker even when different frequency bands of the target speech were presented with different ITDs. It was suggested that this indicated that the auditory system is able to exploit differences in ITD between the target and the masker within each frequency channel independently. However, while the results of experiment 2 suggest that the auditory system is not constrained to select information at a particular ITD, the result was inconclusive in other respects. First, it was difficult to determine whether different frequency regions of a target sound presented with different ITDs contribute to binaural unmasking simultaneously or whether their contributions are pooled together over time. Second, experiment 2 did not consider what role the ITD of the masking sound might have had in the unmasking process. Consequently, experiment 3 was designed to test whether a common ITD could be used to drive either: (i) an attention-switching mechanism for selecting target speech presented with different ITDs at different frequencies, or (ii) a mechanism that cancels at a fixed internal delay rather than selecting the target speech.

Speech intelligibility was measured for a *swapped* ITD configuration (i.e., the ITD of the target at low frequencies matched that of the masker at high frequencies and vice versa). When the target and masker have their ITDs in the high-frequency and low-frequency regions swapped, it should not be possible to integrate information across frequency at a common ITD without recovering a mixture of target and masker. No amount of attention switching in this condition will remove the presence of the masker. Furthermore, it should be impossible to selectively cancel out the masker across frequency in the swapped condition, as any target speech with the same ITD as the masker will also be canceled. Consequently, if the auditory system is restricted to the exploitation of a common ITD across frequency, then speech intelligibility should suffer in the swapped ITD configuration (i.e., SRTs for the swapped ITD configuration should be markedly higher than those measured for the consistent ITD configuration). However, if the SRTs measured under consistent and swapped conditions are indistinguishable, then a strategy for exploiting within-channel differences in ITD independent of frequency will be supported.

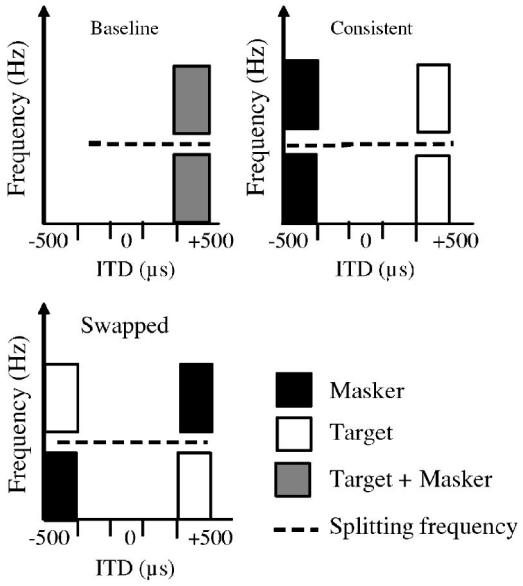

FIG. 5. A schematic illustration of the ITD configurations of experiment 3. Stimuli are represented in high- and low-pass bands presented at distinct ITDs. Target speech bands are depicted in white with black outline, masker bands are depicted in black, and regions that have both target and masker sharing a common ITD are shown in gray. The splitting frequency used to divide the high- and low-pass bands (750, 1500, and 3000 Hz) is shown as a dotted line.

A. Design

In experiment 3, both the target speech and the masker were presented as a pair of high- and low-pass bands separated by splitting frequencies of 750, 1500, or 3000 Hz. SRTs were measured for three configurations (see Fig. 5) of target and masker ITDs: baseline (both target and masker were presented with a $+500$-μs ITD), consistent (the target speech was presented with a $+500$-μs ITD while the masker was presented with a -500-μs ITD), and swapped (the high-frequency target speech band and the low-frequency masker band were presented with a -500-μs ITD while the low-frequency target speech band and the high-frequency masker band were presented with a $+500$-μs ITD). Two new groups of nine listeners took part in this study. SRTs were measured for target speech presented against a Brown-noise masker in experiment 3a and against competing speech in experiment 3b.

B. Results and discussion

Figure 6 shows the mean SRTs for the two groups of listeners in experiment 3. Intelligibility was poorest for the baseline condition, but improved in the consistent and swapped conditions, giving a BILDs of approximately 4 dB for the Brown-noise masker (dashed lines) and competing-speech masker (solid lines) groups. Again, thresholds were lower and more variable (i.e., larger error bars) against competing speech than against Brown noise

A two-way repeated-measures ANOVA was performed on the SRTs of experiment 3a and showed a significant main effect of ITD [$F(2,8)=60.57$, $p<0.001$] and of splitting frequency [$F(2,8)=6.35$, $p<0.05$]. Tukey pairwise tests showed that the following comparisons were not signifi-

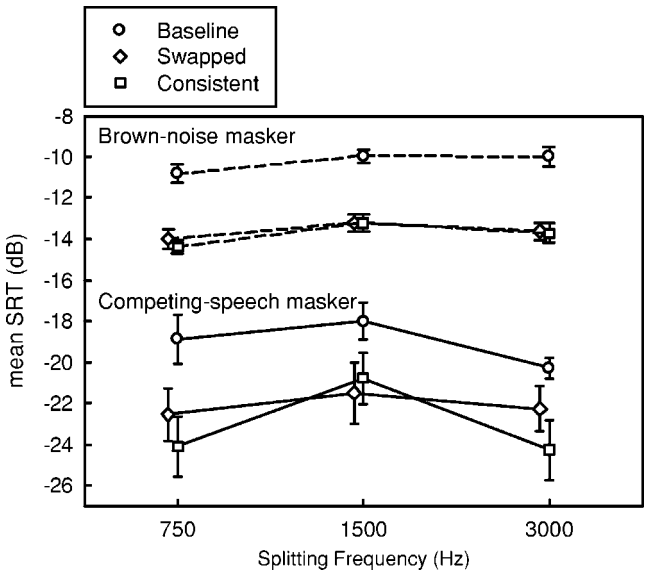

FIG. 6. Mean SRTs of the baseline (circles), swapped (diamonds), and consistent (squares) ITD configurations of experiment 3 for two groups of listeners (Brown-noise masker, dashed lines; competing-speech masker, solid lines). Error bars show standard error. Plots for the swapped condition SRTs are offset along the x axis in order to improve visibility of the error bars.

cantly different: swapped vs consistent, 1500 vs 3000 Hz, and 3000 vs 750 Hz. However, significant differences were found for all other comparisons: baseline vs consistent ($q=13.80$, $p<0.001$), baseline vs swapped ($q=13.14$, $p<0.001$), and 750 vs 1500 Hz ($q=4.95$, $p<0.05$).

Statistical analyses (two-way repeated measures ANOVA) of experiment 3b indicated that there was no effect of splitting frequency. However, ITD configuration yielded a significant effect [$F(2,8)=7.35$, $p<0.05$]. Tukey HSD comparisons showed that the SRTs of the swapped and consistent conditions were not significantly different, but differences were found for baseline vs consistent ($q=5.19$, $p<0.05$) and baseline vs swapped ($q=3.96$, $p<0.05$).

Experiment 3 was designed to test whether listeners simply make use of the best ITD within each frequency channel to segregate a target sentence from its masker or whether they use some strategy that is dependent on the lateralization of sounds (i.e., requiring a common ITD across all frequency channels). The swapped condition was crucial to this test as participants were presented with the target and masker at each ITD. The viability of two strategies for exploiting a common ITD for the segregation of concurrent sounds was evaluated and found lacking. Neither attention switching nor cancellation by common ITD provides a suitable explanation of the data. If participants had employed either of these strategies then the SRTs measured for the swapped condition would have been much higher than those measured in the consistent condition. However, SRTs were found to be equivalent in consistent and swapped conditions, suggesting that listeners make use of differences in ITD between target and masker within each frequency channel independently rather than by selectively grouping or canceling information at one ITD across all frequency channels.

VI. GENERAL DISCUSSION

In this paper we explored the binaural gain in speech intelligibility arising from differences in ITD between target speech and a single concurrent masker. Three experiments were conducted to test whether the segregation of spatially separated sounds is dependent on the consistency of ITD across different frequency bands; in particular, whether or not the binaural gain in speech intelligibility was constrained to the exploitation of a single ITD across frequency. Participants were presented with high- and low-frequency regions of target speech and a masker of either Brown noise or competing speech under a number of binaural configurations. It was found that as long as the target and masker had a different ITD in each frequency channel, the size of the BILD was unaffected.

A. Within-channel processing of ITD

The primary aim of this investigation was to determine how ITD is exploited by the binaural system in order to segregate target speech from a concurrent masker. This issue was addressed in experiments 2 and 3. These experiments were designed to test which of a number of strategies for segregating spatially separated sounds best described the SRTs measured for high- and low-frequency regions of target speech presented in a number of binaural configurations. In particular, we were interested in determining (i) whether the segregation of target speech from a concurrent but spatially separated masker was dependent on the exploitation of a common ITD for selecting or canceling sound elements across frequency, or (ii) whether the auditory system was free to choose the best ITD within each frequency channel in order to improve the audibility of the target.

In experiment 2, the target speech was split into high- and low-frequency regions each with a different ITD. It established that binaural advantage could be achieved even when the high- and low-frequency regions of the target speech were given ITDs of equal but opposite magnitude. This suggests that the auditory system is not constrained to select information at a particular ITD across frequency, as doing so would have resulted in a BILD based on the contribution of only the high frequencies or only the low frequencies. We suggested that the most likely interpretation was that listeners were able to exploit the difference in ITD between the target and masker for both the high frequencies and the low frequencies simultaneously. However, at least two other alternatives exist.

First, it is possible for the BILDs of experiment 2 to be explained by the exploitation of a common ITD in order to cancel the masker rather than select the target. The recovery of target speech from a concurrent masker is often implemented in computational models of spatial unmasking by subtracting the masking sound from the compound waveform (e.g., Durlach's E-C model and beamforming techniques for automatic speech recognition). A similar procedure has been proposed to describe the existence of the pitch percept(s) that listeners experience when presented with dichotically delayed noises (Bilsen and Goldstein, 1974).

Second, this experiment did not rule out the possibility that listeners might be able to switch the focus of their atten-

tion from one moment to the next (i.e., in order to piece together the contributions of the high- and low-frequency bands of target speech over time). Peissig and Kollmeier (1997) discussed the possibility of an attention-switching strategy as a mechanism for improving speech intelligibility against multiple masking sounds. However, rather than suggesting that this mechanism selects target speech, they suggested that the binaural system employs this strategy for canceling multiple maskers. Because the waveform of speech is modulated, when multiple voices are presented concurrently there will be, at any time, instantaneous differences between these envelopes that produce differences in the signal-to-noise ratio. They suggested that the auditory system is able to exploit these spectro-temporal gaps in order to cancel the most intense competing voice at a given point in time. By doing so, this process is able to produce gains in the intelligibility of the target speech presented in a stimulus containing multiple speech sources arriving from different directions. However, Hawley *et al.* (2004) recently cast doubt upon the effectiveness of this attention-switching strategy by investigating the effects of speech-spectrum-shaped noises modulated by the temporal envelope of the target on the BILD. Such maskers provided listeners with the same opportunities for exploiting spectro-temporal gaps as a competing-speech masker. If attention switching is a viable strategy for canceling the masker(s) in such a situation, then one might expect the intelligibility of target speech heard against each type of masker to be comparable, but this was not the case. SRTs indicated that listeners received greater benefit from spatial separation when either speech or reversed-speech maskers were used rather than speech-shaped or speech-modulated noise maskers.

Experiment 3 was designed to address the three questions left open in experiment 2. First, whether a common ITD is used to cancel the masker across frequency. Second, whether listeners can exploit different ITDs at different moments in time (i.e., attention switching). Third, whether the auditory system is free to exploit the best ITD within each frequency channel. In order to test for the importance of these strategies to the BILD, we devised a condition (i.e., swapped ITD) in which support for either of the first two strategies would result in a detriment in speech intelligibility, while if the BILD was unaffected by such a binaural configuration this would provide support for the third proposition (i.e., a within-channel mechanism). As the SRT for this swapped-ITD condition was indistinguishable from that of the consistent condition, we suggest that the auditory system is free to choose the best ITD within each frequency channel in order to maximize the audibility of target speech against a concurrent masker. Consequently, this result appears to support Culling and Summerfield's (1995) mE-C model. At the same time, this experiment also supports the dissociation between perceived location and the effects of spatial separation on speech intelligibility (e.g., Licklider, 1948; Carhart *et al.*, 1967, 1968; 1969). Previously, the relationship between perceived location and spatial unmasking was confounded by the fact that, while one of the sounds was diffusely located, the other was clearly localized. That being the case, one might argue that full binaural advantage can be achieved under such conditions by either selecting a clearly localized target or by canceling a clearly localized masker (i.e., the perceived location of the other sound is largely irrelevant). Experiment 3, on the other hand, provided a control for the dissociation of perceived location and spatial unmasking. By ensuring that different portions of target speech and masker were presented with the same ITD, it was not possible to extract information residing at one ITD (i.e., at one spatial location) across frequency in order to either select the target or cancel the masking sound.

B. Informational masking

A number of studies have attempted to distinguish the effects of different types of sounds as maskers. In particular, a distinction has been made between energetic maskers and informational maskers (Pollack, 1975; Watson *et al.*, 1976) depending on which stage in the segregation process the interference takes place (Kidd *et al.*, 1994). Interference at peripheral stages of processing is described as energetic masking (i.e., the target and masker both contain energy at the same critical bands). On the other hand, informational maskers cause interference at some higher level of processing (i.e., uncertainty at the decision stage prevents the target and masker from being perceptually segregated). Consequently, it has been suggested that informational masking can produce an excess of masking (i.e., in addition to any energetic masking caused by the interfering sound). Furthermore, it has been suggested that the spatial separation or apparent spatial separation of two sounds can provide a release from informational masking (Freyman *et al.*, 1999; Brungart, 2001; Brungart *et al.*, 2001; Freyman *et al.*, 2001, 2004).

It is possible to consider both competing speech and Brown noise as energetic maskers. Competing speech can also be considered to be an informational masker, as it might produce interference at a number of levels other than at the peripheral level (e.g., semantically, syntactically, or similarity of pitch). Given that all three of the experiments reported in this paper were conducted against both a Brown-noise masker and competing speech, one might expect to see some evidence for informational masking or release from informational masking in the SRTs that we measured. In particular, one might expect some additional improvements in speech intelligibility against the competing-speech masker due to spatial separation that are not evident in the thresholds measured for target speech presented against Brown noise. However, while these experiments certainly demonstrate a difference in the amount of masking produced by Brown noise and competing speech, it is difficult to describe this effect in terms of informational masking for two reasons.

First, the SRTs measured against competing speech were consistently lower (in the region of 12 dB) than those measured for target speech heard against the Brown-noise masker. Furthermore, the difference between competing-speech and Brown-noise interference was probably underestimated here because Brown noise has much of its energy at very low frequencies which might have limited the degree to which it masked the target speech. This effect likely reflects the difference in energetic masking afforded by each of the

maskers. Brown noise is a purely energetic masker, while the competing-speech materials contained natural pauses and spectro-temporal gaps which might have reduced the amount of energetic masking produced. Whether or not this effect also reflects any informational masking is difficult to determine. What remains clear, however, is that the competing-speech maskers were less effective than a purely energetic Brown noise.

Second, there was no masker-dependent additional release from masking due to the perceived spatial separation of the target speech from the masking sound. While there was greater variance in SRTs measured against the competing-speech masker than against the Brown-noise masker, the BILDs for the corresponding conditions do not provide any direct evidence for informational masking. The difference between consistent and baseline condition SRTs was roughly the same for both speech and noise maskers. However, it is possible that the effects of informational masking on speech intelligibility in these experiments were confounded by other factors that also contribute to the SRT (e.g., pitch differences between the two voices) and no doubt warrant further investigation in order to control for these effects. Nonetheless, it is difficult to conclude that there is any evidence of informational masking or release from informational masking due to spatial separation from these data.

C. Conclusion

While the exploitation of a common ITD might be necessary for sound localization/lateralization (Stern *et al.*, 1988; Shackleton *et al.*, 1992), the results of the experiments described in this paper suggest that this is not the case for binaural unmasking. Here, we have demonstrated that the masked threshold of speech cannot be explained by selecting or canceling information at a common ITD across frequency. Rather, the process responsible for binaural unmasking appears to exploit ITD independently within each frequency channel. Consequently, this result supports previous accounts of the BILD that suggest binaural unmasking is indifferent to the perceived direction of sounds (Carhart *et al.*, 1968; Edmonds and Culling, in press).

ACKNOWLEDGMENT

Work supported by the UK EPSRC (Grants GR/M96155 and GR/S11794).

[1]If the target and masking sounds have different ILDs equalization can also be achieved by applying an internal level adjustment.

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**, 1135–1148.

Bilsen, F. A., and Goldstein, J. L. (**1974**). "Pitch of dichotically delayed noise and its possible spectral basis," J. Acoust. Soc. Am. **55**, 292–296.

Blauert, J. (**1983**). *Spatial Hearing—The Psychophysics of Human Sound Source Localization* (MIT Press, Cambridge).

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Bregman, A. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge).

Broadbent, D. E. (**1954**). "The role of auditory localization in attention and memory span," J. Exp. Psychol. **47**, 191–196.

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. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

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

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Release from multiple maskers: Effects of interaural time disparities," J. Acoust. Soc. Am. **45**, 411–418.

Carhart, R., Tillman, T. W., and Johnson, K. R. (**1967**). "Release of masking for speech through interaural time delay," J. Acoust. Soc. Am. **42**, 124–138.

Carhart, R., Tillman, T. W., and Johnson, K. R. (**1968**). "Effects of interaural time delays on masking by two competing signals," J. Acoust. Soc. Am. **43**, 1223–1230.

Colburn, H. S., and Durlach, N. I. (**1978**). "Models of binaural interaction," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), pp. 467–518.

Culling, J. F. (**1998**). "Dichotic pitches as illusions of binaural unmasking. II. The Fourcin pitch and the dichotic repetition pitch," J. Acoust. Soc. Am. **103**, 3527–3539.

Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1998**). "Dichotic pitches as illusions of binaural unmasking: I. Huggins' pitch and the 'binaural edge pitch,"' J. Acoust. Soc. Am. **103**, 3509–3526.

Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.

Darwin, C. J., and Hukin, R. W. (**2000**). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," J. Acoust. Soc. Am. **107**, 970–977.

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

Durlach, N. I. (**1960**). "Note on the equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am. **32**, 1075–1076.

Durlach, N. I. (**1963**). "Equalization and cancellation model of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), pp. 369–462.

Edmonds, B. A., and Culling, J. F. (**in press**). "The role of head-related time and level cues in the unmasking of speech in noise and competing speech," Acta Acust. Acust.: special issue on spatial and binaural hearing.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

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.

Hafter, E. R., Bourbon, W. T., Blocker, A. S., and Tucker, A. (**1969**). "A direct comparison between lateralization and detection under conditions of antiphasic masking," J. Acoust. Soc. Am. **46**, 1452–1457.

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

Hirsh, I. J. (**1948**). "The influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.

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

Hukin, R. W., and Darwin, C. J. (**1995**). "Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel," J. Acoust. Soc. Am. **98**, 1380–1387.

IEEE (**1969**). ''IEEE recommended practice for speech quality measurements,'' IEEE Trans. Audio Electroacoust. **17**, 225–246.
Jeffress, L. A. (**1972**). ''Binaural signal detection: Vector theory,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York).
Kidd, G. J., Mason, C. R., Deliwala, P. S., and Woods, W. S. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**, 3475–3480.
Kubovy, M., and Van Valkenburg, D. (**2001**). ''Auditory and visual objects,'' Cognition **80**, 97–126.
Levitt, H., and Rabiner, L. R. (**1967a**). ''Binaural release from masking for speech and gain in intelligibility,'' J. Acoust. Soc. Am. **42**, 601–608.
Levitt, H., and Rabiner, L. R. (**1967b**). ''Predicting binaural gain in intelligibility and release from masking for speech,'' J. Acoust. Soc. Am. **42**, 820–829.
Licklider, J. C. R. (**1948**). ''The influence of interaural phase relations upon the masking of speech by white noise,'' J. Acoust. Soc. Am. **20**, 150–159.
Moore, B. C., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitatory patterns,'' J. Acoust. Soc. Am. **74**, 750–753.
Neuhoff, J. G. (**2003**). ''Pitch variation is unnecessary (and sometimes insufficient) for the formation of auditory objects,'' Cognition **87**, 219–224.
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. (**1979**). ''Improving the reliability of testing the speech-reception threshold for sentences,'' Audiology **18**, 43–52.
Pollack, I. (**1975**). ''Auditory informational masking,'' J. Acoust. Soc. Am. **57**, S5.
Rayleigh, L. (**1876**). ''On perception of the direction of a source of sound,'' Nature (London) **14**, 32–33.
Rayleigh, L. (**1907**). ''On our perception of sound direction,'' Philos. Mag. **8**, 214–232.
Schubert, E. D. (**1956**). ''Some preliminary experiments on binaural time delay and intelligibility,'' J. Acoust. Soc. Am. **28**, 895–901.
Schubert, E. D., and Schultz, M. C. (**1962**). ''Some aspects of binaural signal selection,'' J. Acoust. Soc. Am. **34**, 844–849.
Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1992**). ''Across-frequency integration in a model of lateralization,'' J. Acoust. Soc. Am. **91**, 2276–2279.
Stern, M. R., Zeiberg, A. S., and Trahiotis, C. (**1988**). ''Lateralization of complex binaural stimuli: A weighted image model,'' J. Acoust. Soc. Am. **84**, 156–165.
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1186.
Zurek, P. M. (**1992**). ''Binaural advantages and directional effects in speech intelligibility,'' in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.