

Extracting binaural information from simultaneous targets and distractors: Effects of amplitude modulation and asynchronous envelopes

Mark A. Stellmack, Andrew J. Byrne, and Neal F. Viemeister

Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 24 July 2009; revised 20 June 2010; accepted 24 June 2010)

When different components of a stimulus carry different binaural information, processing of binaural information in a target component is often affected. The present experiments examine whether such interference is affected by amplitude modulation and the relative phase of modulation of the target and distractors. In all experiments, listeners attempted to discriminate interaural time differences of a target stimulus in the presence of distractor stimuli with ITD=0. In Experiment 1, modulation of the distractors but not the target reduced interference between components. In Experiment 2, synthesized musical notes exhibited little binaural interference when there were slight asynchronies between different streams of notes (31 or 62 ms). The remaining experiments suggested that the reduction in binaural interference in the previous experiments was due neither to the complex spectra of the synthesized notes nor to greater detectability of the target in the presence of modulated distractors. These data suggest that this interference is reduced when components are modulated in ways that result in the target appearing briefly in isolation, not because of segregation cues. These data also suggest that modulation and asynchronies between modulators that might be encountered in real-world listening situations are adequate to reduce binaural interference to inconsequential levels. © 2010 Acoustical Society of America. [DOI: 10.1121/1.3466868]

PACS number(s): 43.66.Pn, 43.66.Mk [MAA]

Pages: 1235–1244

I. INTRODUCTION

When attempting to identify the source location of a particular sound, a listener typically does so in the presence of other sounds emanating from different locations. The listener's task then amounts to associating the correct source location with the spectral components of interest. It has been shown that when numerous spectral components with different interaural time differences (ITDs) are presented simultaneously over headphones, listeners have a great deal of difficulty in discriminating the ITD of a target subset of the components. This effect has been shown to occur when the spacing of the components is sufficiently close in frequency such that there might be within-channel interactions (e.g., [Henning, 1980](#); [Dye, 1990](#); [Stellmack and Dye, 1993](#)) as well as when the components are spaced so widely that the interference represents an across-channel effect (e.g., [McFadden and Pasanen, 1976](#); [Bernstein and Trahiotis, 1995](#); [Dreyer and Oxenham, 2008](#); [Bernstein and Trahiotis, 2008](#)). Similar effects have been demonstrated with respect to interaural level differences (ILDs; [Bernstein and Trahiotis, 1995](#); [Dye, 1997](#)). Comparable interference effects also have been shown for sound sources in the free field ([Perrott, 1984](#)). [McFadden and Pasanen \(1976\)](#) first described their effects as “interference,” and the term “binaural interference” most often is used to refer specifically to such across-channel effects rather than within-channel effects, in which energetic masking plays a role. However, effects which may involve within-channel interactions also have been described as “binaural interference” on occasion ([Stellmack and Dye, 1993](#); [Best et al., 2007](#), in their review of the literature). In the present paper, for convenience, we use the term “interference” to

refer generally to the increase in threshold ITD for a target component when distractors are present, such as when distractor sound sources interfere with a listener's ability to localize a target sound source.

Presumably, in order to accurately extract the binaural information in a subset of spectral components of a complex stimulus, a listener must identify and perceptually segregate the spectral components of interest from other components that are present. Consequently, one might expect that binaural information extraction is affected by monaural cues that can lead to the perceptual segregation and fusion of spectral components in a complex sound. Several studies have shown mixed effects of *simultaneous* grouping cues, such as varying the harmonic relationship between target and distractor components and introducing onset asynchronies between targets and distractors ([Buell and Hafter, 1991](#); [Hill and Darwin, 1996](#); [Stellmack and Dye, 1993](#); [Woods and Colburn, 1992](#)). In an experiment looking at the effects of *sequential* grouping cues, [Best et al. \(2007\)](#) showed that binaural interference by a 500-Hz distractor on a 4-kHz target (with both target and distractor amplitude modulated at 250 Hz) could be reduced when the distractor was temporally flanked by identical pulses designed to capture the distractor into a distinct perceptual stream, thus segregating the distractor from the target. In another study that looked at the effect of simultaneous grouping cues on binaural interference, one that has particular relevance to the present experiments, [Stellmack \(1994\)](#) showed that the interference between components could be substantially reduced when the target component is presented very briefly in isolation, such as when the distractor components are turned off for as little as 25 ms. One

possibility is that gating off the distractors briefly provided a monaural grouping cue to the segregation of the target and distractors to reduce the interference by the distractors. [Stellmack and Lutfi \(1996\)](#) showed that the interaction of binaural information across frequencies could be modeled in terms of differential weighting of the information in different components. Perhaps monaural cues to the segregation of stimulus components serve to alter the perceptual weight that listeners give to the binaural information in different components of a stimulus.

Given these observations, one might ask what role the interaction of binaural information from different sources plays in everyday listening. Real-world sound sources usually produce complex stimuli (rather than pure tones) that contain fluctuations in amplitude and frequency. In addition, different sound sources rarely produce sounds that overlap exactly in time from onset to offset even when they are intended to do so. For example, when trained musicians perform ensemble music and they attempt to play notes simultaneously, the onsets of simultaneous notes in the musical score typically vary by 30–50 ms when they are actually played ([Rasch, 1979](#)). The asynchronous onsets and amplitude modulations in real-world sounds can result in particular sounds appearing in relative isolation for brief periods. The musical performance asynchronies noted by [Rasch \(1979\)](#) or the amplitude modulation of 20 Hz or less that is typical in speech ([Drullman, 2006](#)) can produce temporal gaps that exceed the 25 ms that [Stellmack \(1994\)](#) found was effective in reducing interference in the processing of ITDs. As a result, one might expect binaural interference to be less problematic in the case of such real-world sound sources.

In an attempt to more closely approximate the conditions that might occur in real-world listening, the present experiments examined the effects of amplitude modulation on the ability of listeners to discriminate the ITD of a target in the presence of distractors. One might imagine two ways in which amplitude modulation can have an effect: (1) modulation might act as a simultaneous grouping cue that leads to perceptual segregation of the target and distractors through perceptual grouping mechanisms, which in turn might lead to an increased ability to process the target ITD separately from that of the distractor, or (2) modulation of the distractors may offer brief “looks” at the target ITD in isolation, which might enable more accurate processing of the target ITD. If (1) is correct, one would expect that either the modulation of the target *or* the distractor could act as a grouping cue that would facilitate perceptual segregation of the target and, in turn, processing of its ITD. [It should be noted that mixed results have been obtained in past research on the effectiveness of amplitude modulation as a monaural grouping cue, as discussed in [Sheft \(2008\)](#).] On the other hand, if (2) is correct, only modulation of the distractor should be effective in reducing interference between the target and distractor because when the distractor is modulated, an unmodulated target will appear in isolation during the envelope minima or dips in the distractor modulation. When only the target is modulated, it will never appear in isolation during an unmodulated distractor; as a result, one might expect substantial interference between the target and distractor to occur.

To test these predictions, in the present series of experiments, the just-discriminable change in ITD was measured for a target in the presence of diotic distractors. In Experiment 1, the targets and distractors were pure tones. In separate conditions, either the target or distractors were amplitude modulated, or the distractors contained a brief temporal gap such that the target appeared briefly in isolation. The threshold ITD of the target was lower when the distractor was amplitude modulated in such a way that the target appeared briefly in isolation during the listening interval. In Experiment 2, the target and distractors were synthesized sounds of plucked guitar strings. The target and distractor notes repeated at a rate of 8 Hz. It was found that binaural interference between the target and distractors was reduced substantially when the target and distractor notes were played slightly asynchronously. Experiment 3 produced results similar to those of Experiment 2 but with pure tones, suggesting that the complex spectra of synthesized musical notes is not necessary for the reduction in binaural interference to occur with asynchronous modulation. Experiment 4 showed that the decrease in binaural interference that occurred when the distractors were modulated was not due to increased detectability of the target component. As a whole, these results suggest that the release from binaural interference does not occur as a result of the modulation serving as a grouping cue that permits perceptual segregation of the target, but rather that asynchronous modulation of the distractor envelope provides brief looks at the target in relative isolation, which facilitates processing of the target ITD.

II. EXPERIMENT 1: EFFECTS OF AMPLITUDE MODULATION ON BINAURAL INFORMATION SEGREGATION

In this experiment, ITD-discrimination thresholds were measured for a pure-tone target in isolation and in the presence of diotic pure-tone distractors in order to establish the baseline elevation of thresholds for the target with distractors. Thresholds then were measured with either the target or distractors amplitude modulated at 4 or 8 Hz in order to determine whether this potential perceptual grouping cue reduces the elevation in thresholds. ITD-discrimination thresholds also were measured for the target in the presence of distractors with a temporal gap as a comparison to previously published results ([Stellmack, 1994](#)). The stimuli are depicted schematically in Fig. 1.

A. Method

1. Stimuli

The stimulus in each listening interval was 500 ms in duration and was windowed with 20-ms raised cosine on-off ramps. The three listening intervals of each trial (see below) were separated by 350 ms of silence. The target component was a 753-Hz pure tone and the distractors were pure tones with frequencies of 553, 653, 853, and 953 Hz (as in some of the stimuli of [Stellmack, 1994](#)). All components were presented at 70 dB SPL (prior to modulation, if modulated). All

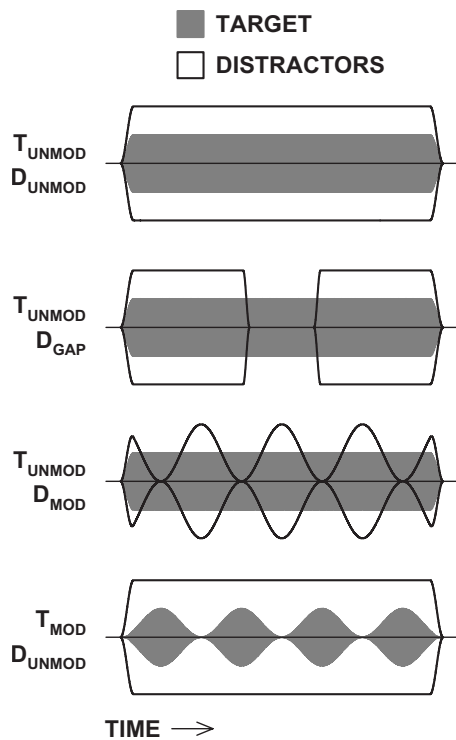


FIG. 1. A schematic illustration of the envelopes of the stimuli of Experiment 1. The envelope of the target is represented by the gray outline and the envelope of the distractors is represented by the black outline. The target carrier was a 753-Hz pure tone. The distractors prior to modulation were pure tones with frequencies of 553, 653, 853, and 953 Hz.

stimuli were generated in Matlab and presented to listeners seated in a sound-attenuating chamber over Sony MDR-V6 headphones.

The starting phases of the components were chosen randomly and independently in each interval. In the two non-signal intervals of each trial, the target and distractors were presented diotically. In the signal interval, the distractors were presented diotically and the target was presented with an ITD that led to the left ear. Thus, the task could be considered a “left-center” discrimination task in which the listener chose the interval containing the target that was interaurally delayed such that it likely would appear to be to the left of the listener’s midline if presented in isolation. The ITD of the target was produced by shifting the phase of the target in one ear. The target was gated on and off simultaneously in the two ears.

The stimulus conditions are depicted in Fig. 1. In order to measure the baseline interference produced by the distractors, ITD discrimination thresholds were measured for the 753-Hz target alone and in the presence of the four distractor frequencies, with all components unmodulated and gated on and off simultaneously, termed $[T_{UNMOD}, D_{UNMOD}]$. In order to replicate some of the conditions of Stellmack (1994), ITD discrimination thresholds also were measured with gaps in the temporal center of the distractors in each listening interval, termed $[T_{UNMOD}, D_{GAP}]$. The gap durations were 25, 50, and 100 ms. At the beginning and end of each gap, the distractors were gated off and on with 10-ms raised-cosine ramps. The gap durations did not include the onset-offset ramps.

In additional conditions, only the target was amplitude modulated, termed $[T_{MOD}, D_{UNMOD}]$, or only the distractors were amplitude modulated, termed $[T_{UNMOD}, D_{MOD}]$. In these conditions, the appropriate components were sinusoidally amplitude modulated as follows:

$$x(t) = 0.5[1 + m \cos(2\pi f_m t)]\cos(2\pi f_c t), \quad (1)$$

where f_c was the component frequency and f_m was the modulation frequency (either 4 or 8 Hz). The modulation index m was 1.0. The 0.5 scaling factor in Eq. (1) served to hold the peak amplitude of the modulated components equal to that of the unmodulated components and to that of the distractors in the $[T_{UNMOD}, D_{GAP}]$ conditions. Threshold ITDs also were measured for the modulated target in isolation at the two modulation rates, termed $[T_{MOD}]$.

2. Procedure

All thresholds were measured using a three-interval, three-alternative forced-choice procedure in which one interval contained the target with a nonzero ITD. The target ITD was varied using a 2-down, 1-up adaptive procedure that tracked to the 70.7-percent correct point on the psychometric function (Levitt, 1971). The target ITD was varied in geometric steps by multiplying or dividing the target ITD by 1.25 for the first 4 reversals and 1.12 for subsequent reversals. The target ITD was set to a well-detectable level that was 3–4 steps above the expected threshold, based on previous experience. The adaptive procedure was limited such that the program would terminate if the target ITD exceeded $664 \mu\text{s}$ (half of a period of the 753-Hz target), which would cause the target to lead in phase to the opposite ear. This occurred occasionally for some listeners in the $[T_{UNMOD}, D_{UNMOD}]$ condition and is discussed below. Each adaptive run continued until 12 reversals were obtained, with the geometric mean of the final eight reversals taken as threshold for that run; geometric means were computed because geometric steps were used in the adaptive tracking procedure to estimate thresholds. For each condition, a total of four runs were completed and the geometric mean of the four threshold estimates for each condition was taken as the final threshold estimate for that condition. All of the conditions were run in a pseudo-random order that was chosen by the experimenters.

Subjects initiated each run and made responses by pressing keys on the PC keyboard. At the start of each trial, a “ready” light was illuminated on the PC monitor for 250 ms followed by a 350-ms pause and then the first listening interval. Each interval was marked with a visual marker on the PC monitor and subjects received visual correct-answer feedback after each trial.

3. Subjects

Five subjects provided data in all conditions. One of the subjects was the first author. The remaining subjects were four undergraduate students (three female and one male) from the University of Minnesota who were paid to participate in the study. All listeners had pure-tone thresholds of 15 dB HL or better at octave frequencies from 250–8000 Hz.

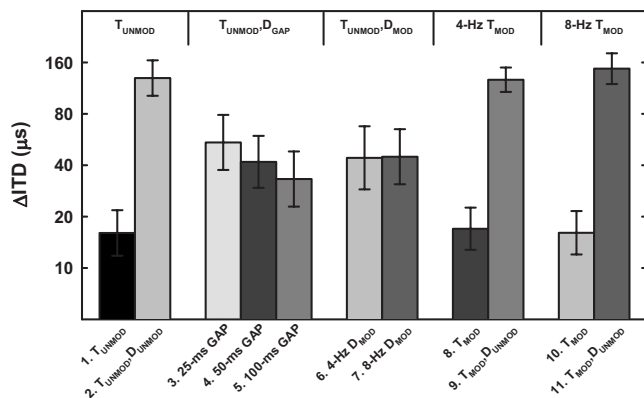


FIG. 2. The results of Experiment 1 averaged across 5 listeners. Threshold Δ ITDs are plotted in μ s on a log axis for each condition. The numbers on the labels of the abscissa are used in the text to refer to the conditions. The error bars represent standard errors of the mean computed across subjects.

All subjects had previous experience in other psychoacoustical tasks and performed several practice runs of the unmodulated target alone condition and the $[T_{UNMOD}, D_{UNMOD}]$ condition prior to data collection in order to familiarize them with the task and to reach asymptotic performance.

B. Results and discussion

Although there were individual differences between subjects in terms of overall sensitivity, the pattern of thresholds was similar across listeners. Therefore, the geometric means of the threshold Δ ITDs (the ITD of the target in the signal interval at threshold) averaged across listeners are plotted in Fig. 2. Threshold Δ ITDs are plotted in microseconds on a logarithmic axis.

The two leftmost bars (bars 1 and 2) show the thresholds for the target-alone condition and the $[T_{UNMOD}, D_{UNMOD}]$ condition. The presence of the diotic distractors greatly elevated thresholds. For the four listeners other than the first author, the adaptive procedure occasionally exceeded the maximum allowable target ITD in the $[T_{UNMOD}, D_{UNMOD}]$ condition and the run was terminated. In those cases, additional runs were performed until the procedure terminated without the target ITD exceeding the maximum. Thus, the plotted thresholds for the $[T_{UNMOD}, D_{UNMOD}]$ condition (bar 2) potentially underestimates the true thresholds. Bars 3–5 show the threshold Δ ITD for the $[T_{UNMOD}, D_{GAP}]$ conditions. Consistent with the data of Stellmack (1994), it can be seen that thresholds were substantially lower when the distractors were turned off briefly during the course of the stimulus. The next two bars (bars 6 and 7), representing the $[T_{UNMOD}, D_{MOD}]$ conditions, show that thresholds when the distractors were amplitude modulated were comparable to those in the $[T_{UNMOD}, D_{GAP}]$ conditions.

The four rightmost bars represent thresholds for the target-modulation conditions for the two modulation rates, without and with distractors. Thresholds for the modulated targets in isolation (bars 8 and 10) were similar to those for the unmodulated target in isolation (bar 1). Thresholds for the $[T_{MOD}, D_{UNMOD}]$ conditions (bars 9 and 11) were similar to those for the $T_{UNMOD}-D_{UNMOD}$ condition (bar 2). These results indicate that the interference between components

was reduced when only the distractors were modulated but not when only the target was modulated. This was true for the individual subjects as well as the mean data.

A repeated-measures ANOVA showed that significant differences existed among the 11 conditions shown in Fig. 2, $F(10, 40) = 46.80$, $p < 0.001$. In order to perform post hoc comparisons of the conditions of particular interest, individual subjects' means were computed across the conditions with distractor modulation (6 and 7) and, separately, with the $[T_{MOD}, D_{UNMOD}]$ conditions (9 and 11). Post hoc comparisons were performed with the Bonferroni correction for multiple comparisons and showed that mean thresholds with only the distractors modulated (6 and 7 pooled) were significantly different from thresholds with only the target modulated (9 and 11 pooled), $t(4) = 4.78$, $p = 0.009$. Mean thresholds with modulated distractors (6 and 7 pooled) were significantly different from thresholds with the target alone (condition 1), $t(4) = 4.279$, $p = 0.013$. These results indicate that some binaural interference was produced by the modulated distractors with an unmodulated target, but less than the amount of binaural interference produced by unmodulated distractors on a modulated target.

As a whole, these results suggest that modulation of the distractors reduces interference with the target because it provides brief looks at the target ITD in isolation, similar to the situation for temporal gaps in the distractors. Because substantial interference occurs (thresholds remain high) when only the target component is modulated, these results do not support the idea that modulation introduces segregation cues that facilitate processing of the target ITD.

III. EXPERIMENT 2: SEGREGATION OF BINAURAL INFORMATION IN SYNTHESIZED MUSICAL SOUNDS

Playing a series of notes on a musical instrument effectively produces a stream of notes with envelope modulation at a relatively low rate due to the timing of successive notes. This experiment examined the role of the low-rate envelope modulations associated with playing a series of notes in permitting the segregation of binaural information in multiple simultaneous streams of musical notes.

In this experiment, the target and distractor stimuli consisted of sounds of plucked acoustic guitar strings produced by a keyboard synthesizer. The target was a repeating G5 (with a fundamental frequency of about 784 Hz). The diotic distractors were a repeating C5 and B6 (with fundamental frequencies of about 523 and 988 Hz, respectively). The target and distractors were presented simultaneously to simulate two guitars playing different sequences of notes at slightly different locations. The listener's task was to discriminate ITDs in the target stream of notes as a function of the degree of synchrony between the target and distractors. Given the results of Experiment 1, it was expected that asynchronies that occur between the different streams of notes would afford opportunities to extract the binaural information in the target stream.

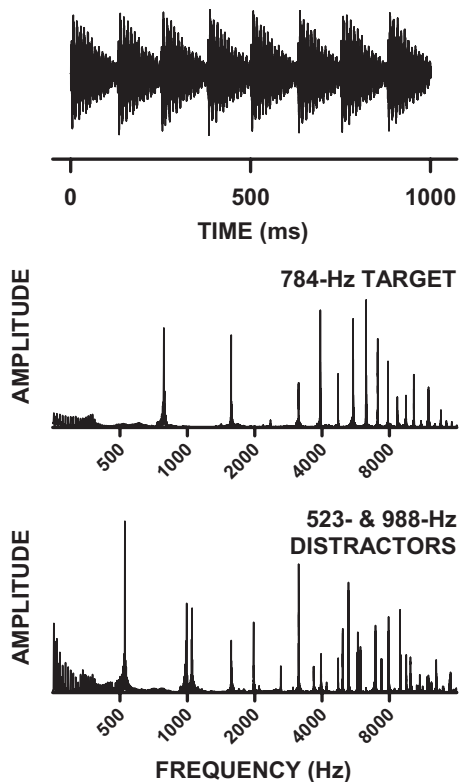


FIG. 3. Upper panel: The time-domain waveform of a synthesized plucked guitar string used as a target in Experiment 2. Center panel: The spectrum of the same target, which had a 784-Hz fundamental. Lower panel: The spectrum of the distractor in Experiment 2, the sum of two synthesized plucked guitar strings with fundamental frequencies of 523 and 988 Hz.

A. Method

1. Stimuli

The stimuli were produced by playing notes on a keyboard synthesizer (Casio CTK-900). All target and distractor notes were performed using the same voice (tone 080, PurA-coGt, the synthesized plucked guitar string sound as identified on the synthesizer). Each target and distractor stream of notes consisted of a single note repeated at a rate of eight notes per second for a duration of 1 s, resulting in an 8-Hz modulation rate. The target and distractor streams of notes were recorded on the synthesizer's internal sequencer. The timing of the notes was controlled by the sequencer by setting the sequencer's metronome and the note durations to nominally produce eight notes per second. The contents of the sequencer were played back and sampled on a PC (using Audacity sound-editing software) at a sampling rate of 44.1 kHz. The sampled sequences were inspected visually to determine that the desired timing was achieved. Each sequence of repeating notes (one sequence containing the target and another containing the two distractors) was sampled separately and the samples were combined with the appropriate relative phases between streams of notes in Matlab. The time-domain waveform of a sampled sequence of target notes and the spectra of the target and distractors are shown in Fig. 3. The effective modulation index of the stimulus in the top panel of Fig. 3 was estimated by extracting its Hilbert envelope, low-pass filtering it at 20 Hz, and finding the maximum and minimum amplitudes of the result, which

yielded an estimated modulation index (m) of 0.95. Recall that for the sinusoidally amplitude-modulated stimuli of Experiment 1, $m=1.0$.

In order to produce ITDs of the target, one channel of the target was interaurally delayed by multiplying the spectrum of the sampled guitar sound with that of a delayed impulse in Matlab. The time-domain waveform was reconstructed from the resulting spectrum via inverse fast Fourier transform. As in Experiment 1, thresholds were measured in a three-interval, three-alternative forced-choice task.

Experiment 1 utilized a left-center task in which the signal interval contained an interaurally delayed target component and diotic distractors while in the non-signal interval, all components were diotic. It is possible that the interaural correlation or perceived width of the stimulus served as a cue to solving the task rather than the perceived intracranial location of the target component. In contrast, in Experiment 2, the target phase led to the left ear in the signal interval and to the right ear by the same amount in the non-signal intervals, and the distractors were always presented diotically. Thus the task in Experiment 2 amounted to a left-right task, where the listener chose the interval in which the target appeared to be to the left of the midline. Because the target was interaurally delayed by the same amount to the left or right ear in each interval, the interaural correlation and presumably the perceived width of the stimulus was equal in all intervals of each trial and could not serve as a cue to performing the task.

The onset delay of the target notes relative to the distractor notes was set to either 0 ms (the onsets of all notes were simultaneous), 31 ms, or 62 ms, which correspond to delays of zero, 1/4, and 1/2 of a modulation period, given the 125-ms repetition rate of the notes. Thus, in the 0-ms delay condition, the onsets of the target and distractor notes occurred simultaneously, while in the 62-ms delay condition, the onsets of the target and distractor notes essentially were perfectly interleaved. In the 0- and 31-ms conditions, eight target notes were played while in the 62-ms delay condition, only seven target notes were played so that each note was temporally flanked by distractor notes. In all conditions, the two distractor notes (C5 and B6) were played synchronously. The target stream of notes was presented at an overall level of 72 dB SPL and the distractor stream of notes, consisting of both C5 and B6 played together, was also presented at an overall level of 72 dB SPL. Thus, the target notes were slightly more intense than either individual distractor note in all conditions.

All stimulus manipulations were performed in Matlab. The stimuli were presented over Sony MDR-V6 headphones to listeners seated in a sound-attenuating chamber, as in Experiment 1.

2. Procedure

Thresholds were measured using the same adaptive procedure as in Experiment 1. Thresholds were measured in a random order for all four conditions. Final threshold estimates were based on the mean threshold of four adaptive runs in each condition.

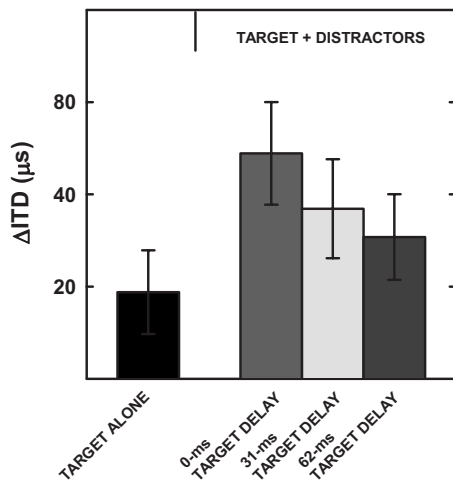


FIG. 4. The results of Experiment 2 averaged across four listeners. Threshold Δ ITDs for the synthesized plucked guitar string target are plotted in μ s on a log axis. The error bars represent standard errors of the mean computed across subjects.

3. Subjects

Three of the listeners in this experiment, the first author and a male and female, also participated in Experiment 1. The fourth listener was the second author, who had pure-tone thresholds of 15 dB HL or better at octave frequencies from 250–8000 Hz. The second author had substantial experience in other similar psychoacoustical tasks and required no training prior to data collection.

B. Results and discussion

Thresholds varied across listeners but the pattern of thresholds across conditions was similar, therefore thresholds were averaged geometrically across listeners and are plotted in Fig. 4. Threshold Δ ITDs are plotted as the difference in target ITD in μ s between the signal and non-signal intervals at threshold (i.e., twice the target ITD in the signal interval). The leftmost bar represents the threshold for the target alone and the three remaining bars represent thresholds in the target-delay conditions.

A repeated-measures ANOVA indicated that significant differences existed across the four conditions shown in Fig. 4, $F(3,9)=12.25$, $p=0.02$. Post hoc tests using the Bonferroni correction for multiple comparisons were performed to compare selected individual conditions. The mean of the target-alone condition was significantly different from that of the 0-ms target-delay condition [$t(3)=7.70$, $p=0.005$] but not that of the 31-ms target-delay condition [$t(3)=5.09$, $p=0.015$] or the 62-ms target-delay condition [$t(3)=3.09$, $p=0.054$]. However, the mean of the 0-ms target-delay condition was not significantly different from that of the 31-ms target-delay condition [$t(3)=1.70$, $p=0.19$] or the 62-ms target-delay condition [$t(3)=2.41$, $p=0.095$]. Although it failed to reach statistical significance for the small number of subjects used here, there was a trend for decreasing thresholds with increasing target delay.

Comparing the target-alone and 0-ms target-delay thresholds, the presence of synchronous distractors elevated thresholds for these synthesized guitar sounds, although the

elevation was not as great as for the pure-tone stimuli of Experiment 1. Delaying the entire target waveform relative to the distractor waveform by 1/4 or 1/2 period of modulation reduced thresholds to nearly that for the target alone. Note that the 31-ms target delay is on the order of the average asynchrony that is observed between musicians in performed music (Rasch, 1979). Thus, for asynchronies comparable to those that are typically exhibited in musical performance, there appears to be little interference between the binaural information in the different streams of musical notes presented here.

IV. EXPERIMENT 3: SEGREGATION OF BINAURAL INFORMATION IN ASYNCHRONOUSLY MODULATED TONES

Experiment 1 showed that the binaural information in a target is more readily available when distractor components are turned off briefly or amplitude modulated in order to permit brief looks at the target in relative isolation. Experiment 2 showed that for synthesized musical sounds, binaural information in a target series of repeating notes is perceived more accurately when distractor notes are presented asynchronously. It is possible that the asynchronous modulation between the target and distractor notes in Experiment 2 effectively introduced brief isolated glimpses at the target notes, which in turn facilitated processing of the ITD of the target as in Experiment 1. However, the stimuli in Experiment 1 (pure tones and modulated pure tones) were much more spectrally simple than those in Experiment 2. Perhaps the complexity of the stimuli in Experiment 2 accounts for the facilitation of ITD processing in the presence of modulation. Furthermore, different stimulus durations were used in Experiments 1 and 2 and in some conditions of Experiment 2, the target and distractors were both modulated with equivalent temporal envelopes, which was not the case in Experiment 1. Experiment 3 linked the observations of Experiments 1 and 2 by presenting conditions that were similar to those of Experiment 2 using pure tones and amplitude-modulated pure tones.

A. Method

As in Experiment 2, threshold Δ ITDs were measured for a target component in the presence of two distractors. Unlike Experiment 2, the target and distractors in this experiment were pure tones. The target frequency was 784 Hz and the distractor frequencies were 523 and 988 Hz, which were the fundamental frequencies of the musical notes of Experiment 2.

Threshold Δ ITDs for the target were measured in the following conditions: target and distractors unmodulated [T_{UNMOD}, D_{UNMOD}], only the target modulated [T_{MOD}, D_{UNMOD}], and only the distractors modulated [T_{UNMOD}, D_{MOD}]. Equation (1) was applied to modulate the appropriate components, with $f_m=8$ Hz and $m=1.0$. The stimulus duration was 500 ms in the [T_{UNMOD}, D_{UNMOD}] condition, while the duration was 1000 ms in the remaining conditions. The target and distractors were gated on and off

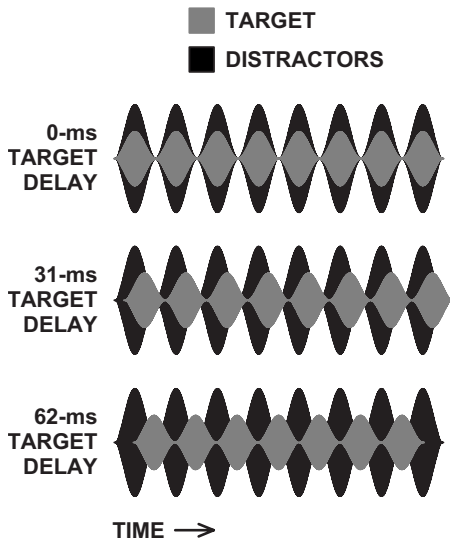


FIG. 5. A schematic illustration of the relative timing of the target and distractor envelopes in the three target delay conditions of Experiment 3.

simultaneously in these conditions. These conditions are similar to those of Experiment 1, although with different frequencies and durations.

Unlike Experiment 1, three conditions also were run in which *both* the target and distractors were modulated $[T_{MOD}, D_{MOD}]$ with several onset delays of the target, as was done with the musical sounds in Experiment 2. For these conditions, the target and distractor were modulated as follows:

$$x(t) = 0.5[1 - m \cos(2\pi f_m t)] \cos(2\pi f_c t). \quad (2)$$

Note that the phase of the modulator in Eq. (2) is inverted relative to Eq. (1), meaning that the target and distractors began and ended during envelope minima. As in Eq. (1), the 0.5 scaling factor in Eq. (2) held the peak amplitude of the modulated stimulus equal to that of the unmodulated stimulus.

The envelopes of the target and distractors are illustrated in Fig. 5. In the $[T_{MOD}, D_{MOD}]$ conditions, the target onset was delayed relative to the onset of the distractors by 0, 31, or 62 ms, the latter two delays corresponding to approximately 1/4 and 1/2 period of the 8-Hz modulator. In the 0- and 31-ms target-delay conditions, the target and distractors were 1000 ms in duration. Thus, in the 0-ms delay condition, the target and distractors were presented synchronously and were modulated in phase. In the 31-ms target-delay condition, the entire 1000-ms target stimulus was delayed such that both the onset and the offset of the target occurred 31 ms after the distractor onset and offset. In the 62-ms delay condition, the target was reduced to 7 cycles of modulation such that the onset of the target occurred after that of the distractors but the offset of the target preceded the offset of the distractors. In the 62-ms target-delay condition, peaks of the target modulation occurred essentially in the temporal center of the dips of the distractor modulation.

Each target and distractor component was windowed with 20-ms raised cosine on-off ramps in all conditions. All components were presented at 70 dB SPL prior to modula-

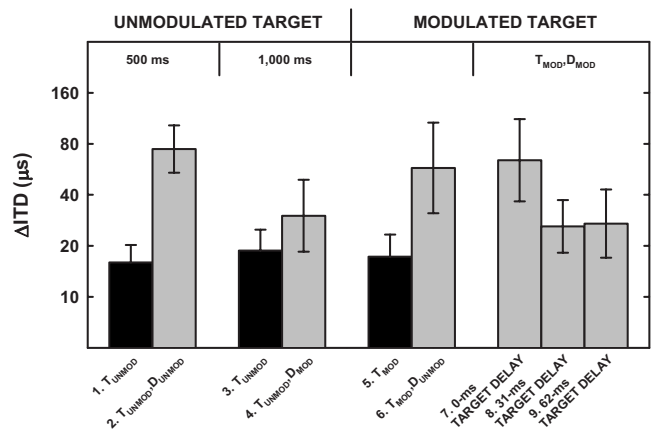


FIG. 6. The results of Experiment 3 averaged across four listeners. Threshold Δ ITDs for the target are plotted in μ s on a log axis. The numbers on the labels of the abscissa are used in the text to refer to the conditions. The error bars represent standard errors of the mean computed across subjects.

tion. This is in contrast to Experiment 2, in which the target level was set equal to the overall level of the combined distractors.

Thresholds were measured using a three-interval, three-alternative forced-choice procedure. In one interval, the target component carried an ITD leading to the left ear and in the remaining two intervals the target ITD led to the right ear. Thus, the listener's task was a left-right task such that the listener selected the interval in which the target appeared to be to the left of midline, as in Experiment 2.

Thresholds were measured using the same adaptive procedure as in Experiments 1 and 2. Thresholds were measured for all conditions in a random order. The four listeners from Experiment 2 participated in this experiment.

B. Results and discussion

Thresholds averaged across listeners are plotted in Fig. 6. The threshold Δ ITDs are plotted as the difference between the ITD of the target in the signal and non-signal intervals at threshold, that is, twice the ITD of the target in the signal interval at threshold in μ s. The black bars represent threshold Δ ITDs for the target in isolation (either modulated or unmodulated, depending on the condition) and the gray bars represent thresholds for the same target in the presence of the distractors.

Comparing the black bars across conditions, threshold Δ ITDs of the target were essentially equal across the two durations (500 and 1000 ms) independent of target modulation. These results match those of Tobias and Zerlin (1959), who found that threshold ITDs for noise bursts decreased with increasing duration, reaching an asymptote at a duration of about 700 ms, but thresholds were nearly equal for durations of 500 and 1000 ms.

As shown in Fig. 6, thresholds were elevated by different amounts by the distractors in the different conditions. Generally, thresholds were elevated most by the distractors when the target was overlapped completely by the distractors, i.e., in the $[T_{UNMOD}, D_{UNMOD}]$, $[T_{MOD}, D_{UNMOD}]$, and 0-ms target-delay conditions (bars 2, 6, and 7, respectively). Thresholds were elevated much less by the distractors when

the stimulus was such that the target appeared in relative isolation during a brief portion of the stimulus, i.e., in the $[T_{\text{UNMOD}}, D_{\text{MOD}}]$ condition (bar 4), and 31- and 62-ms target delay conditions (bars 8 and 9). These data support the idea that the binaural information in a target can be extracted reasonably well from a complex stimulus when the target appears very briefly in isolation due to asynchronous modulation of components. Complex target and distractor spectra, such as those of the musical notes in Experiment 2, are not necessary to produce this effect. This might be expected given that the higher harmonics of the guitar sounds are near or above the frequency at which ITDs can be discriminated on the basis of fine structure (i.e., interaural phase differences; Yost, 1974), so they may not contribute substantially to the lateralization of these sounds on the basis of ITD.

In terms of statistical analysis of the data, a repeated-measures analysis of variance (ANOVA) comparing the means shown in Fig. 6 was significant [$F(8, 24)=14.78$, $p < 0.001$], indicating that two or more of the means in that group were significantly different from one another. In order to identify which pairs were significantly different, post hoc comparisons were performed using the Bonferroni correction for multiple comparisons. These post hoc analyses showed that the mean of the 0-ms target delay condition (bar 7) was significantly different from the mean of the 62-ms target delay condition [bar 9, $t(3)=6.73$, $p=0.007$], but not significantly different from the mean of the 31-ms target delay condition [bar 8, $t(3)=3.31$, $p=0.04$]. Furthermore, the mean of the modulated target-alone condition (bar 5) was not significantly different from the means of the 0-ms target delay condition [bar 7, $t(3)=4.41$, $p=0.022$], the 31-ms target delay condition [bar 8, $t(3)=5.63$, $p=0.01$], or the 62-ms target delay condition [bar 9, $t(3)=2.50$, $p=0.09$]. Although some of the larger mean differences failed to reach significance due to the small sample size, the trends in the data were consistent across listeners and support the conclusions described above.

V. EXPERIMENT 4: EFFECT OF SENSATION LEVEL ON TARGET LATERALIZATION WITH MODULATED DISTRACTORS

Experiments 1 and 3 showed that threshold Δ ITDs for the target are lower in the presence of distractors that are amplitude modulated compared to unmodulated distractors. One potential explanation for this result is that the target is at a higher sensation level and may be more detectable in the presence of modulated distractors than among unmodulated distractors, given that a pure tone is more detectable in a modulated masker than in an unmodulated masker (the so-called modulated-unmodulated difference, or MUD; Carlyon *et al.*, 1989; Bacon *et al.*, 1997). Perhaps increasing the detectability of the target (and so increasing its sensation level) in the presence of distractors makes its ITD more separable and reduces thresholds. This experiment examined the effect of target level on threshold target Δ ITD in the presence of modulated distractors. Specifically, in this experiment, thresholds were measured for the target in the presence of modulated distractors when the target was at the same sensation level as the target among unmodulated distractors. By

reducing the sensation level of the target when in the presence of modulated distractors, we are not eliminating the potential for listeners to benefit from listening to the target during dips of the distractor modulation, rather we are counteracting the increase in sensation level that occurs when a masker is modulated.

A. Method

The target and distractors were identical in frequency and duration to those in Experiment 3 in the $[T_{\text{UNMOD}}, D_{\text{MOD}}]$ condition. In order to establish the detectability of the target in the presence of the distractors, detection thresholds were measured binaurally for a diotic 784-Hz target in the presence of unmodulated and modulated distractors (523 and 988 Hz) that also were presented diotically. The distractors, when modulated, were modulated sinusoidally using Eq. (1) at $f_m=8$ Hz, $m=1$. The distractors were presented at 70 dB SPL prior to modulation. Stimuli were presented in a three-alternative forced-choice task in which the three intervals of each trial were 1 s in duration, separated by 350 ms of silence. Detection thresholds were measured using a two-down-one-up adaptive procedure in which the target level was varied to estimate threshold. The step size was 4 dB for the first four reversals and 2 dB thereafter. Each run was terminated after 12 reversals and the target levels at the final eight reversals were averaged to yield threshold. Four such runs were completed and the mean threshold of those four runs was taken as the final detection threshold.

The difference between the thresholds in the presence of the modulated and unmodulated maskers was computed; this is the value of the MUD. The target level in the $[T_{\text{UNMOD}}, D_{\text{MOD}}]$ condition of Experiment 3 (70 dB SPL) was then reduced by the MUD for each individual subject and threshold ITD was measured again for the unmodulated target in the presence of modulated distractors. (Note that even when the threshold was reduced by the MUD, the target was still well above its detection threshold.) Thresholds were measured using the same procedure as in Experiment 3.

The three listeners in this experiment participated in the previous experiments. They were the first and second authors (S1 and S3, respectively) and a female undergraduate student from the University of Minnesota (S2).

B. Results and discussion

Threshold Δ ITDs for the three individual listeners are shown in Fig. 7. The black bars in the figure show the threshold for each listener measured in the $[T_{\text{UNMOD}}, D_{\text{MOD}}]$ condition of Experiment 3 with the target at 70 dB SPL. The gray bars show the threshold measured in the present experiment with the target level adjusted to the same sensation level as in the $[T_{\text{UNMOD}}, D_{\text{UNMOD}}]$ condition of Experiment 3. The levels of the target for S1, S2, and S3 in these conditions were, respectively, 55, 41, and 50 dB SPL (corresponding to sensation levels of 29, 20, and 32 dB). It can be seen that thresholds were not influenced by target SL over this range of target levels. Independent-samples t-tests using the four individual threshold estimates that contributed to each

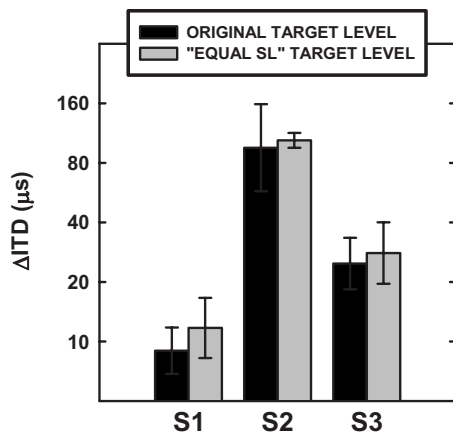


FIG. 7. The results of Experiment 4 for three individual listeners. Threshold Δ ITDs for the target are plotted in μ s on a log axis. Thresholds represented by the black bars were measured with the pure-tone target at 70 dB SPL. Thresholds represented by the gray bars were measured with the pure-tone target at the SPL that makes the target equally detectable to the unmodulated target in the presence of unmodulated distractors. The error bars represent standard deviations of the four threshold estimates for each individual listener.

mean showed no significant differences between the means represented by each pair of bars in Fig. 7 [S1: $t(6) = 1.20$, $p = 0.28$; S2: $t(6) = 0.34$, $p = 0.75$; S3: $t(6) = 0.52$, $p = 0.62$]. Therefore, the lower threshold Δ ITDs in the $[T_{UNMOD}, D_{MOD}]$ condition of Experiment 3 relative to the $[T_{UNMOD}, D_{UNMOD}]$ condition cannot be attributed to the increased SL of the target. This is consistent with the idea that modulation of only the distractors affords brief looks at the target in isolation that result in more precise discrimination of the target ITD.

VI. GENERAL DISCUSSION

As described in the introduction, previous research has shown that when multiple spectral components carry different ITDs, thresholds for the discrimination of the ITD of a target subset of those components are elevated relative to when those components are presented in isolation. The present data suggest that the threshold elevation is diminished greatly when asynchronies exist between components such that the target components to be discriminated appear very briefly in isolation by either asynchronous modulation of the components or by turning off the distractors. The types of asynchronies used in these experiments are similar in quality and timing to those that often occur in real-world situations, such as in musical performance.

In the presence of unmodulated distractors, threshold Δ ITDs for the target are nearly as high when the target is modulated as when it is unmodulated. This suggests that the reduction in thresholds does not stem from segregation cues that might result when a subset of the stimulus components are modulated (although perceptual segregation per se was not measured in the present experiments). Furthermore, Experiment 4 showed that the greater sensitivity to target ITDs does not occur as a result of changes in the SL of the target that might occur with modulation of the distractor components. Rather, the critical feature for enhanced segregation of binaural information of the target is that the modulation and

relative phases must be such that the target appears in relative isolation for at least a brief period during the stimulus. This is consistent with explanations of monaural detection phenomena that favor “dip-listening” over explanations of those phenomena that depend on modulation-based grouping cues across frequency (Moore *et al.*, 1990; Borrill and Moore, 2002).

The present results stand in contrast to the results of Best *et al.* (2007), who found that grouping cues did reduce the interference between target and distractor components. However, as noted in the introduction, Best *et al.* used much more widely spaced components than in the present experiments (500 Hz and 4 kHz) and they studied the effects of sequential grouping cues that occur across time. Most of the experiments in the present series utilized pure tones that were more closely spaced and examined the effects of simultaneous grouping cues, namely, amplitude modulation. It is possible that asynchronous modulation could serve as a simultaneous grouping cue that permits segregation of binaural information in a target from that of distractors for wider frequency spacings, but it should be noted that the original binaural interference effects reported by McFadden and Pasanen (1976) were between a 230-Hz wide target band of noise centered at 4 kHz and a 50-Hz-wide distractor band of noise centered at 500 Hz. Although the target and distractor bands had different temporal envelopes, there was still ample binaural interference between bands. The 50-Hz bandwidth of the distractor band would provide only very short duration looks at the target band in relative isolation (presumably around 10 ms, on average), which clearly was insufficient to eliminate binaural interference.

The present results support the idea that binaural interference produced by distractors is greatly reduced when a target appears very briefly in isolation. When listeners are asked to describe the apparent lateral position of a target and distractors that are presented synchronously but with different ITDs, listeners often report that the complex stimulus is perceived at some intermediate position that seems to be based roughly on the average of the individual component ITDs (e.g., Dye, 1990). When the distractors in such a stimulus are briefly turned off, the remaining ITD is that of the continuing target component. If the listener can lateralize the target when the distractors are turned off for 25 ms nearly as well as when the target is in complete isolation for over 100 ms, this implies that the binaural system can react very quickly to changes in binaural information. However, numerous studies have shown that the binaural system responds gradually to changes in the binaural information in an ongoing stimulus, over about 100 ms after the change occurs (e.g., Grantham and Wightman, 1978; Kollmeier and Gilkey, 1990). Given the apparent sluggishness of the binaural system, it is more likely that the ITD of the target is individually accessible to the binaural system and turning off the distractors causes the target to be perceived as a separate auditory event, which results in its ITD being quickly segregated from that of the distractors. This is consistent with the idea that binaural interference represents a weighted combination of binaural information and that cues to source segregation might alter the perceptual weight given to the binaural infor-

mation of the individual components (Stellmack and Lutfi, 1996). Stellmack (1994) discussed these general issues more extensively.

Although the interference effects produced by simultaneous frequency components bearing conflicting binaural information might yield important insights into the operation of the binaural system, the present results suggest that this type of interference may be of little consequence in everyday listening. For example, in the asynchronous conditions of Experiment 3, the average threshold Δ ITD for the modulated target alone (i.e., with no distractors) was about 17 μ s while the average threshold in the condition in which only the distractor was modulated (bar 4 in Fig. 6) and the two nonzero target delay conditions (the asynchronous distractor conditions, bars 8 and 9 in Fig. 6) was about 28 μ s. Using a simple spherical head model to compute the corresponding changes in azimuthal angle for real sound sources (Woodworth and Schlosberg, 1954), those threshold Δ ITDs correspond to differences in azimuth (which might be analogous to minimum audible angles) of about 1.9 degrees for the no distractor condition and 3.1 degrees for the asynchronous distractor conditions. In contrast, the average threshold Δ ITD for the $T_{\text{MOD}}-D_{\text{UNMOD}}$ and 0-ms target delay conditions (the conditions in which the distractors overlapped the target completely in time and maximum interference occurred) was 61 μ s, or about 6.8 degrees of azimuth using the spherical head model, a substantial increase over the asynchronous conditions. Thus, the estimated decrease in localization accuracy is substantially smaller, and perhaps of little practical significance, in those situations in which the target is not overlapped completely by the distractor. Furthermore, this release from binaural interference occurred when the ITDs of the target and distractors (at or near 0 μ s) were consistent with sources that were in close spatial proximity. One might expect interference between components to be less evident when the sources have greater spatial separation, as suggested by the data of Hill and Darwin (1996), which is far more likely in real-world listening situations.

ACKNOWLEDGMENTS

This work was supported by Research Grant No. R01 DC 00683 from the National Institute on Deafness and Communication Disorders, National Institutes of Health. We thank Dr. Michael Akeroyd and two anonymous reviewers for comments that helped us improve the manuscript. We also thank Dr. Ewan Macpherson for his assistance in generating stimuli.

Bacon, S. P., Lee, J., Peterson, D. N., and Rainey, D. (1997). "Masking by modulated and unmodulated noise: Effects of bandwidth, modulation rate, signal frequency, and masker level," *J. Acoust. Soc. Am.* **101**, 1600–1610.

Bernstein, L. R., and Trahiotis, C. (1995). "Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms," *J. Acoust. Soc. Am.* **98**, 155–163.

Bernstein, L. R., and Trahiotis, C. (2008). "Discrimination of interaural temporal disparities conveyed by high-frequency sinusoidally amplitude-modulated tones and high-frequency transposed tones: Effects of spec-

trally flanking noises," *J. Acoust. Soc. Am.* **124**, 3088–3094.

Best, V., Gallun, F. J., Carlile, S., and Shinn-Cunningham, B. G. (2007). "Binaural interference and auditory grouping," *J. Acoust. Soc. Am.* **121**, 1070–1076.

Borrill, S. J., and Moore, B. C. J. (2002). "Evidence that comodulation detection differences depend on within-channel mechanisms," *J. Acoust. Soc. Am.* **111**, 309–319.

Buell, T. N., and Hafter, E. R. (1991). "Combination of binaural information across frequency bands," *J. Acoust. Soc. Am.* **90**, 1894–1900.

Carlyon, R. P., Buus, S., and Florentine, M. (1989). "Comodulation masking release for three types of modulator as a function of modulator rate," *Hear. Res.* **42**, 37–45.

Dreyer, A. A., and Oxenham, A. J. (2008). "Effects of level and background noise on interaural time difference discrimination for transposed stimuli," *J. Acoust. Soc. Am.* **123**, EL1–EL7.

Drullman, R. (2006). "The significance of temporal modulation frequencies for speech intelligibility," in *Listening to Speech: An Auditory Perspective*, edited by S. Greenberg and W. A. Ainsworth (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 39–47.

Dye, R. H. (1990). "The combination of interaural information across frequencies: Lateralization on the basis of interaural delay," *J. Acoust. Soc. Am.* **88**, 2159–2170.

Dye, R. H., Jr. (1997). "The relative contributions of targets and distractors in judgments of laterality based on interaural differences of level," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 151–168.

Grantham, D. W., and Wightman, F. L. (1978). "Detectability of varying interaural temporal differences," *J. Acoust. Soc. Am.* **63**, 511–523.

Henning, G. B. (1980). "Some observations on the lateralization of complex waveforms," *J. Acoust. Soc. Am.* **68**, 446–454.

Hill, N. I., and Darwin, C. J. (1996). "Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning," *J. Acoust. Soc. Am.* **100**, 2352–2364.

Kollmeier, B., and Gilkey, R. H. (1990). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," *J. Acoust. Soc. Am.* **87**, 1709–1719.

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

McFadden, D., and Pasanen, E. G. (1976). "Lateralization at high frequencies based on interaural time differences," *J. Acoust. Soc. Am.* **59**, 634–639.

Moore, B. C. J., Glasberg, B. R., and Schooneveldt, G. P. (1990). "Across-channel masking and comodulation masking release," *J. Acoust. Soc. Am.* **87**, 1683–1694.

Perrott, D. R. (1984). "Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity," *J. Acoust. Soc. Am.* **75**, 1201–1206.

Rasch, R. A. (1979). "Synchronization in performed ensemble music," *Acustica* **43**, 121–131.

Sheft, S. (2008). "Envelope processing and sound-source perception," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 233–279.

Stellmack, M. A. (1994). "The reduction of binaural interference by the temporal nonoverlap of components," *J. Acoust. Soc. Am.* **96**, 1465–1470.

Stellmack, M. A., and Dye, R. H., Jr. (1993). "The combination of interaural information across frequencies: The effects of number and spacing of components, onset asynchrony, and harmonicity," *J. Acoust. Soc. Am.* **93**, 2933–2947.

Stellmack, M. A., and Lutfi, R. A. (1996). "Observer weighting of concurrent binaural information," *J. Acoust. Soc. Am.* **99**, 579–587.

Tobias, J. V., and Zerlin, S. (1959). "Lateralization threshold as a function of stimulus duration," *J. Acoust. Soc. Am.* **31**, 1591–1594.

Woods, W. S., and Colburn, H. S. (1992). "Test of a model of auditory object formation using intensity and ITD discrimination," *J. Acoust. Soc. Am.* **91**, 2894–2902.

Woodworth, R. S., and Schlosberg, H. (1954). *Experimental Psychology*, revised ed. (Holt, Rinehart, and Winston, New York), p. 351.

Yost, W. A. (1974). "Discriminations of interaural phase differences," *J. Acoust. Soc. Am.* **55**, 1299–1303.