Effects of level and background noise on interaural time difference discrimination for transposed stimuli

Anna A. Dreyer
Harvard University/MIT Division of Health Sciences and Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139
adreyer@alum.mit.edu.

Andrew J. Oxenham
Department of Psychology, University of Minnesota, Elliott Hall, 75 East River Parkway, Minneapolis, Minnesota 55455
oxenham@umn.edu.

Abstract: Just-noticeable interaural time differences were measured for low-frequency pure tones, high-frequency sinusoidally amplitude-modulated (SAM) tones, and high-frequency transposed stimuli, at multiple levels with or without a spectrally notched diotic noise to prevent spread of excitation. Performance with transposed stimuli and pure tones was similar in quiet; however, in noise, performance was poorer for transposed stimuli than for pure tones. Performance with SAM tones was always poorest. In all conditions, performance improved slightly with increasing level. The results suggest that the equivalence postulated between transposed stimuli and pure tones is not valid in the presence of a spectrally notched background noise.

PACS numbers: 43.66.Pn [Q-JF]
Date Received: July 12, 2007 **Date Accepted:** October 23, 2007

1. Introduction

Binaural hearing allows humans to localize sound sources in space and can assist in the perceptual segregation of competing sound sources. A dominant cue for determining the azimuth of a sound source is the interaural time difference (ITD) (e.g., Macpherson and Middlebrooks, 2002). Sensitivity to ITD is found for long-duration pure tones only below about 1500 Hz; however, when complex stimuli with time-varying temporal envelopes are used, some ITD sensitivity is also observed for sounds containing only high frequencies (e.g., Klumpp and Eady, 1956; Henning, 1974; McFadden and Pasanen, 1976). It has been suggested that any remaining differences between low- and high-frequency processing of ITDs might be attributed to stimulus characteristics, rather than any inherent deficit in temporal sensitivity at high frequencies (Colburn and Esquissaud, 1976). Van de Par and Kohlrausch (1997) developed so-called "transposed stimuli" to specifically test this hypothesis. Transposed stimuli are designed to produce temporal discharge patterns in the auditory nerve in response to high-frequency modulated stimuli that resemble the discharge patterns in response to low-frequency pure tones. Transposed stimuli are constructed by modulating a high-frequency pure-tone carrier with a halfwave-rectified low-frequency sinusoid.

Using transposed stimuli, Bernstein (2001) and Bernstein and Trahiotis (2002) showed that ITD-based discrimination and lateralization of transposed stimuli was as good as, or better than, that for pure tones at frequencies (or modulation rates) up to 150 Hz. Furthermore, transposed stimuli produced markedly better performance than sinusoidally amplitude-modulated (SAM) tones. However, as (modulation) frequency was increased, lateralization performance using transposed stimuli fell below that using pure tones, suggesting that additional factors beyond peripheral representation also play a role.

The psychophysical data suggest that transposed stimuli are successful to some extent in enhancing the temporal representation of the modulation frequency. However, the question of how well transposed stimuli actually simulate pure-tone temporal discharge patterns in the auditory nerve was tested only recently. Contrary to the idealized model used for motivating transposed stimuli, Dreyer and Delgutte (2006) found that the phase locking to pure tones and transposed stimuli is only similar near rate threshold in the cat auditory nerve. At levels greater than 10 dB above neural threshold, the synchronization index in response to transposed (and SAM) tones was found to decrease rapidly with increasing stimulus level, whereas the synchronization index for pure tones remained more stable with level. Most psychophysical studies using transposed stimuli have been performed well above absolute threshold, where many neural units should be well above their threshold, suggesting a possible discrepancy between the neural and behavioral measures. In other words, based on the synchrony indices from single units in the auditory nerve of cats, one might not expect psychophysical performance using transposed tones to be as good as that using pure tones, in contrast to the available data (Bernstein and Trahiotis, 2002).

One reason for the apparent discrepancy between the neural and psychophysical data may be that information from off-frequency neurons (i.e., neurons with best frequencies that do not match that of the stimulus) determines performance and that, when the stimuli are presented in quiet, there will always be some neurons responding to the stimulus that are close to their rate threshold (e.g., Siebert, 1968). In this way, "off-frequency listening" may play a role in producing equivalent ITD performance for pure-tone and transposed stimuli, even at levels far above threshold.

We tested this hypothesis by presenting the stimuli in a background noise, which was designed to limit the ability of listeners to use off-frequency information, while minimizing the amount of direct interference (i.e., masking) produced by the background noise. Just-noticeable differences in ITD were measured as a function of level for pure tones, transposed stimuli, and SAM tones (as in Bernstein and Trahiotis, 2002). Thresholds were measured in quiet (unrestricted listening) and in the presence of a background noise, spectrally shaped to restrict the spectral region over which information was available, thereby more accurately reflecting the conditions tested in the physiological experiments.

2. Methods

Detection of ongoing ITDs was measured behaviorally in humans as a function of level in unrestricted and restricted listening conditions using low-frequency pure tones, 100% modulated SAM tones, and transposed stimuli. The frequency of the pure tones and the modulation rate of SAM tones and transposed stimuli was always 125 Hz, which corresponds to a frequency at which ITD performance using pure-tone and transposed stimuli has been found to be equivalent (Bernstein and Trahiotis, 2002; Oxenham *et al.*, 2004). All stimuli were presented at levels between 30 and 70 dB sound pressure level (SPL) in 10 dB steps. The SAM tones were generated by modulating a 4 kHz sinusoid with a 125 Hz sinusoid. The transposed tones were generated by multiplying a 4 kHz sinusoid with a 125 Hz sinusoid that had been half-wave rectified and lowpass filtered (Butterworth fourth order) at 800 Hz (0.2 times the carrier frequency). The carrier frequency of 4 kHz is above the assumed range of optimal phase locking by individual auditory nerve fibers (ANFs). Pure tones, SAM tones and transposed stimuli were all 500 ms in length, including 100 ms onset and offset ramps to prevent spectral splatter and to de-emphasize the salience of the on- and offsets. For the high-frequency stimuli, only the ongoing portion of the temporal envelope was delayed.

The pure tones, SAM tones, and transposed stimuli were presented either with or without a background noise designed to limit off-frequency listening (e.g., O'Loughlin and Moore, 1981). For the SAM and transposed stimuli, the noise contained a stop-band between 0.9 and $1.1f_c$ (3600–4400 Hz), so that the noise energy was always at least 10% of the center frequency away from the stimulus (see Fig. 1(a)). For the pure-tone stimuli, a bandpass noise was used, which extended from 150 to 350 Hz, resulting in the noise being no less than 20% of the pure-tone frequency away from the tone. A larger spectral gap was chosen to compensate for the

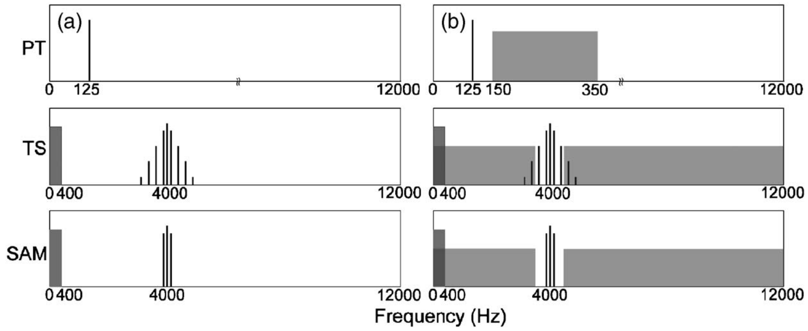

Fig. 1. Schematic illustration of the experimental stimuli. The unrestricted listening condition (a) includes only the low-pass background noise for the high-frequency stimuli (0–400 Hz, shown in dark gray), while the restricted listening condition (b) includes high-pass noise for pure-tone stimuli, and notched noise for SAM tones and transposed stimuli (shown in light gray). The frequency scales for the pure-tone and high-frequency stimuli differ. The frequency axis for the pure-tone stimuli is expanded from 0–400 Hz with a break in the pure-tone frequency axis at 400 Hz, while no such expansion or break is used for representing the high-frequency stimuli.

relatively broader auditory filters at low frequencies (e.g., Glasberg and Moore, 1990), and no noise was added below the pure-tone frequency, because it was deemed unlikely that off-frequency listening would play a role below 100 Hz (e.g., Sek and Moore, 1994). Interaurally correlated noise was used, as informal pilot tests suggested that listeners found it difficult to reliably lateralize transposed stimuli in uncorrelated notched noise within physically realizable ITD limits. Greater interference from uncorrelated noise has also been observed by Ito *et al.* (1982); however, because the task involved distinguishing a right- from a left-leading ITD, correlated noise should not provide additional lateralization cues (e.g., Nuetzel and Hafter, 1976).

For all presentations of SAM tones and transposed stimuli, an additional uncorrelated noise, low-pass filtered with a cutoff frequency of 400 Hz (see Fig. 1(b)), was presented at equal levels to both ears to prevent additional lateralization cues from auditory distortion products. The spectrum level of the low-pass-filtered noise ranged from 29 to 69 dB (re: 20 μPa) for the 30 and 70 dB SPL probe tone, respectively. Adequate masking of distortion products by background noise was confirmed for the 70 dB SPL tone, using the method described by Vliegen and Oxenham (1999).

All noise, when present, was gated on 400 ms prior to the first stimulus interval, and was gated off 200 ms after the second interval, within each two-interval trial, for a total duration of 2.1 s. Both the onset and offset of the noise were shaped with 100 ms raised-cosine ramps.

The stimuli were digitally generated and played through a LynxStudio LynxOne soundcard with 24 bit resolution at a 32 kHz sampling rate. The stimuli were presented to the subjects via Sennheiser HD 580 headphones after being passed though a TDT PA4 programmable attenuator and a TDT HB6 headphone buffer. All testing was conducted in a double-walled, sound-attenuating chamber.

Four normal-hearing subjects (thresholds below 20 dB HL at octave frequencies between 125 and 8000 Hz) served as listeners. Each subject began with a training phase with feedback until they reached asymptotic performance. Training for each subject lasted between 4 and 6 h. Feedback was also provided during the data collection phase. Subjects were paid an hourly wage for their services.

A preliminary masking experiment was performed to determine the levels of the background noise necessary to adequately limit off-frequency listening, while not directly masking the target sounds. Detection thresholds for the stimuli in quiet and in the notched (4 kHz carrier) or bandpass (125 Hz tone) noise were measured using a two-interval forced-choice procedure with a two-down one-up adaptive tracking rule (Levitt, 1971) at spectrum levels in the noise passband of 10, 20, 30, and 40 dB (re: 20 μPa). The masking noise had the same duration spectral characteristics as the noise in the main discrimination study. Subjects' diotic detection

Table 1. Background noise spectrum levels in dB (re: 20 μPa) used for the high-frequency (SAM and transposed) and pure-tone stimuli. Noise levels were the same for all subjects in the presence of the pure tones, but differed for each subject in the presence of the high-frequency stimuli.

		Stimulus level (dB SPL)			
		40	50	60	70
SAM and Transposed	Subject 1	4.2	12.1	19.9	27.8
	Subject 2	2.1	9.5	16.9	24.4
	Subject 3	1.8	10.2	18.5	26.8
	Subject 4	4.1	13.3	22.5	31.6
	Mean (S.D.)	*3.1 (1.3)*	*11.3 (1.7)*	*19.5 (2.3)*	*27.7 (3.0)*
Pure tone	All Subjects	*18.2*	*29.0*	*39.9*	*50.8*

thresholds in quiet and in the two noise conditions were measured for all listeners. Individual signal thresholds were plotted as a function of noise spectrum level and a linear regression was performed to determine the threshold level as a function of the noise level for each subject. The noise level needed to just mask a tone of a particular level was estimated from this regression. The subjects' diotic detection thresholds for pure tones were similar and therefore the background noise levels were chosen by averaging the detection thresholds for all the subjects. Because the detection thresholds differed somewhat between subjects for the high-frequency stimuli, the noise levels were selected individually for each subject. These thresholds were then used to set the band-restricting noise in the ITD discrimination experiment at a level 20 dB below that needed to mask the signal for each subject individually. The noise spectrum levels used in the discrimination experiments are shown in Table 1.

In the lateralization tasks, discrimination thresholds were measured using a two-interval, two-alternative forced-choice method with a two-down, one-up adaptive procedure that tracks the 71% point on the psychometric function (Levitt, 1971). The listeners' task was to decide whether the right ear led in the first or second presentation. Thresholds were obtained in each listening condition in random order and each condition was run three times. The initial ITD was 500 μs (i.e., ±250 μs) and the initial step size was a factor of 2, which was reduced to 1.4 after two reversals and to 1.18 after another two reversals. For each run, testing continued until four reversals at the smallest step size occurred and the threshold was taken as the mean of these four reversals. Final thresholds were geometrically averaged for each listener and stimulus condition over the three adaptive runs.

3. Results

One of the four subjects had some difficulty detecting any ITDs when the tone was at a level of 30 dB SPL and, because of these missing data, only thresholds between 40 and 70 dB SPL are shown. The pattern of results was rather similar across the four subjects, and so only the (geometric) mean results from the ITD discrimination experiment are shown in Fig. 2. Consider first the results without the additional masking noise (Panel A). Thresholds for the pure tones (diamonds) and transposed stimuli (squares) are very similar, in line with previous results showing similar pure-tone and transposed-tone thresholds at a frequency of 125 Hz (e.g., Bernstein and Trahiotis, 2002). Thresholds using SAM tones are elevated relative to those for pure and transposed tones by a factor of between 2 and 3, again broadly consistent with previous studies (Bernstein and Trahiotis, 2002). A repeated-measures analysis of variance (RMANOVA, with Huynh-Feldt correction for sphericity, where appropriate), carried out on the log-transformed data in Panel A confirmed these observations. There was a main effect of condition ($F_{2,6}=20.7, P=0.002$) and a main effect of level ($F_{3,9}=6.54, P=0.043$), with no significant interaction between the factors ($F_{6,18}<1$, n.s.). A contrast analysis revealed a significant linear trend for thresholds to decrease with increasing level ($F_{1,3}=447.0, P<0.001$). Finally, thresholds in

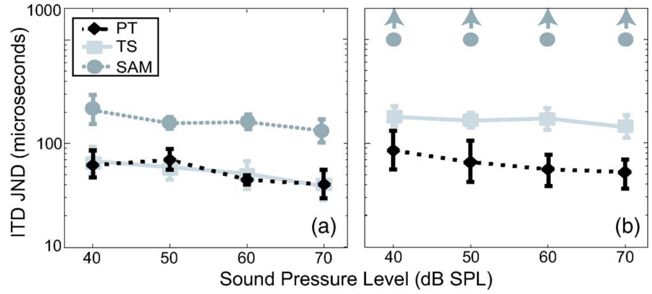

Fig. 2. (Color online) Interaural temporal just noticeable differences measured in an adaptive listening task for pure tones (PT; diamonds), transposed stimuli (TS; squares) and SAM tones (circles) without (a) and with (b) noise added to restrict the listening band. Upward arrows represent SAM tone conditions where the ITD thresholds exceeded 1000 μs. Error bars represent ±1 standard error of the mean.

the SAM condition were shown to be significantly higher than those in the pure-tone and transposed-tone conditions ($P<0.05$ in both cases), which were not significantly different from each other ($F_{1,3}<1$, n.s.).

Consider next the conditions in which the listening band was restricted by noise (Fig. 2(b)). Consistent with results from the unrestricted listening conditions, performance was poorest with the SAM tones. In fact, in this case listeners could not consistently perform the task at all for ITDs of 1000 μs or less, as indicated by the upward-pointing arrows in Fig. 2(b). In contrast to the results without the additional masking noise, performance with the transposed stimuli was consistently poorer than with the pure tones. The difference between pure-tone and transposed thresholds in noise was observed with all four subjects to varying degrees. One subject had thresholds in the transposed conditions that were on average only a factor of 1.2 higher than in the pure-tone conditions, whereas the others showed larger effects, with thresholds being higher by a factor of between 2.7 and 4.5. Averaged across subjects and levels, thresholds in the transposed conditions were about a factor of 3 higher than in the pure-tone conditions. There also remained a trend for improved thresholds with increasing level, which is particularly apparent in the pure-tone condition. A RMANOVA, carried out on just the pure-tone and transposed-tone data with noise, confirmed these observations: significant effects of level ($F_{3,9}=4.4, P=0.036$) and condition ($F_{1,3}=12.7, P=0.038$) were found, again with no interaction ($F_{3,9}<1$, n.s.).

A comparison of performance in the pure-tone and transposed-tone conditions with and without noise showed a significant effect of noise for the transposed-tone condition ($F_{1,3}=49.1, P=0.006$), but not for the pure-tone condition ($F_{1,3}<1$, n.s.). In both cases, there was no interaction between the effects of noise and level ($F_{3,9}<1$, n.s.). Overall, therefore, the additional noise had a substantial effect on thresholds in the transposed and SAM conditions, but not in the pure-tone condition.

4. Discussion

A comparison of ITD discrimination using pure tones and transposed stimuli found similar performance in the absence of background noise, consistent with earlier studies. However, when background noise was added, thresholds in the transposed-tone conditions deteriorated by a factor of about 3 on average, whereas pure-tone thresholds remained very similar. In all conditions, small improvements in performance were found with increasing stimulus level. The results show that the equivalence of transposed and pure tones at low (modulation) frequencies (Bernstein and Trahiotis, 2002) does not hold in the presence of noise. Caution should therefore be exercised in treating the information provided by the two types of stimuli as equivalent.

The question remains as to why the noise affects thresholds for the transposed and SAM stimuli more than for the pure-tone stimuli. The noise in our experiment was designed to restrict off-frequency listening and, as such, the results are consistent with the physiological findings that the auditory-nerve synchrony to transposed tones is similar to that found for pure tones only when the firing rate of the fibers is close to threshold. As mentioned earlier, in the absence of noise, there will always be some off-frequency fibers whose rates are close to threshold, even at high stimulation levels. Thus, our results are consistent with the idea that the noise restricts the "listening band" to those fibers whose characteristic frequencies are close to that of the stimulus, leading to poor performance in the high-frequency conditions, where off-frequency listening is necessary to maintain performance.

Other interpretations are possible, however. For instance, a number of studies have shown that the presence of one or more simultaneous interfering stimuli can reduce sensitivity to target ITDs (e.g., Buell and Hafter, 1991; Woods and Colburn, 1992; Stellmack and Dye, 1993). However, it has been shown that a continuous diotic background noise in spectrally flanking regions produces little or no interference when it is presented continuously, instead of being gated with the target (Trahiotis and Bernstein, 1990). Furthermore, a recent study using transposed stimuli found no significant effects of interference by low-frequency noise, even when the noise was gated synchronously with the target (Bernstein and Trahiotis, 2004). Thus, while it remains a possibility that the diotic noise produced more spectrally remote interference for the transposed stimuli than for the pure-tone stimuli, this is made less likely by the fact that our noise was presented in a quasi-continuous manner (starting 400 ms before the first target and ending 200 ms after the second target). Another possibility is that the noise produced some direct masking effects. This may also explain our informal finding in pilot runs that listeners found the task much more difficult in a background of interaurally uncorrelated noise. Ito *et al.* (1982) also found that an interaurally uncorrelated broadband background noise with a spectrum level 10 dB below that of the narrowband target noise could increase ITD thresholds substantially. However, in their condition that was most comparable to ours (diotic noise masker), the amount of interference was relatively small for most subjects. Furthermore, our targets were presented 20 dB above masked threshold and were always separated from the noise by a spectral gap.

Finally, it is worth noting that even in the restricted listening cases, performance did not degrade with increases in level between 40 and 70 dB SPL. In fact a small but significant improvement with level was found. These findings do not mirror the physiological results of Dreyer and Delgutte (2006), who found that the synchronization index of the auditory-nerve discharge pattern in cats decreased with increasing level, particularly with SAM and transposed tones. This may be a species difference, or may reflect additional (perhaps efferent) influences, not found in the auditory nerve of an anesthetized cat, which help to maintain good representations of the temporal envelope across a wide dynamic range. Another possibility is that the noise itself shifts the operating point of neurons to higher levels, thereby effectively increasing the dynamic range (Palmer and Evans, 1982). Similar apparent discrepancies between auditory-nerve data and psychophysical performance have also been noted for other tasks, such as intensity discrimination (e.g., Viemeister, 1983).

Acknowledgments

This work was supported by NIH Grant Nos. T32 DC 00038 (AAD) and R01 DC 03909 (AJO). Leslie Bernstein, Christophe Micheyl, Bertrand Delgutte, Steve Colburn, and one anonymous reviewer provided helpful comments on earlier versions of this manuscript.

References and links

Bernstein, L. R. (**2001**). "Auditory processing of interaural timing information: New insights," J. Neurosci. Res. **66**, 1035–1046.

Bernstein, L. R., and Trahiotis, C. (**2002**). "Transposed stimuli improve sensitivity to envelope-based interaural timing information for stimuli having center frequencies of up to 10 kHz," J. Acoust. Soc. Am. **111**, 2466–2466.

Bernstein, L. R., and Trahiotis, C. (**2004**). "The apparent immunity of high-frequency "transposed" stimuli to low-frequency binaural interference," J. Acoust. Soc. Am. **116**, 3062–3069.

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

Colburn, H. S., and Esquissaud, P. (**1976**). "An auditory-nerve model for interaural time discrimination of high-

frequency complex stimuli," J. Acoust. Soc. Am. **59**, S23.
Dreyer, A. A., and Delgutte, B. (**2006**). "Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: Implications for sound localization," J. Neurophysiol. **96**, 2327–2341.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.
Ito, Y., Colburn, H. S., and Thompson, C. L. (**1982**). "Masked discrimination of interaural time delays with narrow-band signal," J. Acoust. Soc. Am. **72**, 1821–1826.
Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.
McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.
Nuetzel, J. M., and Hafter, E. R. (**1976**). "Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration," J. Acoust. Soc. Am. **60**, 1339–1345.
O'Loughlin, B. J., and Moore, B. C. J. (**1981**). "Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking," J. Acoust. Soc. Am. **69**, 1119–1125.
Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.
Palmer, A. R., and Evans, E. F. (**1982**). "Intensity coding in the auditory periphery of the cat: Responses of cochlear nerve and cochlear nucleus neurons to signals in the presence of bandstop masking noise," Hear. Res. **7**, 305–323.
Sek, A., and Moore, B. C. J. (**1994**). "The critical modulation frequency and its relationship to auditory filtering at low frequencies," J. Acoust. Soc. Am. **95**, 2606–2615.
Siebert, W. M. (**1968**). "Stimulus transformations in the peripheral auditory system," in *Recognizing Patterns*, edited by P. A. Kolers, and M. Eden (MIT Press, Cambridge, MA.), pp. 104–133.
Stellmack, M. A., and Dye, R. H. (**1993**). "The combination of interaural information across frequencies: The effects of number and spacing of the components, onset asynchrony, and harmonicity," J. Acoust. Soc. Am. **93**, 2933–2947.
Trahiotis, C., and Bernstein, L. R. (**1990**). "Detectability of interaural delays over select spectral regions: Effects of flanking noise," J. Acoust. Soc. Am. **87**, 810–813.
Van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.
Viemeister, N. F. (**1983**). "Auditory intensity discrimination at high frequencies in the presence of noise," Science **221**, 1206–1208.
Vliegen, J., and Oxenham, A. J. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am. **105**, 339–346.
Woods, W. S., and Colburn, H. S. (**1992**). "Test of a model of auditory object formation using intensity and interaural time difference discrimination," J. Acoust. Soc. Am. **91**, 2894–2902.