ISSN 1063-7710, Acoustical Physics, 2006, Vol. 52, No. 1, pp. 87–92. © Pleiades Publishing, Inc., 2006.
Published in Russian in Akusticheskiĭ Zhurnal, 2006, Vol. 52, No. 1, pp. 98–104.

# Effects of Interaural Decorrelation and Acoustic Spectrum on Detecting the Motion of an Auditory Target[1]

**K. Saberi* and A. Petrosyan**

*Department of Cognitive Sciences, University of California, Irvine, CA, 92697 United States*

**e-mail: saberi@uci.edu*

Received October 29, 2004

**Abstract**—The ability to detect the motion of an auditory target based on dynamic changes in interaural time differences was measured as a function of interaural correlation and acoustic spectrum in a single-interval forced-choice design. Three subjects listened to headphone-presented noise containing a dynamic linear change in interaural delay (500 μs/s). The stimulus spectrum was broadband (0.1 to 10 kHz), highpass above 1.5 kHz, or lowpass below 1.2 kHz, and interaural correlation ranged from 0.1 to 1.0. Subject performance was nearly identical for the broadband and lowpass conditions, with near perfect detection for interaural correlations of 0.5 or greater, and above threshold ($d' > 1$) detection for a correlation of 0.3. Performance was near random when the correlation was 0.1. In the highpass condition, performance rapidly deteriorated from substantially above threshold ($d' > 2$) to random level as interaural correlation was reduced from 1.0 to 0.7. This rapid decline in performance at high frequencies may be explained in terms of interaural envelope decorrelation.

PACS numbers: 43.80.+p

**DOI:** 10.1134/S106377100601012X

## INTRODUCTION

This is a brief report on the effects of interaural decorrelation and stimulus spectrum on the detection of auditory motion based on dynamic changes in interaural time differences (ITD). Naturally occurring sounds are often not perfectly correlated at the two ears due to a number of factors including extraneous signals [1, 2] and differential filtering effects of the pinna and reflections off the torso (head-related transfer functions [3–8]). In addition, the neural representations of sounds to the two ears are further decorrelated as a result of additive independent neural noise in the ascending auditory tracts prior to binaural convergence in brainstem nuclei [9–14].

While a number of prior studies have investigated how auditory motion detection is affected by such stimulus parameters as bandwidth [15, 16], velocity [17–23], acceleration [24], trajectory [18, 25, 26], and interaural cue [27, 28], none has examined the effects of interaural correlation. Investigating how stimulus correlation at the two ears affects processing of dynamic ITDs is important not only because natural signals are often dynamic and partly decorrelated, but also for what it can theoretically reveal about signal processing by the binaural system, which computes spatial position via neural cross-correlation [29–37].

Here we describe experiments on motion detection based on dynamic ITDs for broadband and filtered noise (highpass >1500 Hz and lowpass <1200 Hz) as a function of interaural correlation ($r$ = 0.1, 0.3, 0.5, 0.7, and 1.0). We show that, for broadband noise, detection remains above threshold ($d' > 1$) for correlations as low as 0.3, comparable to performance previously reported for lateralization of partially correlated stationary sounds [38, 39]. Furthermore, we show that nearly all the ability to detect dynamic ITDs of partially correlated waveforms is based on low-frequency cues. Even slight decorrelation (from 1.0 to 0.7) is sufficient to reduce detectability of dynamic ITDs in highpass noise from considerably above threshold ($d' > 2$) to near random performance ($d' < 0.4$). The implications of these findings for mechanisms of binaural interaction are discussed.

## METHOD

### *Stimulus Generation and Calibration*

All stimuli were generated digitally and presented via 16-bit digital-to-analog converters (Sound Blaster Live, –120 dB noise floor, Milpitas, CA). The sampling rate was 44.1 kHz, and the analog output was filtered through 20-kHz antialiasing filters. Stimulus generation and presentation was controlled via software running on a PC workstation. Stimulus levels were calibrated to 70 dB (A-weighting) using a 6cc coupler, 0.5" microphone (B&K, Model 4189), and a modular precision sound analyzer (B&K, Type 2260). The lowpass and highpass stimuli were ramped with 20-ms lin-

[1] The text was submitted by the authors in English.

ear rise–decay times to avoid spectral splatter. The ramps were simultaneous and synchronized at the two channels and thus provided no cues for resolving the task. The timing and levels between left and right channels were checked for accuracy using a dual-channel digital storage oscilloscope (Tektronix, Model TDS210). All calibrations were conducted in a double-walled steel acoustically isolated chamber (Industrial Acoustics Company).

Broadband stimuli were Gaussian noisebursts filtered between 0.1 and 10 kHz. The highpass and lowpass stimuli were also Gaussian, but filtered in the frequency domain with cutoffs of 1500 Hz (highpass) and 1200 Hz (lowpass). Generation of motion (dynamic ITD) stimuli was a two-step process, with both steps occurring in "real-time" between trials. In the first step, a dichotic Gaussian noise waveform was produced with a dynamic linear change in interaural delay, and in the second step, the noise waveforms to the two channels were partially decorrelated (when required) by the addition of independent Gaussian noise to each channel. This added noise had the same spectrum as the dynamic ITD stimulus.

To generate a noise waveform with a dynamic linear shift in interaural delay (i.e., motion), we first generated a Gaussian noise burst for one of the two audio channels in the frequency domain with amplitudes sampled from a Rayleigh distribution and phases from a uniform $(0, 2\pi)$ distribution. The spacing between frequency components $(\Delta\omega)$ of this waveform is dependent on duration $(\Delta\omega = 1/T)$, and the Discrete Fourier Transform (DFT) will only have the measured energy at the harmonics of $T$ [40]. To generate the stimulus for the second channel, we selected a duration $(T_2)$ for the waveform to channel 2 that produces a DFT array with component spacing $\Delta\omega_2$, such that each frequency component of channel 2 would be shifted relative to the corresponding component in channel 1 by a proportion required to produce a binaural-beat waveform with a linear shift in interaural delay. This array is then filled with complex numbers (amplitudes and phases) from channel 1. For example, a 0.1-s stimulus will have DFT component spacing of 10 Hz at {0, …, 500, 510, …, 1000, 1010, …} Hz. To produce a motion velocity of 2000 μs/s, channel 2 will require a frequency spacing of $\Delta\omega_2 = 10.02$ Hz, a duration $T_2 = 1/\Delta\omega_2 = 99.800399$ ms, and thus component frequencies of {0, …, 501, 511.02, …, 1002, 1012.02, …} Hz.[2] The inverse FFT of the stimuli at channels 1 and 2 provides the time waveforms for these channels. We have used this procedure to generate linear interaural-delay-based motion for a number of complex sounds, including noise waveforms, natural sounds, and speech sentences. The perceived image is of a smoothly moving intracranial image along the interaural axis (see [41] for details).

Even though the interaural delay of such waveforms is dynamic, we label this dichotic condition as having an interaural correlation of 1, consistent with binaural studies of stationary sounds [11, 38, 39] and because such a stimulus is perceived as a punctuate auditory image (for a motion velocity of zero, the measured correlation is unity). To produce stimuli with an interaural correlation of $r$, we added independent bursts of Gaussian noise to each channel [42]:

$$\begin{cases} X_L = kX_1 + X_{mL} \\ X_R = kX_2 + X_{mR}, \end{cases}$$

where $X_L$ and $X_R$ are the left and right channel waveforms, respectively, $k = \sqrt{\frac{1}{r} - 1}$ is a scalar for independent noise samples $X_1$ and $X_2$, and $X_{mL}$ and $X_{mR}$ are the left- and right-ear motion stimuli generated in the previous step. By definition, when $r$ is 0, two independent noise bursts are used. The final levels of $X_L$ and $X_R$ were calibrated to 70 dB SPL.

### *Procedure*

Three normal-hearing subjects participated in the experiments. Each subject practiced on the various experimental conditions for two hours prior to data collection. Subjects listened through Sennheizer HD-400 headphones in a double-walled acoustically isolated steel chamber (IAC). There were 15 experimental conditions and each subject completed two runs of 100-trials per condition. Five of the conditions were broadband, five were highpass, and five were lowpass conditions. The five conditions in each category corresponded to five interaural correlation values of 0.1, 0.3, 0.5, 0.7, and 1.0. The experiments were run in a block design with the correlation and spectral content held constant within a run. The order in which subjects ran the various conditions was completely randomized.

The experimental design was single-interval forced-choice, in which the subject had to indicate whether the sound moved or was stationary. On a given trial, either a single stationary stimulus (constant ITD) or a single motion stimulus (dynamic ITD) was presented with equal prior probabilities. The stimuli on stationary trials had the same spectral content as the dynamic stimulus for that run (i.e., if the motion condition was highpass, the stationary trials in that run contained highpass stimuli). When the trial was a motion trial, the direction

[2] The duration is rounded to the nearest digital sampling period, which for the current study is ~23 μs. Thus, the maximum quantization error in estimating duration is 11.5 μs, which for a 1-s stimulus (the duration used in the current experiments) results in a maximum frequency quantization error of less than 0.00002 Hz.

of motion (i.e., the direction of change in ITD) was randomly selected toward the right or left along the interaural axis. The motion stimulus had a duration of 1 s and a maximum change in interaural delay of 500 μs (i.e., "distance") and thus a velocity of 500 μs/s. This velocity was chosen because previous studies have shown that it produces a strong perception of intracranial motion as well as low motion-detection thresholds [23, 43]. A random offset interaural delay was used to reduce absolute position cues. The offset allowed the dynamic interaural delay to cover a 500-μs distance selected from the range of –500 to 500 μs. For example, the motion stimulus could have had a starting interaural delay of –300 μs and a terminating interaural delay of +200 μs, or a starting delay of +50 μs and an ending delay of –450 μs. If the trial was a "stationary" trial, a single 1-s stimulus was presented which had an interaural delay randomly selected from the range of –500 to 500 μs.

## RESULTS

Figure 1 shows results from all three bandwidth conditions. Data are from 3 subjects (symbols) and their mean performance (solid lines). The abscissa represents interaural correlation and the ordinate represents the index of detectability $d'$ calculated from hit and false-alarm rates associated with 200 trials per condition per subject. A hit (H) was defined as a "motion" response when the stimulus had a dynamic interaural delay, and a false-alarm (FA) was defined as a "motion" response when the stimulus interaural delay was constant (stationary). Motion detection was then calculated from $d' = z(H) - z(FA)$, where the $z$ transform is the inverse of the normal distribution function [44, 45]. A ceiling value of $d' = 3.29$ was imposed on performance by assuming a 2% inattention rate ($P(c) = 0.99$) to reduce estimation problems associated with the high variance of large $d'$ given limited sample sizes on which the $d'$ estimate is based [45–47].

As is evident from Fig. 1, all subjects performed well in the broadband and lowpass conditions when the interaural correlation was equal to or greater than 0.3. Even at a correlation value of 0.3, all subjects performed above threshold ($d' > 1$). Near-perfect performance is observed for correlation values of 0.5 and higher. It is clear, however, that although subjects can detect dynamic ITDs in highpass stimuli when the interaural correlation is 1.0, even a small reduction in correlation (from 1.0 to 0.7), which has no evident effect in the lowpass or broadband conditions, results in a complete failure of subjects to detect motion based on dynamic ITDs. The average $d'$ at a correlation of 0.7 for the highpass condition is 0.33. It thus appears that the ability to detect a dynamic ITD in partially correlated broadband stimuli is primarily determined by information contained in low-frequency bands. This result, however, should be qualified by the observation that subjects performed at ceiling levels in the lowpass and broadband conditions when the correlation was above 0.5; thus, a small drop in performance between correlations of 1.0 and 0.7 may not have been measurable given the excellent performance at these correlations. Nonetheless, subjects reported a very clear perception of motion for both the 1.0 and 0.7 correlations and no noticeable subjective interference of the slight decorrelation at $r = 0.7$ in the lowpass and broadband conditions. The dashed line in the lower panel of Fig. 1 shows average performance for the broadband motion condition (from top panel) and is replotted here to facilitate visual comparison. Note that the abilities to detect dynamic ITDs in broadband and lowpass filtered noise, as a function of interaural correlation, are nearly identical. A two-way repeated-measures analysis of variance on the data of Fig. 1 showed a statistically significant

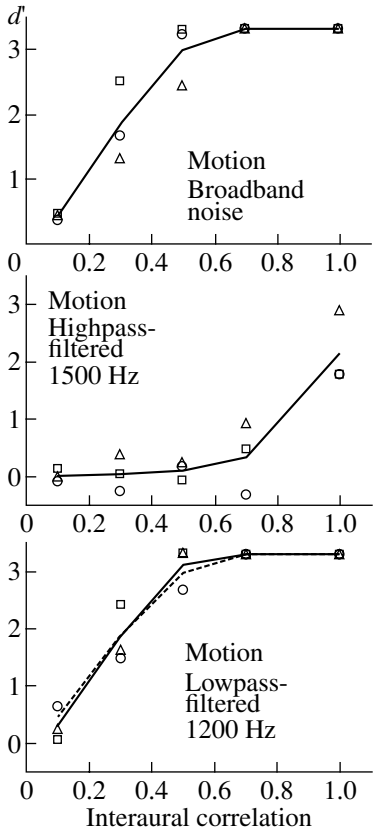


**Fig. 1.** Motion detection as a function of interaural correlation and bandwidth. Each panel shows results from one bandwidth condition. Abscissa represents interaural correlation and ordinate represents index of detectability $d'$. Different symbols represent different subjects and the solid lines show mean performance. The dashed line in the bottom panel is the same as the solid line in the top panel and is replotted to allow comparison between the broadband and lowpass conditions. Motion velocity was 500 μs/s.

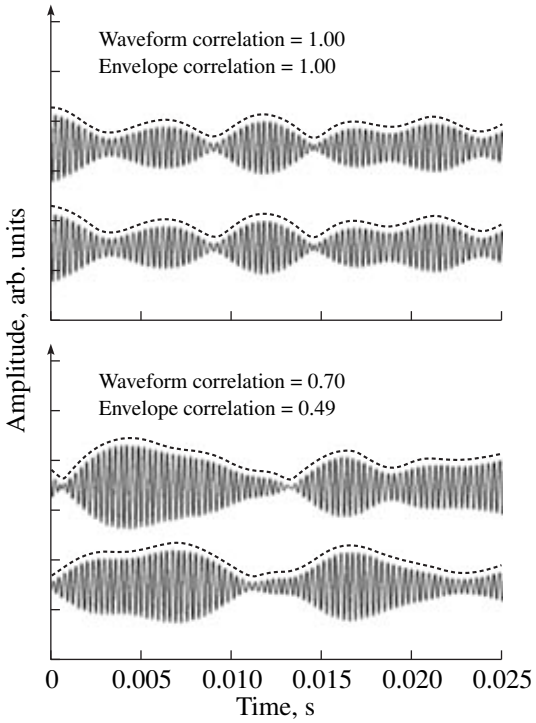


**Fig. 2.** Effects of waveform decorrelation on envelope correlation. Note from the bottom panel that when the waveform correlation is 0.7, the envelope correlation is considerably lower at 0.49. This additional reduction in envelope correlation may underlie the rapid drop in motion detection as a function of decorrelation for high-frequency stimuli.

effect of interaural correlation ($F_{4,8} = 71.36$, $p < 0.001$) a significant effect of bandwidth ($F_{2,4} = 59.85$, $p < 0.001$) and a significant interaction between bandwidth and correlation ($F_{8,16} = 14.32$, $p < 0.001$).[3]

## DISCUSSION

A number of previous studies have examined lateralization of stationary sounds for partially correlated dichotic noise stimuli. These studies have shown that interaural delay thresholds remain fairly high for correlation values as low as 0.2–0.3, consistent with our findings on the detection of dynamic ITDs for broadband and lowpass stimuli. In a classic study by [38], subjects were instructed to center the perceived position of an intracranial image by adjusting a noiseband's ITD using a keypad. The variance of position adjustments was then taken as a measure of the detectability of interaural delays. Performance remained relatively constant for correlation values from 1.0 to 0.2, but decreased precipitously for correlations less than 0.2. Neurophysiological studies have also shown results consistent with these observations. Optic tectum neurons in owls display identifiable spatial receptive fields for interaural correlations as low as 0.4 [11] and nucleus laminaris neurons as low as 0.2–0.4 [10]. Furthermore, owls also show behavioral sensitivity to ITDs even when interaural correlation is as low as 0.2–0.3 [11]. Although the owl's dominant frequency region for ITD coding is considerably higher than that for humans, these findings are, in principle, comparable, since ITD-sensitive neurons in the owl phase-lock to binaural time-difference cues in waveform carriers as high as 9 kHz. It is also worth noting that interaural decorrelation does not significantly affect interaural-level coding either neurophysiologically [48] or behaviorally [48, 49] even when interaural correlation is zero.

The current study also showed that the ability to detect dynamic ITDs in partially correlated waveforms is a low-frequency phenomenon, since even a small change from perfect correlation results in a failure to detect dynamic ITDs in high-frequency noisebands. This finding, that low-frequency ITD processing is relatively immune to decorrelation, is consistent with a recent neuroimaging study of human cortical activation in response to interaural decorrelation, which reported a nearly linear relationship between local cortical activation levels and interaural correlation in the primary auditory cortex located bilaterally at the lateral extent of Heschl's gyrus, a primarily low-frequency region of the auditory cortex. Activation patterns for correlation values near unity were more distinct than those near zero. Furthermore, activation levels associated with interaural correlations of 0.33 and 0.60 were distinguishable, but those associated with 0.33 and 0.00 were not distinguishable [50]. Since it is well documented that lateralization at high frequencies is based on envelope cues and not on fine-structure interaural delays [51–56], one may speculate from our findings that a reduction in interaural correlation at high frequencies will have a relatively large impact on waveform envelope correlation. Figure 2 shows that this is in fact the case. The top panel of this figure shows two samples of narrowband noise waveforms with a correlation of 1.0, and the bottom panel shows two noise samples with a waveform correlation of 0.7. All waveforms in this figure were bandpass-filtered between 3 and 3.3 kHz, roughly equivalent to the bandwidth of an auditory filter at this frequency range [57]. The envelopes, derived from the Hilbert Transform of the original waveforms, are shown as dashed curves and slightly displaced vertically to facilitate visual inspection. The correlations shown are the measured waveform and envelope correlations. Although only a small part of the waveform is shown in this figure, the actual correlations were measured for 1-s-duration waveforms at a sampling rate of 44.1 kHz. Note that when two waveforms have a correlation of 0.70, their envelopes are substantially less correlated (0.49). We have examined envelope correlations

[3] Parts of the broadband data have been used elsewhere in developing a model of motion detection.

for a number of waveform correlations and have found that the envelope correlation is always lower than waveform correlation, and is especially lower when waveform correlation is between 0.3 and 0.7.

In summary, we found that the auditory system can process dynamic ITDs in highly corrupted signals. Above-threshold performance is observed even when interaural correlation is reduced to as low as 0.3. This finding is in line with that reported for lateralization of partially correlated stationary sounds, as well as with findings from animal neurophysiology. Furthermore, we found that the ability to detect dynamic ITDs in partially correlated sounds is primarily a low-frequency phenomenon. Our analysis showed that decorrelation of high-frequency waveforms has a substantial effect on interaural envelope correlation, which likely partially underlies the rapid decline in performance as a function of decorrelation at high frequencies.

## ACKNOWLEDGMENTS

We thank the anonymous reviewer for helpful comments. This work was supported by grant no. BCS-0417984 from the National Science Foundation, United States.

## REFERENCES

1. W. A. Yost, R. H. Dye, and S. A. Sheft, Percept. Psychophys. **58**, 1026 (1996).
2. W. A. Yost, *The Cocktail Party Effect: 40 Years Later, in Binaural and Spatial Hearing in Real and Virtual Environments*, Ed. by R. Gilkey and T. Anderson (Lawrence Erlbaum Associates Publishers, Mahwah, 1997), Chap. 17, pp. 329–348.
3. Lord Rayleigh, Philos. Mag. **13** (6), 214 (1907).
4. R. A. Butler, *Influence of the External and Middle Ear on Auditory Discrimination. Handbook of Sensory Physiology*, Vol. V/2: *Auditory System*, Ed. by W. D. Keidel and W. D. Neff (Springer, Berlin, 1975), Chap. 6, pp. 247–260.
5. R. A. Butler, J. Acoust. Soc. Am. **61**, 1264 (1977).
6. E. A. G. Shaw, J. Acoust. Soc. Am. **39**, 465 (1965).
7. E. A. G. Shaw, J. Acoust. Soc. Am. **55**, 1848 (1974).
8. G. F. Kuhn, in *Directional Hearing*, Ed. by W. A. Yost and G. Gourevitch (Springer, New York, 1987), pp. 3–25.
9. K. J. Diercks and L. A. Jeffress, J. Acoust. Soc. Am. **34**, 981 (1962).
10. Y. Albeck and M. Konishi, J. Neurophysiol., **74**, 1689 (1995).
11. K. Saberi, Y. Takahashi, M. Konishi, *et al.*, Neuron, No. 4, 789 (1998).
12. T. C. T. Yin and J. C. K. Chan, J. Neurophysiol., No. 2, 465 (1990).
13. C. E. Carr and M. Konishi, Proc. Natl. Acad. Sci. USA **85** (21), 8311 (1988).
14. C. E. Carr and M. Konishi, J. Neurosci. **10**, 3227 (1990).
15. D. W. Chandler and D. W. Grantham, J. Acoust. Soc. Am. **91**, 1624 (1992).
16. T. Z. Strybel, and M. L. Menges, Perception **27**, 483 (1998).
17. J. A. Airman and D. V. Viskov, J. Acoust. Soc. Am. **61**, 816 (1977).
18. M. Yu. Agaeva and A. Ya. Al'tman, Akust. Zh. **50**, 337 (2004) [Acoust. Phys. **50**, 278 (2004)].
19. J. A. Airman and I. G. Andreeva, J. Audiol. **43**, 227 (2004).
20. W. Waugh, T. Z. Strybel, and D. R. Perrott, J. Aud. Res. **19**, 103 (1979).
21. D. R. Perrott and J. Tucker, J. Acoust. Soc. Am. **83**, 1522 (1988).
22. T. Z. Strybel, S. A. Span, and A. M. Witty, Percept. Psychophys. **60**, 1441 (1998).
23. K. Saberi, P. Tirtabudi, A. Petrosyan, *et al.*, Acust. Acta Acust. **89**, 333 (2003).
24. D. R. Perrott, B. Costantino, and J. Ball, J. Acoust. Soc. Am. **93**, 1053 (1993).
25. K. Saberi and D. R. Perrott, J. Acoust. Soc. Am. **88**, 2639 (1990).
26. D. W. Grantham, B. W. Y. Hornsby, and E. A. Erpenbeck, J. Acoust. Soc. Am. **114**, 1009 (2003).
27. D. W. Grantham and F. L. Wightman, J. Acoust. Soc. Am. **63**, 511 (1978).
28. J. A. Altaian, O. V. Variaguina, N. I. Nikitin, and E. A. Radionova, J. Acoust. Soc. Am. **105**, 366 (1999).
29. L. A. Jeffress, J. Comp. Physiol. Psychol. **41**, 35 (1948).
30. B. Sayers and E. Cherry, J. Acoust. Soc. Am. **29**, 973 (1957).
31. J. Blauert and W. Cobben, Acustica **39**, 96 (1978).
32. K. Saberi, J. Acoust. Soc. Am. **98**, 3146 (1995).
33. K. Saberi, Psychol. Rev. **103**, 137 (1996).
34. P. Zurek and K. Saberi, Percept. Psychophys. **65**, 95 (2003).
35. K. Saberi, T. Takahashi, R. Egnor, *et al.*, J. Assoc. Res. Otolaryngol. **3** (1), 80 (2002).
36. K. Saberi, J. V. Antonio, and A. Petrosyan, Hear. Res. **191** (1–2), 1 (2004).
37. K. Saberi and A. Petrosyan, J. Theor. Biol. **235** (1), 45 (2005).
38. L. A. Jeffress, B. H. Deatherage, and H. C. Blodgett, J. Acoust. Soc. Am. **34**, 1122 (1962).
39. D. E. Robinson and L. A. Jeffress, J. Acoust. Soc. Am. **35**, 1947 (1963).
40. L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, 1975).
41. K. Saberi, Behav. Res. Methods Instrum. Comput. **36** (4), 585 (2004).
42. L. A. Jeffress and D. E. Robinson, J. Acoust. Soc. Am. **34**, 1658 (1962).
43. K. Saberi, P. Tirtabudi, A. Petrosyan, *et al.*, Hear. Res. **174** (1–2), 149 (2002).

44. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
45. N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge, 1991).
46. K. Saberi and D. M. Green, Percept. Psychophys. **59**, 867 (1997).
47. A. B. Watson and D. G. Pelli, Percept. Psychophys. **33**, 113 (1971).
48. S. E. R. Egnor, J. Comp. Physiol. A **187** (8), 589 (2001).
49. W. M. Hartmann and Z. A. Constan, J. Acoust. Soc. Am. **112**, 1037 (2002).
50. T. W. Budd, D. A. Hall, M. S. Goncalves, *et al.*, Neuroimage **20**, 1783 (2003).
51. G. B. Henning, J. Acoust. Soc. Am. **55**, 84 (1974).
52. G. B. Henning, J. Acoust. Soc. Am. **68**, 446 (1980).
53. J. M. Nuetzel and E. R. Hafter, J. Acoust. Soc. Am. **60**, 1339 (1976).
54. J. M. Nuetzel and E. R. Hafter, J. Acoust. Soc. Am. **69**, 1112 (1981).
55. W. A. Yost and G. Gourevitch, *Directional Hearing* (Springer, New York, 1987).
56. W. A. Yost and E. R. Hafter, *Lateralization in Directional Hearing*, Ed. by W. A. Yost and G. Gourevitch (Springer, New York, 1987), Chap. 3, pp. 49–84.
57. B. R. Glasberg and B. C. J. Moore, Hear. Res. **47** (1–2), 103 (1990).