Assessing the processing of interaural temporal disparities within high frequency stimuli via manipulations of the temporal signatures of their envelopes

Leslie R. Bernstein and Constantine Trahiotis

Departments of Neuroscience and Surgery (Otolaryngology), University of Connecticut Health Center, Farmington, CT, USA

During the past several years, we have investigated the processing of interaural temporal disparities (ITDs) conveyed within high-frequency auditory channels. Historically, ITD-processing at high frequencies has been found to be less efficient than that measured at low frequencies. Using “transposed” stimuli, we have reported that ITD-processing at high frequencies can be enhanced in terms of resolution of ITDs, extents of ITD-based laterality and resistance to the binaural interference found with conventional high-frequency stimuli. Notably and of theoretical import, transposed stimuli provide envelope-based binaural information within high-frequency channels similar to that provided by the waveform within low-frequency channels. More recently, we have utilized “raised-sine” high-frequency stimuli to investigate which particular features of the envelopes of high-frequency waveforms foster enhanced ITD processing. Such raised-sine stimuli permit independent variation of the modulation frequency, modulation depth, and “dead-time/relative peakedness” of the envelope of a high-frequency waveform, while also suitably restricting its spectral content. It will be seen that an interaural correlation-based model including stages mimicking peripheral auditory processing can explain much of the patterning of the results.

INTRODUCTION

One of the fundamental historical tenets concerning the localization of sources of sound is the “duplex theory” (Rayleigh, 1907). The theory arose from studies employing pure-tone stimuli and holds that the localization of low-frequency sounds is based upon interaural time (or phase) differences (ITDs) while the localization of high-frequency sounds is based upon interaural intensitive differences (IIDs). This general statement was incorrectly overgeneralized to mean that ITDs could not be appreciated or processed via the high-frequency channels of the auditory system. Decades later, temporally and spectrally complex high-frequency stimuli, such as sinusoidally amplitude modulated (SAM) and bands of noise were employed in experiments measuring resolution of ITDs and extents of laterality. Many such studies demonstrated that the envelopes of complex, high-frequency stimuli could convey useful ITD-information. High-frequency-envelope-based ITDs were, however, found to be much less “potent” than those conveyed by the fine-structure of low-frequency

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

waveforms. The relatively poorer processing of ITDs at high frequencies had been observed in two different types of experiments. First, high-frequency threshold-ITDs were found to be larger (e.g., Klumpp and Eady, 1956; Zwislocki and Feldman, 1956; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1976; Henning, 1980; Bernstein and Trahiotis, 1982; 1994; Blauert, 1983) and, second, high-frequency ITD-based extents of laterality were found to be smaller (e.g., Blauert, 1983; Bernstein and Trahiotis, 1985). Those outcomes were almost universally interpreted to mean that there was something intrinsically lacking or "wrong" with the high-frequency channels themselves. That interpretation also meshed well with Buell and Hafter's (1991) position that the integration of ITD information across frequency gives greater weight to the low-frequency channels which they characterized as being inherently less "noisy" than the high-frequency channels. Their characterization stemmed directly from the fact that low-frequency threshold-ITDs were typically substantially smaller than those measured at high frequencies.

A very different interpretation for the disparities observed between low-frequency and high-frequency ITDs was offered by Colburn and Esquissaud (1976). They postulated that the disparities reflected differences in the neural information that serves as input to the low-frequency and high-frequency binaural portions of the auditory system. Specifically, they suggested that frequency-related differences in sensitivity to ongoing ITDs could result from the rectification and low-pass filtering that occurs as a natural part of monaural, peripheral processing. For low frequency stimuli, such processing would result in neural impulses synchronized to the whole waveform (i.e., both the fine-structure and the envelope). For high-frequency stimuli, such processing would result in neural impulses synchronized to only the envelope of the waveform. This point of view suggests that, given a suitable choice of stimuli, the high-frequency channels would not be found to be inherently inefficient conveyers of ITD information. It also holds open the possibility that envelope-based ITD-processing in high-frequency channels could be enhanced and, perhaps, be made to be as efficient as that found for low-frequency channels.

In fact, one could characterize much of the research conducted in our laboratory during the last decade as being directed toward elucidating how best to understand ITD-processing in high-frequency channels. We believe that such an understanding is both theoretically and practically of great importance and is fundamental to an understanding of binaural processing in general. In addition, having such knowledge promises to influence the design and utility of prosthetic devices, especially with regard to improvements in the processing of envelope-based information conveyed by high-frequency channels.

In order to begin to address these issues, we conducted experiments using "transposed" stimuli that were created with the goal of providing high-frequency channels of the binaural processor with envelope-based inputs that, other things being equal, would essentially mimic waveform-based inputs normally available only in low-frequency channels. Such stimuli were generated by capitalizing on a technique originally described by van de Par and Kohlrausch (1997) which involves multiplying halfwave rectified, low-pass filtered versions of low-frequency stimuli by high-frequency sinusoidal carriers.

Our first experiments (Bernstein and Trahiotis, 2002) demonstrated that threshold-ITDs obtained with transposed tones centered at 4 kHz were smaller than those obtained with conventional SAM tones and, in some cases, were equal to or smaller than threshold-ITDs obtained with low-frequency pure tones (see Fig. 1a).

Large enhancements in sensitivity to ITD were even observed for transposed stimuli centered at 6 kHz and 10 kHz, center frequencies for which ITD-processing with conventional stimuli is usually found to be extremely poor. Quantitative analyses revealed that threshold-ITDs obtained with both conventional and transposed stimuli could be well accounted for via an interaural-correlation-based model incorporating an initial stage of gammatone-based bandpass filtering at 4 kHz (see Patterson *et al.*, 1995), "envelope compression" (exponent = 0.23), square-law rectification, and low-pass filtering at 425 Hz to capture the loss of neural synchrony to the fine-structure of the stimuli that occurs as the center frequency is increased (Weiss and Rose, 1988).

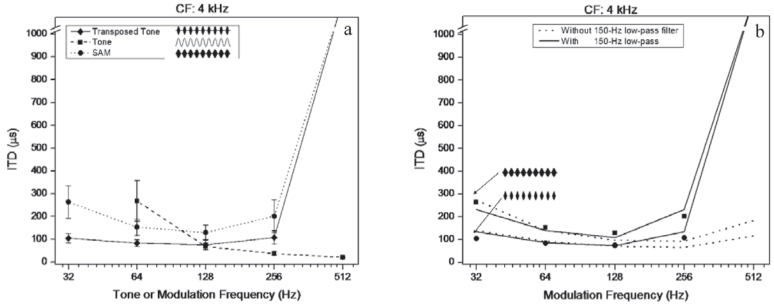

Fig. 1: Panel a: Threshold ITDs averaged across four listeners as a function of the modulation or pure-tone frequency. The center frequency of the high-frequency SAM and transposed stimuli was 4 kHz. The parameter of the plot is the type of stimulus employed. The error bars represent ± one standard error of the mean. The "broken" ordinate and "broken" lines through the data indicate conditions for which average threshold ITDs could not be computed because, for a subset of the listeners, thresholds could not be determined even for ITDs of up to 1 ms. Panel b: Threshold ITDs for the SAM (squares) and transposed (circles) stimuli. The dotted lines represent predictions based on a constant criterion change in the normalized correlation computed subsequent to compression, rectification, and low-pass filtering at 425 Hz. The solid lines represent predictions obtained when the peripheral processing was supplemented by an additional 150-Hz low-pass filter. (Adapted from Bernstein and Trahiotis, 2002.)

In order to account, quantitatively, for data obtained with high modulation frequencies (i.e., those above about 150 Hz), it proved necessary to include within the model a special stage of low-pass filtering of the envelopes within the high-frequency channels of each ear. Those channels serve as inputs to the binaural processor. As evident in Fig. 1b, the model, when augmented with the 150-Hz low-pass filter, was even highly accurate (accounting for 86% of the variance) and even predicted the stimuli for which listeners would be unable to perform the task.

Next, we (Bernstein and Trahiotis, 2003) found that ITD-based extents of laterality obtained with transposed stimuli were much greater than those obtained with bands of high-frequency Gaussian noise. Furthermore, low-frequency Gaussian noises centered at 125 Hz and their high-frequency transposed counterparts produced extents of laterality that were highly similar and, in many cases, essentially equivalent. That is, we found that ongoing ITDs conveyed by the envelopes of relatively narrow band high-frequency stimuli could produce extents of laterality like those measured with ITDs conveyed by the fine-structure of low-frequency stimuli. The overall patterning of the data was fairly well accounted for by a cross-correlation based model. Figure 2 is an example of our findings when the bandwidth 125-Hz-centered Gaussian noise and that of the 4-kHz-centered Gaussian was 25 Hz.

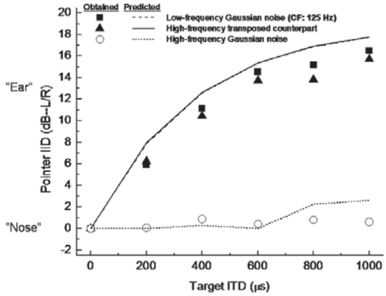

Fig. 2: IID of an acoustic "pointer" (in dB) required to match the intracranial position of the target as a function of the ITD (left ear leading) of the target. The data points represent the mean values computed across four listeners. Positive values along the ordinate indicate IIDs favoring the left (leading) ear. The parameter within the plot is the type of stimulus employed. Squares represent data obtained with a 25-Hz-wide band of Gaussian noise centered at 125 Hz; triangles represent data obtained when that low-frequency noise was transposed to 4 kHz; circles represent data obtained with a 25-Hz-wide band of Gaussian noise centered at 4 kHz. The lines represent predictions from the cross-correlation-based model described in the text. (Adapted from Bernstein and Trahiotis, 2003.)

What surprised us at first was the model's correct prediction that 25-Hz-wide bands of conventional Gaussian noise centered at 4 kHz would be heard at midline even when they carried an ITD as large as 1 ms. Analyses revealed that the stage of peripheral compression included within the model was responsible. When compression was omitted, the lateral position of a 25-Hz-wide band of high-frequency Gaussian noise was incorrectly predicted to be perceived far toward the leading ear.

In summary, our findings with transposed stimuli suggest that high-frequency "central" binaural channels are not inherently inferior to low-frequency channels vis-a-vis the processing of ITDs. In addition, the findings make clear that ITD-processing can only be understood and improved by consideration of the stimuli *as processed* by the peripheral auditory system, be it normal or abnormal, such as is the case for listeners dependent on of hearing-aids and/or cochlear implants.

Raised sine stimuli

Knowing that transposed stimuli yield enhanced ITD-processing, we began a series of experiments (see Bernstein and Trahiotis, 2009) with the general goal of determining which aspects of the envelopes of high-frequency temporal waveforms are sufficient for the processing of ITDs to be enhanced as compared to that measured with conventional stimuli (e.g., SAM tones and bands of Gaussian noise). We capitalized on a method described recently by John *et al.* (2002) which allows one to vary, independently, the frequency of modulation, the depth of modulation, and the exponent of the raised sine (the exponent affects the "peakedness" or "sharpness" of the envelope). In addition, the use of raised-sine stimuli allows one to generate high-frequency signals having envelopes with temporal features that "fall in between" those of SAM tones and those of transposed stimuli while having spectral content restricted to a relatively narrow range.

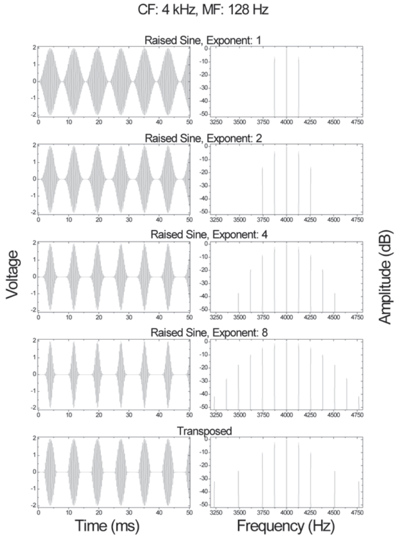

Fig. 3: Raised sine and transposed stimuli. (From Bernstein and Trahiotis, 2009.)

The generation of raised-sine stimuli entails raising a DC-shifted sine-wave to a power greater than or equal to 1.0 prior to multiplication with a carrier. The equation used to generate such stimuli is:

$$y(t) = (\sin(2\pi f_c t))(2m(((1+\sin(2\pi f_m t))/2)^n - 0.5)+1) \qquad \text{(Eq. 1)}$$

where f_c is the frequency of the carrier, f_m is the frequency of the modulator, m is the modulation index, and n is the exponent denoting the power to which the DC-shifted modulator is raised.

The left side of Fig. 3 depicts the time-waveforms for cases in which a 128-Hz modulating tone was raised using exponents of 1, 2, 4, or 8 prior to multiplication with a 4-kHz carrier. In all cases, m=1.0. The bottom row of the figure depicts a 128-Hz tone transposed to 4 kHz. Note that an exponent of 1.0 yields a conventional SAM waveform. Examination of the figure reveals that the peakedness or sharpness of the envelope increases directly with the value of the exponent to which the modulator is raised. Simultaneously, for these 100%-modulated signals, the “dead-time” or “off-time” between individual lobes of the envelope also increases with increasing values of the exponent. The right side of the figure displays the long-term spectrum of each stimulus. Note that, increasing the value of the exponent also increases the number of “sidebands” and their spectral extent. It is important to note that, for each of the stimuli depicted, the vast majority of its energy falls within the approximately 500-Hz wide auditory filter centered at 4 kHz (see Moore, 1997).

The data in Fig. 4 represent mean “normalized” threshold ITDs, calculated across four listeners, for 4-kHz-centered raised sine stimuli. The rate of modulation was varied parametrically with variations in the exponent of the raised sines. Normalized thresholds were calculated in order to remove differences in absolute sensitivity to ITD across listeners commonly found with high-frequency, complex stimuli. The normalization was accomplished by dividing an individual listener’s threshold ITDs

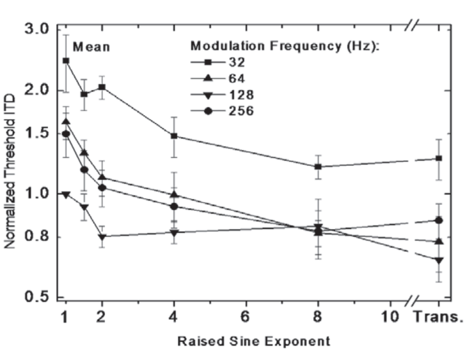

Fig: 4. Mean normalized threshold ITDs, calculated across four listeners as a function of the raised-sine exponent. Normalized threshold ITDs (see text) obtained with the transposed stimuli are plotted at the far right. The parameter of the plot is the frequency of modulation. Error bars represent ± one standard error of the mean normalized thresholds. (From Bernstein and Trahiotis, 2009.)

by that listener's threshold ITD obtained with a SAM tone (raised-sine exponent equal to 1.0) having a frequency of modulation of 128 Hz. The error bars represent ± one standard error of the mean.

Note that, for all rates of modulation, threshold ITDs decrease with increases in the exponent of the raised-sine and approximate threshold ITDs obtained with transposed stimuli (plotted at the far-right in Fig. 4) when the exponent is 8.0. In general, the data show that graded changes in the exponent (affecting the peakedness of the envelope) lead to graded changes in sensitivity to ITD. Also note that, threshold ITDs decrease with increases in rate of modulation from 32 to 128 Hz and then increase slightly when the rate of modulation is increased to 256 Hz. The same trend is observed when frequency of modulation of SAM and transposed tones is varied over the same range (see Fig. 1).

Let us now use the data in Fig. 4 to illustrate how we evaluate which aspects of the stimuli as processed by the auditory periphery determine sensitivity to ITDs. Here, we consider two candidate metrics: "peakedness" of the envelope and interaural correlation of the envelope. As mentioned earlier, the peakedness of the envelope varies directly with the magnitude of the exponent of the raised sine. It is also the case that the peakedness of the envelope varies directly with the rate of modulation of the raised sine. For all of the stimuli represented in Fig. 3, we computed the peakedness of the envelope after passing the stimuli through the "peripheral" stages of our model. The "peakwidth" was defined as the time during which the magnitude of an individual lobe of the envelope was 80% or more of its peak value (referred to as the "80% peakwidth")

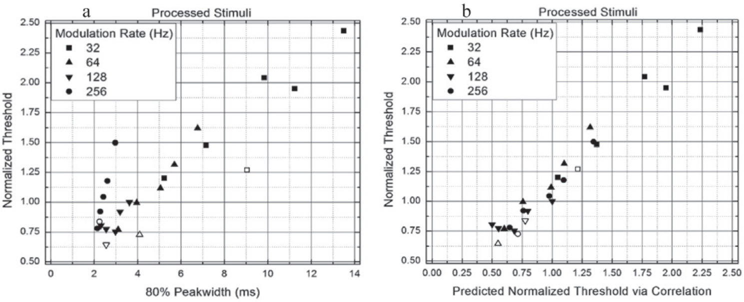

Fig. 5: Panel a: Normalized threshold ITDs averaged across four listeners as a function of the 80% peakwidth of the envelope for all of the stimuli in Fig. 4. Peakwidth was computed after passing the stimuli through the "peripheral" stages of our model (see text). The parameter of the plot is the rate of modulation. For *each* rate of modulation, the five solid symbols represent data obtained with the five values of raised sine exponent and the open symbol represents data obtained with a transposed tone. Panel b: Same as panel a, except that the data are plotted as a function of the normalized threshold ITD necessary to reduce the normalized interaural correlation by a constant criterion value.

Figure 5a plots normalized threshold against the 80% peakwidth for all of the stimuli in Fig. 4. The parameter of the plot is the rate of modulation. For *each* rate of modulation, the five solid symbols represent data obtained with the five values of raised sine exponent and the open symbol represents data obtained with a transposed tone. Note that, for raised sine stimuli of a given rate of modulation, threshold ITDs increase with increases in peakwidth. Nevertheless, across the entire set of stimuli, peakwidth, per se, is not a good predictor of threshold ITD. That is, given values of peakwidth do not necessarily produce similar, let alone essentially constant, values of threshold ITD.

In order to evaluate the predictive power of interaural correlation for the same set of data, we calculated, via the model, the ITD necessary to reduce the normalized interaural correlation by a constant criterion value. Those predictions are plotted in Fig 5b and clearly indicate that threshold ITDs accurately predicted independent of rate of modulation, value of raised sine exponent and type of stimulus (raised sine or transposed sine).

SUMMARY AND CONCLUSION

This short overview has highlighted two important types of findings concerning the processing of ITDs. First, it appears that the almost ubiquitous finding that sensitivity to ITDs at high-frequencies is poorer than that observed at low frequencies does not reflect a relative deficit of central, high-frequency, binaural channels, per se. Rather, that finding appears to be a manifestation of the interaction between conventional auditory stimuli (e.g., SAM tones and bands of noise) and peripheral auditory mechanisms. As shown, when ITDs are conveyed by transposed tones and raised-sine stimuli, sensitivity to ITD for high-frequency stimuli can approach and even rival that typically observed with low-frequency stimuli.

Second, these new findings can be explained quite well via the same general interaural-correlation-based model that has proven to be successful in a variety of binaural detection and discrimination experiments employing conventional low- and high-frequency stimuli (e.g., Bernstein and Trahiotis, 1996; 2002; 2003, Bernstein *et al.*, 1999). More specifically, the differential effects on ITD-processing produced by alterations of the temporal signatures of the envelopes that conveyed the ITDs in these experiments can be understood by considering the interaural correlation of the stimuli, *as processed*.

These successes notwithstanding, it must be stressed that our research has been conducted with normal-hearing listeners. At an empirical level, it is not clear whether and to what degree similar improvements in the processing of ITDs can be found with hearing-impaired listeners and/or listeners relying on cochlear implants. At a theoretical level, what constitutes “as processed” for the modeling of data collected re the processing of ITDs for such populations will be very different and, perhaps, unique to the individual patient.

More optimistically, it may be the case that this kind of information will be useful to those who seek to improve and to understand envelope-based cues presented via cochlear implants. For example, Majdak *et al.* (2006) lament the fact that many bilateral implants provide unsynchronized fine-structure-based pulses across the two ears. This leaves only the envelope as a viable vehicle for conveying ITDs. Thus, it seems essential to find effective transformations of the stimuli such that information can be efficiently conveyed to presumably normal central auditory system.

ACKNOWLEDGMENTS

Portions of this chapter are adapted from some of our prior work. References are included to facilitate the reader's ability to find more detail concerning of the several topics covered. This research was supported by research grants NIH DC-04147 and DC-04073 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

REFERENCES

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). "The normalized correlation: Accounting for NoSπ thresholds obtained with Gaussian and "low-noise" masking noise," J. Acoust. Soc. Am. **106**, 870-876.

Bernstein, L. R., and Trahiotis, C. (**1982**). "Detection of interaural delay in high-frequency noise," J. Acoust. Soc. Am. **71**, 147-152.

Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of sinusoidally-amplitude-modulated tones: Effects of spectral locus and temporal variation," J. Acoust. Soc. Am. **78**, 514-523.

Bernstein, L. R., and Trahiotis, C. (**1994**). "Detection of interaural delay in high-frequency SAM tones, two-tone complexes, and bands of noise," J. Acoust. Soc. Am. **95**, 3561-3567.

Bernstein, L. R., and Trahiotis, C. (**1996**). "The normalized correlation: Accounting for binaural detection across center frequency," J. Acoust. Soc. Am. **100**, 3774-3784.

Bernstein, L. R., and Trahiotis, C. (**2002**). "Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli," J. Acoust. Soc. Am. **112**, 1026-1036.

Bernstein, L. R., and Trahiotis, C. (**2003**). "Enhancing interaural-delay-based extents of laterality at high frequencies by using 'transposed stimuli'," J. Acoust. Soc. Am. **113**, 3335-3347.

Bernstein, L. R., and Trahiotis, C. (**2009**). "How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli," J. Acoust. Soc. Am. **125**, 3234-3242.

Blauert, J. (1983). *Spatial Hearing,* (MIT Press).

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** Suppl. 1, S23.

Henning, G. B. (**1980**). "Some observations on the lateralization of complex waveforms," J. Acoust. Soc. Am. **68**, 446-453.

John, M. S., Dimitrijevic, A., and Picton, T. (**2002**). "Auditory steady-state responses to exponential modulation envelopes," Ear and Hearing **23**, 106-117.

Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859-860.

Majdak, P, Bernhard, L., and Wolf-Dieter, B. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing'," J. Acoust. Soc. Am. **120**, 2190-2201.

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. (**1997**). "Frequency analysis and pitch perception," in *Handbook of Acoustics,* edited by Malcome Crocker (John Wiley and Sons, New York), V. III, pp. 1447-1460.

Nuetzel, J. M. and Hafter, E. R. (**1976**). "Lateralization of complex waveforms: Effects of fine-structure, amplitude, and duration," J. Acoust. Soc. Am. **60**, 1339-1346.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). "Time domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc Am. **98**, 1890 1894.

Rayleigh (**1907**). "On our perception of sound direction," Philos. Mag. **13**: 214-232.

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.

Weiss, T. F., and Rose, C. (**1988**). "A comparison of synchronization filters in different auditory receptor organs," Hear. Res. **33**, 175-180.

Zwislocki, J., and Feldman, R. S. (**1956**). "Just noticeable differences in dichotic phase," J. Acoust. Soc Am. **28**, 860-864.