The role of envelope waveform in the processing of high-frequency interaural timing disparities

Mathias Dietz, Martin Klein-Hennig, Volker Hohmann, and Stephan D. Ewert
Medical Physics, Department of Physics, University of Oldenburg, D-26129 Oldenburg, Germany

Interaural timing disparities provide an important cue for lateralization in the human auditory system. The auditory system is sensitive to timing disparities in the fine-structure and the envelope of sounds. In the normal-hearing system, only envelope disparities can be exploited at high frequencies due to the lack of phase-locking to the fine-structure. Similarly, in cochlea implant users, interaural timing disparities of the envelope are of importance. It is, however, still unclear which specific envelope waveform properties promote the most stable features for lateralization. In this study, psychophysical measurements were conducted with customized envelope waveforms in order to investigate the isolated effect of attack and decay times, as well as pause and hold durations in the envelope waveform on lateralization. For high-frequency tones centred at 4 kHz with systematic envelope modifications, the just noticeable differences of ongoing interaural time differences in the envelope were measured. The results indicate that attack times and pause durations prior to the attack are the most important envelope features. The results are compared to predictions of binaural auditory models with different adaptation mechanisms prior to their binaural stage. Consequences of the different adaptation mechanism for monaural processing and for processing of cochlea-implant-like stimulation are discussed.

INTRODUCTION

Interaural time differences are an important binaural cue for the human auditory system, helping to localize sound sources in space. At low frequencies, the fine structure of a sound is preserved and interaural time difference (ITD) can be used for a variety of stimuli including pure tones. At high frequencies, above about 1500 Hz, however, the lack of phase-locking causes the effective extraction of the stimulus' envelope in the auditory system. The envelope can convey binaural information in the form of onsets (and offsets) of a sound (referred to as "onset" or "transient" interaural delay, e.g., McFadden and Pasanen, 1976) and in the form of "ongoing" ITDs which occur for sounds with time-varying temporal envelopes. In the latter case, ITD sensitivity on the basis of interaural delays of the envelope is observed (e.g., Klumpp and Eady, 1956; Henning, 1974; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1976, 1981).

A number of recent studies have demonstrated that the processing of ongoing envelope ITDs for high-frequency stimuli depends on the envelope waveform and

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.

the degree of fluctuations in the envelope. Certain envelope waveforms as produced by "transposed" stimuli (van de Par and Kohlrausch, 1997) which provide high-frequency auditory channels with envelope-based information mimicking the fine-structure based information typically available in low-frequency channels, show enhanced processing of the interaural disparities in the high-frequency region (e.g., van de Par and Kohlrausch, 1997; Bernstein, 2001; Bernstein and Trahiotis, 2002). Earlier, the degree to which the envelopes of the stimuli fluctuate was suggested to describe binaural processing performance of ITDs conveyed by high-frequency stimuli (Dye Niemiec and Stellmack, 1994). They quantified the degree of envelope fluctuations by the normalized fourth moment of the envelope (Hartmann, 1987; Hartmann and Pumplin, 1988). Bernstein and Trahiotis (2007), however, demonstrated that for transposed stimuli the "internal" interaural envelope correlation accounts for binaural processing performance rather than the normalized fourth moment of the envelope as a measure of the degree of envelope fluctuations.

For arbitrary high-frequency stimuli, it still remains an open question which specific envelope features determine binaural processing performance for ongoing envelope ITDs. The use of analytically describable features such as overall level, sine exponent, modulation rate or modulation depth have the drawback that a modification necessarily causes a co-variation of the "fundamental" parameters like, e.g., attack and decay times or pause and hold durations (off-/on-times). These fundamental parameters appear to also be potentially suited features for the characterization of binaural auditory function and may additionally close the gap between onsets (and offsets) and ongoing temporal disparities in a generalized view of envelope features.

In the present study, psychoacoustic measurements were conducted employing 4-kHz tones with systematic modifications to their envelope waveform. The fundamental parameters of the envelope were varied independently and the just noticeable difference (JND) in the interaural time difference was measured for each condition.

Simulations were performed with different model approaches: The normalized cross-correlation coefficient (NCC, Bernstein and Trahiotis, 2002, 2007) and three models that determine the maximal difference between the left and right channel after different stages of neural adaptation.

EXPERIMENTS

Method

Five normal-hearing listeners ranging in age from 25-29 years participated in the experiments. Two of the subjects were the authors MD and MKH. All subjects received several hours of listening experience prior to the final data collection.

Subjects were seated in a double-walled sound attenuating booth in front of a computer keyboard and monitor. Subjects listened via Sennheiser HD580 headphones driven by a Tucker Davis HB7 headphone buffer. Signal generation and presentation during the experiments were computer controlled using the AFC software package for MATLAB, developed at the University of Oldenburg. The stimuli were digitally generated at a sampling rate of 48 kHz and converted to analog signals by a high-quality 24-bit sound card and external digital-to-analog converter (RME DIGI96/8 PAD, RME ADI-8 PRO).

Amplitude modulated 4-kHz pure tone stimuli were used to measure the just noticeable ITD. The envelopes were periodic at a modulation rate of 35-100 Hz. The lower waveform in Panel (a) of Fig. 1 represents a single period of the envelopes used in the present study. Each envelope period was comprised of a pause segment, an attack segment (raised-cosine ramp), a hold segment, and a decay segment (raised-cosine ramp). The durations of the segments were varied independently to study their individual importance. A sinusoidal amplitude modulation (SAM) resulted for pause and hold duration of zero ms. Otherwise, a "square-wave" modulation (SWM) with variable duty cycle and attack/decay times resulted. The shortest attack/decay durations used in this study were 1.25 ms to control for spectral broadening of the stimuli. The ITD was applied to the envelope waveform as a whole (indicated by the dotted line) or to the attack only as indicated by the dashed line in Fig. 1a. Control experiments with ITDs in the decay segment only indicated that the ITD of the decay had virtually no influence on the lateralization. Stimuli were 500 ms in duration and were gated with 125-ms raised-cosine ramps to minimize the salience of onset cues. A low-pass filtered noise (5th-order Butterworth at 1000 Hz, white spectrum up to 200 Hz, -3 dB per octave slope up to 1000 Hz) was simultaneously presented to preclude the listeners' use of any information at low frequencies (e.g., Nuetzel and Hafter, 1976, 1981; Bernstein and Trahiotis, 2007). The low-pass noise was 600 ms in duration resulting in 50 ms temporal fringes during which it was gated with 50-ms raised-cosine ramps and had a level of 45 dB SPL. The level of the SAM stimuli was 60 dB SPL. All other stimuli had the same peak amplitude as the SAM stimuli, except for one stimulus with a doubled peak level (66 dB SPL). Additionally, in the "Offset" condition (reduced modulation index of 0.43), the peak amplitude was scaled by a factor of 1.67 as shown by the upper waveform trace in Fig.1a.

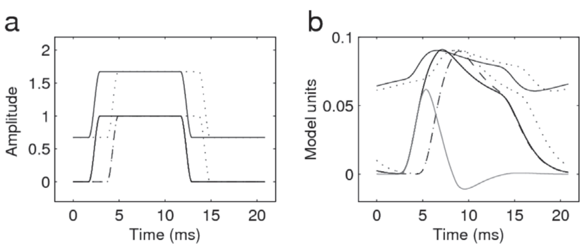

Fig. 1: Panel (a): Schematic representation of a single stimulus envelope period (solid line) with an ITD applied to the attack only (dashed line) and the whole waveform (dotted line). Note that for the ITD waveform, the dotted line coincides with the dashed line during the attack. The upper traces are for the "Offset" condition. Panel (b): Adapted signals (FFA model) of the respective envelope conditions in panel (a). The grey line at the bottom indicates the interaural difference which is used in the detector.

An adaptive, two-interval, two-alternative forced-choice (AFC) procedure was used in conjunction with a 1-up 3-down tracking rule to estimate the 79.4% correct point on the psychometric function (Levitt, 1971). Feedback was provided after each trial. Listeners responded via the computer keyboard or mouse. The ITD started at 2 ms and the initial factor by which the ITD was reduced was 2. After the second and fourth reversals the factors were 1.4 and 1.1, respectively. The adaptive run then continued for a further six reversals, and the threshold was defined as the geometric mean of the levels at those last six reversals. Four threshold estimates were obtained and geometrically averaged from each listener in each condition. Final thresholds reported here are the geometric means across listeners.

Results

The experimental data are shown in Fig. 2 as closed triangles. Panel (a) shows the just noticeable ITD as a function of the attack duration of the SWM (with the pause, hold, and decay duration fixed at 8.75, 8.75 and 1.25 ms, respectively). An almost linear dependence between the JND and the attack duration is observed indicating that the ITD sensitivity of the binaural system is highly correlated to the steepness of the rising flank. Panel (b) shows data for different duty cycles of the SWM. The modulation rate was 50 Hz and the attack and decay duration 1.25 ms. The highest JND is observed for the condition with the highest duty-cycle in which the decay is directly followed by the attack without additional pause duration. Otherwise, the data is nearly independent of the duty cycle. Panel (c) shows the effect of level (60 and 66 dB, indicated as condition 60 and 66) and envelope offset for 50-Hz SAM (left) and 50-Hz SWM (right) with a duty cycle of 0.5 and attack and decay durations of 1.25 ms. A decreased sensitivity is observed for the Offset condition. Particularly for the SAM condition, the JND increases by a factor of five in the Offset condition while the steepness of the ramps is identical to condition 60. Additionally, it can be seen that the 6-dB level increase (doubled steepness) in condition 66 causes a slightly increased ITD sensitivity.

In panel (d), the effect of modulation frequency is shown SAM (right) and SWM (left) with a duty cycle of 0.5 and attack and decay durations of 1.25 ms. It can be seen that the JND is decreased by a factor of about 2 when the SAM rate is increased from 50 to 100 Hz. For the SAM stimuli, the increased modulation frequency results in an increased steepness of the flanks which was shown to influence the JND in panel (a) of Fig. 2. In the case of the SWM, the steepness is constant and no effect on the JNDs is observed.

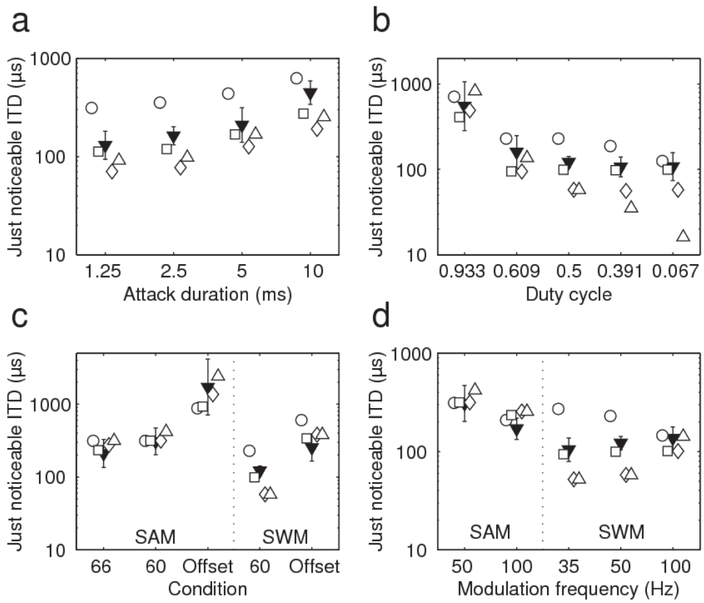

Fig. 2: Just noticeable ITDs for different stimulus conditions. Panel (a): Effect of attack duration for SWM. Panel (b): Effect of duty cycle for SWM. Panel (c): Level and envelope offset variations for SAM and SWM. Panel (d): Effect of modulation frequency for SAM and SWM. Experimental data are shown as closed triangles. Model predictions (open symbols) for the NCC, 5AL, 1AL, and FFA model are indicated as circles, triangles, diamonds, and squares, respectively.

MODEL PREDICTIONS

Models

Four models were used to predict the experimental data. All models shared the (monaural) preprocessing consisting of middle-ear filtering (1st order bandpass with cutoff frequencies of 500 and 8500 Hz), auditory filtering (4-kHz, 4th-order Gammatone filter, Patterson *et al*., 1987), half-wave rectification and low-pass filtering (770-Hz 5th-order, Breebaart *et al*., 2001) and peripheral power-law compression with an exponent of 0.4 (e.g., Dietz *et al*., 2008).

The normalized cross-correlation (NCC) model (Bernstein and Trahiotis, 2002, 2007) operated directly on the output of the preprocessed signals at both ears. The model was fit to correctly predict psychoacoustic JND for a reference condition (50-Hz SAM).

In the adaptation loop (5AL) model, the preprocessed signals were passed to a series of five adaptation loops (Püschel, 1988) as used in, e.g., Dau *et al*. (1996) and Jespen *et al*. (2008). JNDs were determined by the ITD at which the maximum difference of the adapted signals in the left and right channel exceeded a threshold value. Figure 1b shows a schematic plot of the adapted signals of the corresponding envelope conditions in panel (a). The upper traces are for the Offset condition. The grey line in panel (b) indicates the interaural difference. The maximum difference was calculated from a 100-ms steady-state part of the adapted signals (300 to 400 ms after stimulus onset). The threshold value was set individually for the 5AL, 1AL, and FFA model to match the psychoacoustic JND for the 50-Hz SAM condition.

The single adaptation loop (1AL) model used only the first (and fastest) adaptation loop of the 5AL model with a 5-ms time constant of the low-pass filter stage in the loop. The same detector stage as in the 5AL model was used.

The adaptation model (FFA) employed a feed-forward mechanism to simulate adaptation instead of a feedback loop. The output of the preprocessing stage was divided by an RMS-normalized and 1st-order low-pass filtered version of it (time constant of 5 ms). The output of the normalized low-pass stage was set to not be less than a threshold value of 0.9. The detector stage was again identical to the 5AL and 1AL model.

All four models include a 1st-order, 150-Hz low-pass filter prior to the binaural stage in order to account for monaural processing limitations of high-frequency envelopes (Kohlrausch *et al*., 2000; Ewert and Dau, 2000).

Model predictions

Model predictions are shown in Fig. 2 (open symbols) together with the psychoacoustic data (closed triangles). The NCC, 5AL, 1AL, and FFA model are indicated by the circles, triangles, diamonds, and squares, respectively. The dependence on the attack duration in panel (a) is generally well described by all models. The 5AL and 1AL models, however, show generally too low JNDs while the NCC model shows too high JNDs. Panel (b) shows that the data of the duty-cycle experiment are best described by the FFA model and with minor deviations by the 1AL model. In contrast to the data, the NCC model predicts an increase in sensitivity when the hold duration is reduced to zero (right-most condition) while particularly the 5AL model overestimates the JND decrease with decreasing duty cycle (increased pause duration). In panel (c) it can be seen that all models account quite well for the increased JNDs at the offset conditions with reduced modulation depth. However, the NCC model is per definition independent of the overall level. This is in contrast to the data of this study and in contrast to the data of Siveke *et al*. (2010). The 5AL and 1AL models also tend to underestimate the level dependence. The predictions of all models describe the trends in the data for different SAM frequencies in the left part of panel (d) well. As mentioned in the psychoacoustic results, the JNDs decrease with increasing steepness of the attacks. However, for the SWM stimuli in the right part of panel (d), the NCC model does also predict decreasing JNDs with increasing modulation frequency which is not observed in the data. The FFA model shows slightly increasing JNDs as in the data, caused by the slightly shorter pause duration at higher modulation frequencies. Again, the 5AL and 1AL models overestimate the importance of the pause duration: With decreasing modulation frequency, the JND predictions of these models decrease too much.

DISCUSSION

The psychoacoustic results in Fig. 2 demonstrate an important role of the attack durations (rise time) and pause durations preceding the attack for the sensitivity to envelope ITD. In case of the pause durations the main effect is already reached for pause durations as short as about 5 ms (Fig. 2b). Beyond 5 ms, further increase of the pause duration has little effect on the just noticeable ITD. A reduction of the modulation index to m = 0.43 by replacement of the pause (silence) with a region of reduced carrier intensity (Offset condition) leads to an increase in JND (Fig. 2c). A comparable effect was found in Stellmack *et al*. (2005) for their 128-Hz condition, while no difference between m = 1 and m > 0.3 was found for 300 Hz in Stellmack *et al*. (2005) and Nuetzel and Hafter (1981). The slight increase in sensitivity with level observed in Fig. 2c is in line with psychoacoustic data by Kohlrausch *et al*. (2000) where lower detection thresholds for monaural SAM were found and with Dietz *et al*. (2009) where an increased salience of envelope ITDs with increasing level was observed. Physiologic data on level dependence is inhomogeneous. In line with the current psychoacoustic data, Siveke *et al*. (2010) found a more precise tuning of binaural LSO neurons with increasing level. However, Dreyer and Delgutte (2006) found that the synchronization index of the auditory-nerve discharge pattern decreased with increasing level for both SAM and transposed tones.

The fact that the increase of modulation frequency does only result in a reduction of the just noticeable ITD for SAM and not for SWM (fixed attack duration) indicates that it is not generally a certain period fraction of the envelope which determines the just noticeable (ongoing) ITD as implied by, e.g., the NCC model (Bernstein and Trahiotis, 2002). Only in case of the SAM, the potential rise-time feature is directly correlated to the period duration. The stimuli used here might also be useful to test potential limitations of the IPD model (Dietz *et al*., 2009), where only the width of the modulation filters can account for differences between SAM and SWM. Furthermore, the current data support the hypothesis that the high sensitivity to ITDs in transposed tones might be less related to their special design to mimic the auditory nerve pattern of low-frequency pure-tones but rather to their pronounced pause regions and attacks when compared to SAM. With longer pause durations and steeper attacks (e.g. Fig. 2b), JNDs are even lower than for transposed tones in this study (not shown). Taken together, the data indicate that the slew rate of the attacks, additionally pronounced by preceding gaps, is the dominant temporal envelope feature for ongoing ITDs. Such behaviour is expected if the binaural system operates on an internal signal after a form of neural adaptation.

The model results suggest that adaptation prior to the binaural stage is generally suited to account for the data. The 5AL model includes too long time constants (up to 500 ms) which are not appropriate to be assumed prior to the binaural stage (see Fig.

2b) as in, e.g., Breebaart *et al.* (2001). With respect to the integration time constant responsible for the time course of the adaptation, there is no dramatic difference between the FFA (feed-forward) and the 1AL (feedback) model which both use the same time constant of 5 ms. Differences between the models are related to the strong "overshoot" limitation integrated in the FFA model (by means of a threshold value for the output of the low-pass filter stage). Here, the 1AL model shows a much more pronounced overshoot resulting in an excessive slew-rate boost of the attacks in the internal representation. The slew rate is reduced after adaptation by the 150-Hz low-pass filter, however, it is also plausible that there is an absolute slew-rate limitation in the neural responses independent of overall level. Such a limitation is not accounted for by the current model approaches. The missing effect of modulation index for $m > 0.3$ on just noticeable ITDs for the rather high envelope rates of 300 Hz in Stellmack *et al.* (2005) and Nuetzel and Hafter (1981) is in line with limitations for envelope processing beyond 150 Hz prior to the binaural stage.

The NCC model shows a modulation rate dependent prediction in case of the SWM (Fig. 1b) and is too insensitive in Fig. 1a. Additionally, the effect of duty cycle cannot be predicted correctly. The NCC model is by design capable of explaining the results as long as the just noticeable ITD is inversely related to the spectral bandwidth of the stimuli.

In general, the feed-forward adaptation (FFA) model shows the best agreement with the data. The major difference to the 5AL model is the more severe overshot limitation in the FFA model. All three adaptation models tested here share the same binaural difference detector which was developed as a functional concept. However, the concept is also physiologically plausible: The binaural difference detector can be realized with excitatory-inhibitory cells, which are typical cells in the lateral superior olive, the region which is assumed to encode temporal disparities in the stimulus envelope (e.g., Joris and Yin, 1995).

ACKNOWLEDGLEMENTS

This work was supported by the German Research Foundation (Deutsche Forschungsgemein-schaft) SFB/TRR 31.

REFERENCES

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**). "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. (**2007**). "Why do transposed stimuli enhance binaural processing?: Interaural envelope correlation vs envelope normalized fourth moment," J. Acoust. Soc. Am. **121**, EL23.

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

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). “A quantitative model of the effective signal processing in the auditory system. I. Model structure,” J. Acoust. Soc. Am. **99**, 3615-3622.

Dietz, M., Ewert, S. D., Hohmann, V., and Kollmeier, B. (**2008**). “Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences,” Brain. Res. **1220**, 234-245.

Dietz, M., Ewert, S. D., and Hohmann, V. (**2009**). “Lateralization of stimuli with independent fine-structure and envelope based temporal disparities,” J. Acoust. Soc. Am. **125**, 1622-1635.

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.

Dye, R. H. Jr, Niemiec, A. J., and Stellmack, M. A. (**1994**). “Discrimination of interaural envelope delays: The effect of randomizing component starting phase,” J. Acoust. Soc. Am. **95**, 463-470.

Ewert, S.D., and Dau, T. (**2000**). “Characterizing frequency selectivity for envelope fluctuations,” J. Acoust. Soc. Am. **108**, 1181-1196.

Hartmann, W. M. (**1987**). “Temporal fluctuations and the discrimination of spectrally dense signals by human listeners,” in *Auditory Processing of Complex Sound*, edited by Y. A. Yose, and C. S. Watson (Erlbaum, Hillsdale, New Jersey).

Hartmann, W. M., and Pumplin, J. (**1988**). “Noise power fluctuations and the masking of sine signals,” J. Acoust. Soc. Am. **83**, 2277-2289.

Henning, G. B. (**1974**). “Detectability of interaural delay in high-frequency complex waveforms,” J. Acoust. Soc. Am. **55**, 84-90.

Jepsen, M. L., Ewert, S.D., and Dau, T. (**2008**). “Modeling spectral and temporal masking in the human auditory system,” J. Acoust. Soc. Am. **124**, 422-438.

Joris, P.X., and Yin, T. C. (**1995**). “Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences,” J. Neurophysiol. **73**, 1043-1062.

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

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). “The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,” J. Acoust. Soc. Am. **108**, 723-734.

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.

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.

Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112-1118.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," paper presented at a meeting of the IOC Speech Group on Auditory Modeling at RSRE, 14-15 December.
Püschel, D. (**1988**). *Prinzipien der zeitlichen Analyse beim Hören* (Principles of Temporal Processing in Hearing), Ph.D. thesis (University of Göttingen).
Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2005**). "Discrimination of interaural phase differences in the envelopes of sinusoidally amplitude-modulated 4-kHz tones as a function of modulation depth," J. Acoust. Soc. Am. **118**, 346–352.
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.
Siveke, I., Leibold, C., Kaiser, K., Grothe, B., and Wiegrebe, L. (**2010**). "Adaptation of interaural-time-difference analysis to sound level", to appear in *Advances in auditory research: physiology, psychophysics and models*, edited by E. A. Lopez-Poveda, A. R. Palmer and R. Meddis (Springer, New York).