The Effect of Envelope Waveform on Lateralization

M. Klein-Hennig, M. Dietz, S. D. Ewert, V. Hohmann

Universität Oldenburg, 26129 Oldenburg, Germany, Email: martin.klein.hennig@uni-oldenburg.de

Introduction

Interaural time differences (ITDs) are important cues for lateralization in the human auditory system. The auditory system is sensitive to timing disparities in the fine-structure and the envelope of sounds. Due to the lack of phase-locking to the fine-structure at frequencies above about 1500 Hz, only envelope disparities can be exploited. In previously published data (e.g. [1]) parameters like the modulation frequency or the power of a modulating sinusoid have been altered in order to investigate the role of the envelope waveform on lateralization. However, these "classical" parameters simultaneously change several "secondary" envelope parameters like attack and decay times, as well as pause and hold durations, which might conceal the fundamental cues exploited by the auditory system [2]. In this study, psychophysical measurements were conducted with customized envelope waveforms in order to investigate the effect of isolated secondary parameters on lateralization. The results indicate that attack times and pause durations prior to the attack are the most important envelope features. The results are compared to the predictions of binaural auditory models.

Experiments

Method

Amplitude modulated pure tone stimuli with a carrier frequency of 4 kHz were used to measure the just noticeable envelope ITD (abbreviated JND: just noticeable difference). The modulation rate was 35 to 100 Hz. The waveform in Figure 1 shows a single period of a characteristic envelope used in this study. Each period consisted of a pause segment, an attack segment (raised cosine ramp), a hold segment and a decay segment (raised cosine ramp). To study their individual influence on the JND, the durations of the segments were varied. Setting the pause and hold durations to zero resulted in a sinusoidal amplitude modulation (SAM), while larger hold and pause durations led to a "square-wave" modulation (SWM) with variable duty cycle and attack/decay times. The minimal attack/decay durations were limited to 1.25 ms in order to control for spectral broadening of the stimuli. The ITD was applied either to the whole envelope waveform or to the attack only, as indicated by the dashed and dotted lines in Figure 1. Control experiments with ITDs applied merely to the decay segment indicated that the ITD of the decay had virtually no influence on the lateralization. The stimuli had a duration of 500 ms and were gated with 125-ms raised cosine ramps to minimize the influence of onset cues. A low-pass filtered noise (5^{th}-order Butterworth at 1000 Hz, white spectrum up to 200 Hz, -3dB per octave slope up to 1 kHz) was simultaneously presented to suppress binaural information at low frequencies. The low-pass noise had a duration of 600 ms and a level of 45 dB SPL. The stimuli were centred in the noise. The level of the SAM stimulus was 60 dB SPL. All other stimuli had the same peak amplitude as the SAM stimuli, with the exception of three additional SAM stimuli that had 36, 48 and 66 dB SPL. For the 36 and 48 dB SPL stimuli, the noise level was set to 20 and 30 dB SPL, respectively. Additional "offset" conditions with a modulation index of 0.43 were tested. Their peak amplitude was scaled by 1.67 as shown in Figure 1.

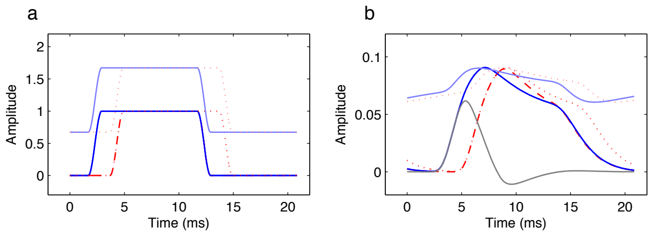

Figure 1: (a) Schematic representation of a single envelope period (blue solid line) with an ITD applied to the attack only (red dashed line) and the whole waveform (red dotted line). The upper curve shows the "offset" condition. (b) Adapted signal (AD model) of the respective envelope conditions in panel (a). The grey line at the bottom indicates the interaural difference used in the detector for a signal with onset delay.

The JNDs were measured in an adaptive two-interval, two-alternative force choice (AFC) procedure with a 1-up 3-down tracking rule. After each trial, feedback was provided. Listeners responded using a computer keyboard or mouse. The ITD started at 2 ms and was reduced by an initial factor of 2. After the second and fourth reversals, the factors were 1.4 and 1.1. A single run then continued for another six reversals. Since the ITD is varied in relative stepsizes, the threshold was defined as the geometric mean of the ITD at those last six reversals. Four thresholds were geometrically averaged from each listener in each condition. All thresholds reported in this study are the geometric means across listeners. The test subjects were five normal-hearing listeners aged 25 to 29.

Results

Figure 2 shows the experimental data as closed triangles. In Panel (a), the JND are shown as a function of the attack duration of the SWM (with pause, hold and decay duration fixed at 8.75, 8.75 and 1.25 ms, respectively). There is an almost linear dependence between the JND and the attack duration. Panel (b) shows JND for different duty cycles of the SWM. The modulation rate was fixed at 50 Hz and the attack and decay duration at 1.25 ms. The highest JND is observed for the condition with the highest duty cycle, in which the decay is followed immediately by the attack without additional pause duration. Otherwise, the data shows little dependence on the duty cycle. The dependence on level

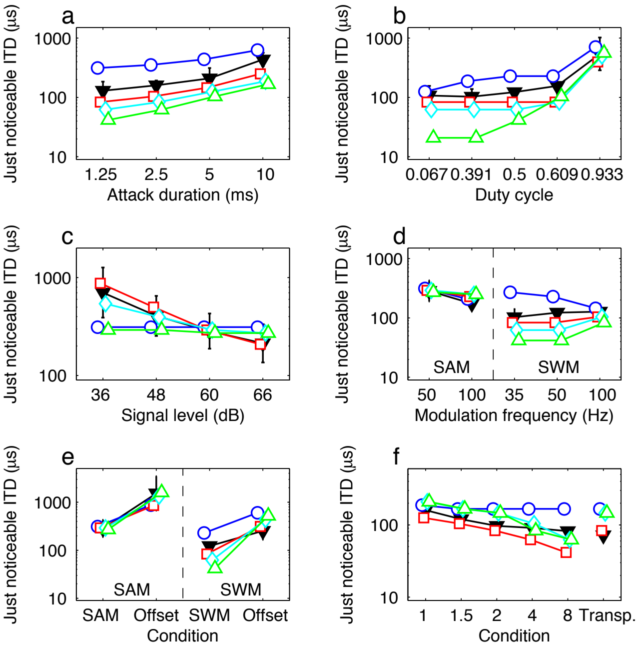

Figure 2: Just noticeable ITDs for different stimulus conditions. Panel (a): Effect of attack duration for signals with square wave modulation SWM. Panel (b): Effect of duty cycle for SWM. Panel (c): Effect of signal level for SAM. Panel (d): Effect of modulation frequency for SAM and SWM. Panel (e): Effect of envelope offset for SAM and SWM. Panel (f): Psychoacoustic data from [3] and model predictions. Experimental data are shown as black triangles. Model predictions (open symbols) for the NCC, ADL, SADL and AD model are indicated as circles, triangles, diamonds and squares, respectively.

is shown for SAM in panel (c). The 6-dB level increase from 60 to 66 dB shows a slight increase in the ITD sensitivity. The 12 dB steps from 60 to 48 dB and from 48 to 36 dB almost double the JND. Panel (d) shows the effect of modulation frequency for SAM (left) and SWM (right) with a duty cycle of 0.5 and attack and decay durations of 1.25 ms. Increasing the SAM rate from 50 to 100 Hz decreased the JND by a factor of about 2. 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 Figure 2. For the SWM stimuli, the steepness is constant and no effect on the JND is observed. In panel (e), the effect of envelope offset for 50-Hz SAM and 50-Hz SWM with a duty cycle of 0.5 and attack and decay durations of 1.25 ms is shown. A decreased sensitivity is observed for the offset condition. For the SAM condition, the JND increases by a factor of about five in the offset condition. The steepness of the flanks is identical to the SAM without offset. Panel (f) shows JNDs for raised-sine and transposed tone stimuli (f_m= 128 Hz, f_c=4 kHz) from [3]. For larger exponents, the raised-sine JNDs decrease, approaching the transposed tone JND. These results were included to perform model predictions on external data.

Model Predictions

Models

Four different models were used to predict the experimental data. The monaural preprocessing stage was the same for all models with adaptation. It consisted of middle-ear filtering (1^{st} order bandpass with cutoff frequencies 500 and 8500 Hz), auditory filtering (4-kHz 4^{th} order gammatone filter), half-wave rectification, low-pass filtering (770-Hz 5^{th} order) and peripheral power-law compression with an exponent of 0.4 (e.g. [4]). Prior to the binaural stage, all models include a 1^{st}-order, 150 Hz low-pass filter to account for monaural processing limitations of high-frequency envelopes ([5]).

The normalized cross-correlation (NCC) model ([1], [6]) operated directly on both channels of the preprocessed signal. The preprocessing stage was implemented as described in [2]. The model was fit to correctly predict the psychoacoustic JND for a reference condition (50-Hz 60-dB SAM).

The adaptation loop model (ADL) passes the preprocessed signals through a series of five adaptation loops ([7]). The ITD for which the maximum difference between left and right channel of the adapted signal exceeded a threshold value was predicted as the JND. In order to model a representative envelope period the maximum difference was calculated from a 100-ms steady-state part of the adapted signals (300 to 400 ms after stimulus onset). This interval covers at least one full envelope period and thus provides a stable output independent on minor variations of the detection interval. The single adaptation loop (SADL) model used only the first adaptation loop (5-ms time constant) of the ADL model, with the same detector stage.

The adaptation model (AD) used a feed-forward mechanism instead of a feedback loop to simulate adaptation. The preprocessed signals were divided by an rms-normalized and 1^{st} order low-pass filtered (5-ms time constant) version of themselves. The output of this normalized low-pass stage was set to be no less than a threshold value of 0.9. The detector stage was identical to the ADL and SADL model.

Model predictions

Figure 2 shows the model predictions along with the psychoacoustic data (closed triangles). The NCC, ADL, SADL and AD model predictions are indicated by circles, triangles, diamonds and squares. The dependence on attack duration in panel (a) is generally well described by all models. The ADL and SADL model, however, predict too low JNDs whereas the NCC model predicts too high JNDs. The data of the duty cycle experiment shown in panel (b) are best described by the AD model and with minor deviations by the SADL model. The NCC model predicts an increase in sensitivity when the hold duration is reduced. The ADL model overestimates the JND decrease with increased pause duration. Panel (c) shows data for SAM tones with decreasing overall level. The NCC and ADL models are independent of the overall level by design. The SADL model overestimates the level dependence for lower levels. All models can account quite well for the increase in modulation rate (and steepness) of the SAM tones as shown in panel (d) on the left side. For the SWM stimuli in the right part of panel (d), the NCC model predicts decreasing JND with increasing modulation frequency which is not observed in the data. The AD model shows slightly increasing JNDs, caused by the slightly shorter pause duration at higher modulation frequencies. The ADL and SADL models overestimate the importance of the pause duration, their JND predictions decrease too much with decreasing modulation frequency. In panel (e) it can be seen that all models can predict the increased JNDs at the offset conditions with reduced modulation depth quite well. For the SWM stimuli, the NCC model underestimates the influence of stimulus offset. Panel (f) shows the model predictions for raised-sine stimuli and a transposed tone along with psychoacoustic data from [3]. The NCC model shows almost no dependence on the exponent of the raised sine tones. The ADL model predicts generally too low JNDs whereas the SADL model is not sensitive enough for exponents of up to n=2 and too sensitive at n=8. The AD model predicts all raised-sine JNDs too low, but performs better than the ADL model. For the transposed tone, the ADL, SADL and NCC models are too insensitive, the ADL model predicts the transposed tone JND quite well.

Discussion

The psychoacoustic results shown in Figure 2 emphasize the important role of the attack duration (rise time) and pause duration preceding the attack for the sensitivity to envelope ITD. For pause durations, the main effect is reached for durations as short as 5 ms (Figure 2b). Beyond 5 ms, further increase of the pause duration has little effect on the JND. Reducing the modulation index to m=0.43 by replacing the pause (silence) with a region of reduced carrier intensity (offset condition) leads to an increase in JND. The increase in sensitivity with level observed in Figure 2c is in line with psychoacoustic data found in [5], where lower detection thresholds were found for monaural SAM, and with [4] where an increased salience of envelope ITDs with increasing level was observed. The fact that the increase of modulation frequency does only result in a reduction of the JND for SAM and not for SWM (fixed attack and decay duration) indicates that the just noticeable (ongoing) ITD is not generally determined by a certain period fraction of the envelope. With longer pause durations and steeper attacks (e.g., Figure 2b), JNDs are even lower than for transposed tones used in this study (data not shown). In conclusion, the data indicate that the slew rate of the attacks, additionally pronounced by preceding gaps, is the dominant temporal envelope feature for ongoing ITDs. This behaviour can be 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 suitable to account for the psychoacoustic data. The ADL model includes too long time constants (up to 500 ms). Regarding the integration time constant being responsible for the time course of the adaptation, the AD (feed forward) and SADL (feedback) models, both using a 5 ms time constant, do not differ dramatically. Differences between the AD and SADL models are related to the strong “overshoot” limitation in the AD model due to a threshold value for the output of the low-pass filter stage. The SADL model shows a much more pronounced overshoot leading to an excessive slew rate boost of the attacks in interaural representation. The 150-Hz low-pass filter after the adaptation reduces the slew rate, but it is also possible that there is an absolute slew-rate limitation in the neural responses independent of overall level. This limitation is not accounted for by the current model approaches.

The NCC model is too insensitive in Figure 2a and shows a modulation rate dependence in case of the SWM (Figure 2d). Additionally, the effect of duty cycle can not be predicted correctly. By design, the NCC model can explain the results as long as the JND is inversely related to the spectral bandwidth of the stimuli.

In general, the feed-forward adaptation (AD) model shows the best agreement with the data. For all three adaptation models tested here, the same binaural difference detector was used. The concept of a difference detector is physiologically plausible: it can be realized with excitatory-inhibitory cells, which are typical cells in the lateral superior olive, which is assumed to be the region in which temporal disparities in the stimulus envelope are encoded ([8]).

Acknowledgements

This study was supported by the DFG (SFB/TRR 31 „The active auditory system“).

References

[1] Bernstein L.R., Trahiotis C. Why do transposed stimuli enhance binaural processing? Interaural envelope correlation vs envelope normalized fourth moment. J. Acoust. Soc. Am. **121** 2007, EL23

[2] Heil, P. Coding of temporal onset envelope in the auditory system. Speech communication **41** 2003, 123-134

[3] Bernstein L.R., Trahiotis C. Enhanced processing of interaural temporal disparities at high-frequencies: Beyond transposed stimuli. Hearing – from Sensory Processing to Perception, Springer 2007.

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

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

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

[7] Dau T., Püschel D., Kohlrausch A. A quantitative model of the effective signal processing in the auditory system. I. Model Structure. J. Acoust. Soc. Am. **99** 1996, 3615-3622

[8] Joris P.X., Yin T.C. Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences. J. Neurophysiol. **73** 1995, 1043-1062