Technical University of Denmark

The importance of binaural cues for the perception of apparent source width at different sound pressure levels

Käsbach, Johannes; May, Tobias; Le Goff, Nicolas; Dau, Torsten

Published in:
Proceedings of DAGA 2014

Publication date:
2014

Document Version
Early version, also known as pre-print

The importance of binaural cues for the perception of apparent source width at different sound pressure levels

Johannes Käsbach, Tobias May, Nicolas Le Goff and Torsten Dau
Centre for Applied Hearing Research, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, Email: johk@elektro.dtu.dk

Introduction

Regarding room perception, apparent source width (ASW) is an essential measure that describes the perceived spatial extent of a sound source. Mainly three important factors have been mentioned that contribute to this percept: (i) The degree of correlation of the two ear signals (IC_{ears}), whereby a decorrelated signal causes a high ASW, (ii) the frequency-content, i.e. low-frequency sounds are perceived as being wider than high-frequency sounds for a fixed IC_{ears} value [1] and (iii) the sound pressure level (SPL) at low frequencies, i.e. ASW increases with increasing SPL [2]. However, the exact contributions of these individual cues to the complex percept are still unclear. This study extends our previous work [3] and presents a psychoacoustic evaluation where the three parameters IC_{ears}, frequency-content and sound pressure level were varied in order to quantify their influence on ASW under controlled conditions. Two binaural models were considered to predict ASW: a complex, nonlinear auditory model as described in [4] and a simple, linear computational model. Their results are compared to the psychoacoustic data.

Method

The experiment was conducted in the Spacelab at the Technical University of Denmark, with a reverberation time of $T_{60} = 0.2s$. A loudspeaker-ring with 11 loudspeakers was placed in the horizontal plane, ranging from $-75°$ to $75°$ with a spacing of $15°$ to each other (see Fig. 1). All loudspeakers were of type Dynaudio BM6. Only the two loudspeakers at $\alpha = 30°$ (typical stereo-setup) were used to play back the stimuli. One additional loudspeaker at $0°$ was used for a reference condition. Room impulse responses (RIR) were measured for the two stereo-loudspeakers with a B&K head and torso simulator (HATS) of Type 4100 placed at the listener's position at a distance of $d = 1.8$ m. The signals were convolved with the RIRs and the binaural analysis was based on the resulting signals.

The stimuli were bandlimited noises with center frequencies of $f_c = 0.25$ kHz, 1 kHz and 4 kHz and a bandwidth of $bw = 2$ octaves. In addition, a highpass filtered noise signal with cut-off frequency at $f_c = 8$ kHz was used. The bandpass and highpass filters were both digital Butterworth filters of eighth and fourth order, respectively. The two loudspeaker signals had a duration of 2 s and were generated from a 2-dimensional normal distribution where the covariance was adjusted to yield inter-channel correlation values (IC_{LS}) of 0, 0.3, 0.6, 0.8 and 1. Due to the influence of the head, cross-talk and the listening room, the corresponding interaural cross-correlation values (IC_{ears}) at the ears of the HATS deviated from IC_{LS} and are listed in Table 1. The stimuli were presented at two sound pressure levels (SPL) of 50 and 70 dB within limits of ± 2 dB variations among IC_{LS} values.

Table 1: Measured IC_{ears} values for corresponding IC_{LS} values at all center frequencies (cf).

cf [kHz]	IC_{LS}				
	0	0,3	0,6	0,8	1
0.25	0,63	0,75	0,83	0,89	0,93
1	0,30	0,41	0,57	0,67	0,77
4	0,27	0,32	0,53	0,66	0,80
8 (HP)	0,30	0,29	0,44	0,58	0,73

In order to measure ASW, subjects were asked to indicate the apparent opening angle in degrees on a scale with a $5°$ resolution of each presented stimulus via a touchscreen as indicated in Fig. 1. The stimuli were presented in random order and repeated three times. A reference was available to the listener which was a broadband noise signal at 45 dB SPL presented over the loudspeaker in the center. This produced a very narrow source with an opening angle of $0°$. Both, stimulus and reference, could be played back as often as desired by the listener. Thirteen normal-hearing listeners participated, of which seven evaluated the additional condition with the highpass filtered noise signal. All listeners were made familiar with the evaluation procedure in a short training and the entire evaluation procedure lasted about one hour per subject.

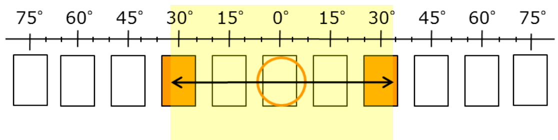

Figure 1: Sketch of the experimental set-up and procedure. The loudspeakers at $\pm 30°$ generate a phantom source at $0°$. Subjects were asked to indicate ASW in degree on the given scale.

Model predictions of ASW

In rooms, the decorrelation at the ears of a listener is primarily caused by room reflections which lead to temporal fluctuations of the binaural cues, namely interaural time and level differences [5]. Two binaural models that analyse these fluctuations were used to predict ASW for the stimuli presented in the psychoacoustic measurement. The first model is based on Breebaart's equalisation-cancellation (EC) approach with excitation-inhibition (EI)-type elements to simulate binaural interaction [4]. It has a complex peripheral processing consisting of a 4th-order gammatone filter bank, hair-cell transduction, absolute threshold of hearing (ATH) and adaptation loops to simulate input level sensitivity. Its decision device (P_{ASW}) is based on a linear combination of the standard deviation of ITDs in the frequency bands from 387 Hz to 1.84 kHz and the monaural output level at the low-frequency bands from 168 to 387 Hz. The second model is a simple linear, computational model that is based on a cross-correlation estimation. It consists of a 4th- order gammatone filterbank, with 1 ERB-wide filters around center frequencies ranging from 80 to 18197 Hz. The ITDs and ILDs were extracted for frames of 20 ms duration with 50% overlap. The variability of each binaural cue over time was calculated from the differences between the 90th and 10th percentile of the respective disributions per frequency channel. The mean value across frequency of each binaural distribution was calculated, called ITD_{width} and ILD_{width}, and was used as an estimate of ASW.

Results and Discussion

The psychoacoustic data and the model predictions are shown in Fig. 2 for the two sound pressure levels 50 dB (left) and 70 dB (right). The abscissa shows the measured IC_{ears} values obtained with the HATS. The ordinate shows the measured ASW (in degrees) in the upper panel and the model outputs in the respective model units in the second to fourth panel. The mean values with their standard errors are shown. The psychoacoustic data reveal that ASW decreases monotonically with increasing IC_{ears} for all tested frequencies, represented by the different colours, as expected from literature. It is important to note that even for stimuli with frequency components above 2 kHz, subjects were able to discriminate ASW. Two frequency-dependent effects can be seen: First, considering a fixed IC_{ears}, for instance at $IC_{ears} \approx 0.6$, ASW decreases with increasing center frequency. Second, at 50 dB SPL, the dynamic range of ASW, representing the difference between the largest and smallest values, also decreases with center frequency. Increasing the sound pressure level to 70 dB resulted in a higher ASW: At stimuli with $f_c = 0.25$ kHz, all ASW results are shifted upwards, such that the dynamic range of ASW is maintained. For higher frequencies, the higher SPL resulted in a larger dynamic range of ASW. These results suggest that the increase of SPL affects ASW in a broad frequency range, which complements the findings in [2] where this effect was only observed at low frequencies.
The output of the more complex model, P_{ASW}, shows a frequency-dependency that is similar to that in the psychoacoustic data. The model accounts for the relationship between SPL and ASW at $f_c = 0.25$ and 1 kHz due to its nonlinear components. However, the model's sensitivity to IC_{ears} is strongly reduced leading to a smaller dynamic range compared to the data. This might be caused by an overemphasis of the sound pressure level component in the model relative to the ITD statistics component. For the two high-frequency stimuli, i.e. at $f_c = 4$ kHz and the highpass filtered noise, a flat curve remains across IC_{ears} and leads to a non-monotonic function at the highpass filtered stimuli (at 70 dB SPL). The reduced model output is presumably caused by limiting the analysis of the ITD statistics to frequencies below 2 kHz.
The ASW predictions of the simple model are based on both ITD and ILD estimates, measured in ms and dB, respectively. Note that the sound pressure level does not have an effect on both estimates, since no SPL dependent components were included. Both binaural cues show a large dynamic range across all stimuli. Considering the ITD-based predictions, both the dependency on IC_{ears} and on frequency are comparable to the patterns in the psychoacoustic data. However, predictions for the stimuli at $f_c = 0.25$ kHz showed a compressive behaviour and were overestimated for the highpass filtered stimuli for $IC_{ears} < 0.5$ compared to the psychoacoustic data. Similar results were obtained with the ILD-based predictions that show a compressive behaviour across all frequencies.

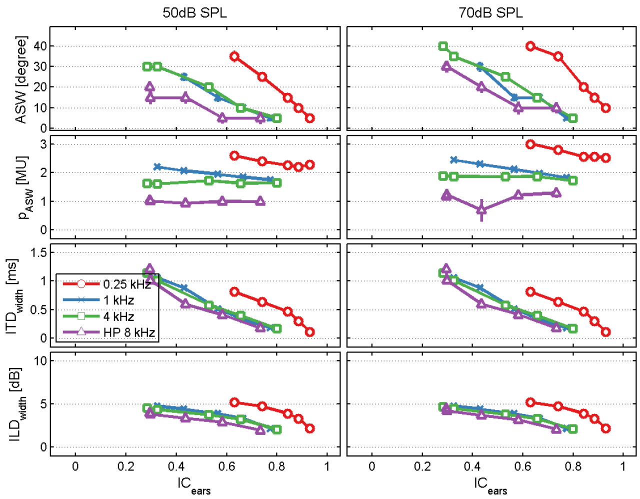

Figure 2: ASW data measured in degrees and the corresponding model predictions as a function of IC_{ears} for two different SPLs. Shown are the mean values and standard errors for the bandlimited noise signals with center frequencies $fc = 0.25$, 1 and 4 kHz and a highpass (HP) filtered noise signal at 8 kHz.

Summary and Conclusions

The psychoacoustic data showed ASW as a function of IC_{ears}, frequency and sound pressure level. Listeners were able to discriminate ASW even for stimuli with a frequency content above 2 kHz. Increasing the sound pressure level showed a frequency-dependent increase in ASW. The complex binaural model was shown to account for the frequency and sound pressure level effects seen in the data, but revealed an insensitivity regarding IC_{ears}. The predictions with this model were restricted to the analysis of ITD fluctuations below 2 kHz. The predictions obtained with the simple model showed that besides ITDs and the monaural sound pressure level, also ILD statistics contributed to ASW estimates. This model cannot, however, account for changes of the sound pressure level since it does not include any nonlinear components. The ideal model would need both, non-linear components and an appropriate weighting of information across ITDs and ILDs, in order to account for the psychoacoustic data. Future studies will further investigate the role of the different auditory processes on ASW, also including more complex stimuli like e.g. speech and music.

References

[1] Blauert, J. et al.: Spatial mapping of intracranial auditory events for various degrees of interaural coherence. JASA, 1986.

[2] Okano, T. et al.: Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls. JASA, 1998.

[3] Käsbach, J. et al.: The relation between perceived apparent source width and interaural cross-correlation in sound reproduction spaces with low reverberation. Proc. of DAGA, Italy, 2013.

[4] van Dorp Schuitman, J. et al.: Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model. JASA, 2013.

[5] Blauert, J. et al.: Auditory spaciousness: Some further psychoacoustic analyses. JASA, 1986.