Technical University of Denmark

The relation between perceived apparent source width and interaural cross-correlation in sound reproduction spaces with low reverberation

Käsbach, Johannes; Marschall, Marton; Epp, Bastian; Dau, Torsten

Published in:
Proceedings of DAGA 2013

Publication date:
2013

The relation between perceived apparent source width and interaural cross-correlation in sound reproduction spaces with low reverberation

Johannes Käsbach, Marton Marschall, Bastian Epp, Torsten Dau

Centre for Applied Hearing Research, Technical University of Denmark (DTU), Email: johk@elektro.dtu.dk

Introduction

Apparent source width (ASW) is a perceptual attribute that describes the perceived width of a sound image. The interaural cross-correlation (IACC) is commonly used in room acoustics as an objective measure for ASW. Early reflections in a room cause a decorrelation of the two ear signals, i.e., a reduction of IACC, which leads to a larger ASW. In a dichotic listening condition using headphones, the ASW changes when varying the correlation between the two channels. Blauert and Lindemann [1] showed for bandlimited noise signals that the ASW increases with decreasing inter-channel correlation (IC). For a given bandwidth, the ASW was found to increase with decreasing center frequency of the noise while keeping the IC constant. In such conditions, the sound image is internalised, i.e. perceived inside the head. Since ASW relates to the spatial extent of a sound source, this measure becomes more meaningful for externalised sources, such as produced by two loudspeakers. A phantom source is then perceived in the middle of the two loudspeakers ([2]). In contrast to the headphone (HP) presentation, the loudspeaker (LS) presentation produces an interaural cross-talk (CT) referring to the cross paths between left LS and right ear and vice versa. The present study investigates the influence of the interaural cross-talk on the perception of ASW in rooms with low reverberation that are typically used for sound reproduction systems and virtual sound environments.

Method

Three experiments were performed: 1) LS presentation in a listening room; 2) HP presentation of the same listening room using head and torso simulator (HATS) recordings and 3) HP presentation of an anechoic listening condition using a HRTF database.

Apparatus

The experiments were conducted in the Immersive Presence Lab at CIRMMT, McGill University, with a reverberation time of $T_{60} = 0.2s$. A typical stereo-setup with an opening angle of $\alpha = 30^\circ$ was installed using Genelec 8030A loudspeakers. The listener was seated at a distance of $d = 2.3$m (Experiment 1). Room impulse responses (IR) were measured for each loudspeaker in the same room with a B&K head and torso simulator (HATS) of Type 4100 placed at the listener's position. The signals were convolved with the IRs and presented to the listener via headphones. The resulting stimuli were then either presented including or excluding the interaural cross- talk (Experiment 2). For the anechoic listening condition, the CIPIC HRTF database [3] was used (Experiment 3).

Stimuli and experimental procedure

The stimuli were bandlimited noises with center frequencies of $f_c = 0.25$kHz or 1 kHz and a bandwidth of $\Delta f = 1.5$ octaves. The signals were generated from a 2-dimensional multivariate normal distribution where each dimension corresponded to one loudspeaker channel. The bandpass filter was a digital 4th-order Butterworth filter with 24 dB/octave roll-off. The signals were amplitude modulated with a modulation depth of 60% and a modulation frequency of 8 Hz. All stimuli were presented with constant spectral density of 35dB/Hz (at IC = 0) and had a duration of 4s.

The experiments were separately tested using a Multi Stimulus test with hidden reference and anchor (MUSHRA) [4], excluding anchor and reference. In Experiment 1, the ASW was measured for three inter-channel correlation (IC) values of 0, 0.6 and 1 at the two center frequencies of 0.25 and 1 kHz. In Experiments 2 and 3, the two conditions with and without interaural cross-talk were considered for the same noise signals and ICs. Subjects were asked to rate the stimuli relative to each other in terms of ASW on a scale from 0 (narrow) to 100 % (wide). In addition, the task was to identify the narrowest and the widest stimulus with 0 and 100, respectively. The stimuli were presented in random order and repeated six times. Thirteen normal-hearing subjects participated in Experiments 1 and 2, and six of these participated in Experiment 3. The results were analysed with a 3-way ANOVA with the null hypothesis that all stimuli produced the same ASW. The three factors were stimuli, subjects and repetitions. The post hoc analysis were based on Least Significant Difference (LSD) bars. Differences were considered significant for $p < 0.05$.

IACC-based predictions of ASW

According to [5], the IACC describes the correlation between the left-ear signal, $p_l(t)$, and the right-ear signal, $p_r(t)$, normalised with their rms values. The resulting IACC coefficient corresponds to the maximum of the cross- correlation function $\rho_{lr}(\tau)$, calculated with a delay time interval of $|\tau| \leq 1$ms and using a time window of $t_2 - t_1$. It takes values between zero and one:

$$\rho_{lr}(\tau) = \frac{\int_{t_1}^{t_2} p_l(t)p_r(t+\tau)\mathrm{d}t}{\sqrt{\int_{t_1}^{t_2} p_l^2(t)\mathrm{d}t \int_{t_1}^{t_2} p_r^2(t)\mathrm{d}t}} \quad (1)$$

$$IACC = max|\rho_{lr}(\tau)| \quad (2)$$

The IACC was calculated for the same stimuli as presented to the listeners, i.e. the binaural room impulse responses, measured with the HATS, were convolved with the noise signals. The integration time window $t_2 - t_1$ was chosen to be the duration of the stimuli.

Results

The results of the three experiments are shown in Figures 1 to 3. The measured ASW scores (open symbols) are indicated as median values and 25th and 75th percentiles for the three considered IC values (0, 0.6 and 1). The corresponding interaural cross-correlation coefficients were calculated and are represented as 1-IACC. They are indicated as filled symbols connected by solid lines in the figures for comparison with the measured ASW scores.

The results from Experiment 1 are shown in Figure 1. The measured ASW scores were similar for the two noise bands centered at 0.25 kHz (open circles) and 1 kHz (open squares). At both frequencies, ASW decreased with increasing IC. At both frequencies, a comparable dynamic range (of $0.8 - 0.9$) was obtained, representing the difference between the largest and smallest values. In contrast to the data, the calculated IACC values were different for the two bands. Larger IACC values were produced at 1 kHz than at 0.25 kHz. Furthermore, a reduced dynamic range of only 0.2 was obtained for the low-frequency band, compared to the high- frequency band with a dynamic range of 0.4. The statistical analysis revealed no significant differences for the two frequency bands for IC = 0 and 0.6, but the difference for IC = 1 was significant.

Figure 2 shows the measured data from Experiment 2 obtained in the conditions with cross-talk (open squares) and without cross-talk (open circles) using the HRTFs recorded in the CIRMMT lab. The results for 0.25 kHz are shown in the top panel and the results for 1 kHz are represented in the lower panel. The data show that, in the case of cross-talk, lower ASW scores were obtained than in the absence of cross-talk, for all specified IC values. The statistical analysis revealed that there was a significant difference between the two conditions at both frequencies. The median of the ASW scores was found to be shifted downwards by the same amount of about 0.4 in both conditions. Thus, the dynamic range for the ASW values as a function of IC was similar with and without cross-talk which was the case for both frequency bands. The only exception was the high-frequency band

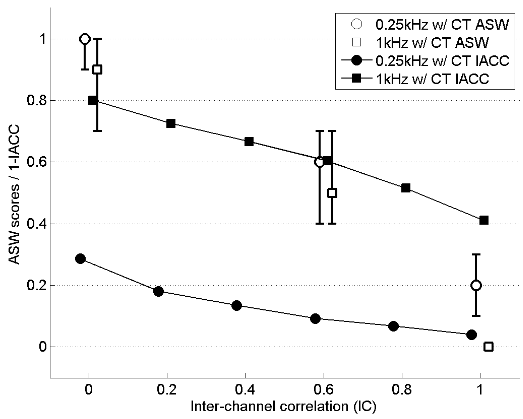

Figure 1: ASW scores and calculated 1-IACC values as a function of the inter-channel correlation in Experiment 1. Two frequency bands centered at $f_c = 0.25$ and 1 kHz were considered in the CIRMMT lab, presented via two loudspeakers. The subjective results are represented by their median and 25th and 75th percentiles.

for $IC = 1$ where a similar ASW score was observed with and without cross talk. In contrast, the IACC predictions, indicated by the filled symbols connected by solid lines, resulted in a shallower curve in the condition with cross-talk (filled squares) than in the condition without cross-talk (filled circles). At 1 kHz, this resulted in an intersection of the two IACC curves at IC = 0.4.

Figure 3 shows the results from Experiment 3 for the two conditions with and without cross-talk in the anechoic listening environment. The results were similar to those obtained in Experiment 2 and the statistical analysis revealed that there was a significant difference between the two conditions at both frequencies. Lower ASW scores were obtained when the cross-talk was present. The difference between the two conditions was, however, slightly smaller than in Experiment 2. The IACC predictions for the 1-kHz band provided similar values for ICs between 0.2 and 0.6 and decreasing values for outside this IC range. This is in contrast to the measured ASW values that showed a monotonic decrease with increasing IC.

Discussion

The results demonstrated that the IACC predicted a reduced dynamic range of values in the cross-talk conditions whereas the measured ASW scores showed a similar dynamic range in the conditions with and without cross-talk and a downward shift in the cross-talk condition. This discrepancy from the data can be explained by considering the effect of cross-talk on the cross-correlation function $\rho_{lr}(\tau)$ from Eqn. (1) as illustrated in Figure 4 (left panel). $\rho_{lr}(\tau)$ is shown for a broadband signal for five inter-channel correlation values. Besides the peak at zero lag which varies with IC, two additional constant peaks occur at time lags corresponding to the angular offset of the loudspeakers from the median plane. Due to this offset, the signal from the left LS will reach the left ear before the right ear, and vice versa for the

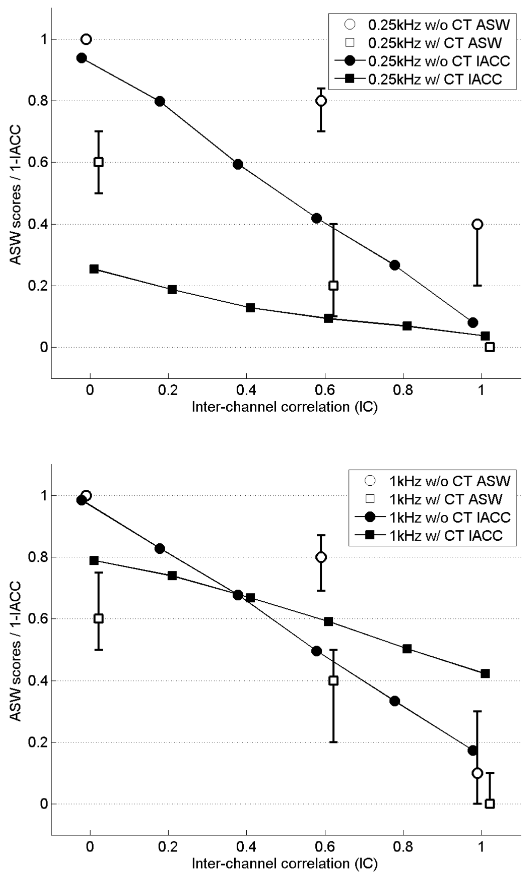

Figure 2: ASW scores and 1-IACC predictions as a function of IC in Experiment 2. The two conditions with and without cross-talk were presented via headphones in the CIRMMT lab using HATS recordings, for two frequency bands. The measured data are represented by their medians and 25th and 75th percentiles.

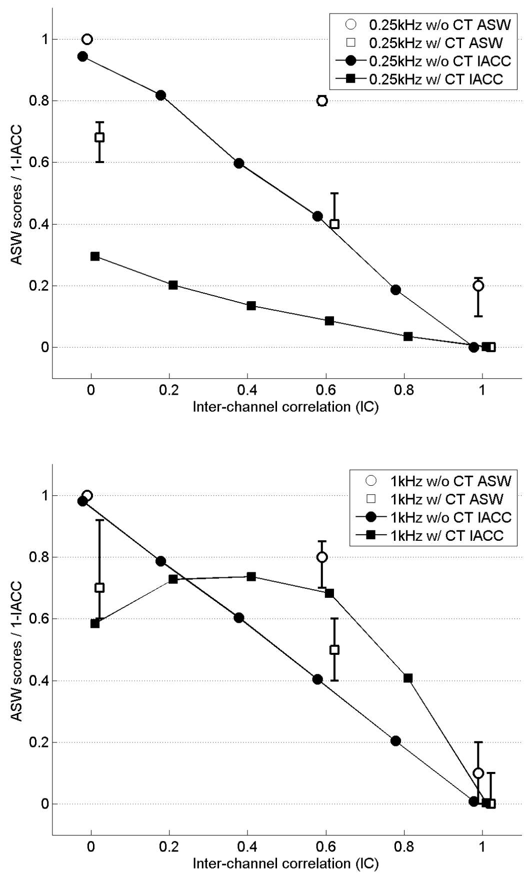

Figure 3: ASW scores and 1-IACC results as a function of IC in Experiment 3. The two conditions with and without cross-talk were presented via headphones in an anechoic listening environment using the CIPIC database [3], for two frequency bands. The measured data are represented by their medians and 25th and 75th percentiles.

right LS. For low correlation values between the two LS signals (from IC = 0 to IC = 0.4), the peaks caused by the delay are larger than the actual specified IC value (peak at zero lag). According to Eqn. (1), the IACC coefficient corresponds to the maximum of $\rho_{lr}(\tau)$, which is equal to the value of the side peaks unless the peak at zero lag becomes larger. A monotonically decreasing 1-IACC with increasing IC is thus only resulting in the absence of cross-talk. When bandlimiting the signal, the peakwidths increase and interfere with each other. The middle panel of Figure 4 shows $\rho_{lr}(\tau)$ for the 1-kHz frequency band considered in Experiment 3. The pattern shows a minimum at IC = 0 for a delay time of 0ms and a roughly constant maximum value (defined by the two side peaks) for ICs between 0.2 and 0.6, consistent with the corresponding 1-IACC function shown in Figure 3 (filled squares in the lower panel). For the low-frequency band at 0.25 kHz (right panel of Figure 4) from Experiment 2, the width of the peaks increase such that they are not distinguishable from each other and result in one large peak. This explains the reduced range 1-IACC values shown in Figure 2 (filled squares in the upper panel).

The outlined differences between data and predictions based on the IACC suggest that other cues might be available to a listener. More advanced models, such as the one proposed in [6], suggest that the stimulus energy below 500 Hz and the variation of interaural time differences above 500 Hz provide important cues for the perception of ASW. The analysis of these cues was outside the scope of the present study. However, the listeners mentioned a difference in timbre between stimuli with different ASW. A stimulus with a "brighter" timbre was often associated with a narrower sound source, such as in conditions with high IC values, particularly in the headphone-based Experiments 2 and 3. Furthermore, image splitting was reported which refers to the perception of two or multiple sources ([1]), in contrast to one fused sound image. While in the LS-based experiment (Experiment 1) only one listener reported image splitting, six listeners in Experiment 2 and two listeners in Experiment 3 reported this effect. In case of image splitting, subjects based their judgement on the overall ASW as they reported after the experimental procedure. Most subjects mentioned that the task of evaluating ASW was easiest for the LS presentation in Experiment 1. Nevertheless, it should be noted that the measured data showed remarkably stable results obtained with the chosen method.

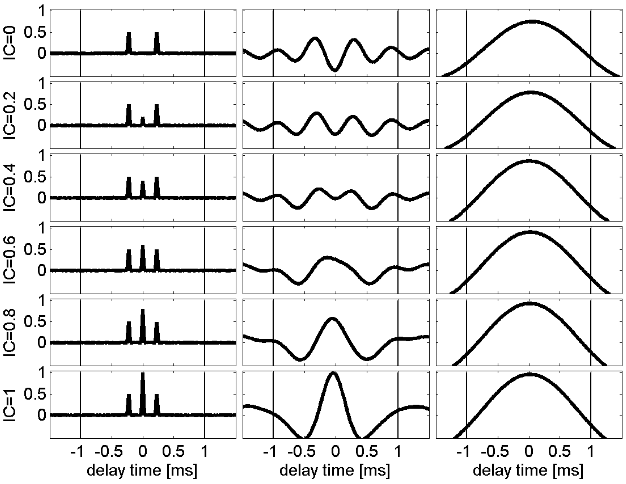

Figure 4: The normalised cross-correlation function $\rho_{lr}(\tau)$ in the conditions with cross-talk, for five values of IC and three stimulus conditions. Left: for a broadband signal; middle: for the noise band centered at 1 kHz from Experiment 3; and right: for the noise band centered at 0.25 kHz from Experiment 2.The maximal delay time interval of $|\tau| \leq 1$ms is indicated by the vertical lines.

Conclusions

A stereo set-up was used to study the perception of ASW for partially correlated bandlimited noise signals between the two loudspeakers and to test the performance of the IACC that is commonly used to predict ASW. The following results were obtained: 1) In a listening environment with low reverberation, ASW was found to be frequency independent for ICs of 0 and 0.6, in contrast to the case of dichotic HP presentations as investigated in [1], whereas a slight frequency effect was observed for IC = 1; 2) With and without cross-talk, the ASW decreased monotonically with increasing IC. The cross-talk generally caused a smaller ASW than the conditions without cross-talk, but the dynamic range of the ASW ratings was preserved. This was the case both for the listening room with a low reverberation time and the anechoic listening condition and was found to be independent of frequency. In the absence of cross-talk, the predictions based on the IACC were consistent with the measured data. However, in the presence of cross-talk, the IACC did not describe the data correctly. The change of ASW with IC was strongly underestimated at 0.25 kHz. At 1 kHz, the IACC based prediction did not account for the monotonic decrease of ASW with increasing IC. The discrepancies are caused by the interferences in the cross-correlation function introduced by the cross-talk delays. More advanced predictors are needed to correctly account for the measured data.

References

[1] Blauert, J. and Lindemann, W.: "Spatial Mapping of intracranical auditory events for various degrees of interaural coherence".
J.Acoust.Soc.Am. 79 (3), 1986

[2] Plenge, G.: "Über das Problem der Im-Kopf-Lokalisation".
Acustica, Vol.26, Heft 5, 1972

[3] Reference to the CIPIC HRTF database
URL: `http://interface.cipic.ucdavis.edu/sound/hrtf.html`

[4] Recommendation ITU-R BS.1534-1. "Method for the subjective assessment of intermediate quality level of coding systems".
The ITU Radiocommunication Assembly, 2001-2003.

[5] de Vries, Diemer, Hulsebos, Edo M. and Baan, Jan. "Spatial fluctuations in measures for spaciousness".
J. Acoust. Soc. Am. 110 (2), Aug. 2001

[6] van Dorp Schuitmann J.. "Auditory modelling for assessing room acoustics". PhD Thesis, Laboratory of Acoustical Imaging and Sound Control, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands, 2011.