The dependence of the spatial impression of sound sources in rooms on interaural cross-correlation and the level of early reflections.

Stefan Klockgether, Steven van de Par
Hearing4all cluster of excellence, Dept. of Medical Physics and Acoustics, Acoustics Group,
Carl von Ossietzky University, 26111 Oldenburg, Germany,
stefan.klockgether@uni-ol.de, steven.van.de.par@uni-ol.de

Introduction

The perceived sound of a source in a room is always influenced by the acoustical properties of that room. This influence of a room can be very disturbing when e.g. reverberation decreases the speech inteligibility in a class room, but it can also be very pleasant. Concert halls are built in way that they have a positive effect on the sound which is presented on their stages. The reflexions of the sound in a concert hall lead to a decorrelation of the sound between both ears of the listener. This reduction in interaural coherence leads to sources which are perceived broader than they phisically are [2] and to a stronger perception of envelopment by the sound in the concert hall [1]. Both the broader source and the stronger envelopment are perceived as very pleasant.
The physical acoustical quality of rooms is measured with room acoustical parameters like the reverberation time or the clarity index. Both values give a hint on the perceptual quality of a room. However, the connection between the perceived quality of a room and physical room acoustical parameters is not completely understood.
This study investigates the influence of direct manipulations to the physically measured room acoustic on the perceived acoustic of a room. Binaural room impulse responses (BRIR), which represent the physical acoustical information of a room, are used to determine which physical parameters affect the perception in rooms and how strong these effects are. BRIRs can be divided in two parts. A direction depending first part which includes the direct sound and the early reflections and a diffuse, reverberant tail which still has spatial but no perceivable directional information [4]. The first part is linked to the perceptual qualities of the source ("source presence") and the reverberant tail is linked to the room presence [3]. The cross-over between both parts is the perceptual mixing time (PMT) [6] with a lower threshold, where the diffuse tail starts to dominate in the impulse response and an upper threshold after which no directional information is perceivable.
Two different manipulations of the BRIR are investigated. The interaural cross-correlation (IACC) of the BRIRs is increased by cross-mixing the left and the right channel of the BRIR. The second manipulation is done by in- or decreasing the level of the early reflections in the BRIR. The influence on the perceived room acoustic is investigated in a psychoacoustical experiment. Therefore the manipulated BRIRs were convolved with anechoic music signals and subjects had to assess the apparent source width (ASW), the listener envelopment (LEV) and the presence of the source (PRE) of these stimuli. The subjective results of the different manipulations are compared among each other and to expectations from literature.

Method

Manipulations

The manipulation of the interaural cross-correlation was done by cross-mixing the left L and right R channel of the unmanipulated BRIR as it is shown in equation (1).

$$L' = \sum_f (L_f + \alpha R_f) \cdot \frac{\mathrm{RMS}(L_f)}{\mathrm{RMS}(L_f + \alpha R_f)} \,,$$
$$R' = \sum_f (R_f + \alpha L_f) \cdot \frac{\mathrm{RMS}(R_f)}{\mathrm{RMS}(R_f + \alpha L_f)} \,. \qquad (1)$$

L' and R' represent the manipulated left and right ear signals of the manipulated BRIR and f stands for ERB-sized frequency bands. The mixing parameter α controled the strength of the manipulation. After cross-mixing, a level normalization was made within each critical band to avoid coloration in the manipulated BRIRs. Please note that due to the level normalization a very small decorrelation may remain, even when α is equal to one.
The cross-correlation manipulation was applied only on the first part of the impulse response, only on the reverberant tail, or on the complete BRIR. A schematic of the manipulated parts is shown in figure 1. The border between the early part and the reverberant tail of the BRIR was chosen at the upper threshold of the PMT and the change-over was realised with cos^2-shaped ramps with a length of 4 ms. The used values for the mixing parameter were 0, 0.6, 0.8, 0.9, 0.95 and 1, whereas 0 would mean an unmanipulated signal and a mixing parameter of 1 would imply a fully correlated signal.
As second manipulation the level of the early reflections was either in- or decreased. Figure 2 shows a schematic where the manipulated parts are colored in blue and orange. The level of the BRIR was manipulated from 4 ms either up to the upper (blue) or lower (orange) threshold of the PMT. The direct sound and the reverberant tail were not manipulated. The change-over between the unmanipulated an the manipulated parts of the BRIR was realised with cos^2-shaped ramps with a length of 4 ms. The level of the early reflections was in- or respectively

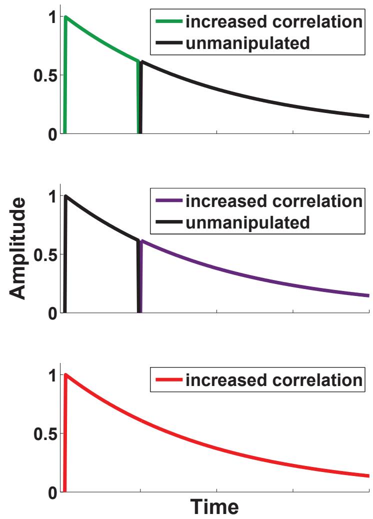

Figure 1: Schematic of the cross-correlation manipulation of the binaural room impulse responses (BRIR). In the colored parts, the cross-correlation between left and rigth ear channel was increased (see Eq. (1)). The first part, the reverberant tail or the complete length of the BRIR was manipulated. The changeover between the first part and the reverberant tail was chosen at the upper threshold of the perceptual mixing time [6].

decreased by 1, 3, or 10 dB and a condition with no manipulation was added as a reference to the experiment. It has to be mentioned, that no equalization of the overall level of the BRIR was done after the manipulation.

Stimuli and subjects

The stimuli used in the psychacoustical experiment were BRIRs convolved with anechoic music signals. A two second long random excerpt of a guitar-, violin or snare drum play was used for the convolution. The BRIRs were recorded with the FABIAN dummy head in a lecture hall (T_{60} = 1.7 s) and a seminar room (T_{60} = 0.8 s) of the TU Berlin and in a concert hall, the "Gewandhaus" in Leipzig, Germany (T_{60} = 2.3 s). The perceptual mixing times of the three rooms were determined with the model proposed by Lindau et al. 2012 [6] and their lower and upper thresholds were 60 and 120 ms in the lecture hall, 50 and 100 ms in the seminar room and 120 and 240 ms in the concert hall.
The stimuli were presented via headphones (Sennheiser HD 650) in a listening booth. The subjects had to rate the stimuli due to the perceived width of the source (ASW), the perceived envelopment by the sound (LEV) and the perceived presence of the source (PRE) with a slider. The slider was a grafical user interface in Matlab and it was operated with a computer mouse. The output of the slider were values between 0 (small source width, no envelopment, weak source) and 1 (wide source width, complete envelopment, dominant source). Only one perceptual attribute was evaluated during each run of the experiment. Within a run all conditions were repeated with all three instruments and the conditions were presented in a random order. Each run started with a short training phase which contained the conditions with the

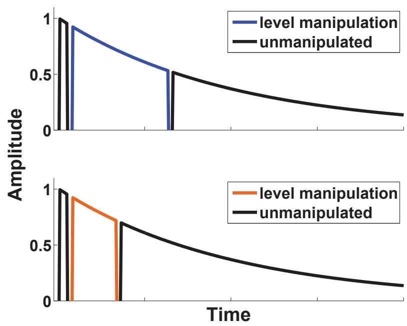

Figure 2: Schematic of the level manipulations of the BRIRs. The colored parts show the BRIR from 4 ms up to the upper or lower threshold of the PMT [6]. Within these parts the early reflections dominate the BRIR. To investigate the influence of the early reflections in a room on the perception, the level of the colored parts was either in- or decreased. The direct sound and the reverberant tail of the BRIR were not manipulated.

strongest manipulations. Nine normal-hearing subjects participated in the experiment and each subject repeated the experiment three times.

Results

The results of the subjective ratings are shown in the figures 3, 4 and 5. The results are averaged over the three runs and all three instruments. The graphics show mean values of all subjects with standard errors. The left panels show the results for the lecture hall, the middle panels for seminar room and the right panels for the concert hall. The upper panels always show the ratings for the cross-correlation manipulation as a function of the mixing parameter α. Results for the manipulation of only the first part are shown in green △, for the reverberant tail in purple ▽ and for the complete BRIR in red ○. The lower panels show the ratings for the manipulated level of the early reflections as a function of the strength of the level manipulation. The ratings for level manipulations up to the upper threshold of the PMT are shown in blue □ and in orange ◊ for the manipulations up to the lower threshold of the PMT. The red ○ in the lower panels corresponds to the value at a mixing parameter of 1 and a full length manipulation of the interaural cross-correlation in the upper panels. The values with a mixing parameter of 0 in the upper panels of course correspond to the values at a level manipulation of 0 dB in the lower panels.
The results for the apparent source width (ASW) (see Fig. 3) show that the perceived source width decreases for increasing cross-correlation. The manipulation of the IACC on the complete BRIR has the biggest impact on the ASW. The ASW also decreases when only the early part of the BRIR is manipulated but not as strong as for the full length manipulation. The correlation manipulation on only the reverberant tail of the BRIR has hardly any influence on the ASW. The manipulated level of the early reflections hardly affects the ASW, except for the 10 dB enhancement, which leads to a bigger ASW.

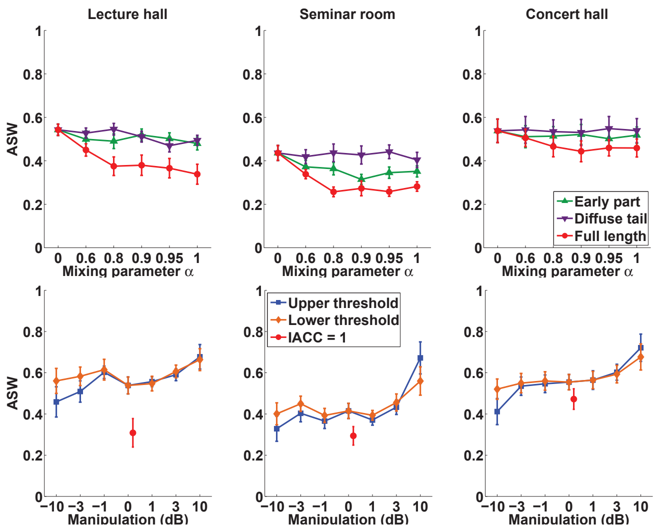

Figure 3: Mean values of the subjective ratings for the apparent source width (ASW) of all subjects with standard errors for the three different rooms. The upper panel shows the ratings for the manipulated IACC of the first part △, the reverberant tail ▽ or the complete ○ BRIR as a function of the mixing parameter α (see Eq. (1)). The lower panel shows the ratings for the in- or decreased level of the early reflections up to the upper □ or lower ◊ threshold of the PMT. The ○ in the lower panels corresponds to the value at $\alpha = 1$ on the full length BRIR in the particular upper panel.

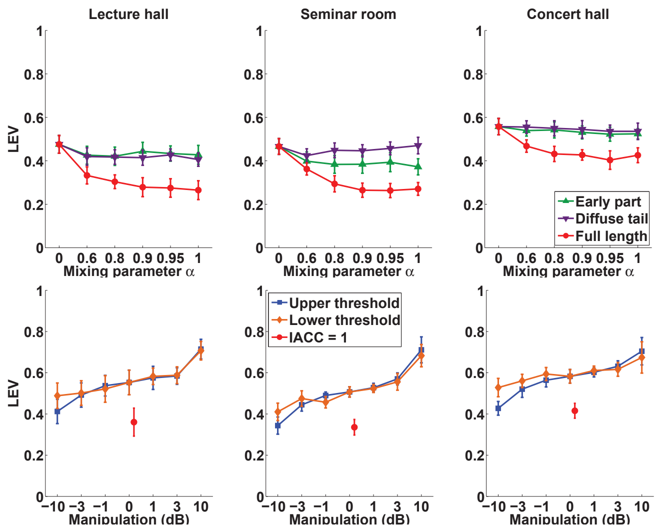

Figure 4: Mean values of the subjective ratings for the listener envelopment (LEV) of all subjects with standard errors for the three different rooms. The upper panel shows the ratings for the manipulated IACC of the first part △, the reverberant tail ▽ or the complete ○ BRIR as a function of the mixing parameter α (see Eq. (1)). The lower panel shows the ratings for the in- or decreased level of the early reflections up to the upper □ or lower ◊ threshold of the PMT. The ○ in the lower panels corresponds to the value at $\alpha = 1$ on the full length BRIR in the particular upper panel.

The results for the perceived listener envelopment (LEV) (see Fig. 4) show, that the LEV decreases with increasing cross-correlation. As for the ASW, the manipulation of the complete BRIR has by far the biggest effect on th LEV. If only the reverberant tail is manipulated, the LEV shows only a small decrease. The manipulation of the IACC of the early part seems to have the same or even a bigger effect on the LEV. The level of the early reflections seems to affect the LEV in a way that a higher level leads to a bigger LEV and a lower level of the early reflections reduces the LEV, though it seems not to matter for the LEV if the level is increased up to the lower or the upper threshold of the PMT. In both the upper and the lower panels, it can be seen, that the LEV in the concert hall is rated higher than the LEV in the two other rooms.

The results for the perceived presence of the source (PRE) (see Fig. 5) show, that the IACC has only a small effect on the PRE. The level of the early reflections however, strongly affects the PRE. Even small increases of the level lead to a more dominant source. A decrease of the level of the early reflections leds to a lower PRE-rating. As for the LEV-results, there seems to be no difference in the PRE, when the level manipulation is applied up to the upper threshold of the PMT than to the lower threshold. The PRE in the lecture hall is lower for all manipulations compared to the two other rooms. Both manipulations show a smaller effect on all three evaluated perceptual attributes in the concert hall than in the seminar room and the lecture hall.

Discussion

The reults of the subjective ratings of the perceptual attributes ASW and LEV agree with conclusions of Hidaka et al. (1995) [2] and Beranek (2008) [1]. Both perceptual attributes decrease with increasing IACC. However the manipulation of the complete impulse response has by far the biggest impact. The ASW and LEV were hardly affected when the IACC was only increased on the early part or the reverberant tail of the impulse response. The human auditory system seems not to be capable of distinguishing clearly between the IACC of the early part or the reverberant tail of a BRIR [5]. Figure 6 shows the resulting interaural coherence of the manipulated BRIRs of an exemplary violin stimulus used in the experiment as a function of the mixing parameter α. Manipulating only the reverberant tail (purple) only leads to a small increase in overall correlation, which may explain that the LEV and ASW was hardly affected by this manipulation. The manipulated early part (green) leads to almost the same increase of the interaural coherence as the complete manipulation does, but the sensitivity to correlation changes is very high at high correlation levels and decreases strongly for lower correlation values [7]. This could explain the small effect of the manipulated early part on the perceived ASW. The inability of the human auditory system to separately perceive the interaural coherence of different parts of a BRIR might also explain why ASW and LEV are often perceived similar in real acoustical environments.

The results show that there is still a small amount of perceived source width and envelopment, even when the

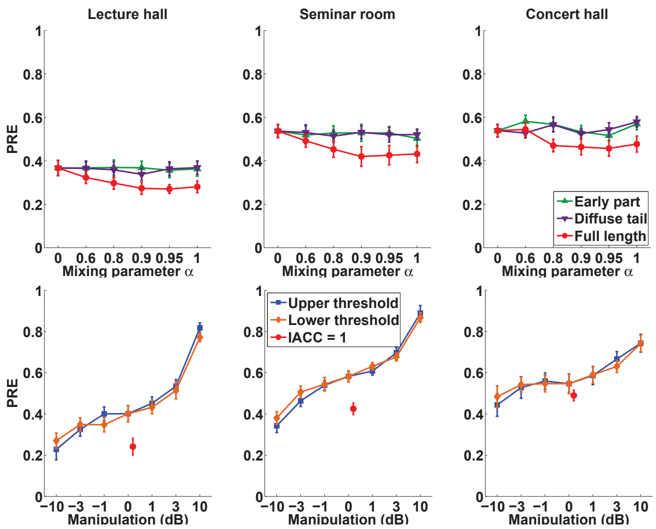

Figure 5: Mean values of the subjective ratings for the presence of the source (PRE) of all subjects with standard errors for the three different rooms. The upper panel shows the ratings for the manipulated IACC of the first part △, the reverberant tail ▽ or the complete ○ BRIR as a function of the mixing parameter α (see Eq. (1)). The lower panel shows the ratings for the in- or decreased level of the early reflections up to the upper □ or lower ◊ threshold of the PMT. The ○ in the lower panels corresponds to the value at $\alpha = 1$ on the full length BRIR in the particular upper panel.

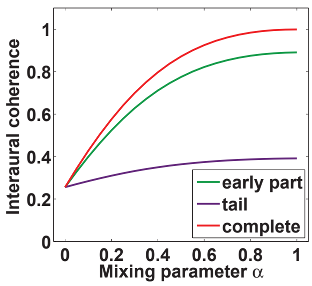

Figure 6: Interaural coherence of an exemplary violin stimulus from the experiment as function of the mixing parameter α. The correlation of either the first part, the reverberant tail or the complete BRIR was increased (see Eq. (1)).

signal is virtually fully correlated, which indicates that ASW and LEV are not only depending on the IACC. An additional cue for the estimation of ASW and LEV might be the spectral coloration in rooms due to reflections.
In addition the results show that LEV seems not only to depend on the reverberant tail of an BRIR but may also be affected by parts of the early part of the impulse response. The early reflections help listeners to estimate the size of a room which might indirectly or even directly affect the perceived envelopment.
The in- and decreased level of the early reflections has a big impact on the PRE, which may be explained by the direct to reverberant ratio of the BRIR. Since the early part of the impulse response is linked to the source, the manipulation of the level of the early reflections strongly changed the energy ratio between the source and the room. Therefore the source seemed to be much more dominant in the room, when the level of the early reflections was increased and the room was more dominant when the direct to reverberant ratio was decreased.

Conclusion

The human auditory system seems to be incapable to separate the interaural coherences of the early part from the interaural coherence of the reverberant tail of a BRIR. Both parts seem to mask each other which might explain, why apparent source width and listener envelopment are often perceived similar in real acoustical environments.
The perception of the presence of the source is strongly affected by the level of the early reflections. More energy in the early part of an impulse response, which is linked to the perception of the source, leads to a more dominat source in a room.

References

[1] Beranek, L. L.: Concert hall acoustics. J. Audio Engin. Soc. 56 (2008), 532-544

[2] Hidaka, T., Beranek, L. L., Okano, T.: Interaural crosscorrelation, lateral fraction, and low-frequency and highfrequency sound levels as measures of acoustical quality in concert-halls. J. Acoust. Soc. Am. 98 (1995), 988-1007

[3] Kahle, E.: Room acoustical quality of concert halls: perceptual factors and acoustic criteria - return from experience. International Symposium on Room Acoustics (ISRA 2013), Paper No. P074

[4] Klockgether, S., van de Par, S.: Limits for the perception of directional dependence of the reverberant tail in Binaural Room Impulse Responses (BRIR). Fortschritte der Akustik (2012), 231-232

[5] Klockgether, S., van de Par, S.: A model for the prediction of room acoustical perception based on the just noticeable differences of spatial perception. Acta Acust. united Ac. 100 (2014), 964-971

[6] Lindau, A., Kosanke, L., Weinzierl, S.: Perceptual Evaluation of Model- and Signal-Based Predictors of the Mixing Time in Binaural Room Impulse Responses. J. Audio Eng. Soc. 60 (2012), 887-898

[7] Walther A., Faller C.: Interaural correlation discrimination from diffuse field reference correlations. J. Acoust. Soc. Am. 133 (2013), 1496-1502