Bibliographical reference

A Model for the Prediction of Room Acoustical Perception Based on the Just Noticeable Differences of Spatial Perception, S. Klockgether and S. Van De Par, Acta Acustica **vol. 100** (Number 5), 2014, pp. 964-971

DOI
https://doi.org/10.3813/AAA.918776

Acta Acustica united with Acustica
Vol. 100 (2014) 964 – 971

DOI 10.3813/AAA.918776

A Model for the Prediction of Room Acoustical Perception Based on the Just Noticeable Differences of Spatial Perception

S. Klockgether, S. van de Par
Hearing4all cluster of excellence, Dept. of Medical Physics and Acoustics, Acoustics Group,
University of Oldenburg, Carl-von-Ossietzky-Str. 9-11, 26129 Oldenburg, Germany.
stefan.klockgether@uni-oldenburg.de

Summary

The accurate physical simulation of room acoustics is a very complex task. It is known that the human auditory system is limited in its ability to distinguish between subtle differences in the acoustics of a room. This is for example demonstrated by the observation that certain room acoustical parameters can be highly variable on very short distances while the human auditory system seems to be insensitive to the variations in the associated perception of room acoustical attributes. Therefore it seems attractive, to better understand the human auditory perception of room acoustics and to better know what properties of room acoustics are perceptually relevant. This would help to determine how accurate a room-acoustical simulation needs to be, i.e. to sound the same as the real room that is simulated.

In this study the influence of the interaural cross-correlation of the early part and the reverberant tail of a binaural room impulse response on spatial perception is surveyed by directly manipulating the impulse response. The subjective results are compared to the prediction by an auditory model.

A two-stage binaural psychoacoustic model is presented, to predict the perception of the acoustics of a room. The first stage extracts the binaural cues and is based on the limits of spatial perception. The just noticeable differences of the interaural cross correlation is used to tune the model. The second stage predicts several perceptual attributes which characterize the perception of room acoustical attributes. The model uses the information of binaural room impulse responses for the prediction of the apparent width of sources and the perceived listener envelopment.

PACS no. 43.55.Br, 43.55.Hy, 43.60.-c, 43.60.Jn, 43.64.-q, 43.64.Bt

1. Introduction

To characterize the acoustics of a room, the acoustical properties of a room can be measured physically. These room acoustical parameters are used to determine the acoustical quality of a room and allow inferences about the perception of sounds in the room. In real rooms the measurements of room acoustical parameters show high variations on very short distances [1], even on distances smaller than the human head size, room acoustical parameters can vary strongly. The perception of the acoustics of a room, however, does not seem to be varying a lot on short distances. Similarly, based on the variations in room acoustical parameters with small head movements one may expect a major impact on the perception of a sound in a room, but the auditory system seems to be very insensitive to such variations. Therefore the prediction of room acoustical perception needs to take this insensitivity of the auditory system into account.

The insensitivities are also of great interest for the fields of virtual acoustics and auralisation where it is tried to simulate the acoustics of rooms in a virtual acoustic reality. These simulations are very complex to perform and the better the reproduction should be, the more computational costs are required. Therefore it is interesting to know, where efforts can be saved without a perceivable loss in quality. A better understanding of the perception of room acoustics by the human auditory system helps to determine the perceptual relevance of the room acoustical properties.

Room acoustical perception is commonly assessed based on room acoustical perceptual attributes. Specifically spatial attributes have been shown to be important for the quality of room acoustics [2]. Important spatial parameters are apparent source width (ASW) and listener envelopement (LEV). For the evaluation of room acoustics a connection between the perceptual attributes and physically measurable parameters is of high interest. Mostly different perceptual attributes are linked to one room acoustical parameter. The direct-to-diffuse ratio of a sound for example affects not only the reverberance and clarity [3]

Received 19 May 2014,
accepted 05 July 2014.

of a sound, but also the distance perception [4]. Interaural time and level differences affect the azimuth location of a sound source and their variations also influence the apparent source width. The interaural cross-correlation is linked to the spaciousness of a sound by affecting its perceived source width [5], as well as the envelopment of the listener by the sound [6].

This correlations between perceptual attributes and the measurable room acoustical parameters are not always linear. While the human auditory system is highly sensitive to small variations in interaural cross-correlation for high correlations, it is very insensitive at low correlations. The perception of room acoustics depends on the combination of room acoustical parameters and multiple corresponding perceptual attributes. A prediction of the perceptual attributes needs to be adjusted to the linked room acoustical parameters. Van Dorp Schuitman *et al.* [3] suggested a model which is based on the separation of the signal into a direct and a reverberant stream. The model does a strict separation of both streams, extracts monaural and binaural room acoustical parameters from room impulse responses and predicts listener envelopment, apparent source width and perceptual equivalents to reverberance and clarity.

Kahle [7] suggests to group the perceptual attributes used to describe the acoustical quality of a room in two categories: “Source presence” which would be linked to the early part of the impulse response of a room and “room presence” which is linked to the late part the impulse response, the reverberant tail. The perception seems to be segregated in a source- and a room stream. The perception of apparent source width is linked to the early part of the impulse response and is one main attribute subjects link to source presence [8], whereas the perceived envelopment is linked to the late part [6], the room presence.

The results of a psychoacoustical experiment of this survey show that perceptual attributes which are linked to the reverberant part of a room impulse response also may be affected by the direct part and vice versa. This may indicate that listeners have difficulties to segregate the source stream from the room stream in real sounds. This paper suggests a model to predict the influence of interaural cross-correlation on the spatial impression, which includes a smooth changeover between direct sound and reverberation. The model uses the interaural cross-correlation and also takes the insensitivity of the human auditory system to changes in the interaual cross-correlation into account [9], by being motivated by just noticeable differences of cues for spatial perception. The analysis of the interaural cross-correlation is an established mechanism in binaural modeling and for example used by Stern and Colburn [10] to predict subjective lateral positions.

2. Method

2.1. Model

The suggested model shown in Figure 1 consists of two main stages. The first stage extracts binaural cues from a stereo input signal. It is framed by the dotted line in the figure. The second stage at the bottom of the figure predicts perceptual attributes which characterize the perception of room acoustical parameters.

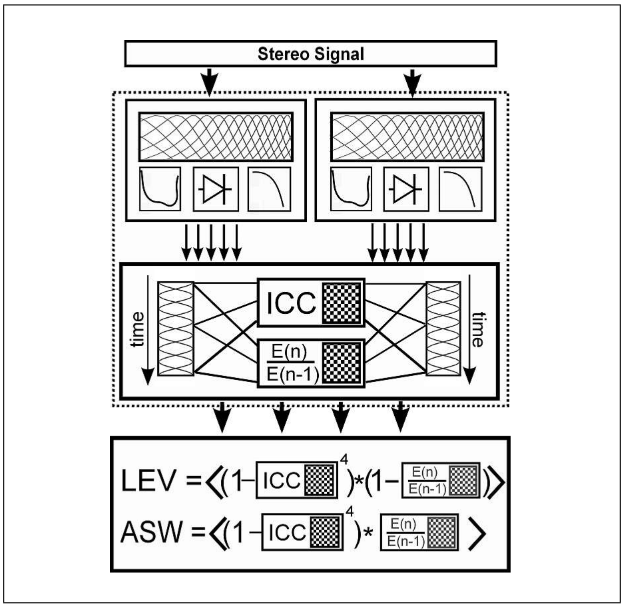

Figure 1. Schematic diagram of the structure of the suggested model. The dotted line frames the first part which extracts the cues of spatial perception from the input signal. The model extracts the interaural cross-correlation (ICC) and the ratio of the energy of each time segment divided by the energy of the previous segment. The second part at the bottom shows the prediction stage, which predicts the perceptual attributes listener envelopment (LEV) and apparent source width (ASW). The energy-ratio determines direct and diffuse parts of the signal and is used for weighting the ICC.

2.1.1. Extraction of cues

The first stage contains stages representing processing in the auditory periphery for the left and right ear channel including the frequency processing of the inner ear and the binaural processing stage. Auditory filters are realised with 42-channel filterbanks of 4th order gammatone filters followed by a half-wave rectification with a slightly increased lower threshold, simulating the inner hair-cells. The threshold cuts of all signal components below the lowest perceiveable level and is constant for all frequency channels. Each frequency channel is filtered by a 4th-order low-pass filter with a cut-off frequency of 1.5 kHz to just extract envelopes at high frequencies, which resembles the loss of phase locking above 1.5 kHz.

The left and right ear monaural stages each transfer the processed data to the binaural processor, which divides the signal in segments of 40 ms. The segments are multiplied with a Hann-window and overlap for 20ms with each adjacent segment. Each two corresponding segments from the left and right ear input are compared. The RMS level of both segments is calculated and their mean is stored as the level of that segment. The normalised cross-correlation function of the corresponding segments for left and right ear is calculated. A weighting function suggested by Stern and Shear [11] is used, which guarantees that the cross-correlation is extracted within the limits of the perception of realistic acoustical signals. The maximum of the weighted cross-correlation function is stored as the interaural cross-correlation coefficient of that segment. The binaural processing stage transfers a matrix with cross-correlation coefficients ($\mathrm{ICC}(n, c)$) and a matrix with levels fo r each frequency channel (c) and each time segment (n) to the prediction stage.

2.1.2. Prediction stage

The prediction stage consists of three parts. The first part calculates a weighting matrix (P) with the likelihoods of each channel (c) and time segment (n), containing direct sound. For that purpose the level of each segment (E_n) is divided by the level of the previous segment (E_{n-1}). The resulting ratio determines whether the segment contains direct sound or is dominated by reverberation. This is done within each frequency channel. The estimation of the amount of direct sound and reverberation in a time segment is simply the ratio of the level of the current time segment divided by the level of its previous one. The equation (1) shows the calculation for each entry ($P(n, c)$) of the weighting matrix

$$P(n, c) = \begin{cases} 1 & \text{if } E_n \geq E_{n-1}, \\ \sin^2\left(\tilde{E}_n \cdot \frac{\pi}{2}\right) & \text{if } E_{0.2} < E_n < E_{n-1}, \\ 0 & \text{if } E_n \leq E_{0.2}, \end{cases} \tag{1}$$

with

$$\tilde{E}_n = \frac{\frac{E_n}{E_{n-1}} - E_{0.2}}{1 - E_{0.2}}.$$

A ratio of one or higher implies a likelihood of one for containing direct sound and a likelihood of zero for containing only reverberation (cf. equation (1) first case). The ratios will be small, when a segment contains no direct sound and the smallest values will indicate the diffuse to dircet ratio of the sound in the room. Therefore the mean value of the twenty percent of lowest ratios ($E_{0.2}$) is calculated and used as the upper threshold for containing only reverberation. Every segment with a ratio lower than that threshold has a likelihood of one for containing only reverberation and a likelihood of zero for containing direct sound (cf. equation (1) third case). The smooth cross-over between a likelihood of one and a likelihood of zero is realised with a $\sin^2$-slope for both predictions of the segment containing direct sound or only reverberation (cf. equation (1) second case). The sum of both likelihoods for each segment is always one. The $\sin^2$-slope is motivated by the slope of psychometric functions and respects the reliability of the prediction for each time segment. The likelihood of the first segment of each frequency channel for containing direct sound is set to one, thus the likelihood for containing only reverberation is zero. The resulting matrixes consist of values between one and zero and are used for weighting specific parts of the signal when predicting perceptual attributes.

The model makes a prediction for the perceived width of a sound source (Apparent Source Width, ASW, cf. equation (3)) and the perceived envelopment by the sound (Listener Envelopment, LEV, cf equation 2). Both predictors are using the weighting matrix ($P(n, c)$, cf. equation 1) with likelihoods for each frequency channel and time segment of a signal containing direct sound or only reverberation. The apparent source width as well as listener envelopment prediction are based on the interaural cross-correlation of a sound signal. Both perceptual attributes are rated higher, when the interaural correlation of a sound signal is low. The matrix with the interaural cross-correlation coefficients of the binaural processing stage is used for the calculation of the prediction. The values in the matrix are raised to the power of four to take the high sensivity of the human auditory system at high correlation values and the insensitivity at low correlation values into account. Without this modification to the cross-correlation matrix, the model would be equal sensitive to correlation changes at all correlation levels. That would overestimate the sensitivity at low correlation levels and underestimate the sensitivity at high correlation levels. Equations (2) and (3) show the calculation of the predictions for ASW and LEV,

$$\mathrm{LEV} = 1 - \frac{\sum_{n,c} \mathrm{ICC}^4(n, c) \cdot (1 - P(n, c))}{\sum_{n,c} (1 - P(n, c))}, \tag{2}$$

$$\mathrm{ASW} = 1 - \frac{\sum_{n,c} \mathrm{ICC}^4(n, c) \cdot P(n, c)}{\sum_{n,c} P(n, c)}. \tag{3}$$

The perception of envelopment is linked to the reverberance of a sound signal. Thus the segments containing a lot of reverberation are weighted stronger compared to the ones containing mostly direct sound. The modified cross-correlation matrix ($\mathrm{ICC}^4(n, c)$) is multiplied with the weighting matrix ($1 - P(n, c)$), which contains high values for reverberant segments. The predicted listener envelopment is one minus the weighted mean value of the resulting matrix.

Apparent source width is linked to the earlier parts of a room impulse response. The segments with a high likelihood of containing direct sound are weighted stronger. The modified cross-correlation matrix ($\mathrm{ICC}^4(n, c)$) is multiplied with the weighting matrix ($P(n, c)$) containing high values for segments, which contain direct sound. The prediction of the perceived width of the source is calculated as one minus the weighted mean value of the resulting matrix.

Segments which contain only zeros due to no present information of the stimulus, would have a correlation value of zero and are not excluded from the average determination. This leads to a higher prediction of LEV and ASW values since it decreases the average of the correlation matrix.

2.2. Experiment

A psychoacoustic experiment has been conducted to investigate the influence of the interaural cross-correlation on the perception of envelopment and source width. As stimuli, binaural room impulse responses (BRIR), which were recorded in a seminar room and a lecture hall of the TU Berlin, have been convolved with anechoic music signals of a guitar, a violin and a snare drum excerpt. The lecture hall had a volume of 5179 m^3 and an average reverberation time of $T_{60} = 1.67s$. The seminar room had a volume of 182 m^3 and an average reverberation time of $T_{60} = 0.79s$ (cf. [12]). A three-way dodecahedron loudspeaker was used as source and the impulse responses have been recorded with a FABIAN dummy head. The distance between source and receiver was chosen to be twice the critical distance [13].

For investigating the effect of manipulating the interaural cross-correlation, cross-mixing of the left and right ear signal was used for each ERB band independently as shown in equation (4),

$$L' = \sum_f (L_f + \alpha R_f) \frac{\mathrm{RMS}(L_f)}{\mathrm{RMS}(L_f + \alpha R_f)},$$

$$R' = \sum_f (R_f + \alpha L_f) \frac{\mathrm{RMS}(R_f)}{\mathrm{RMS}(R_f + \alpha L_f)}. \qquad (4)$$

L and R are the left and right signal of the unmanipulated impulse response. L' and R' are the manipulated left and right ear signals of the impulse response and f stands for ERB-sized frequency bands. The mixing parameter α controls the strength of the manipulation and its values were chosen to be 0, 0.6, 0.8, 0.9, 0.95 and 1. The value $\alpha = 1$ would imply a fully correlated signal, whereas $\alpha = 0$ matches the unmanipulated signal. After cross-mixing, a level normalization was made per critical band to make sure that the spectral envelopes of the manipulated BRIRs were the same as the original BRIRs. Please note that due to the level normalization that is applied to avoid coloration of the left and right ear signals, a small decorrelation remains, even when α is equal to one. This decorrelation is, however, very small.

For the experiment either the early part up to the perceptual mixing time, the reverberant tail after the perceptual mixing time or the complete impulse response was manipulated. The early part of a binaural room impulse response contains strong directional information, whereas the reverberant tail still contains spatial but only diffuse information [14]. The perceptual mixing time marks the changeover between both parts [12]. Based on the results of a previous experiment [14] the mixing times were set at 120 ms for the lecture hall and at 100 ms for the seminar room. If just one part was manipulated, the change-over between both parts was realised with cosine shaped ramps, which had a length of 4 ms.

The resulting interaural cross-correlation of the complete stimuli can be seen exemplary in Figure 2 as a function of the mixing parameter α. The graphic shows that manipulating the complete impulse response (continuous line) results in a very high correlation which gets very close to one for high mixing parameters (≈ 0.998). Manipulating the early part (dotted line) strongly increases the overall cross-correlation of the stimulus whereas a manipulation of the reverberant tail (dashed line) has only a small impact on the overall cross-correlation of the stimulus.

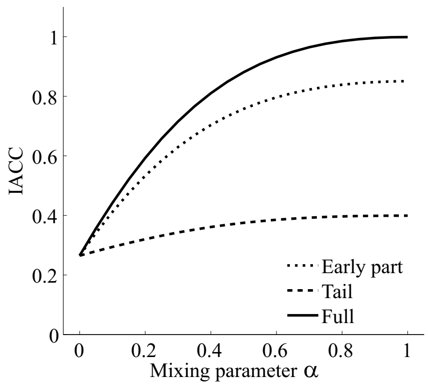

Figure 2. Interaural cross-correlation coefficient (IACC) resulting of an exemplary violin stimuli from the experiment as function of the mixing parameter α. The correlation of either the early part (dotted line), the reverberant tail (dashed line) or the complete (continuous line) binaural room impulse response was manipulated.

For the experiment all possible combinations of rooms, manipulations, instruments and manipulation strength were investigated, resulting in 96 different conditions which were rated during each run of the experiment in a random order. At the beginning of each run a set of training sounds was presented. It contained the unmanipulated condition and the three conditions with a mixing parameter of one for all three instruments. The subjects had to rate the signals concerning listener envelopment and apparent source width. For the rating a Matlab graphical user interface with a slider was used. The subjects have been asked: “How much do you feel enveloped by the sound” with a simple slider and the extrema “Not at all” and “Completely”, or “How wide was the source” with a mirrored double slider and the extrema “Small” and “Wide”. The double slider was an additional visual feedback to facilitate the impression of source width.

The subjects were allowed to listen to each condition as often as they wanted to. The sound signals were random excerpts from longer excerpts of musical plays and were four seconds long. The random excerpts where convolved with the BRIRs and the stimulus was cut after the four seconds. Therefore the reverberant tail was never presented in such a way that it can be heared solely until it became inaudible. Random excerpts where chosen to avoid that the specific properties of the selected excerpt would bias the results. All stimuli were faded in and out with Hann-ramps of 100 ms length.

During a run, either the listener envelopment or the apparent source width was rated and each subject did three runs for each perceptual attribute. For this study it is assumed that the concepts of listener envelopment and apparent source width, which were developed to describe the acoustics of concert halls, can also be used to determine the perception of the acoustics of more common rooms.

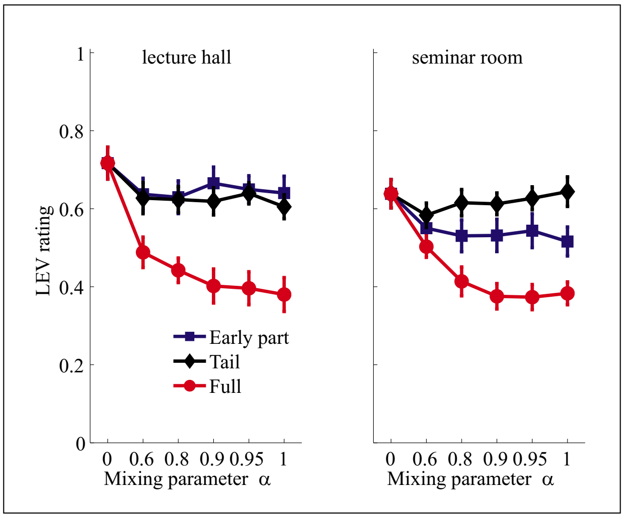

Figure 3. Perceived listener envelopment (LEV) of music signals in two different rooms as a function of the strength of a manipulation increasing the interaural cross-correlation (ICC) of the rooms binaural room impulse responses (BRIR). Squares show the rating for a manipulation of the early part of the BRIR, diamonds for manipulating the reverberant tail and circles for increasing the ICC for the full BRIR. Error bars show standard errors.

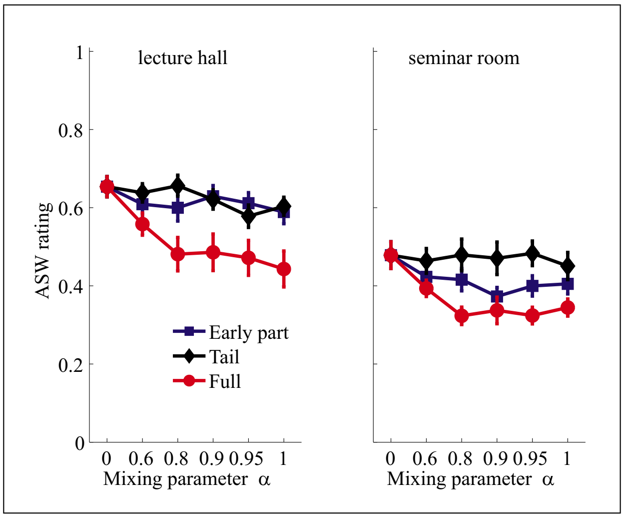

Figure 4. Perceived width of source (ASW) of music signals in two different rooms as a function of the strength of a manipulation increasing the interaural cross-correlation (ICC) of the rooms binaural room impulse responses (BRIR). Squares show the rating for a manipulation of the early part of the BRIR, diamonds for manipulating the reverberant tail and circles for increasing the ICC for the full BRIR. Error bars show standard errors.

This assumption has also been made by Van Dorp Schuitman *et al.* [3]. Eleven normal hearing subjects participated in the experiment. The sounds were presented with a open headphone (Sennheiser HD 650) in a hearing booth at a level of 65 dB SPL. The signals were processed in Matlab and sound playback was done by a USB Soundcard (RME Fireface UC).

3. Results

3.1. Psychoacoustic experiment

The results of the experiment are shown in the Figures 3 and 4. The mean results of all subjects, instruments and runs are plotted with standard errors as a function of the mixing parameter α. With increasing α the strength of the manipulation and thus the cross-correlation increases. A mixing parameter of zero means no manipulation and the values are used as references. On the left side of each figure the ratings for the lecture hall are shown and on the right side the ones for the seminar room. The blue squares show the results for the manipulation of the early part of the impulse response, the black diamonds for the reverberant tail. The results for the manipulation of the full length of the room impulse response are shown by the red circles.

Figure 3 shows the rating of the perceived listener envelopment for the different conditions. The higher the value, the higher was the perceived envelopment. A rating of one implies a complete envelopment and a rating of zero no perceived envelopment at all. Results show that the manipulation of the full length of the impulse response reduces the perceived envelopment in both rooms as a monotone function of the manipulation strength. Compared to the reference, the rating of the full manipulated condition is up to 0.4 lower in the lecture hall and up to 0.3 lower in the seminar room. The subjects still rate the perceived envelopment of stimuli, which left and right channel were completely mixed with values between 0.3 and 0.4. The manipulation of the early part leads to a small decrease in listener envelopment ratings. The tail manipulation only seems to decrease the perception of envelopment in the lecture hall. Statistical significant differences are found between the ratings of the reference and the ratings of the manipulations on the full length of the room impulse response in the lecture hall data as well as in the seminar room data.

Figure 4 shows the rating of the perceived width of source for the different conditions. The higher the rating, the wider the source was perceived. A rating of one would imply a source which is perceived in the complete frontal hemisphere. A rating of zero would imply a source, which perceived width is only a few degrees. The results show a general 0.2 higher rating for the source width in the lecture hall than in the seminar room. It can also be seen that an increase in the interaural cross-correlation on the full length of the impulse response decreases the perceived width of a source. A small decrease in apparent source width may also be seen for the manipulation of only the early part, but not for the tail manipulation in the seminar room.

3.2. Model predictions

The proposed model (see section 2.1) was used to predict the perceptual attributes listener envelopment and apparent source width. The predictions of the ratings are shown in the figures 5 and 6. The mean values of three runs and the three instruments are plotted with standard errors as a

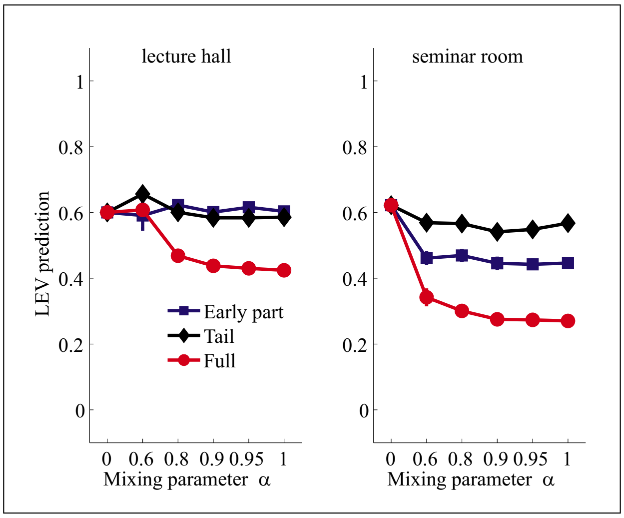

Figure 5. Predicted listener envelopment (LEV) of music signals in two different rooms as a function of the strength of a manipulation increasing the interaural cross-correlation (ICC) of the rooms binaural room impulse responses (BRIR). Squares show the rating for a manipulation of the early part of the BRIR, diamonds for manipulating the reverberant tail and circles for increasing the ICC for the full BRIR. Error bars show standard errors.

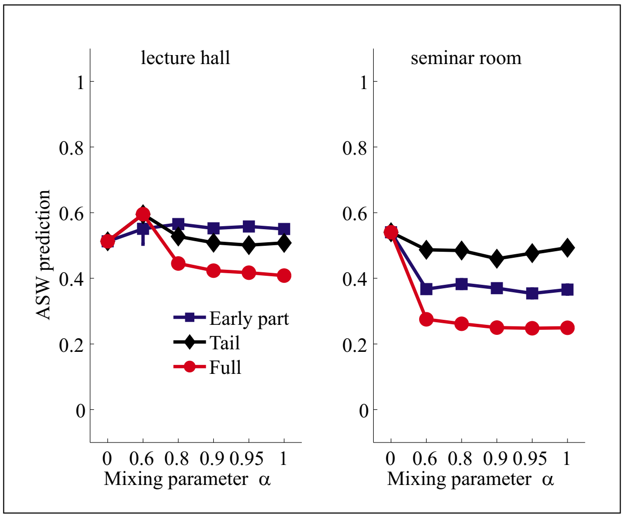

Figure 6. Prediction of the perceived source width (ASW) of music signals in two different rooms as a function of the strength of a manipulation increasing the interaural cross-correlation (ICC) of the rooms binaural room impulse responses (BRIR). Squares show the rating for a manipulation of the early part of the BRIR, diamonds for manipulating the reverberant tail and circles for increasing the ICC for the full BRIR. Error bars show standard errors.

function of the mixing parameter α. With increasing α the strength of the manipulation and thus the cross-correlation increases. A mixing parameter of zero means no manipulation and the values are used as references. On the left side of each figure the predictions for the lecture hall are shown and on the right side the ones for the seminar room. The blue squares show the predictions for the manipulation of the early part of the impulse response, the black diamonds for the reverberant tail. The predictions for the manipulation of the full length of the room impulse response is shown by the red circles.

Figure 5 shows the predictions for the perceived listener envelopment. A high rating corresponds to a high predicted envelopment. The model predicts up to 0.4 lower ratings for the full length manipulation compared to the unmanipulated condition. A stronger manipulation leads to a lower predicted rating. The tail manipulation seems to have no impact on the prediction as well as the early part manipulation in the lecture hall, but the early part manipulation in the seminar room decreases the predicted envelopment rating for about 0.15 compared to the reference. The lowest predicted values for listener envelopment are 0.4 for the lectur e hall and 0.3 for the seminar room.

Figure 6 shows the predicted rating for apparent source width. A high value corresponds to the prediction of a wide source in azimuth. The general predicted rating is about 0.1 to 0.2 higher in the lecture hall than in the seminar room. The predictions show a strong decrease in the rating up to 0.2 in the lecture hall and up to 0.4 in the seminar room with increasing manipulation strength for the full length manipulation. The early part manipulation also decreases the predicted apparent source width in the seminar room.

The calculations shown in the equations 2 and 3 would lead to predicted values of zero for LEV and ASW for a fully correlated BRIR but for the inclusion of segments which contain no information. These segments have a correlation value of zero and decrease the average of the correlation matrix. Therefore even a fully correlated BRIR would not lead to an average correlation value of one.

4. Discussion

The results of the psychoacoustic experiment (Figure 3 and Figure 4) confirm that the spatial impression of a sound in a room is influenced by the interaural cross-correlation. With increasing cross-correlation the ratings for listener envelopment as well as for apparent source width decrease. An increase of the correlation on the full length of the binaural room impulse response has, as expected, the strongest influence on the perception. The subjects still seem to feel a little enveloped by a sound, even when the intearaural cross-correlation of the complete signal is almost one, which would match a mono-signal. This might be caused by the reverberation, which still suggests a spatial impression of the room.

The perceived listener envelopment shows just a small decrease in the lecture hall and no decrease in the seminar room when only the reverberant tail is manipulated although listener envelopment has been suggested to be linked to the later part of an impulse response [6]. Interestingly a manipulation of the early part also leads to a decreased perception of envelopment. One explanation for this could be that in the running stimulus that is presented, the reverberant tail is virtually never present in isolation, and always mixed with direct sound and early reflections.

The auditory system may not be able to separate these different parts of the binaural room impulse response very well. In addition, the results that were found could also be influenced by the chosen perceptual mixing times, which separate the directional information containing early part of the impulse response from the reverberant tail. Compared to a model to predict perceptual mixing times proposed by Lindau *et al.* [12]), the mixing times were chosen in the range of the higher prediction that is proposed in that study. Thus the early part of the reverberant tail is contained in the first 100 respectively 120 ms of the used impulse responses. Since the lecture hall is more reverberant than the seminar room, it has also a higher ratio of energy in the reverberant tail divided by the energy of the early part of the impulse response than the seminar room. This may also explain the stronger decrease in listener envelopment rating of the blue graph for the early part manipulation compared to the black graph for the tail manipulation in the seminar room (Figure 3). A follow up study which invetigates the perception of listener envelopment as a function of the mixing time could clarify the hypothesis.

The apparent source width is less influenced by the manipulations than the listener envelopment. This supports the thesis that perceived source width is not only based on interaural cross-correlation [15]. In the lecture hall the manipulation of the early part of the room impulse response, including early and late reflections, has hardly any effect on the perceived source width (Figure 4). An influence to the apparent source width rating can also be seen for the early part manipulation in the seminar room but not in the lecture hall. The low interaural correlation of the tail may mask the higher correlation of the early part. This effect would be stronger in a more reverberant room like the lecture hall, which has a higher reverberation time than the seminar room and thus more energy in the reverberant tail. The generally higher rating of the apparent source width in the lecture hall could be caused by a smaller direct to diffuse ratio which makes localization more difficult. In addition the bigger distance from the listener to the source in the bigger room might increase that effect.

The predictions of the model, which uses just the interaural cross-correlation and a direct-diffuse weighting show generally the same tendencies as the experimental results. The predicted ratings for the listener envelopment are matching quite well with the ratings of the subjects. The range of the results and also the difference between the different manipulations are predicted correctly. The influence of a correlation increase due to a medium high mixing parameter (0.6) is underestimated in the lecture hall and overestimated in the seminar room. Tested with a simple covariance analysis, the modeling approach is able to account for about 78% of the variance in the lecture hall and about 94% in the seminar room. The predictions of the apparent source width deviate stronger from the experimental results. Correct tendencies are visible but especially the prediction in the seminar room resembles more to the listener envelopment data. The model approach can resolve 88% percent of the variance in the seminar room and only about 62% in the lecture hall. This is also a hint that the apparent source width cannot be predicted by only looking at the interaural cross-correlation. The direct-diffuse ratio has to be taken into account. Another idea would be to include the strength of the early reflections to predict ASW.

The influence of the ICC manipulation of the late part of the BRIR on the ASW and of the manipulation of the early part of the BRIR on the LEV, might be related to the fact that early and late parts of the BRIR tend to be strongly overlapping for running music signals. As a consequence, the auditory system would not be able to separetely process the ICC of the early and late part of the BRIR very well.

5. Summary and conclusion

This study investigated the connection between measureable room acoustical parameters and perceptual attributes. This is important to characterize the perception of room acoustics, especially with regard to the quality of virtual acoustics. A listening experiment was conducted to investigate the influence of an increase of the interaural cross-correlation on the perception of listener envelopment and apparent source width. Subjects had to rate the listener envelopment and source width of different music plays in reverberant rooms. Different parts of the room impulse responses have been manipulated for the experiment. With increasing cross-correlation the ratings of the perceptual attributes listener envelopment and apparent source width decreased. The increase of the cross-correlation on the full length of a BRIR induces the strongest decrease in the ratings. The results of the experiment also show that the manipulation of the reverberant tail of an impulse response also affects the perception of source width and that late and maybe also early reflections have an impact on the perceived listener envelopment.

The paper also presented a model to predict listener envelopment and apparent source width. The model extracts the interaural cross-correlation on a perceptually motivated time-frequency grid and uses this to predict the spatial impression. It also estimates the likelihood of each part of the impulse response to be direct or diffuse sound. This likelihood is used to weight the cross-correlation coefficients. Direct sound is weighted strong for the prediction of apparent source width and the diffuse parts are weighted stronger when the listener envelopment is predicted. The model is able to predict the tendencies of the influence of cross-correlation on the two surveyed perceptual attributes correctly. Especially the prediction of listener envelopment matches quite well the experimental results. The prediction of apparent source width deviates partially from the experimental data. This could be a hint for other dependencies of the apparent source width in addition to the interaural cross-correlation.

Acknowledgement

This work was supported by the DFG Research Unit SEACEN.

The authors would like to thank the Audio communication group of the TU Berlin for the recordings of the binaural room impulse responses.

References

[1] D. de Vries, E. M. Hulsebos, J. Baan: Spatial fluctuations in measures for spaciousness. Journal of the Acoustical Society of America **110** (2001) 947–954.

[2] M. R. Schroeder, D. Gottlob, K.F. Siebrass: Comparative study of european concert halls - correlation of subjective preference with geometric and acoustic parameters. Journal of the Acoustical Society of America **56** (1974) 1195–1201.

[3] J. van Dorp Schuitman, D. de Vries, A. Lindau: Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model. Journal of the Acoustical Society of America **133** (2013) 1572–1578.

[4] P. Zahorik, D. S. Brungart, A. W. Bronkhorst: Auditory distance perception in humans: A summary of past and present research. Acta Acustica United With Acustica **91** (2005) 409–420.

[5] T. Hidaka, L. L. Beranek, T. Okano: Interaural cross-correlation, lateral fraction, and low-frequency and high-frequency sound levels as measures of acoustical quality in concert-halls. Journal of the Acoustical Society of America **98** (1995) 988–1007.

[6] L. L. Beranek: Concert hall acoustics - 2008. Journal of the Audio Engineering Society **56** (July 2008) 532–544.

[7] E. Kahle: Room acoustical quality of concert halls: perceptual factors and acoustic criteria – return from experience. International Symposium on Room Acoustics, Toronto, Canada (2013).

[8] A. Haapaniemi, T. Lokki: Identifying concert halls from source presence vs room presence. Journal of the Acoustical Society of America **135** EL311 (2014).

[9] S. Klockgether, J. van Dorp Schuitman, S. van de Par: Perceptual limits for detecting interaural-cue manipulations measured in reverberant settings, POMA **19** (2013) 015004.

[10] R. M. Stern, H. S. Colburn: Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position. Journal of the Acoustical Society of America **64** (1978) 127-140.

[11] R. M. Stern, G. D. Shear: Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay. Journal of the Acoustical Society of America **100** (1996) 2278–2288.

[12] A. Lindau, L. Kosanke, S. Weinzierl: Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses. Journal of the Audio Engineering Society **60** (2012) 887–898.

[13] H. Kuttruff: On the audibility of phase distortions in rooms and its significance for sound reproduction and digital-simulation in room acoustics. Acustica **74** (1991) 3–7.

[14] S. Klockgether, S. van de Par: Limits for the perception of directional dependence of the reverberant tail in binaural room impulse responses (BRIR). Fortschritte der Akustik (2012) 1621–1623.

[15] T. Okano, L. L. Beranek, T. Hidaka: Relations among interaural cross-correlation coefficient (IACC(E)), lateral fraction (LFE), and apparent source width (ASW) in concert halls. Journal of the Acoustical Society of America **104** (1998) 255–265.