Auralizing Listener Position Shifts of Measured Room Impulse Responses

Christoph Pörschmann, Philipp Stade

Technische Hochschule Köln, Institut f. Nachrichtentechnik, 50679 Köln,
Email: Christoph.Poerschmann@th-koeln.de; Philipp.Stade@th-koeln.de

Introduction

The auralization of acoustic environments based on dynamic binaural synthesis can be used for various applications. This approach can be applied in areas like audio engineering, telecommunication, or architectural acoustics to create a natural and plausible room impression. For a headphone based auralization, circular sets of binaural room impulse responses (BRIRs) can be acquired by performing measurements with a rotated dummy head. By considering the listener's head movements in the auralization, localization accuracy increases and externalization of virtual sound sources improves.

Several commercial or scientific rendering engines are available which allow adapting the sound field presented via headphones according to the head movements in real time (e.g. [1]). However, applications using this technology in everyday life are still rare. One reason is the complex procedure for the acquisition of spatial sound fields which cannot be conducted with standard equipment and techniques. This effort increases even more when dealing with different listener positions as one specific set of BRIRs needs to be measured for every position.

Recently, we presented an approach to synthesize circular sets of BRIRs based on a single omnidirectional room impulse response (RIR) [2]. Direct sound, early reflections and diffuse reverberation are extracted from the RIR and treated separately. Spatial information is added according to generic geometric assumptions and by exploiting certain psychoacoustic effects. In [3] this approach has been evaluated in a series of listening experiments and it was shown that this approach is sufficient to obtain a plausible representation of the room.

In this study an enhanced model is presented which allows considering the listener position modifications. Based on a simple estimated geometric model, direct sound and reflections will be modified with the result that the listener position can be shifted freely. This becomes possible because relevant parts of the measured impulse response are transferred into a parameter-based description of the room.

Generally speaking, a parametrization of measured room impulse responses offers various possibilities. As described in detail in this paper the parametrization can be used to shift listener positions or to add spatial information to an omnidirectional RIR. Other prospects of such a parametric description are modifications of the room's acoustic properties regarding single reflections or the diffuse reverberation. Finally the parametrization has the potential of reducing the data size of a BRIR or a circular set of BRIRs.

Basic idea of the approach

The aim of the algorithm described here is the synthesis of circular sets of BRIRs, allowing the adaptation of the listener position in the room. We determined all sets of BRIRs from one single omnidirectionally measured RIR. Therefore, based on geometric acoustics predictable information from a measured RIR is extracted and is used for the synthesis. For the reconstruction of the diffuse reverberation, knowledge regarding the perception of diffuse sound fields is applied.

The basic approach is as follows: The measured RIR is split into different parts. In the early part the temporal structure of the direct sound and the first reflections is detected by analyzing peaks and dips of the omnidirectional RIR. However, directions of incidence cannot be estimated based on the omnidirectional RIR and need to be guessed. Then the room reflections are modified in order to apply distance modifications between listener and receiver. Therefore, as shown in Figure 1, a simple model of the determined reflections based on the mirror image method is created. Amplitudes, directions of incidence and times of incidence are recalculated according to the modified listener position. Thus a binaural synthesis of the early geometric reflected part can be applied.

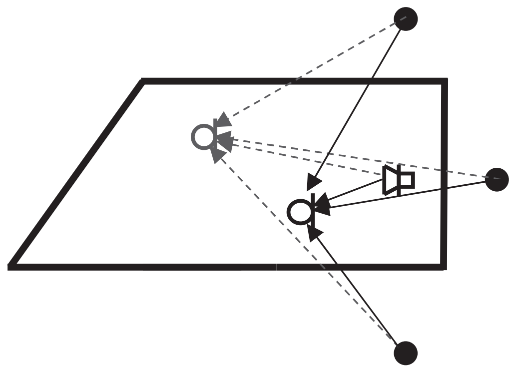

Figure 1: Basic principle of modification of the early reflections applying mirror images. The figure demonstrates the modification of the direct sound and the early reflections. The receiver is moved from an initial position (grey) to a modified position (black). By this the amplitudes and the temporal structure of direct sound and reflections are changed.

The diffuse reverberation part is not affected by a modification of the listener position and is considered to reach the listener temporarily and spatially equally distributed. Several studies have shown that the part of the measured impulse response after a perceptual mixing time can be simplified [4-6]. Even though an exact reconstruction of the temporal and spatial structure is ideally not required [4], small perceptual differences might remain [6]. Several studies have shown that the mixing time is room dependent and can be chosen according to data-based predictors or geometric room properties [5]. In this investigation, a smooth crossover between the early reflection part and the diffuse part was chosen. Binaural properties were added to the diffuse reverberation of the omnidirectional RIR.

The approach presented here holds a number of sources of inaccuracies. Because information about the binaural sound field is absent in the omnidirectional RIR, spatial properties as for example the source width cannot be rebuilt correctly. Furthermore, the directions of incidence of the synthesized early reflections are not in line with the original ones. This leads to perceptual differences between the synthesized BRIR and a corresponding measured one [3].

Implementation

The algorithm without a consideration of listener shifts has already been described in [2, 3]. As input data the omnidirectional RIR as well as basic geometric information about room volume and the direct sound's direction and time of incidence are required by the algorithm. Furthermore, the algorithm accesses a set of head related impulse responses (HRIRs) and a pre-processed sequence of binaural noise. Both were acquired based on measurements with a Neumann KU 100 artificial head [7].

The algorithm only applies to frequency components above 200 Hz. For lower frequencies the interaural coherence of a typical binaural impulse response is nearly one and the omnidirectional impulse response can be maintained.

The first relevant maximum in the impulse response is identified as direct sound. The time section from 5 ms to maximal 150 ms is assigned to the early reflections and the transition towards the diffuse reverberation. In order to determine the early reflections, the energy of a sliding window of 8 ms length is calculated and areas are marked which contain high energy. Peaks which are 6 dB above the RMS are determined and assigned to geometric reflections. A smoothly windowed section around each of the peaks is considered as one early reflection. The incidence directions are arbitrarily chosen based on a reflection pattern of a shoebox-shaped room. Thus a simple parametric model of the early reflections is created. As shown in Figure 2, amplitude, delay and envelope of each of the reflections are stored.

In a next step the shift of the listener position is considered. Therefore the parametric description of the room is adapted by applying a simple mirror image room model. For the direct sound and for each of the determined reflections a mirror image sound source is calculated. The distance between the mirror images and the listener is calculated based on the delay of the peak in the impulse response. In a next step a shifted position of the listener is considered and amplitudes (based on the 1/r law), distances and directions of incidence are recalculated. The extracted sections of the direct sound and the respective reflections are convolved with HRIRs of the recalculated directions and are placed in the synthesized BRIR at the corresponding temporal position and with the calculated amplitude.

The diffuse part of the synthesized BRIR is calculated as follows: The omnidirectional RIR, excluding the sections of the direct sound and the geometric reflections is convolved with small sections of 2.7 ms (128 samples at 48 kHz sampling rate) of the binaural noise (overlap 75 %). Thus, the binaural parameters (e.g. interaural coherence) of the noise and the frequency-dependent envelope of the omnidirectional room impulse response are both maintained. This modified method is different to the one described in [2, 3]. While maintaining the fine structure of the omnidirectional RIR in a better way the modified method delivers comparable or even better perceptual results and affords less computational power. Finally all components of the synthesized impulse response are appropriately weighted and summed up. Figure 2 shows the basic concept of the complete algorithm.

The calculation of the synthesized BRIRs is repeated for constant shifts in the azimuth angle (e.g. 1°) for the direct and the reflected sound. Thus, a circular set of BRIRs is created, which can be used for dynamic binaural synthesis.

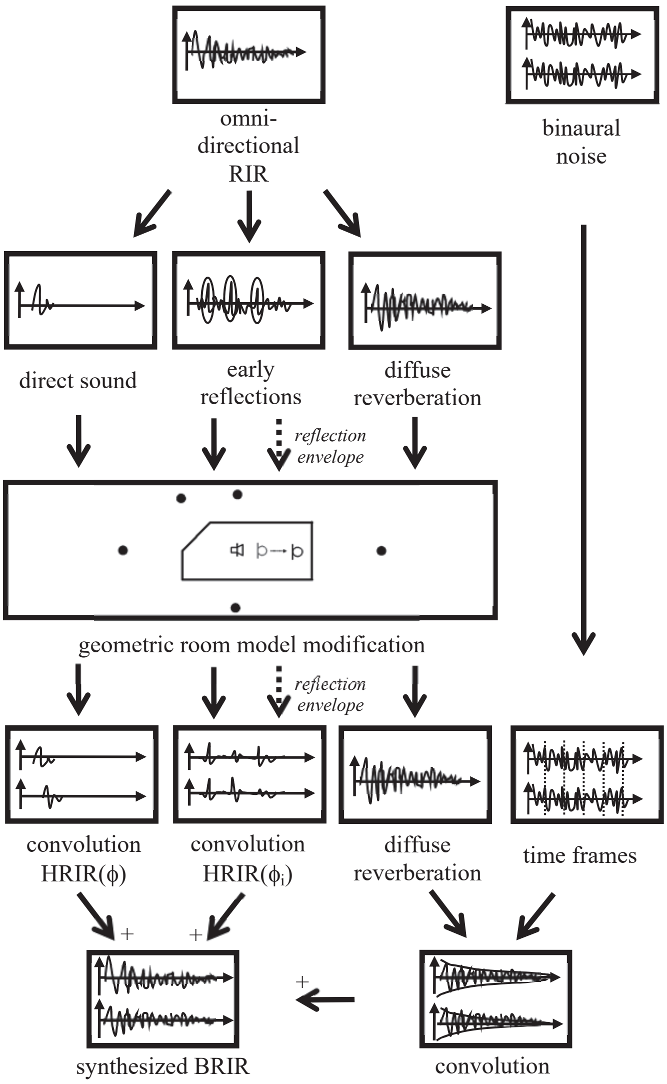

Figure 2: Block diagram of the algorithm for synthesizing a binaural room impulse response based on one single omnidirectional room impulse response.

Measured rooms

To evaluate the algorithm and to perform psychoacoustic experiments, measurements from four different rooms (the first 3 at the WDR in Cologne) were used [8]:

- Control Room 7: Volume 168 m³, base area 60 m², T_{60}(500/1000Hz) < 0.25 s, $\text{Distance}_{\text{SrcRec}}$ = 2.7 m
- Large Broadcast Studio: Volume 6100 m³, base area 579 m², T_{60}(500/1000Hz) = 1.8 s, $\text{Distance}_{\text{SrcRec}}$ = 13.0 m
- Small Broadcast Studio: Volume 1247 m³, base area 220 m², T_{60}(500/1000Hz) = 0.9 s, $\text{Distance}_{\text{SrcRec}}$ = 7.0 m
- TGC Training Room (Cologne): Volume 191 m³, base area 67 m², T_{60}(500/1000Hz) = 2.3 s, $\text{Distance}_{\text{SrcRec}}$ = 6.8 m

The measurements comprised both, a circular reference set of BRIRs and an omnidirectional impulse response, measured at the pivot position of the artificial head.

Psychoacoustic experiments

To evaluate the performance of the algorithm, a listening experiment was applied. The subjects were asked to rate the perceived distance to the auralized sound source on a seven-point category scale in German (sehr weit entfernt, weit entfernt, eher weit entfernt, mittel, eher nah, nah, sehr nah). The scale was displayed on a tablet computer (iPad) and results were given by setting a slider to the appropriate position. The subjects were allowed to rate interim values between the given categories. The equidistance between the categories was underlined by the visual presentation. The scale has been successfully used in earlier experiments [9]. Several test sliders were shown at the same time and the subjects were able to switch between the corresponding stimuli as often as required.

Different distance factors (DFs) which correspond to the quotient between the synthesized stimulus distance and the distance of the measured RIR were presented. In the experiment for each room the following DFs were tested in one table: 0.125; 0.25; 0.5; 1.0; 2.0. The listener position was shifted according to the DFs on a line between sound source and listener. Additionally the measured BRIR – in the following named as Original – was tested in the experiment. Two variants were presented, one set of test tables was loudness normalized according to ITU-R BS. 1770, the other one comprised loudness differences with increasing sound pressure levels towards lower distances as calculated by the geometric model.

The experiment was set up, controlled, and executed running the software SCALE [10]. This software directly accesses the SoundScape Renderer (SSR) [1], which was used as the binaural renderer. The audio signals were presented via an AKG K-601 open headphone. To acquire the head movements, a Polhemus Fastrak tracking system was applied. A looped drum and a guitar sequence with a length of eight and four seconds were used. These source signals have already been presented in previous listening experiments and were regarded being suitable to evaluate distance perception. The playback-level was calibrated to a L_{eq} of 60 dB(A) for the normalized stimuli and for the stimuli with a DF of 1. A maximal L_{eq} of 78 dB (A) was obtained for a DF of 0.125 in the Control Room 7.

The experiment started with a training phase in order to make the subjects familiar with the test procedure and stimuli. The training included stimuli with the expected largest and smallest distance perception and was used for anchoring as well.

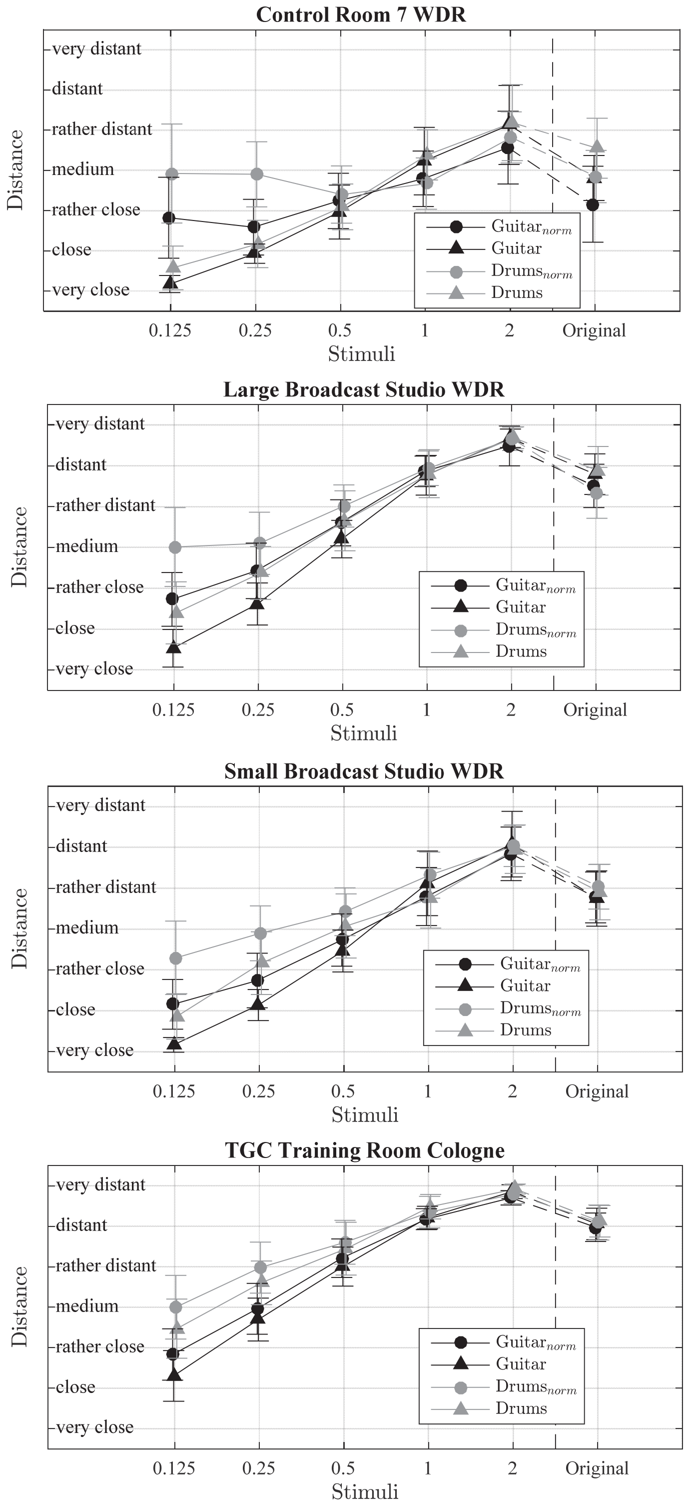

Figure 3: Distance perception (mean values and 95 % confidence intervals) of the test stimuli for Control Room 7 WDR, Large Broadcast Studio WDR, Small Broadcast Studio WDR, TGC Training Room (Cologne).

15 experienced listeners aged between 21 and 46 years (average 28; 14 male; 1 female) participated in the experiments. The subjects stated that the estimation of distance was comparably easy; however some of the subjects stated that a few of the very near stimuli were not perceived being natural. For the statistical analysis a GG-corrected repeated measures ANOVA with the factors "DF", "room", "source signal" and “loudness normalization” was applied. It yielded significant main effects for all factors ($p<.001$). Main factors with highest effect size were “DF” ($\eta^2=.95$) and “room” ($\eta^2=.84$). “Source signal” and “normalization” showed minor effect sizes. For further evaluation, the mean values and the 95% confidence intervals were calculated. For all rooms and source signals a good control of the perceived distances was achieved (see Figure 3). For the loudness-normalized presentations, the distance control was only slightly reduced. However, there is a tendency that differences due to the source signal and influences of the normalization increase towards lower distances. Slight differences in the distance perception of the original BRIR compared to the stimulus with a DF of 1 were observed for all rooms. Similar effects have already been reported in [3].

Parametric model of BRIRs

The modification of the distance between listener and receiver can be regarded as an example in which way measured room impulses can be modified and adapted to room modifications. Thus the studies presented here are embedded in a broader context of describing measured BRIRs with a parametric model. The basic idea of the model is to reduce a BRIR to its main features and to approximate them with mathematical functions or discrete values. Only these parameters have to be stored or broadcasted and they are used to resynthesize a BRIR without any knowledge about the original signal. This approach leads to several advantages. First of all, it is possible to reduce the amount of data of circular BRIR datasets. Furthermore the BRIR synthesis based on parameters enables an easy modification of the BRIRs such as listener position which are presented in this study. Other variations of the BRIRs like changes of the reverberation time or the geometry are conceivable with this parametric model, too. Direct sound, early reflections and diffuse reverberation are treated separately within the model and can be examine particularly. In a recent published paper the synthesis of the diffuse reverberation part is presented and evaluated [11]. The model is scalable so that the resolution of the synthetic BRIRs can be adapted to the used application and desired resolution.

Conclusion

The presented investigations were carried out in the framework of parametrizing and parameter-based modifying measured impulse responses. An algorithm has been presented which allows modifying the sound source distance of a measured RIR based on an underlying estimated simple room acoustic model. The listening experiment showed for different kinds of rooms that an adequate control of the perceived distance can be achieved.

The research presented here has been carried out in the Research Project MoNRa which is funded by the Federal Ministry of Education and Research in Germany. Support Code: 03FH005I3-MoNRa. We appreciate the support.

Static versions of the presented stimuli can be accessed via the following webpage:
http://www.audiogroup.web.fh-koeln.de/DAGA2016.html

Literature

[1] Geier, M., Ahrens, J., Spors, S. (2008). „The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods,” 124th Convention of the Audio Engineering Society, Amsterdam.

[2] Pörschmann, C., Wiefling, S. (2015). „Dynamische Binauralsynthese auf Basis gemessener einkanaliger Raumimpulsantworten,” In: Fortschritte der Akustik – DAGA 2015, DEGA e.V., D – Berlin, pp. 1595 – 1598.

[3] Pörschmann, C., Wiefling, S. (2015). „Perceptual Aspects of Dynamic Binaural Synthesis based on measured Omnidirectional Room Impulse Responses,” In: Proceedings of the International Conference on Spatial Audio - ICSA 2015, Graz, Austria.

[4] Pörschmann, C., Zebisch, A. (2012). „Psychoakustische Untersuchungen zu synthetischem diffusen Nachhall,“ In: Proceedings of the VDT International Convention, Cologne, Germany.

[5] Lindau, A., Kosanke, L., Weinzierl, S. (2010). „Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses,” Proc. of the 128th AES Convention, London.

[6] Stade, P. (2015). “Perzeptive Untersuchung zur Mixing Time und deren Einfluss auf die Auralisation,” In: Fortschritte der Akustik – DAGA 2015, DEGA e.V., D – Berlin, pp. 1103 – 1106.

[7] Bernschütz, B. (2013). „A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU100,” In: Fortschritte der Akustik – AIA-DAGA 2013, DEGA e.V., D – Berlin, pp.592-595.

[8] Stade, P., Bernschütz, B., Rühl, M. (2012). „A spatial audio impulse response compilation captured at the WDR broadcast studios,” In: Proceedings of the VDT International Convention, Cologne, Germany.

[9] Pörschmann, C., Störig, C. (2009). “Investigations Into the Velocity and Distance Perception of Moving Sound Sources,” ACUSTICA/Acta acustica 95, pp. 696-706.

[10] Vazquez Giner, A. (2015). „A Solution for Conducting Psychoacoustic Experiments with Real-time Dynamic Binaural Synthesis,” In: Fortschritte der Akustik – DAGA 2015, DEGA e.V., D – Berlin.

[11] Stade, P. (2016). “A Perception-Based Parametric Model for Synthetic Late Binaural Reverberation,” In: Fortschritte der Akustik – DAGA 2016, DEGA e.V., D-Berlin.