Demonstrator for the auralization and control of the room divergence effect

Maximilian Schaab[1], Verena Dobmeier[2], Stephan Werner[3] and Florian Klein[4]

Technische Universität Ilmenau, 98693 Ilmenau

[1] *maximilian.schaab@tu-ilmenau.de* [2] *verena.dobmeier@tu-ilmenau.de*
[3] *stephan.werner@tu-ilmenau.de* [4] *florian.klein@tu-ilmenau.de*

Abstract

The goal of binaural headphone reproduction is to synthesize a virtual room or to resynthesize the acoustics of a real room. Former research has shown, that the acoustical divergence between the room presented over headphones and the actual listening room can violate the expectations of the listener. In this case, the perceived quality of the synthesized room is degraded despite of a technical correct synthesis of the ear signals. This effect is called room divergence effect and is measured in a reduction of externalization of sound events. This publication describes a demonstrator which auralizes this effect. For this purpose a 5 channel loudspeaker setup is measured with a KEMAR artificial head in two rooms. Additionally three algorithms are implemented to calculate virtual rooms in between the measured rooms. By listening to the unmodified rooms measurements and their modifications differences in externalization are distinguishable. The influence of each algorithm on externalization in a divergent listening scenario is evaluated in a listening test with 14 participants.

Introduction

Increasing technical progress in the fields of audio reproduction and an extensive understanding of psychoacoustical processes allow the development of spatial audio reproduction systems with a high degree of immersion and plausibility. In this context, the aim of binaural synthesis is to synthesize signals directly at the ear drum of the listener containing the auditory cues which also occur in real world sound fields. Ideally, the listener can not distinguish between a synthesized scene and an actual sound field in that case. In reality however, such systems are highly influenced by numerous factors that can have a negative impact on the perceptual quality. Apart from technical parameters, contextual factors can have considerable impact on the plausibility of such systems [1, 2, 4]. It was shown that a divergence between the listening room (i.e. the room in which a binaurally synthesized scene is reproduced) and the ambient part of the synthesized scene can reduce the perceived externalization and thus lessens the systems plausibility [1]. This effect is now called the *room divergence effect* and is at least partially attributable to the listeners expectations to the room acoustics of the listening room [2, 3]. The authors of the present work address this topic by implementing a demonstrator which renders a binaurally synthesized scene and provides different methods for adjusting certain room acoustic parameters of the synthesized scene in order to control the perceptual effects caused by room divergence. For that, two sets of measured binaural room impulse responses (BRIR) need to be provided for a listening room and a synthesized room respectively. Here, the respective BRIRs in different rooms located at Ilmenau University of Technology were measured with a KEMAR artificial head. The user can listen to a synthesized scene with the particular ambient characteristics of the synthesized room. With the help of three algorithms (see next section) certain acoustic parameters of the synthesized scene can be adjusted in real time in order to fade towards the listening room where the listener is located. Thus hopefully minimizing the impact of the room divergence effect. The basic structure is shown in figure 1.

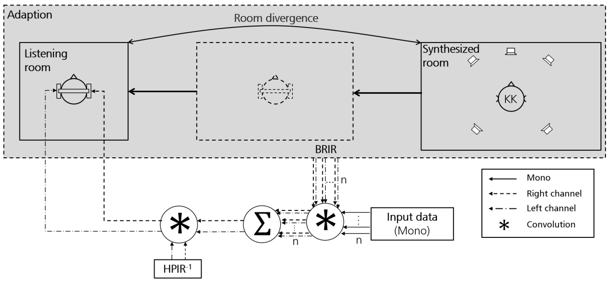


Figure 1: Workflow of the demonstrator. The user controls to which degree the synthesized room is adapted to the listening room resulting in modified binaural room impulse responses (BRIR) which are then convolved with the raw input data. Additionally a headphone equalization stage is included.

Algorithms for adaption

To demonstrate the audible impacts of the room divergence effect three algorithms were implemented to fade between the synthesized room and the listening room. These methods are covered in the following sections.

Interpolation

The first method is based on linear interpolation [7] in time domain between the amplitudes of the sample values of the measured BRIRs of the listening room and the synthesized room.

Equation 1 describes the interpolation between (x_1,y_1) and $(x_{anz+1},\ y_{anz+1})$ with anz interpolated steps. The x-value of the sample at the interpolation index α is labelled with x_α, where $\alpha = 0, 1, ..., anz$.

$$y_t(x_\alpha) = y_t(x_1) + \frac{y_t(x_{anz+1}) - y_t(x_1)}{x_t(x_{anz+1}) - x_t(x_1)}(x_\alpha - x_1) \qquad (1)$$

The interpolation in time domain is a comparatively simple method which requires low computational costs but implies that the reflections of the impulse responses of both the listening and the synthesized room are included in the resulting BRIRs at different temporal positions.

DRR adaption

The second method is based on the modifcation of the *Direct-to-Reverberant-Energy Ratio* (DRR). Apart from monaural and binaural cues for localizing sound, the brain can evaluate further cues when reflections are present (e.g. inside rooms). In such conditions sound is composed of a direct and reverberant part. Depending on the room acoustics and the relative distance of listener to the sound source, the direct and reverberant part of the incoming soundwaves have different energy levels. The ratio between the direct and reverberant energies can thus be a cue for the perception of distance of a sound source and a descriptor of the room acoustic properties [5]. DRR can be calculated directly from (binaural) room impulse responses and is defined as:

$$DRR = 10 \cdot \log \left(\frac{\int_0^T h^2(t)dt}{\int_T^\infty h^2(t)dt} \right) [\text{dB}] \qquad (2)$$

In the upper equation $h(t)$ corresponds to an impulse response and T to a defined limit between direct sound and reverberant sound. In this work T was given a fixed value of 1.5 ms assuming that the direct sound is at position $T = 0\ s$. Thus, while still including monaural and binaural cues, sound caused by reflections on the boundaries of the room are considered as the reverberant part [6]. In the context of the room divergence effect, Werner et al. have shown that modifying the DRR value of a synthesized scene can lead to a perceptual congruence between listening room and synthesized scene [6]. Based on this, a method was implemented which can adapt the DRR value of the synthesized scene stepwise to that of the real listening room. For that, a simple damping or amplification is applied to the reverberant part of the synthesized scene's BRIRs. For each of the ten impulse responses (resulting from the 5 channel setup; left and right ear) the DRR values are calculated for the listening room and the room which is to be synthesized. By matching the corresponding impulse responses between the two rooms the damping/amplification factors for each impulse response are calculated. The synthesized BRIR is then multiplied with a specially designed window function to apply the damping/amplification to the reverberant part. For a detailed description of the window function see [6]. This method accounts only for energetic adjustments of the impulse responses and does not alter the temporal reflection patterns of the synthesized BRIRs.

EDC-Shaping

By applying the third method the so called *Energy-Decay-Curve (EDC)* of the measured BRIRs of the synthesized room gets modified. The EDC describes the remaining energy in a room at any given time following an impulse. It can be calculated as follows [8], using the binaural room impulse response $h(t)$:

$$EDC(t) = \int_t^\infty h^2(\tau)\ d\tau \qquad (3)$$

After normalization to the total energy of the impulse response, the maximum value of the EDC is always equal to 0 dB [9]:

$$EDC(t) = 10 \cdot \log_{10} \left(\frac{\int_t^\infty h^2(\tau)\ d\tau}{\int_0^\infty h^2(\tau)\ d\tau} \right) \text{dB} \qquad (4)$$

To shape an EDC of the synthesized room towards the listening room, an algorithm was implemented which is based on several room acoustic parameters [11, 12]: the reverberation times T_{30} and T_{60}, definition C_{50} and clarity C_{80}.
Füg [9, 10] uses EDC-Shaping to edit the BRIRs of the listening room with the aim of varying the perception of distance of a sound source. In the modified EDC at the temporal positions 50 ms and 80 ms after the direct sound of the impulse response the target values of the parameters C_{50} and C_{80} are inserted. Regarding the desired values of the reverberation times T_{30} and T_{60} the following equations apply (with i_5 as the sample position where a decay of -5 dB of the EDC is reached):

$$EDC_{new}[i_5 + T_{20,target} \cdot f_s] = -25 \text{ dB} \qquad (5)$$

$$EDC_{new}[i_5 + T_{30,target} \cdot f_s] = -35 \text{ dB} \qquad (6)$$

In the context of the room divergence effect a modification of the perception of distance is unintended. Hence Füg's method was adapted to adjust the EDCs of the BRIRs of the synthesized room in direction of the listening room. Therefore the measured values for T_{30}, T_{60}, C_{50} and C_{80} of the listening room were used as target values of the calculation.

Figure 2 shows the EDCs of BRIRs of a large, reverberant seminar room (SR) before and after adaption to an acoustically dry listening lab (LL).
Similar to the DRR adjustment the EDC-Shaping only accounts for energetic adjustments of the impulse responses and does not change the reflection patterns of the synthesized BRIRs. In contrast to the DRR adaption EDC shaping performs a more detailed adjustment of the reverberant part of the impulse responses and is thus expected to create subjectively more plausible results.

Demonstrator

To fade between the synthesized room and the listening room, using the presented three methods, a demonstrator

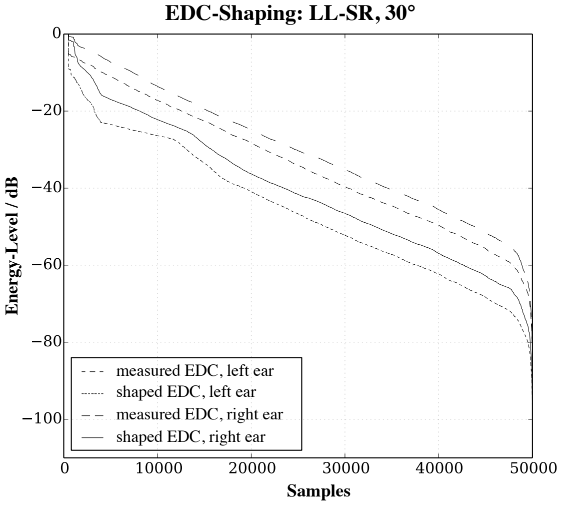
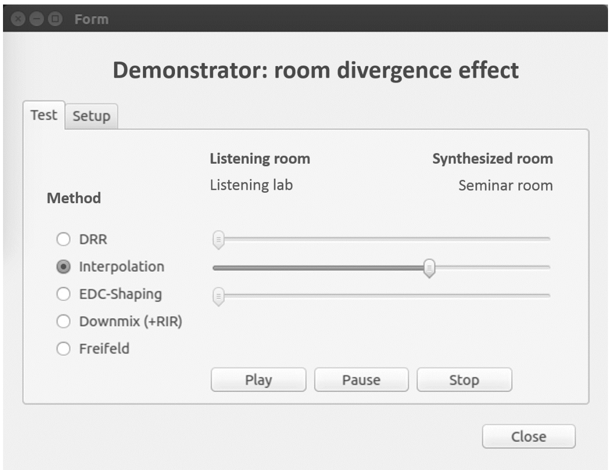


Figure 2: Energy-Decay-Curves of a seminar room (SR) before and after adaption to a listening lab (LL) for an angle of sound incidence of 30 degrees.

was developed which is shown in figure 3. The demonstrator includes two pre-measured sets of BRIRs and can be used with any BRIRs measurements of two rooms with the same measurement setup. This Graphical User Interface enables to adjust each method separately and independently from each other. Furthermore, the user can listen to two references namely a free field option, in which the items are convolved with BRIRs measured in free field conditions and a stereo downmix convolved with a simple monaural room impulse response of the synthesized room.


Form

Demonstrator: room divergence effect

Test | Setup

Listening room: Listening lab

Synthesized room: Seminar room

Method

○ DRR
● Interpolation
○ EDC-Shaping
○ Downmix (+RIR)
○ Freifeld

Figure 3: Graphical User Interface of the demonstrator.

If the slider position is set completely to the right, the scene is rendered using completely unmodified BRIRs measured in the synthesized room. Set to the left (listening room) the demonstrator renders the scene with maximum adaption of the respective parameters to the listening room as described above. The number of steps between the two end positions is set to 30. The demonstrator requires the measured impulse responses of two rooms (listening room and synthesized room) and the respective items as input and operates either in five or two channel mode. Also it requires a headphone transfer function to equalize for the headphone in use.

Evaluation

To subjectively assess the three implemented methods with respect to their ability of reducing the unwanted room divergence effect an evaluation with 14 participants was conducted. The listening tests took place in two acoustically different rooms at Ilmenau University of Technology: an acoustically dry listening lab (T_{60}=0.16s; LL) and a very reverberant seminar room (T_{60}=1.4s; SR). In a room-divergent scenario the first task was to adjust the slider-positions for each method (separately) until the perceived impression was as plausible as possible while listening to 5-channel recordings.
After the adjustment of the sliders the participants' second task was to rate the difference between their selected configurations and a given reference on a five-step impairment scale [13] regarding the attribute *externalization*. As reference the 5-channel recording convolved with the unmodified BRIRs of the listening room was provided.

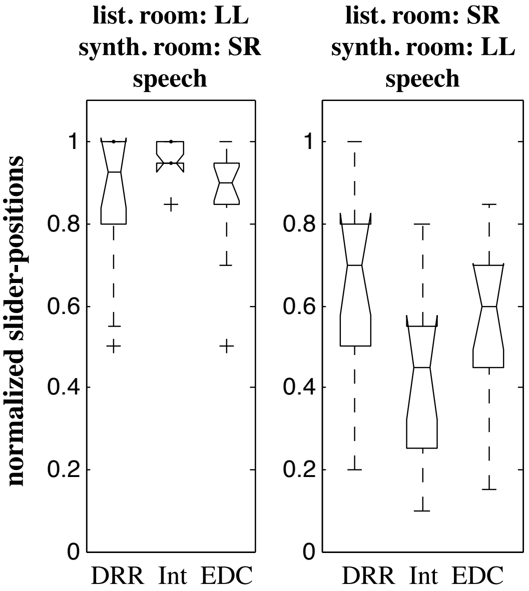


Figure 4: Results of the listening test (task 1): adjusted slider-positions for a speech-signal in the listening rooms listening lab (LL) and seminar room (SR); slider-position 0 = synthesized room, slider-position 1 = listening room.

Figure 4 shows the adjusted slider-positions for task 1 in the listening lab (LL) and the seminar room (SR) as listening rooms. In the acoustically dry LL as listening room the slider-positions were set near the maximum for each method indicating the subjects rated the scene more plausible when adapted to the listening room. In the more reverberant SR as listening room configurations with less reverberation compared to the actual reverberation were chosen. This can be explained by subjects underestimating the degree of reverberation in the seminar room. Furthermore DRR adaption and EDC-Shaping show a considerable drawback in the latter room configuration (reverberant listening room, dry synthesized scene). Since the synthesized impulse responses have distinctively lower energy in the reverberant part, a simple energetic amplification of it can not add natural reverberation but leads inevitably to a amplification of the noise floor. This was supported by subjects commenting that it was significantly harder yet sometimes impossible to find a plausible slider position for DRR and EDC when in this kind of room divergent scenario.

The results of the second part of the listening test are presented in figure 5. Small differences indicating a more convergent scenario. Basically, every method seems to reduce the divergence between the synthesized room and the listening room and thus improves the externalization compared to a room-divergent scenario (would be equal to rating 5). Using interpolation the best results can be achieved. With this method the synthesized room can be adapted completely to the listening room (convergence), because the maximum slider-position in this case means a convolution of the audio signal with the measured BRIRs of the listening room. Hence two identical versions are compared. On the contrary by applying DRR adaption or EDC-Shaping the temporal reflection patterns of the synthesized room are preserved and thus a completely convergent scenario is not possible. But because of the more detailed adjustment of the reverberant part of the BRIRs using EDC-Shaping, a higher congruence between listening room and synthesized room can be achieved with this method compared to DRR adaption. Hence better externalization can be obtained.

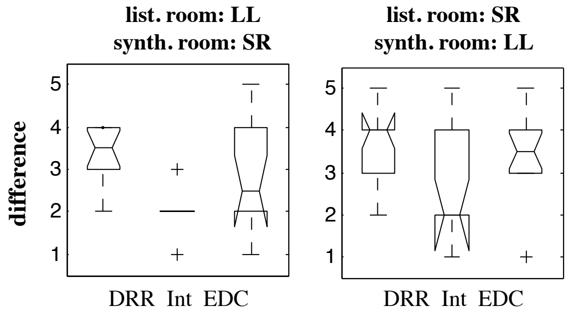


Figure 5: Results of the listening test (task 2): rating of the difference between the plausible adjustment of task 1 and the given reference regarding the attribute externalization; rating 1 = convergent, rating 5 = divergent.

Conclusions

This work addressed the problem of perceiving a lower degree of plausibility when listening to binaurally synthesized scenes with certain ambient characteristics in a listening room with different room acoustic parameters. The authors implemented a software tool to make the room divergence effect audible and up to a certain degree controllable. This is achieved by making use of three different state of the art algorithms which modify the binaural room impulse responses of the synthesized scene. It was shown that all three methods considered are able to reduce the unwanted effects of a room divergence and are thus comparatively simple tools to enhance the plausibility of binaurally synthesized content.

References

[1] Werner, S. ; Klein, F.: Influence of context dependent quality parameters on the perception of externalization and direction of an auditory event. In: AES 55th International Conference. Helsinki, Finland, 2014

[2] Plenge, G.: Über das Problem der Im-Kopf-Lokalisation. In: Acoustica 26, 1972, S. 241–252

[3] Klein, F. ; Werner, S. ; Mayenfels, T.: Influences of Training on Externalization of Binaural Synthesis in Situations of Room Divergence. *Journal of the Audio Engineering Society*, vol. 65, no. 3, 2017

[4] Udesen, J. ; Piechowiak, T. ; Gran, F.: Vision Affects Sound Externalization. In: AES 55th International Conference. Helsinki, Finland, 2014

[5] Blauert, J.: Spatial Hearing - The Psychophysics of Human Sound Localization. Rev. edition. Cambridge : MIT Press, 1997

[6] Werner, S. ; Liebetrau, J: Adjustment of direct-to-reverberant energy ratio and the just-noticable-difference. In: 6th International Workshop on Quality of Multimedia Experience (QoMEX), 2014

[7] Neundorf, W.: Polynome, Interpolation, Splines und Differentation. Preprint No. M 16/04. Technische Universität Ilmenau, 2004

[8] Kahrs, M. ; Brandenburg, K.: Applications of Digital Signal Processing to Audio and Acoustics, Kluwer Academic Publishers, New York, 2008

[9] Füg, S.: Untersuchungen zur Distanzwahrnehmung von Hörereignissen bei Kopfhöhrerwiedergabe, Masterarbeit, Technische Universität Ilmenau, 2012

[10] Füg, S. ; Werner, S.: Controlled Auditory Distance Perception using Binaural Headphone Reproduction – Algorithms and Evaluation. In proceeding of: VDT Int. Convention, 27. Tonmeistertagung, At Cologne, Germany, 11/2012

[11] Weinzierl, S.: Handbuch der Audiotechnik, Springer-Verlag, Berlin Heidelberg 2008

[12] Kuttruff, H.: Akustik – Eine Einführung, S. Hirzel Verlag, Stuttgart, 2004

[13] ITU-R BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, 1994-1997