Investigating continuous adaptation of binaural reproduction to changing listening position

Annika Neidhardt[1], Kai-Peter Jurgeit[1], Afrooz Nasrollahnejad[1], Johannes Nowak[1],[2]

[1] *Technische Universität Ilmenau, 98694 Ilmenau, Deutschland, Email: annika.neidhardt@tu-ilmenau.de*

[2] *Fraunhofer IDMT, 98693 Ilmenau, Deutschland*

Abstract

For virtual and augmented reality using binaural technology it is of interest to reproduce sound scenes based on measurements of real rooms. Furthermore, it has become state of the art to explore a provided scene by walking around. Several methods for interpolation and extrapolation of measured impulse responses have been proposed in the past to provide a convincing continuous binaural reproduction for position changes. However, there is still a lack of experience regarding the practicability and the perceived quality of the various approaches. Only in rare cases, a psychoacoustic evaluation considering interactive listener translation was conducted.

In the presented study, a reproduction area of about 3 m×2.5 m was created from binaural room impulse response (BRIR) measurements from one loudspeaker to a dummy head at three positions with a distance of 1 m in two different rooms. A continuous adaptation of level and direct-to-reverberant energy ratio was applied to achieve a seamless auralization. A listening experiment was conducted to evaluate the continuity of level and timbre as well as sound source stability and plausibility. In the case of pop music the results were satisfying. The reproduction of noise revealed weaknesses.

Binaural auralization based on BRIRs for discrete angles and positions

Dynamic binaural reproduction can be realized by partitioned convolution with direction- and position-dependent filters. The listener's head is tracked in position and orientation and the corresponding filters are selected. If no interpolation is used and switching between filters is realized only with a short cross-fade, a certain minimal angular and positional resolution of the filter set is required to achieve a smooth and plausible reproduction. It has been shown, that with an angular resolution of 2° listeners could not distinguish a binaural auralization from a real scene [1]. Depending on signal and expert level of the listener, a resolution of up to 5° can be sufficient as well [2]. The required positional resolution has not been studied in depth yet. Wefers et al. [4] presented a system for position dynamic binaural reproduction based on room acoustical simulation with ray-tracing. For a quick system response to changes of listener or source position, the BRIR filters are updated section-wise. The direct sound is refreshed for positional differences of a few centimeters, the reverberant part for changes of more than 1 m.

In a previous study [3] an approaching motion towards a virtual sound source was auralized with positional resolutions of 25 cm and 50 cm with male speech and white noise. Fig. 1 shows the ratings for continuity.

Further pretests revealed, that with white noise even a 5 cm resolution still causes audible steps in the reproduction. This suggests, that only a continuous reproduction will provide the desired smooth transition between different positions for all kinds of signals.

Interpolation approaches and their perceptual consequences

In the past a variety of interpolation and extrapolation approaches was proposed to adapt the reproduction to a changing listening position, e.g. positional cross-fading [5] or adjustment of initial time delay gap and energy decay curve [6]. Furthermore, approaches to create BRIRs based on parameterization of room impulse responses measured with an omni-directional microphone [7] or a spherical array, as discussed e.g. by Coleman et al. [8] have been presented. The proposed parameterizations allow to approximately adapt the temporal and spatial structure of the first reflections. In most cases, perceptual evaluation was conducted without considering interactive listener translation.

Neidhardt et al. [3] observed, that a plausible interactive approaching motion towards a virtual loudspeaker could be achieved by adapting only the level of the direct sound of a BRIR according to its measured progress along a given line. This suggests, that besides the adjustment of the direction a continuous adaptation of the level of the direct sound based on few measured reference points should provide a smooth and plausible dynamic reproduction. However, so far this has only been tested for positions within the critical distance of the source in the

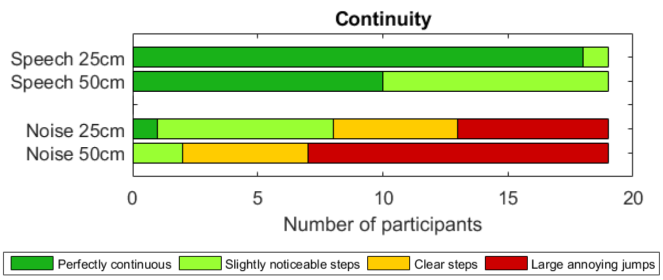

Figure 1: Ratings for the continuity of binaural reproductions with interactive listener translation with two different positional resolutions, taken from [3].

virtual room and for the case of an approaching motion, which does not cause a change of the relative angle to the source. In this work-in-progress paper we present a realization of head-tracked binaural synthesis for interactive position changes within a whole area from BRIRs measured at 3 positions in a row with a spacing of 1 m only by adapting the level of the direct sound.

BRIR sets measured in two rooms

BRIR-sets measured in two different rooms were used for the study. Both sets were created using a G.R.A.S. Kemar 45BA head-and-torso-simulator (HATS) rotated by an electronic turntable Outline ET 250-3D.

- **Listening Laboratory:** Size 8.4 m×7.6 m×2.8 m, V= 179 m^3, T_{60}= 0.25 s, azimuth resolution 4° Loudspeaker Genelec 1030A, data from [9]
- **Office room:** Size 4.84 m×4.68 m×3.22 m V= 73 m^3, T_{60}= 1.15 s, azimuth resolution 5° Loudspeaker Geithain MO-2, data from [10]

For both rooms three measurement positions arranged along a line with distances of 1 m were chosen for this experiment. The relations between source and measurement positions are illustrated in fig. 2 and fig. 3.

Realization of the dynamic reproduction

PyBinSim [11] is a python tool for dynamic binaural reproduction in real-time based on switching between various BRIR filters corresponding to discrete positions and directions.
To avoid artifacts due to a limited positional resolution, a continuous adaption is desired. For this study, the pyBinSim software was extended for continuous reproduction based on selected measured BRIRs. The following paragraphs describe the algorithm in detail.

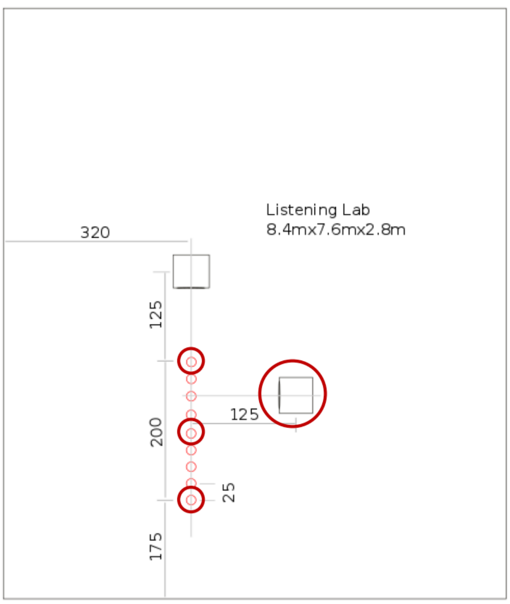

Figure 2: Illustration of source and measurement positions in the listening laboratory with measures in [cm], from [9].

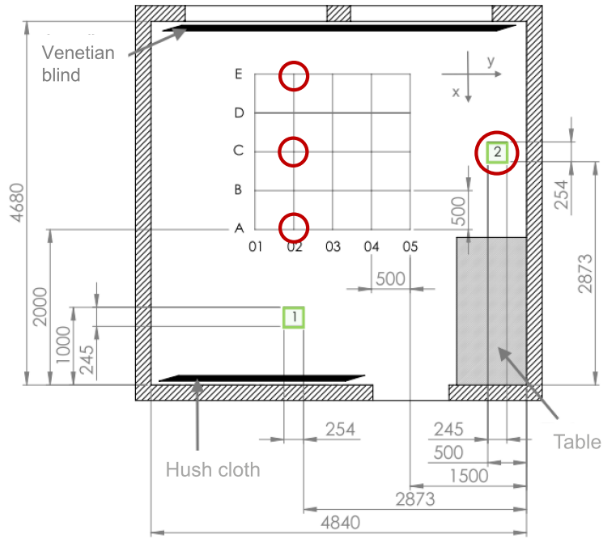

Figure 3: Illustration of source and and measurement positions in the office room with measures in [mm], from [10].

Adaptation of angle and filter selection

If the listener walks past a source, its relative angle to the head changes. The algorithm chooses that BRIR filter of the closest measurement position and with the closest angle match as a fundament for further processing.

Separation of direct and reverberant sound

The detection of the direct sound in each BRIR was realized in an offline process. The estimated sample index of the direct sound was provided to the real-time processing. During the interactive reproduction the direct sound was separated from the reverberation part of the impulse responses by a fully overlapping cosine window at about 1.5 ms after the direct sound.

Adaptation of level

The level of the direct sound is adapted continuously to the new listening position. Two approaches of level adaptation were combined. Based on the inverse distance law the direct sound is weighted by the ratio of the distance between measurement position of active BRIR and the loudspeaker and the distance of the current listening position to the sound source. Due to the directivity of the loudspeaker, the measured direct energy does not only depend on the distance but also on the direction with regard to the sound source. An additional strategy was implemented by considering the measured values in terms of a linear interpolation between the maximum direct energy measured at the three positions. The resulting connecting line built the new, direction dependent reference for the inverse distance law as shown in fig. 4.
If the source is placed within the listening area the listener can get very close. In this case the application of the inverse distance law will lead to a strong amplification. Thus, the gain was limited by a minimum distance condition: If the distance between listening position and source is less than 0.2 m, 0.2 m was taken as distance value. Natural near-field effects are neglected.

Listening experiment

A listening experiment was conducted to provide a perceptual evaluation of the algorithm.

Participants

6 ladies and 12 gentlemen took part in the experiment. The subjects had an average age of 29 years, ranging from 22 to 38. 8 participants had no experiences with dynamic binaural reproduction at all. Among the remaining subjects 3 considered themselves as expert listeners in that field. All stated to have no known hearing impairments.

Test setup

The participants had to wear Oculus Rift VR glasses and SR-202 STAX headphones. Two Oculus tracking sensors captured position and orientation of the subject's head during interactive exploration of the different virtual scenes. They could move around in the listening area with a size of approximately 3 m×2.5 m.
In the display, a neutral virtual grid could be seen, that helped the user to orientate within the scene. The edge of the listening area is marked by an additional blue grid, that appears, if the user gets too close. No visual cues on the position of the virtual source or the properties of the virtual (measured) room were provided.

Test scenes and signals

Test scenes were created on the BRIR sets from both rooms. Both were auralized without any interpolation but direct switching in the center between two measurement positions with a short cross-fade on the one hand. This was chosen as the anchor condition. On the other hand the algorithm described above was used for the auralization. Each scene was provided once with the mono version of a pop song as well as white noise. This results in a total of 8 test scenes.
For the scenes created from the measurement in the listening laboratory, the sound source is placed within the listening area and the whole listening area is located within the critical distance of the sound source. In the case of the office room, the source is situated about 1 m away from the listening area. The reverberant energy of the BRIRs is in average over all directions higher than the direct energy for each of the measurement positions.

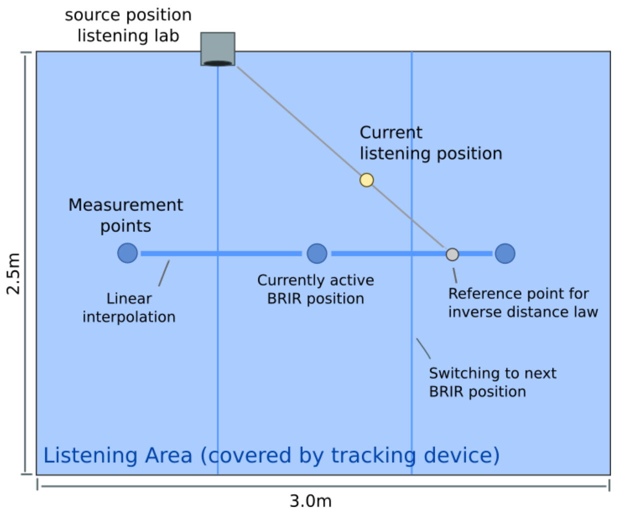

Figure 4: Illustration of the listening area.

Test procedure

The participants could explore each scene interactively by arbitrary translation and rotation within the listening area. Afterwards continuity in loudness and coloration as well as sound source stability and plausibility had to be rated on the 4-point scales shown in fig. 5. The scenes were presented in randomized order.

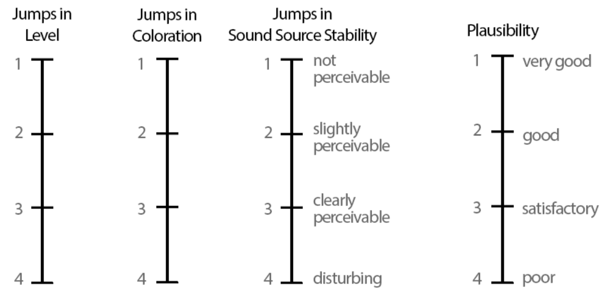

Figure 5: Scales for the evaluation.

Results

Fig. 6 shows the ratings for the four attributes. A Wilcoxon signed-rank test was applied to check for significant differences. Plausibility is rated better for the scenes created with the level adaptation method, however a significant improvement was only found for music in the listening lab and for noise in the office room. Further, it is interesting to notice, that in the office room, direct switching between BRIRs with music was rated better than level interpolation with noise.
Similar results were found for sound source stability and continuity in sound level. The ratings are better for the interpolation algorithm, but the differences are not significant for both signals and rooms.
The results for the continuity of timbre are dominated by the type of signal. The ratings for noise are generally lower than for music, even in the anchor condition. The interpolation did not lead to a significant improvement.
Overall, the scene with level interpolation for music in the office room was rated best. Compared to white noise music is less critical with regard to discontinuities in timbre as well as loudness. Further, the stronger reverberation and the larger distance to the source support the apparent sound source stability.

Observations based on verbal descriptions

The participants reported that in certain areas the reproduction worked very well for them, only in few specific locations they noticed steps in coloration and source position. The subjects indicated, that these steps occurred exactly at those positions, where the reproduction switches to the BRIRs of the next measurement position. However, for the level interpolation with music in the office room, no subject reported discontinuities.

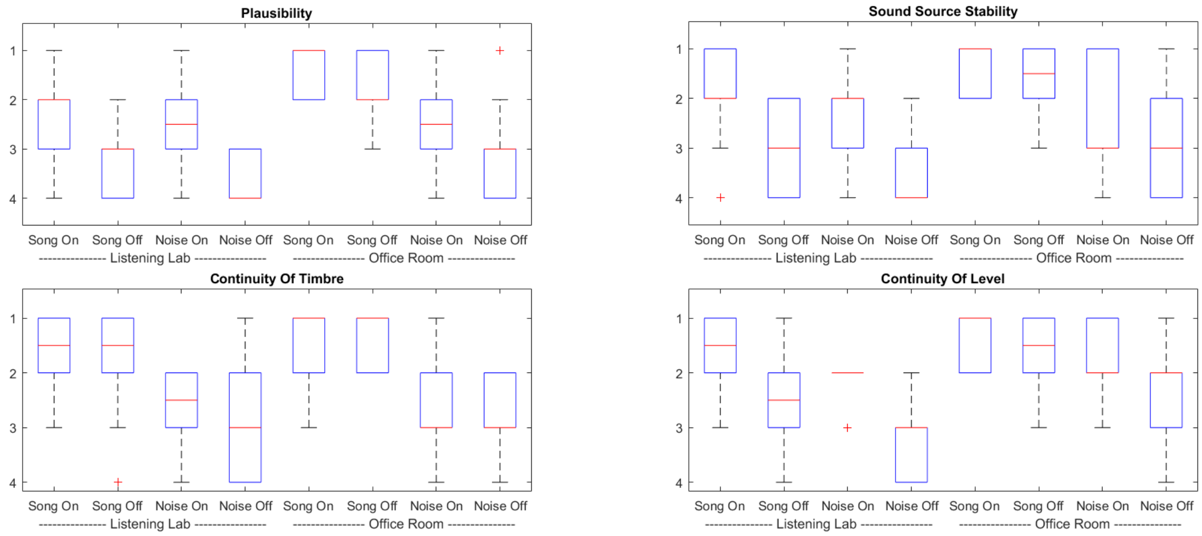

Figure 6: Results of the listening experiment.

Summary and Conclusions

Instead of switching between filters for discrete positions a continuous adaptation of a binaural reproduction to a changing listening position is desirable.

In this work-in-progress-report a first basic implementation is presented. A position dynamic binaural reproduction was created by choosing the BRIRs of the closest measurement position and with best matching angle to the source and adapting only the level of the direct sound in a combination of the inverse distance law and a linear reference point interpolation.

Based on verbal reports of the participants, it can be concluded, that a plausible reproduction was achieved for those areas constructed from only one measurement position. When switching to BRIRs from the next reference point, slight steps in coloration and source position could be noticed in several cases. This is not surprising because no time alignment was applied yet. The creation of a plausible dynamic binaural reproduction for continuous position changes from one measured BRIR has lower requirements than from BRIRs measured at several positions that are 1 m apart from each other and might be possible for the whole listening area. However, both approaches will not represent the natural progress of a soundfield, e.g. with regard to the frequency dependent sound source directivity and the progress of the timbre caused by the position dependent reflection pattern.

Linear level interpolation between the reference points is not optimal, but knowledge of the source directivity is required to implement a better strategy. Furthermore, a continuous adaptation of the first reflections or at least of the ITDG will improve the continuity of the timbre.

An adapted absolute category rating was conducted for verification. In comparison to an anchor condition an improvement of plausibility, sound source stability and continuity of level was observed. However, this test design as well as most other traditional methods cannot capture the perceived local differences in the quality of the reproduction. New methods are required to evaluate position dynamic reproduction systems.

References

[1] F. Brinkmann, A. Lindau, and S. Weinzierl. On the authenticity of individual dynamic binaural synthesis. *J. Acoust. Soc. Am. 142(4), pp. 1784-1795, October*, 2017.

[2] A. Lindau, H.-J. Maempel, and S. Weinzierl. Minimum BRIR grid resolution for dynamic binaural synthesis. In *Acoustics, Paris, France*, 2008.

[3] A. Neidhardt, A.I. Tommy, and A.D. Pereppadan. Plausibility of an interactive approaching motion towards a virtual sound source based on simplified BRIR sets. In *144th Int. AES Conv., Milan, Italy*, 2018. (accepted).

[4] F. Wefers, J. Stienen, S. Pelzer, and M. Vorländer. Interactive acoustic virtual environment using distributed room acoustic simulations. In *EAA Joint Symp. on Auralization and Ambisonics, Berlin, Germany*, 2014.

[5] M. Boerum, B. Martin, R. King, and G. Massenburg. Lateral listener movement on the horizontal plane: Part 2 sensing motion through binaural simulation in a reverberant environment. In *Int. AES Conf. on Audio for Virtual and Augm. Reality, Los Angeles, CA, USA*, 2016.

[6] C. Mittag, S. Werner, and F. Klein. Development and evaluation of methods for the synthesis of binaural room impulse responses based in spatially sparse measurements in real rooms. In *43rd Annual Conference on Acoustics, Kiel, Germany*, 2017.

[7] C. Pörschmann and P. Stade. Auralizing listener position shifts of measured room impulse responses. In *42nd Annual Conference on Acoustics, Aachen, Germany*, 2016.

[8] P. Coleman et al. Object-based reverberation for spatial audio. *J. Audio Eng. Soc., Vol. 65, No. 1/2*, 2017.

[9] A. Neidhardt and N. Knoop. Binaural walk-through scenarios with actual self-walking using an HTC Vive. In *43rd Annual Conf. on Acoustics, Kiel, Germany*, 2017.

[10] C. Mittag, M. Böhme, and S. Werner. Dataset of KEMAR-BRIRs measured at several positions and head orientations in a real room. Technical report, TU Ilmenau, 2016. http://doi.org/10.5281/zenodo.206860.

[11] A. Neidhardt, F. Klein, N. Knoop, and T. Köllmer. Flexible python tool for dynamic binaural synthesis applications. In *142nd Int. AES Conv., Berlin, Germany*, 2017.