Use of Position-Dynamic Binaural Synthesis in an Exemplary Auditory Augmented Reality Installation

Stephan Werner[1], Florian Klein[1], Georg Stolz[1], Stefan Dietrich[1], Matthias Fritz Melzer[1]

[1] *Technische Universität Ilmenau, Electronic Media Technology Group, 98693 Ilmenau Germany, E-Mail:stephan.werner@tu-ilmenau.de*

Introduction and Motivation

In the context of augmented reality spatial audio using a position-dynamic binaural synthesis system is an appropriate approach to fuse virtual audio objects with the real listening environment. This contribution describes a possible technical realization of such a system. Several components are described which help to enable a plausible auditory simulation with a moving listener while keeping the computational effort in mind. The proposed system builds upon measured room impulse responses to avoid possible room divergence between the synthesized scene and real acoustic environment [1,2]. Efficient approaches are presented to synthesize new BRIRs using very few measurements of the listening room and to realize a real-time convolution [3]. These approaches are discussed on an exemplary installation for presenting audio information in the context of a museum exhibition with the aim to improve knowledge transfer.

System Requirements

The requirements for the audio system result from the installation of an application in the context of a photo exhibition about inventors with their inventions. In cooperation with a project partner, an audio-AR installation is to be realized in a room in which virtual audio objects are placed on the positions of six pictures. The audio content to be played back is spoken text about the contents of the pictures. The interest of the project partner lies in investigations of knowledge transfer using spatial audio. The visitors of the exhibition should be able to move freely in the space. Predefined areas of stay in the room shall trigger events in the reproduction of the audio content.

Figure 1 illustrates the geometry of the setup and the room. The walkable area includes the entire room but with a minimum distance from the walls of 0.5 m. The auralization is realized for horizontal movements and for the horizontal head movement of the listener (6-3DoF). Since in the system measured binaural room impulse responses (BRIRs) from one position and for all loudspeakers are used as audio objects, the necessity of a parallax correction arises, especially at small distances of the listening position to the audio object position, in order to be able to realize a localization of the virtual audio object in the picture itself. This is realized by a simple position-dependent adapted selection of corresponding BRIRs for different yaw head poses.

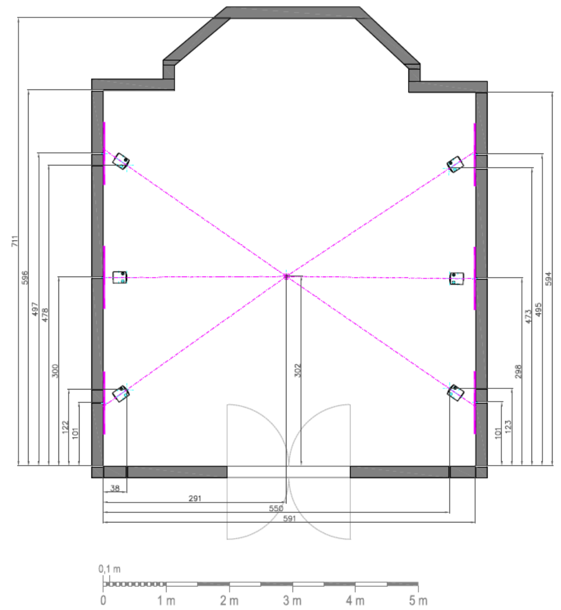

Figure 1: Schematic representation of the setup as well as the room with its dimensions. Measured loudspeaker positions are marked as black squares close to the walls. The intended virtual object positions are marked as pink lines at the walls. The midpoint of the setup is used for measurement of a 360° (5° yaw resolution) BRIRs for all six speaker positions.

General View of the System

A basic scheme of an AAR system is built from several different functional blocks which realize a position-dynamic binaural synthesis. This includes:

1. provision of binaural room filters from measurements or simulations,

2. representation of the scene,

3. creation and/or adaptation of binaural room filters related to the pose and position of the listener, and

4. real-time rendering to make a position-dynamic auralization possible.

The reproduction of an audio object in a reverberant room can be realized using BRIRs. The audio signal of the source is convolved with BRIRs of the current source–receiver position and head orientation of the listener. For the acquisition of binaural room filters, several popular approaches exist. An overview can be found in [3].

The next step is the creation of a scene. Depending on the synthesis approach, the room geometry, source position, and the walkable area have to be defined. Depending on the real-time capability of the succeeding filter creation step, a fine or coarse sampling of the walking area has to be considered. Sub-sampling the area based on psychoacoustic assumptions, for example, by considering just noticeable differences (JND) for direction and distance changes, can avoid the effort for a continuous adaptation of filter coefficients.

When the listener changes his/her pose or position new filters need to be computed (or loaded). The most flexible solution would be a parametric filter creation approach, but the success relies on the quality of the underlying model. Filter shaping approaches rely on a BRIR measurement on one position, and only certain properties of filters are adapted when the listener moves such as energy decay curve (EDC), level of direct and reverberant sound, or initial time delay gap (ITDG). A detailed insight can be found in [3].

The last step is the real-time rendering. The approach used here deliver binaural room impulse responses. These need to be convolved with the desired audio signals. The filters have to be changed each time the position or the pose of the listener changes. As it has become clear, the plausible creation of virtual audio objects fused into a real room is the main challenge for AAR.

Proposal of a Specific System

A block diagram of the proposed audio system is shown in Figure 2. The main blocks 'Spatial Sub-Sampling', 'Filter Creation', and 'Real-Time Processing' are shown in the left part, while single input data and parameters are shown on the right side. A brief insight in the blocks 'Sub-Sampling' and 'Filter Creation' is given below. A detailed description of all blocks, as well as an evaluation is given in [3].

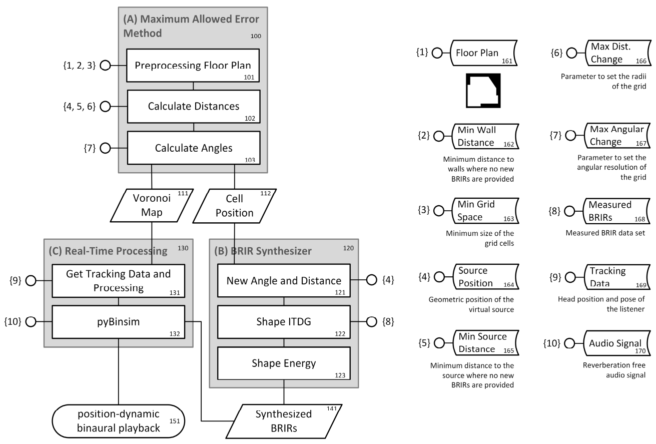

Figure 2: Illustration of the functional blocks of the position-dynamic binaural synthesis system used.

Spatial Sub-Sampling

In the exemplary audio system, the different BRIR data sets, which are required for the listener's movement in the room, are created for predefined discrete areas in the room. These areas are called cells. The totality of all cells is called grid. Depending on the chosen spatial resolution, a BRIR data set is valid for one of these cells. A BRIR data set consists of BRIRs for the left and right ear for all available head orientations and for one audio object. Only in the center of this area, the BRIR data set corresponds to a correct mapping of distance and direction to the sound source. Towards the edges, the deviation of direction and distance cues and thus the reproduction error increases. If the spatial resolution is chosen to be low, perceptible errors will increase. The shapes of these cells can be uniform in the sense of a position-independent shape, for example, quadratic. However, the shapes of the cells can also be nonuniform in the sense of a position-dependent shape.

Filter Creation

A set of BRIRs must be provided for each possible listening position. If a dataset of BRIRs is measured, e.g., with a head-and-torso-simulator at one or few selected positions in a room, then BRIRs for further positions can be generated by interpolation and extrapolation. In our group we pursued two different approaches for that which are briefly presented. The first approach is based on a quite strong simplification and thus allows for an efficient implementation. The second approach manipulates more acoustic details with the goal to provide a better spatial quality.

This first approach is based on the pragmatic approach to keep the late reverberation constant throughout the different positions. To simulate an approaching motion towards a virtual sound source, a simple adjustment of energy of the direct sound according the distance to the sound source can be sufficient to achieve a plausible reproduction [4].

The second approach deals with acoustic shaping of single and/or of combination of distant dependent acoustic parameters [3]. The basic idea is to change the acoustic structure of the early reflections of measured BRIRs based in a shift of the initial time delay gap (ITDG) for example and to adapt the distant dependent energy behavior of direct sound and reverberations. The energy adaptation can also include the change of the direct to reverberation energy ratio (DRR) and of the energy decay curve (EDC).

References

[1] Plenge, G. Über das Problem der Im-Kopf-Lokalisation. Acustica 1972, 26, 241–252.

[2] Werner, S.; Klein, F.; Mayenfels, T.; Brandenburg, K. A Summary on Acoustic Room Divergence and its Effect on Externalization of Auditory Events. In Proceedings of the 8th Int. Conf. on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 2016. doi:10.1109/QoMEX.2016.7498973

[3] Werner, S.; Klein, F.; Neidhardt, A.; Sloma, U.; Schneiderwind, C.; Brandenburg, K. Creation of Auditory Augmented Reality Using a Position-Dynamic Binaural Synthesis System—Technical Components, Psychoacoustic Needs, and Perceptual Evaluation. Appl. Sci. 2021, 11, 1150. https://doi.org/10.3390/app11031150

[4] Neidhardt, A.; Tommy, A.I.; Pereppadan, A.D. Plausibility of an interactive approaching motion towards a virtual sound source. In Proceedings of the 144th International AES Convention, Milan, Italy, 23–26 May 2018.