Efficient synthesis of perceptually plausible binaural room impulse responses

Torben Wendt[1,2,3,4], Steven van de Par[1,3], Stephan D. Ewert[2,3]
[1] *Acoustics Group, University Oldenburg, Germany,* [2] *Medical Physics, University Oldenburg, Germany,*
[3] *Cluster of Excellence "Hearing4all",* [4] *E-mail: torben.wendt@uni-oldenburg.de*

Introduction

Room acoustical simulations are desirable for many purposes, such as developing/testing signal processing algorithms (e. g. for hearing aids), running psychoacoustical experiments in controllable reverberant environments, or to be applied in audio-visual simulation environments (e. g. for training, rehabiliation or computer games).

For headphone auralization, room acoustics is typically represented by binaural room impulse responses (BRIRs). A dry (i. e. reverberation-free) source signal is convolved with the BRIR in order to be perceived as in the respective room. To synthesize BRIRs, in principle various methods exist, ranging from simple artificial reverb algorithms (e. g. [1]) to complex room acoustical simulations based on geometrical acoustics (GA) (e. g. [2, 3, 4, 5]), even for dynamic scenarios (e. g. [6, 7, 8]). Whereas GA based methods enable a high degree of detail, they have a high computational complexity. On the other hand, many of the applications mentioned above require BRIR adaptation in real-time, depending on the source- and/or receiver movement. Rather than a physically correct soundfield rendering, a perceptually convincing simulation that renders correct interaural parameters and room acoustical parameters (e. g. reverberation time, clarity) may be sufficient. To meet these goals, a hybrid BRIR simulation method is suggested, which combines the image source model (GA approach) for the early reflections with a feedback delay network (efficient reverb algorithm) for the late diffuse reflections. The latter was extended in order to be adaptable to room acoustical properties and to create a binaural reverberation. The simulation method was evaluated subjectively and technically.

Simulation method

A hybrid approach was suggested to synthesize BRIRs [9]. The early sound reflections are computed by an image source model (ISM) [2], restricted to a low reflection order because of its high computational complexity. The late reverberation is generated by a binaurally extended feedback delay network (FDN) [1].

Image source model

In the ISM, sound reflections are represented as mirrored versions of the original source, the image sources (IS). The implementation is restricted to a shoebox geometry, representing an important class of rooms, and enabling efficient calculation of IS positions on a regular pattern—in contrast to arbitrary geometries, where visibility tests become necessary for each IS. Nevertheless, for a shoebox room the number of IS up to reflection order N increases as $\mathcal{O}(N^3)$, affecting computational efficiency considerably. An inherent limitation of the ISM is that it assumes only specular instead of diffuse reflections, although they occur in reality.

Besides a different time delay, an IS differs from the original source by an attenuation due to its distance to the receiver and due to wall reflections. In the simulation method, this is implemented directly for each IS, where the wall reflections are applied as the frequency domain product of the reflection coefficients (specified in octave bands) of all walls being involved in the specific reflection path. Finally, the azimuth and elevation angles relative to the receiver's head orientation are applied by HRIRs.

Feedback delay network

The FDN implementation is based on the general multichannel network suggested by [1], consisting basically of a set of parallel delay lines, each with an absorbing filter (controlling the resulting frequency dependent reverberation time), and a mixing feedback matrix, redistributing the outputs of all channels to the inputs in order to increase the pulse density.

With a constant number of $M = 12$ parallel channels, two channels are each associated to one specific shoebox wall, reflected in several parameter choices accounting for the room properties.

The FDN was further extended to create binaural reverberation: Via HRIRs, the output of all channels are mapped to points on a cube surface in such a way that the incidence directions are roughly equally distributed around the listener's head. The cube orientation is aligned with the shoebox room and additionally each channel is weighted with the reflection coefficient of the respectively parallel shoebox wall in order to simulate a direction dependent sound intensity of reverberation.

Combination of ISM and FDN

In order to achieve a straight transition (on a dB scale) between the outputs of ISM and FDN, the energy and initial delay of the FDN input have to be chosen suitably. Here, the n_N ISM pulses of order N (before HRIR filtering) were fed into the FDN as a multichannel signal (as a single channel signal they would lead to undesired comb filter coloration). Because in general $n_N \neq M$, the ith ISM pulse is fed into the FDN channel $[(i-1)\mathrm{mod}M]+1$.

Evaluation

To evaluate the simulation method technically, measured BRIRs of four real rooms (a small, empty and highly reverberant room, "Empty chamber", a medium size laboratory room, a large auditorium, "Aula", and a lecture room) were compared to the corresponding synthesized BRIRs with respect to reverberation time T_{60}, early decay time EDT, definition D, and clarity index C. As an evaluation parameter, the maximum IS order N was varied in $\{0, 1, 2, 3\}$, in order to find an optimum between accuracy and computational effort.

Results

The evaluation results are shown in Fig. 1 (see caption for explanation). For T_{60}, good accordance between measured and all synthesized BRIRs were obtained. Since T_{60} is controlled explicitely by the FDN and the ISM-FDN transition was created suitably, this result could be expected. Exceptions are the curves for the laboratory room (Lab) for high frequencies. This can be attributed to small deviations in the frequency responses of the FDN absorption filters from desired ones, since these filters are applied many times in series.

For EDT, similar results as for T_{60} were obtained, but in contrast, a much larger variation with N occured. This can be explained by the higher weighting of the early reflections during EDT calculation, since especially the shape of the early reflections highly depend on the choice of N. Again, deviations from measurement were largest for the laboratory room in the high frequency range.

Also for D and C, the synthesized BRIRs mostly showed very similar results to the measured ones. An exception was the synthesis with $N = 0$ (direct sound plus FDN), which led to largest deviations for the aula. This can be explained by a small source-receiver distance compared to the large room dimensions ($\approx 19 \times 30 \times 10\,\mathrm{m}^3$), leading to a high definition, which apparently cannot be represented without any geometrically exact early reflections.

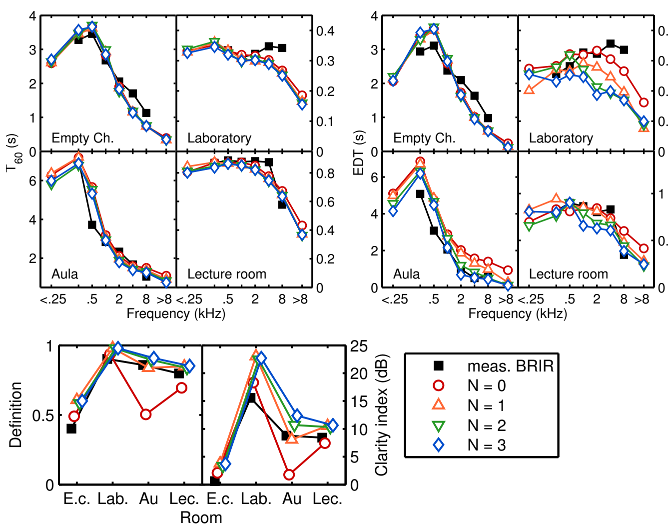

Figure 1: Room acoustical parameters for four different rooms, derived from measured (black) and corresponding synthesized BRIRs with different max. IS orders N (colors): Reverberation time T_{60} (upper left panel), early decay time EDT (upper right panel), definition and clarity (lower panel).

Finally, the results of the subjective evaluation, which are not shown in this paper due to limited space, are summarized: 15 subjects rated (w/o reference), naturalness, reverberance, room size, coloration, metallic character and source width, allowing an indirect comparison between measured and synthesized BRIRs. Differences in these properties between the rooms were clearly conveyed, and mostly good accordances between measured and synthesized BRIRs were found. One exception was coloration.

Summary

A hybrid method for efficient BRIR synthesis, combining the image source model (ISM) (early reflections) with a binaurally extended feedback delay network (reverberation), was presented. Technical evaluations showed its ability to correctly represent room acoustical parameters in comparison to corresponding real rooms, where the maximum reflection order N in the ISM had almost no influence.

Acknowledgements

This work was supported by the Cluster of Excellence EXC 1077/1 "Hearing4all" and the DFG FOR 1732.

References

[1] Jot, J.-M. and Chaigne, A.: Digital delay networks for designing artificial reverberators. 90th AES Convention 1991

[2] Allen, J. B. and Berkley, D. A.: Image method for efficiently simulating small-room acoustics. JASA 66 (1979), 943–950

[3] Krokstad, A., Strøm, S and Sørsdal, S.: Calculating the acoustical room impulse response by the use of a ray tracing technique. J. Sound Vib. 8 (1968), 118–125

[4] Naylor, G. M. and Rindel, J. H.: Predicting Room Acoustical Behaviour with the ODEON Computer Model. 124th ASA meeting New Orleans 1992

[5] Dalenbäck, B.-I.: Engineering principles and techniques in room acoustics prediction. BNAM, Bergen, Norway, 2010

[6] Dalenbäck, B.-I. and Strömberg, M.: Real Time Walkthrough Auralization – The First Year. CATT (Dalenbäck), Valeo Graphics (Strömberg) (2010)

[7] Schröder, D., Wefers, F., Pelzer, S., Rausch, D. S., Vorländer, M. and Kuhlen, T.: Virtual Reality System at RWTH Aachen University. 20th Int. Congress on Acoustics, Sydney 2010

[8] Silzle, A., Novo, P. and Strauss, H.: IKA-SIM: A System to Generate Auditory Virtual Environments. Audio Engineering Society Convention 116 (2004)

[9] Wendt, T., van de Par, S. and Ewert, S. D.: A computational efficient and perceptually plausible algorithm for binaural room impulse response simulation. Subm. to JAES