Required spatial resolution for late reverberation in a 3-dimensional loudspeaker array

Josef Poppitz[1,2], Torben Wendt[1,2], Steven van de Par[1], Stephan Ewert[2]
[1] *Acoustics Group and Cluster of Excellence Hearing4all, University Oldenburg*
[2] *Medical Physics and Cluster of Excellence Hearing4all, University Oldenburg*

Introduction

The simulation and auralization of room acoustics can have numerous applications in interactive evaluation environments (e.g., for the development of hearing aids), architectural acoustics, psychoacoustic studies, rehabilitation and computer games.

For real-time applications computational efficiency is key and it is desirable to render the spatial sound field with an accuracy just sufficient for human perception. For diffuse parts of the sound field, like the late reverberation, a limited number of loudspeakers can be sufficient to reach this goal (e.g.: [1]).

The room acoustics simulator RAZR [2] simulates perceptually plausible binaural room impulse responses (BRIRs) for shoebox-approximations of rooms with a low computational effort. For the early reflections an image source model (ISM) [3] is used up to third reflection order. To generate the diffuse late reverberation, a feedback delay network (FDN) [4] is used which is spatially rendered using a limited number of twelve directions on a 3-dimensional cube around the listener to create a spatially diffuse sound field (see Fig. 1).

In this work, the spatial rendering of RAZR was adapted to loudspeaker arrays. Using perception experiments with normal-hearing listeners, the effect of the number of directions (virtual sources) to render the diffuse late reverberation was assessed.

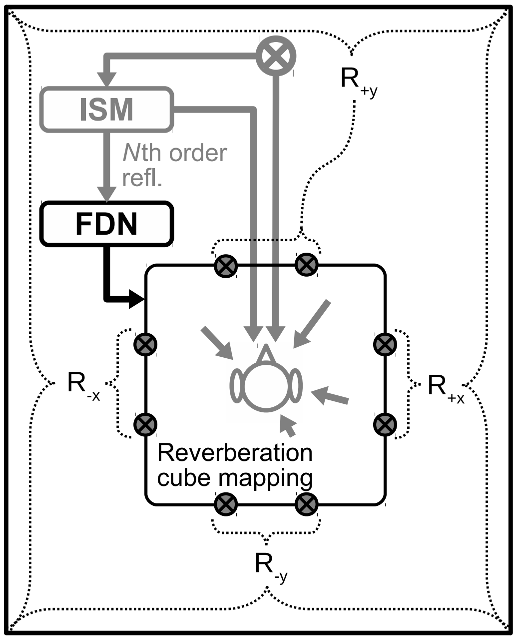

Figure 1: Block diagram of RAZR with sound source (gray ⊗), Nth order ISM, FDN, reverberation cube with two virtual sources on each surface (black ⊗) and the respective reflection coefficient for each wall ($R_{\pm x}, R_{\pm y}$).

Spatial reverberation rendering in RAZR

In RAZR, the spatial rendering of the late reverberation is realized by spatially distributed discrete sound incidence directions to render the individual channels of the multichannel FDN output. In the current setting, the choice of twelve FDN delay lines also yields a number of twelve directions used to spatially sample the reverberant field as illustrated in Fig. 1 for the 2-dimensional horizontal part. The FDN outputs mapped to these twelve directions will be referred to as virtual reverberation sources (VRS). The VRS are placed on a cube that is centered around the listener, two of them positioned on each diagonal of the cube surfaces.

For headphone rendering, the directions of the virtual sources are rendered via head-related transfer functions (HRTFs). In addition, reflection filters are applied in order to account for different reflective properties of the walls. By this the spatial distribution of the late reverberant field can be influenced, which might become relevant when, e.g., one wall is strongly reflecting or the listener is close to a highly absorbing wall.

By limiting the FDN to 12 channels and placing the VRS on the cube, all walls (with possibly unequal reflection coefficients) are weighted equally. The reflection coefficient of each wall is assigned to the corresponding two sources of each cube surface (see reverberation cube mapping in Fig. 1). In the case of elongated rooms, such as corridors, or rooms with widely differing reflection coefficients per wall, there may be an insufficient spatial resolution for rendering the late reverberation, and the effect of wall reflection coefficients on the spatial distribution of the late reverberant field might be incorrect. One way to increase the spatial resolution is the use of more VRS, however, at the cost of computational efficiency.

Reproduction of RIRs in three-dimensional loudspeaker arrays

In general, there are two possible ways to auralize simulated room impulse responses. For headphone auralization, typically HRTFs are applied. Without head tracking and dynamic update of the HRTFs, this rendering method results in sound sources that move in the same way as the listener's head. Another possibility is the use of loudspeaker arrays. In this case no HRTFs are applied and the sound sources stay at fixed positions when the listener is moving the head.

There are several techniques to render a sound source over a set of loudspeakers. In the nearest speaker (NS) approach, only that loudspeaker closest to the desired sound source direction to be presented is used. More accurately, virtual sound sources can be placed between existing loudspeakers using Vector Base Amplitude Panning (VBAP) [5]. Here, similarly to classical stereo panning, the three loudspeakers closest to the virtual sound source are driven with amplitude weightings derived for the specific source direction.

For this study, RAZR was extended to run on loudspeaker arrays. VBAP was used to render the directions of the direct sound, image sources and VRS.

Spatial resolution of late reverberation

In order to adjust the spatial resolution of the late reverberation the number of VRS was modified to be 6, 12, 24, 48 or 96. An even spatial distribution was achieved by using a geometric distribution of the VRS which maximized the sphericity of the resulting polyhedron and allowed an assignment of an equal number of VRS to each of the 6 room surfaces.

The VRS were fed by the output of a fixed number of 96 FDN channels with delays according to sound traveling times corresponding to 24 room directions and diagonals in the shoebox geometry of the room. Extension to 96 delays was done by adding random jitter.

Accounting for room properties

In the current implementation of RAZR each of the VRS is assigned to a single wall and filtered with the associated reflection coefficient (see equation 1 below). This can be problematic for, e.g., elongated rooms and receiver positions close to a wall. Here reflections from walls close to the receiver and walls with a large surface area are more represented (higher proportion of reflection coefficients) than walls far from the receiver and with a small surface area.

An example of such a room configuration is shown in Fig. 2 for a number of 48 VRS in the horizontal plane. Here, in addition to the elongated room geometry, the receiver is placed close to the $+y$ wall (red and blue labels along the black dashed normals mark the direction). The inner circle of red and blue colored $\times$ shows the weighting for the old approach. All walls are represented by the same number of VRS. So far, the indicated $\mathsf{VRS_i}$ is weighted with the reflection coefficient

$$R(\mathsf{VRS_i}) = R_{-y} \tag{1}$$

of the wall $-y$ since it has formally been assigned to that wall. However, from Fig. 2 it is clearly visible that the $\mathsf{VRS_i}$, seen from the listener, should rather be assigned to the wall $+x$. Taken all VRS together, this leads to the same representation of the virtual sources belonging to wall $-y$ and to wall $+y$.

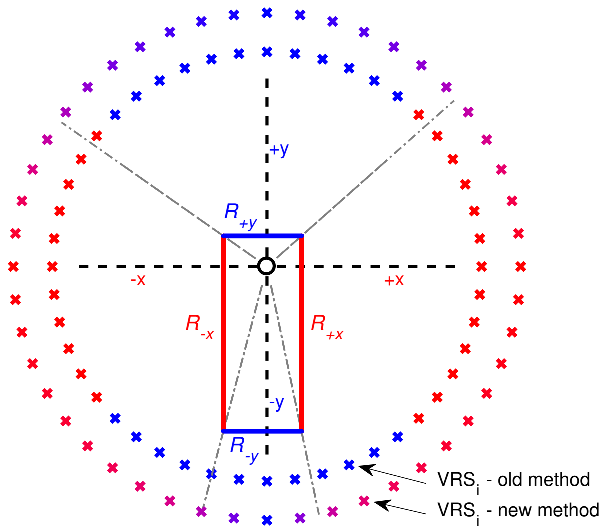

Figure 2: Distribution of VRS (red and blue $\times$) around the listener (gray circle) in a room (red-blue box with corresponding reflection coefficients). Inner circle: weighting according to the old method. Outer circle: weighting according to the new method. Gray dashed dotted lines: intersection of room vertices. Black dashed lines: normal vectors of walls respective to listener position. $\mathsf{VRS_i}$: applied reflection coefficient for one VRS.

This will be audible if, e.g., the $-y$ wall has distinct reflective properties than the other walls. To account for these differences, and to ensure continuity, the mapping of the wall properties to virtual sources was modified such that not only one wall is represented, but also the adjoining ones, depending on the actual VRS position. This is shown in the outer circle of $\times$ in Fig. 2. In the case of $\mathsf{VRS_i}$, the effective reflection coefficient is a mix of R_{-y} and R_{+x}:

$$R(\mathsf{VRS_i}) = \alpha_i \cdot R_{+x} + \beta_i \cdot R_{-y}\,, \tag{2}$$

and is visible by the violet coloration of $\mathsf{VRS_i}$. The coefficients α_i and β_i are calculated via VBAP according to the position of $\mathsf{VRS_i}$. The crossfade area between two walls is between the two normal vectors of each wall (black dashed lines) which have an angle of 90° to each other. However, the point of same weighting of R_{-y} and R_{+x} should not be at the angle of 45°, but at the intersection of the room vertices (gray dashed-dotted lines). To achieve this, the original VRS positions can be warped before calculating the weightings of the reflection coefficients using VBAP (equation 2 for the 2D-case). Analogically, this method is applied to the virtual sources in the actual three-dimensional space.

Perceptual Evaluation

The perceptual evaluation aimed to find answers to the following research questions: What is the required spatial sampling to reproduce a diffuse sound field? And how to account for room geometry for spatial rendering of the late reverberation?

A listening test was performed to clarify these questions.

10 self reported normal-hearing subjects participated in the test in the virtual reality lab at the university of Oldenburg. The virtual reality lab consists of an anechoic chamber ($7 \times 9 \times 7$ m^3) and 86 loudspeakers (Genelec 8030) arranged in a spherical shape with a diameter of 5 m.

In the first experiment, different simulated room shapes were tested: a cube-like room ($4.97 \times 4.12 \times 3$ m^3) with a reverberation time of 0.41 s and a source-receiver distance of 1.7 m, and an elongated room ($12 \times 30 \times 10$ m^3) with a reverberation time of 3.41 s and two source receiver distances of 2.7 m and 13 m, respectively. In the second experiment, a room ($8 \times 24 \times 6$ m^3) with five low absorbing walls and one highly absorbing wall was used. The absorption coefficients for the low absorbing walls were set to [0.019, 0.001, 0.022, 0.047, 0.067, 0.111] and for the highly absorbing wall to [0.999, 0.999, 0.999, 0.999, 0.999, 0.999] for the frequencies [250, 500, 1k, 2k, 4k, 8k] Hz, respectively, resulting in a reverberation time of 1.08 s. The source and receiver were at a fixed distance from each other (4 m) and the distance between source and receiver to the highly absorbing wall was varyed in five steps (0.18 m, 2.31 m, 4 m, 6,93 m and 15 m). The receiver was always oriented towards the source, and both were arranged such that the highly absorbing wall was to the left of the receiver. See Fig. 3 for illustration.

In the first experiment, two test signals were used: A pink weighted impulse, whose spectrum decreases with $1/f$ (being f the frequency) and unreverberated speech sentences of male and female talkers. The test signals were convolved with the multi-channel impulse responses generated by RAZR. Note that for each room and source-receiver condition the direct sound and the ISM were the same, and only the mapping of the FDN to different numbers of VRS was varying. The test procedure was an ABX test, that is, in each trial the signals A and B were either presented in the order ABA or ABB. The subjects had to compare the signals in the three intervals and their task was to tell whether the third interval contained the A or B signal. The assignment of reference or test signal to be A or B was randomized. The reference signal was always the rendering with 96 VRS, and the test signals were varying between the reduced numbers of VRS, 6, 12, 24, and 48. Each condition was repeated 20 times to ensure the subjects do not only guess, thus the chance probability of an ABX test is at 50 %.

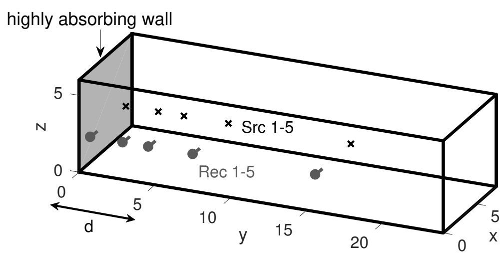

Figure 3: Room with one highly absorbing wall (gray shaded) and five different distances d between source-receiver to this wall. Fixed distance between source and receiver (4 m) and fixed orientation of the receiver towards the source.

The results of the first experiment are shown in Fig. 4. It can be seen that the answers for almost all conditions are within the chance probability indicated by the lower dotted gray line. Only for the $1/f$-pulse and the 6 VRS compared to 96 VRS the subjects were able to perceive a difference. The results for the 6 VRS are clearly above the significance level of 66 % that is marked by the upper gray dashed line in the Figure.

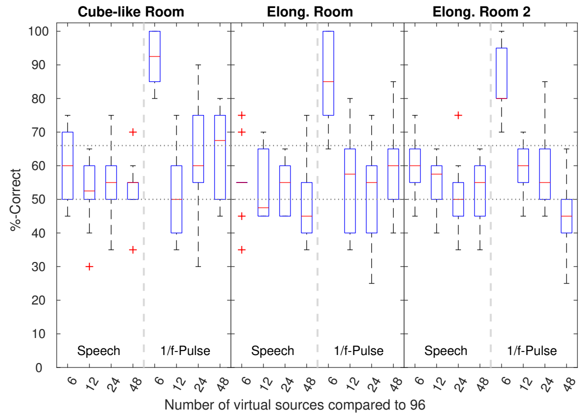

Figure 4: Results of experiment 1: Percentages of correctly detected lower spatial reverberation resolution for the three room conditions and the two test signals, averaged over trials and subjects. Lower gray line: chance probability of 50 %; upper gray line: significance threshold at 66 %.

Figure 5 shows the results of the second experiment where the distance of the source and receiver to a highly absorbing wall was varied. Here, the $1/f$ pulse was used as test signal. As in the first experiment, most of the subjects' answers are within the chance probability. Again, only for the number of 6 VRS the difference to 96 VRS is perceivable, if the distance to the wall is not very small. For the smallest distance chosen, the subjects were able to hear a difference between all reduced numbers of VRS compared to 96 VRS.

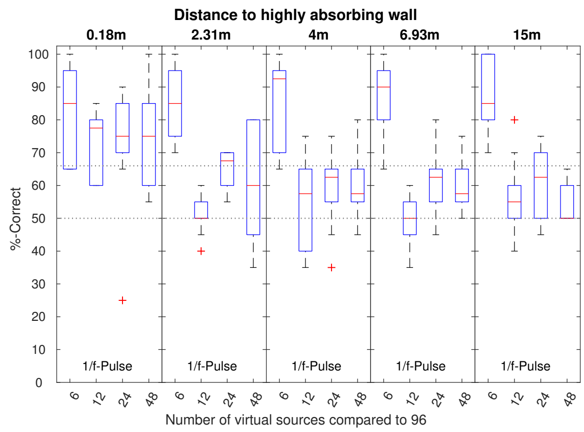

Figure 5: Results of experiment 2: Percentages of correctly detected lower spatial reverberation resolution for the five distances to the highly absorbing wall, averaged over trials and subjects. Lower gray line: chance probability of 50 %; upper gray line: significance threshold at 66 %.

In conclusion, it can be said that already with 12 virtual sources a good spatial representation of the late reverberation is achieved. Only for rooms with, e.g., a highly absorbing wall and receiver positions close to that wall more virtual sources might be necessary to reproduce the spatial properties of the reverberant field perceptually more faithful.

Summary and Conclusions

In this paper a reproduction method for room impulse responses in a three dimensional loudspeaker array was presented and the spatial resolution required for reproduction of the late reverberation was assessed.

A new method to map different reflection coefficients of the walls to VRS directions was introduced. The suggested method can be applied independently from the number and position of virtual reverberation sources.

The perceptual evaluation showed it seems to be sufficient (at least for the tested conditions and the listener position in the center of the loudspeaker array) to have a spatially rather sparse rendering of the reverberant field.

References

[1] Laitinen, M. V., and Pulkki, V.: Binaural reproduction for directional audio coding. In Applications of Signal Processing to Audio and Acoustics. WASPAA'09. IEEE Workshop on (pp. 337–340), 2009

[2] Wendt, T. et al: A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation. J. Audio Eng. Soc., Vol. 62, No. 11, 2014

[3] Allen, J. B. and Berkley, D. A.: Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Am. 65(4), Apr. 1979

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

[5] Pulkki, V.: Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning, J. Audio Eng. Soc., Vol. 49, No. 9, 2001