3D-Reverb for a Wavefield Synthesis System with Ceiling Speakers

Hendrik Bleier[1], Dieter Leckschat[2], Christian Epe[3]
[1] FH-Duesseldorf, University of Applied Sciences, E-Mail: hendrik@tonecraft.de
[2] FH-Duesseldorf, University of Applied Sciences, E-Mail: dieter.leckschat@fh-duesseldorf.de
[3] FH-Duesseldorf, University of Applied Sciences, E-Mail: christian.epe@fh-duesseldorf.de

Introduction

A Wavefield Synthesis System (WFS), as the one located at the University of Applied Sciences Düsseldorf, claims to bear the most realistic physical resemblance and recreation of a given sound field for a listener, independent of his or her listening position. To achieve this goal, as opposed to a stereo or surround setup, a sound is produced through several loudspeakers placed next to each other in a large ring of sound sources. A virtual sound source can then be placed anywhere outside or inside the ring of loudspeakers. It is then easy to identify the directivity of the sound source from almost anywhere in the room. Again, the goal is a realistic reproduction of a real sound field, rendered in real-time. The one thing that has not been taken care of, yet, are the room acoustics.

This paper presents the results of the bachelor thesis of one of the authors (Bleier). It describes the concept and development of a multichannel convolution reverb for a 3D WFS System in order to implement realistic room simulation capabilities within the WFS rendering.

Basic concept

Substantially, most real rooms resemble a shoebox of a given length, width, and height with four opposing walls and a ceiling that each reflect sound. The novel idea in this project is to capture the reflections of each wall separately with super cardioid microphones to then convolve each of the virtual audio sources with the room`s “sampled” walls. Since each virtual sound source has a specific x and z position in the room, we can use the source – wall distance to calculate the delays of the early reflections depending on the source’s position. With the same method we can attenuate the early reflections because the distances to each wall vary and the Inverse Square Law applies. This leads to an even more realistic reproduction of the early reflections. The diffuse or late reverb has by definition equal sound levels from all positions around the listener, so the late reverb will not be attenuated or delayed. To match the late reverb with the overall loudness of the early reflections they can be adjusted depending on the early reflections. In order to use all the algorithms and parameters already available in the WFS Software 2.0, we developed the convolution engine as an upgrade of the 2.0 software to its current 3.0 state.

Room Impulse Response

To obtain a room impulse response (RIR) that fits the software’s requirements we measured a room with five microphones, each representing one wall of the approximated shoebox room - **L**eft, **R**ight, **F**ront, **B**ack and **C**eiling. To gain as much decorrelation as possible, we mounted the super cardioid microphones to a microphone rig which puts them two meters apart from each other (figure 1).

Figure 1: Microphone rig in the St. Dionysius church in Cologne with microphones each pointing to a different wall.

To excite the room we used an exponential sine sweep of 18^{th} order from a dodecahedron at the main point of sound emission, like the altar area or the organ in a church. We achieved a signal-to-noise ratio (SNR) of about 80 to 90dB. Future RIR capturing might achieve better results. The low end of the dodecahedron was for instance not enough to excite the two rooms we captured in such a way that the RIR represents the actual, real room. Although we post processed the RIR to eliminate the nonlinearity of the speaker, the energy emission of the low frequencies still needs to be higher in order to obtain a better SNR.

Position based delay and attenuation of early reflections

The reverb engine is based on the well-known area of convolution. Novel is the position-based delay and attenuation of the early reflections which are only possible because of the given source coordinates from the WFS Software. We basically use the mirror source model [1][2] to calculate the necessary parameters. Each wall is placed on its edge within the virtual room and emits the processed convolved sound as a plane wave. Therefore the WFS algorithm already calculates a delay for the distance from the wall to the listener, so that only the remaining distance has to be evaluated. Figure 2 illustrates the geometrical relationship. We can therefore calculate the remaining distance *s* e.g. to the front wall with *posX* and *posZ* being the virtual sound source’s coordinates and *roomSizeZ* being the room’s virtual dimentions as follows:

$$s = \sqrt{(roomSizeZ + posZ)^2 + (posX)^2} - \frac{roomSizeZ}{2} \quad [\mathrm{m}] \quad (1)$$

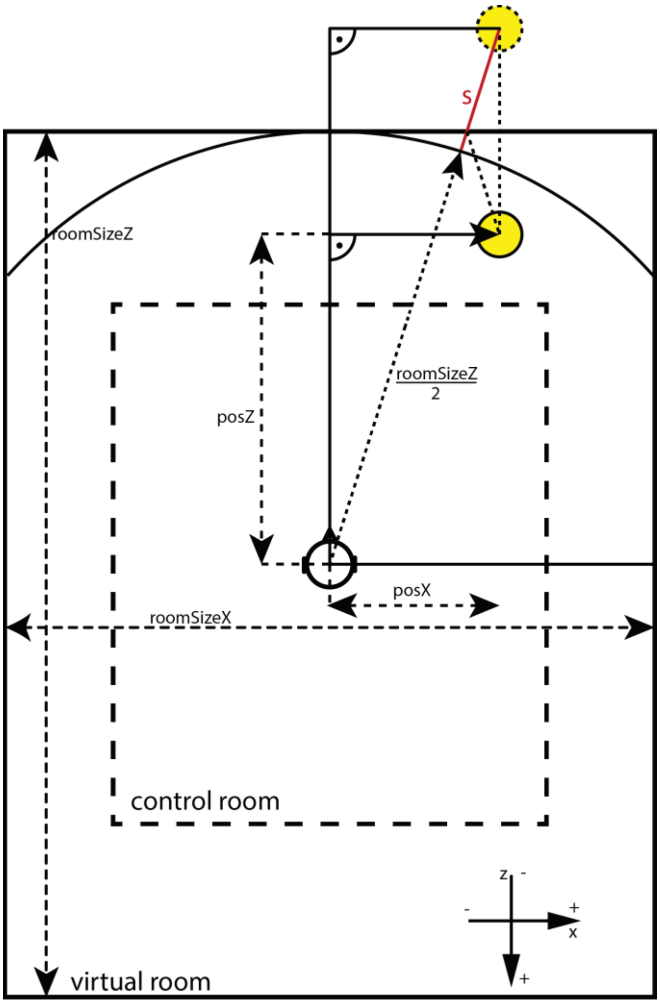

Figure 2: Mirror image source model as used for the delaying of the early reflections part of the reverb.

The calculation of the attenuation follows the same principle but determines the shortest distance from the source position to a wall and calculates each attenuation relative to that distance. For attenuation we use the Inverse Square Law.

Figure 3: Early / Late Parameters for merging the two reverb parts.

It is possible to control the merging point of early reflections and the late reverb (figure 3). The merge knob sets a transition midpoint, X-over can be used to vary the cross fade time and the "Early – Late" knob regulates the ratio between the two reverb parts. The merge midpoint is calculated from the room´s dimensions, like the volume V of a room, as proposed by Polack et. al [3]. Accordingly, the mixing time can be calculated as follows:

$$t_{mixing} \approx \sqrt{V} \qquad [s] \quad (2)$$

Manual regulation of the parameter is still possible, of course. The user can choose the crossover to be linear or logarithmic.

Figure 4: Linear merging of early and late reverb buffers.

Audio Signal Flow

In order to reduce processing load it is not possible to convolve each incoming audio source five times separately. Therefore, as mentioned above, we split the RIR into early reflection and late reverb parts. We copy the incoming audio data to convolve one copy with the early reflection part and the other with the late reverb part of the RIR. We delay and attenuate the audio data that is to be convolved with the early part. We then mix each individual channel to a wall buffer that stores the audio data to be convolved with each wall. This results in ten separate audio buffers, each holding as many channels as audio source input channels exist. The first five buffers hold the audio data for the early part, the last five hold the not-delayed and not-attenuated data for the late part. We then feed the convolver with all ten buffers. Finally, we mix the convolution output to five channels that together resemble the five walls of the room, placed in the virtual room as plane wave audio sources. To the rendering software this lets them appear like incoming audio sources (excluded from additional convolution, of course) so that it renders them as normal plane wave sources. Figure 5 shows the audio signal flow before the convolution.

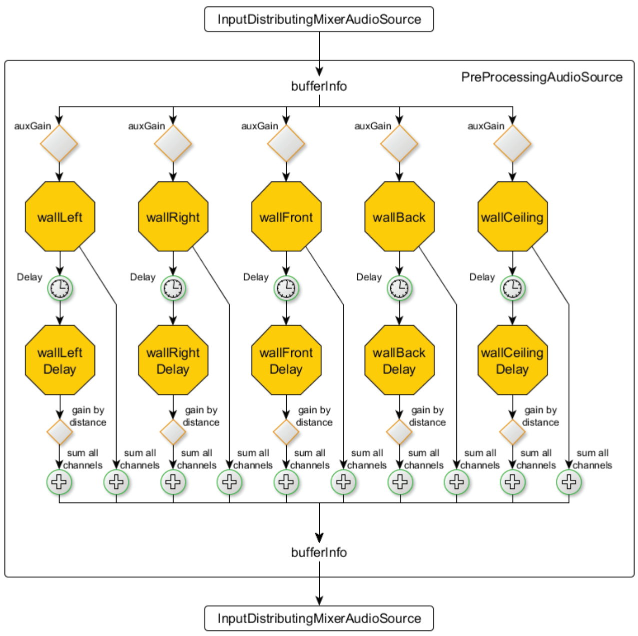

Figure 5: Audio signal flow chart to preprocess the incoming audio data and prepare for convolution.

Room rotation

Because the resulting virtual wall reflections are set up as plane waves with set x and z coordinates in the virtual room, the user can reposition them as needed. The point of each wall follows an elliptical shape through the room in order to maintain relative distances and to meet the requirements of non-cubic rooms. Figure 7 shows the geometrical relationship to calculate the point of each wall in the virtual room. This way, possible standing waves that arise from the geometry of the control room can be reduced with a slight rotation of just a few degrees.

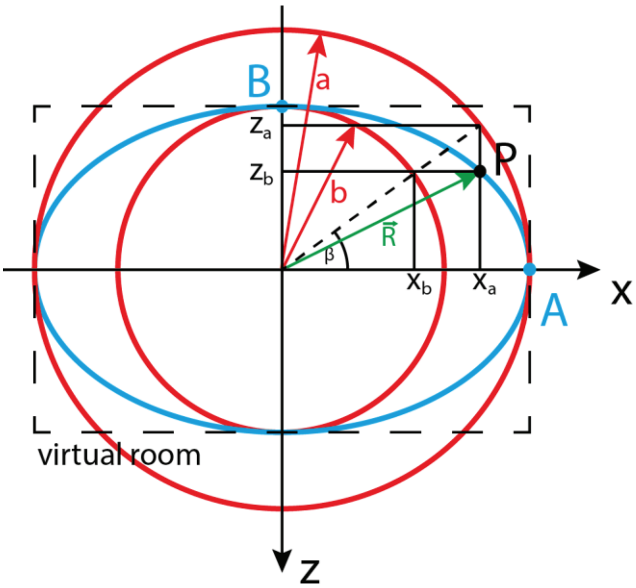

Figure 6: Geometrical relationship to find point P of each virtual wall in the virtual room.

Graphical user interface and usability

The graphical user interface (GUI) is designed to guarantee the user the shortest learning curve possible. The GUI is built like other well-known convolution reverb plugins. Therefore all the usual parameters can be found in the reverb engine of the 3D WFS Software. Predelay, stretch, begin, end, attack and decay parameters manipulate the impulse response directly, while the equalizers filter the output of the reverb engine. The filters can be switched between low / high pass filters or shelving filters. Via the GUI it is also possible to interchange the RIR channels with the convolution channels. Figure 7 shows the GUI of the 3D WFS Software 3.0. The top part shows the WFS input channels, the bottom parts the parameter knobs for manipulating the convolution reverb engine.

Future work and development

The 3D-Reverb upgrade for the WFS Software is ready to use, but not evaluated, yet. First simulations demonstrate the realistic room simulation capabilities to reverb creation of our approach. Listening tests for evaluating the spatiality and dummy head recordings to further substantiate the concept scientifically still need to be done. Additional room impulse responses also need to be generated in order to have a bigger variety for the production workflow.

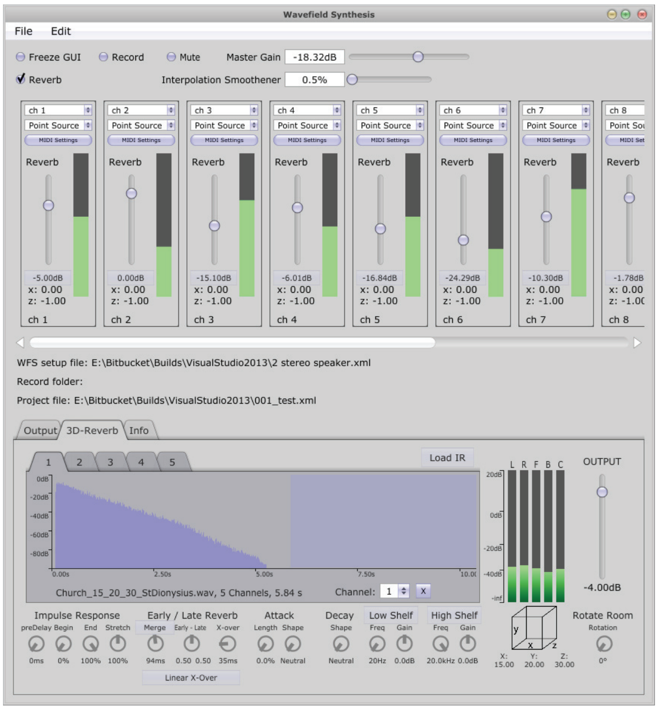

Figure 7: GUI of the new 3D-WFS Software 3.0 with a loaded RIR of a midsized church.

Literature

[1] Boone, M. M., de Vries, D., and Berkhout, A. J., 1994, *Sound Control:* Department of Applied Physics, Delft.

[2] Kuttruff, H., 1991, *Room Acoustics:* Elsevier Science Publishers Ltd., London, 3rd edition.

[3] Polack, J.-D., Defrance, G., 2008, *Measuring the mixing time in auditoria*, Acoustics Paris, Institut Jean le Rond d`Alembert, Paris