On the Generation of Virtual Early Reflections in Wave Field Synthesis

Jens Ahrens
Quality and Usability Lab, Technische Universität Berlin, Email: jens.ahrens@tu-berlin.de

Abstract

In this paper, we present an informal analysis of the impact of spatial aliasing as well as the response of the listening room on the auralization of acoustic room models in wave field synthesis.

Introduction

Wave field synthesis (WFS) aims at creating wave fronts with a controlled curvature by means of dense loudspeaker arrays for the auralization of audio content. Audio practitioners recognized very early in the history of audio presentation that reverberation is a major contributor to a convincing perception [1]. While a vast amount of literature regarding the generation of the direct sound in WFS is available, the generation of reverberation has always been somewhat of a stepchild [2].

A first outline of the process of creating artificial reverberation for WFS (or large-scale loudspeaker arrays in general) can be found in [3] where a two-stage implementation is described. Early reflections are generated using a mirror image model and late reverberation is generated using signals with appropriate statistical parameters.

The process of measuring multipoint room impulse responses for the capture of reverberation for convolution with dry (reverberation-free) source signals in order to obtain the appropriate reverberation for a given virtual sound source in WFS is described in [4]. Due to the large amount of data involved, a parameterization of the captured reverberation based on a plane wave representation and psychoacoustic criteria is proposed. However, no formal perceptual evaluation is provided. An extension to the approach from [4] enabling the manipulation of measured multipoint impulse responses based on a three-dimensional visualization using augmented reality technologies is presented in [5]. The manipulation is performed in time-frequency domain, and its motivation is the provision of more flexibility and artistic freedom to the sound engineer.

In [6], the suitability of WFS to create perceptually diffuse sound fields for the synthesis of late reverberation via a set a plane waves has been proven. Early reverberation was created using the mirror image model, but it was excluded from the evaluation. Input signals for the plane wave components can be obtained, e.g., from microphones distributed in the recording venue as they can deliver sufficiently uncorrelated signals. Other possibilities are discussed in [5]. The evaluation of the approach from [6] was extended to off-center listening positions in [7].

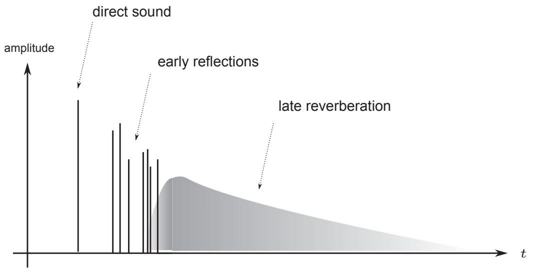

Figure 1: The model of reverberation that is assumed in this paper

For convenience, we assume in this paper that reverberation is composed of discrete early reflections that impinge from various directions, and which become gradually denser in time. Finally, after the so-called mixing time, late reverberation is apparent which exhibits an approximately exponential decay. Refer also to Fig. 1.

In the present contribution, we investigate the effect of spatial aliasing as well as the effect of the listening room on the presentation of virtual early reflections. All simulations presented in this paper were created with/based on the Sound Field Synthesis Toolbox [8].

Spatial Aliasing

Informal experiments by the authors suggest that a straightforward auralization of a given acoustic room model results in reverberation that sounds denser than expected. It was suggested in [9] that this perception is a consequence of the unavoidable spatial aliasing. Spatial aliasing constitutes additional spurious wave fronts that follow the first (wanted) synthesized wave front within a few milliseconds just like reflections in a room. Refer to Fig. 2 for an illustration. Spatial aliasing occurs only above the so-called spatial aliasing frequency, which depends on the loudspeaker spacing and is typically between 1,500 and 2,000 Hz. The loudspeaker array assumed in Fig. 2 and in the remainder of this paper exhibits a spatial aliasing frequency of approx. 1,700 Hz.

It was found in [10] that spatial aliasing reduces the interaural cross-correlation (IACC) and thereby increases the perceived width of a given virtual sound source in an anechoic room. This perceptual dimension is often referred to as *apparent source width* (ASW) in concert hall acoustics. Spatial aliasing was also found to give rise to coloration and smearing of transients [11].

In order to evaluate if the reduction of the IACC due to spatial aliasing has an impact of the perception of virtual reverberation, we conducted the following informal experiment:

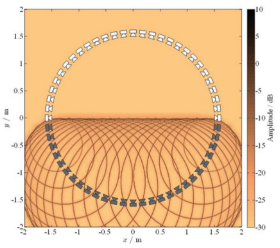

Figure 2: Sound pressure of a sample virtual planar wave front propagating into positive y-direction and carrying a time-domain impulse; the wave is synthesized by a circular 56-channel array; a cross-section through the horizontal plane is depicted; all wave fronts other than the leading straight wave front are spatial aliasing; gray loudspeaker symbols indicate active loudspeakers; white loudspeaker symbols indicate inactive loudspeakers

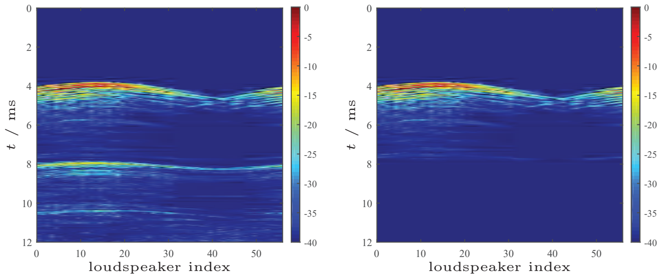

Figure 3: Sample BRIRs (left, listening room included) and HRIRs (right, listening room suppressed) of the 56-channel circular array installed at the Quality and Usability Lab

We measured the binaural room impulse responses (BRIRs) of an manikin in various rooms ranging from a mid-size meeting room (141 m^3) to an opera house (4000 m^3) for one combination of source and receiver positions and different head orientations (1° resolution for head-tracked auralization using the SoundScape Renderer[1]). In order to have a parametric model of the measured BRIRs, we recreated them by manual tuning of a set of filtered head-related impulse responses (HRIRs, for the early reflections) and filtered Gaussian noise (for the late reverberation).

The parametric room model was then auralized in WFS. The WFS auralization was performed over headphones and was created based on measured BRIRs of the loudspeaker array installed at the authors' affiliation to the ear drums of a manikin. The parameters of the array are identical to those of the array assumed in Fig. 2. In order to suppress the response of the room in which the loudspeaker array was installed — a.k.a. the *listening room* —, the array's BRIRs were faded out right before the first room reflection would arrive as illustrated in Fig. 3.

Fig. 4 shows spectrograms of a sample measured and WFS-synthesized ear response for the 141 m^3 meeting room "Sputnik". Informal listening[2] suggests that the WFS-auralized parametric model does indeed sound slightly more spacious that the manikin measurement.

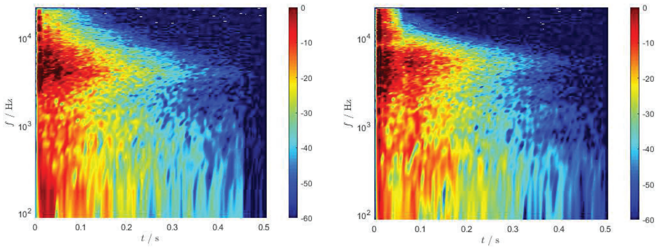

Figure 4: Spectrograms of sample BRIRs of the room "Sputnik" (measured, left image; $t_{\mathrm{mix}} = 80$ ms, $t_{60} = 500$ ms) and of the loudspeaker array from Fig. 2 when auralizing the parametric model of "Sputnik' (listening room response suppressed, right image)

Similar observations were made for the other sample rooms that were investigated. This suggests that the impact of spatial aliasing on the perception of virtual reverberation seems detectable but limited.

The strength of the spatial aliasing can be increased by increasing the spacing between adjacent loudspeakers, for example, by using only every other loudspeaker of the given array, which makes the spatial aliasing frequency drop by a factor of 2. Informal listening suggests that it is difficult to assess the perception of the resulting reverberation independent of the severe impact that the aliasing has on the perception of the direct sound. All attempts of the authors to separate spurious aliasing wave fronts from the desired ones in the BRIRs were not successful.

Interaction of Loudspeaker Array and Listening Room

A potential cause for the observed excessive density of the reverberation created in WFS can be the interaction between the loudspeaker array and the listening room.

A first analysis of the matter based on simulations and measurements of a linear WFS array below the spatial aliasing frequency was presented in [12]. Simulations on the interaction of loudspeaker array and listening room based on the image source method for the entire audible bandwidth in [13] suggested that in the present situation, a room-in-room presentation (a virtual room created inside a real room), the energy decay, particularly the early decay, of the evolving sound field is indeed altered compared to the virtual room model alone. In the remainder of this paper, we investigate the matter based on the measured BRIRs of the loudspeaker array from Fig. 2.

Comparison of the energy decay of a single loudspeaker of the array under investigation and the array when synthesizing a virtual plane wave reveals indeed that the excitation of the room occurs over a longer period in the latter case. This is illustrated in Fig. 5. The actual decay of the room response occurs after a sustained excitation phase of a duration of a few milliseconds (ca. 5-15 ms).

---

[1] `http://spatialaudio.net/ssr/`

[2] Refer to `http://www.soundfieldsynthesis.org/audio_examples/DAGA2016` for audio examples

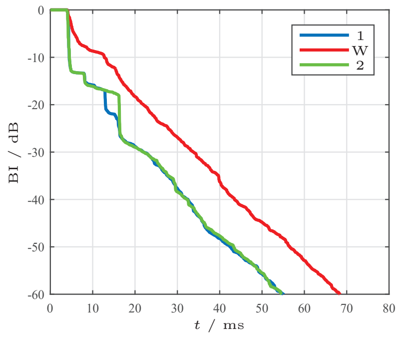

Figure 5: Backward integrated [15] BRIRs of one speaker of the array (blue), two speakers of the array in a stereo configuration (green), and the array synthesizing a virtual plane wave (red)

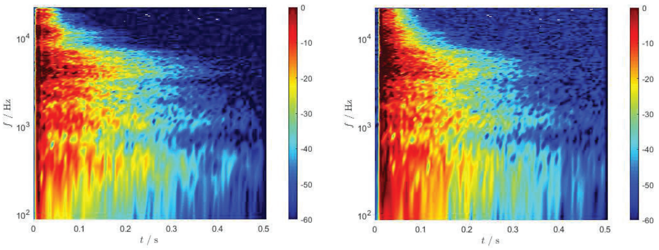

Figure 6: Spectrograms of sample BRIRs of the loudspeaker array from Fig. 3 when auralizing the parametric model of the room "Sputnik" with the listening room response suppressed (left; same picture as Fig. 4, right) and included (right)

Unlike in [14], virtual plane waves were chosen instead of individual loudspeakers to render early reflections because the amplitude decay over distance of a real room reflection is significantly lower than that of an individual loudspeaker. The amplitude decay of a virtual plane wave along its propagation path is roughly 3 dB per each doubling of the distance to the loudspeaker array over a large frequency range and is therefore closer to that of a real room reflection.

Interestingly, the energy decay of two speakers of the array that are positioned in a stereo configuration is similar to the decay of one single speaker (green and blue curves in Fig. 5). We can therefore not transfer the extensive available knowledge on the interaction of stereo speaker systems and listening rooms [16] to the present situation.

With the present loudspeaker array, roughly 20 loudspeakers contribute with significant amplitude to the synthesis of a virtual planar wave front (cf. Fig. 2). The room responses of all active loudspeakers superpose and cause a reflection pattern that is denser than the reflection patter caused a single loudspeaker or by a stereo setup by a factor of 20. Fig. 6 depicts a sample resulting BRIR.

The obvious observation is that the BRIR that includes the room response exhibits higher energy. Although not clearly deducible from the figure, the BRIR with the listening room included is expected to be longer than the anechoic one by the duration of the excitation phase evident in Fig. 5, i.e., by approx. 10 ms. The reflection density cannot be estimated from the spectrogram as the time resolution is too low.

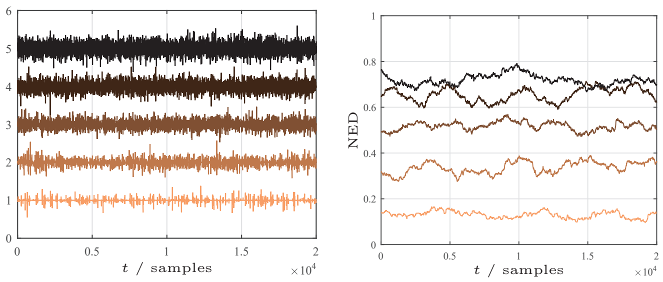

Figure 7: Left: different random signals with pulse densities of 800/s, 3,200/s, 7,200/s, 12,800/s, and 20.000/s (from bottom to top); the time instants of the pulses were determined using a Poisson process; the amplitudes of the pulses were taken from a normal distribution; Right: respective NEDs

Informal listening suggest that the WFS-auralized room models do sound considerably more spacious when the response of the listening room is included compared to when it is suppressed (cf. [2]).

Measurement of the Reflection Density

It would be convenient to have an instrumental means for measuring the actual evoked reflection density in order to, for example, manipulate a given room model such that the reverberation that is finally evoked in the listening area (i.e., the virtual reverberation plus the response of the listening room) is as similar as possible to the original room model. We present an attempt here.

The normalized echo density (NED) [17] appears to be a promising candidate for an instrumental measure of reflection density. It essentially slides a window through a given room impulse response and measures for each position of the window how similar the amplitude distribution of the signal segment in the current window is to a normal distribution. The assumption is that a signal whose amplitude is normally distributed is stochastic (like late reverberation). Originally, NED was created to measure the mixing time of a room instrumentally. It turned out that NED rises during the early part of a room impulse (where the reflection density presumably rises) up to the value 1, which represents a purely statistical signal.

Fig. 7 illustrates the dependency of NED on the pulse density of synthetic random signals. It can indeed be observed that NED rises with pulse density. It is unclear at this point, if this observation holds for general signals.

Fig. 8 compares the NEDs for the parametric room model of "Sputnik" when auralized with WFS with the response of the listening room being suppressed (cf. Fig. 6, left) and with the response of the listening room included (cf. Fig. 6, right).

We would assume that the reflection density when the listening room is included is higher than when the listening room is suppressed by a factor of 20, which we

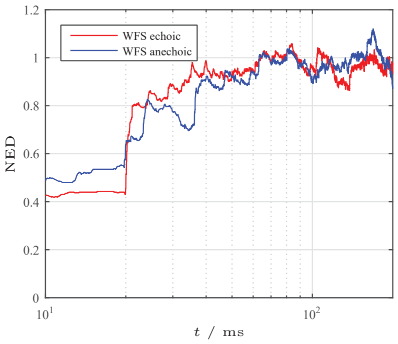

Figure 8: NED of the parametric room model of "Sputnik" auralized with the WFS system from Fig. 2 with the listening room response suppressed (blue) and included (red)

would assume to cause a noticeable increase of NED. Unfortunately, Fig. 8 does not reflect this expectation. The results for all other tested rooms are similar. NED therefore does not seem to be a reliable indicator for the actual density of reflections and the associated spaciousness that is evoked in a human listener.

Conclusions

Our findings suggest that spatial aliasing has a weak but detectable impact on the perception of virtual rooms rendered in wave field synthesis. The energy of spatial aliasing is rather low as it typically occurs only above 1,500 Hz, where most signals exhibit low energy. This might explain the observation that the impact is weak although the aliased wave fronts are distinct and pronounced.

The response of the listening room to the array seems to be more significant. A large number of loudspeakers is usually involved in the synthesis of a synthetic wave front. All speakers radiate a signal with full bandwidth in this case (whereby a highpass filter with a slope of 3 dB/octave is active below the spatial aliasing frequency [2]). Each of the active speakers evokes its own reflection pattern so that the reflection density in the actual synthesized sound field is significantly higher than in the room model that is being rendered. Though, we have not been able to measure the actual resulting reflection density instrumentally. A series of formal user studies is in preparation.

References

[1] Roey Izhaki. *Mixing Audio - Concepts, Practices and Tools*. Focal Press, Oxford, UK, 2007.

[2] Jens Ahrens. *Analytic Methods of Sound Field Synthesis*. Springer-Verlag, Berlin, Heidelberg, 2012.

[3] D. de Vries, A. J. Reijnen, and M. A. Schonewille. The wave field synthesis concept applied to generation of reflections and reverberation. In *96th Convention of the AES*, Amsterdam, The Netherlands, February/March 1994.

[4] Edo Hulsebos. *Auralization using wave field synthesis*. PhD thesis, Delft University of Technology, 2004.

[5] Frank Melchior. *Investigations on spatial sound design based on measured room impulses*. PhD thesis, Delft University of Technology, 2011.

[6] J. J. Sonke. *Variable acoustics by wave field synthesis*. PhD thesis, Delft University of Technology, 2000.

[7] Jens Ahrens. Perceptual evaluation of the diffuseness of synthetic late reverberation created by wave field synthesis at different listening positions. In *DAGA*, Nuremberg, Germany, March 2015.

[8] Hagen Wierstorf and Sascha Spors. Sound field synthesis toolbox. In *132th Convention of the AES*, Budapest, Hungary, May 2012.

[9] Jens Ahrens. Challenges in the Creation of Artificial Reverberation for Sound Field Synthesis: Early Reflections and Room Modes. In *EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany, April 2014.

[10] Evert Walter Start. *Direct Sound Enhancement by Wave Field Synthesis*. PhD thesis, Delft University of Technology, 1997.

[11] Hagen Wierstorf. *Perceptual Assessment of Sound Field Synthesis*. PhD thesis, University of Technology Berlin, 2014.

[12] T. Caulkins, A. Laborie, E. Corteel, R. Bruno, S. Montoya, and O. Warusfel. Use of a high spatial resolution microphone to characterize the early reflections generated by a WFS loudspeaker array. In *28th Conference of the AES*, Piteå, Sweden, June/July 2006.

[13] Vera Erbes, Sascha Spors, and Stefan Weinzierl. Analysis of a spatially discrete sound field synthesis array in a reflective environment. In *Euronoise*, Maastricht, The Netherlands, May/June 2015.

[14] B. U. Seeber, S. Kerber, and E. R. Hafter. A system to simulate and reproduce audio-visual environments for spatial hearing research. *Hearing research*, 260(1):1–10, 2010.

[15] M. R. Schroeder. New method of measuring reverberation time. *Journal of the Acoustical Society of America*, 37, 1965.

[16] Floyd E. Toole. *Sound Reproduction — The Acoustics and Psychoacoustics of Loudspeakers and Rooms*. Focal Press, Oxford, UK, 2008.

[17] J. Abel and P. Huang. Robust Measure of Reverberation Echo Density. In *121st Convention of the AES*, San Francisco, CA, October 2006.