CHALLENGES IN THE CREATION OF ARTIFICIAL REVERBERATION FOR SOUND FIELD SYNTHESIS: EARLY REFLECTIONS AND ROOM MODES

Jens Ahrens

University of Technology Berlin
Ernst-Reuter-Platz 7
10587 Berlin, Germany
jens.ahrens@tu-berlin.de

ABSTRACT

Practical implementations of sound field synthesis evoke considerable artifacts that have to be considered in the creation of artificial reverberation. The most prominent artifact is spatial aliasing, which manifests itself as additional wave fronts that follow the desired synthetic wave front in time. These additional wave fronts propagate into different directions and occur at intervals that are similar to the intervals at which acoustic reflections occur in real rooms. It may be assumed that the human auditory system is not capable of differentiating aliasing artifacts and room reflections so that a synthetic reflection pattern should be designed such that it evokes a plausible pattern together with the aliased wave fronts. Two potential solutions are outlined. Finally, the capability of sound field synthesis of synthesizing room resonances (room modes) is analyzed and the promising results are illustrated based on numerical simulations.

1. INTRODUCTION

Sound field synthesis approaches employ high numbers of loudspeakers in order to synthesize a given desired sound field over an extended area [1]. The two best-known methods are Wave Field Synthesis (WFS) [2] and Near-field Compensated Higher Order Ambisonics (also termed Ambisonics with Distance Coding) [3]. The vast part of the scientific literature so far has focused on the synthesis of the direct sound of virtual sound sources. However, the creation of appropriate reverberation may be considered as important or even more important for achieving a desired spatial impression. Throughout the paper we assume the simple yet effective model of reverberation being composed of discrete early reflections the density of which increases over time and that gradually turn into diffuse late reverberation. The time interval after which the perceptual transition occurs is referred to as *mixing time* [4].

While the perceptual properties of mid-size and large rooms are mostly governed by the later part of the reverberation, small rooms can exhibit distinct early reflection patterns and low-frequency resonances also termed *room modes* [5, 6, 7]. This paper focuses on the creation of appropriate early reflection patterns as well as room modes. Late reverberation is not considered as solutions already exist as discussed below.

A first outline of the process of creating artificial reverberation for WFS can be found in [8] where a two-stage implementation is described. Early reflections are generated using a mirror image model [9] and late reverberation is generated using signals with appropriate statistical parameters. In [10] the capability of WFS of creating perceptually diffuse late reverberation via a set of plane waves is proven. Early reverberation was created using the mirror image model but was excluded from the evaluation. Appropriate input signals for the plane waves can be obtained, e.g., from microphones distributed in the recording venue as they can deliver sufficiently uncorrelated signals.

In [11] a convolution reverb is described that uses multipoint room impulse responses in order to create the proper reverberation for a given virtual sound source in WFS from dry (anechoic) source signals. 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. [12] presents an extension to the approach from [11] that enables the manipulation of measured multipoint impulse responses based on a three-dimensional visualization using augmented reality technologies. The manipulation is performed in time-frequency domain and its motivation is the provision of more flexibility and artistic freedom to the sound engineer.

None of the above mentioned approaches considers the mentioned artifacts that practical implementations of sound field synthesis exhibit. The synthesis of room modes has not been discussed in the literature. The present contribution discusses two approaches for the design of appropriate early reflection patterns considering the unavoidable spatial aliasing artifacts. It then investigates the potential of sound field synthesis of synthesizing room modes. When considering early reflections, we focus on artifacts as they appear in spatially fullband sound field synthesis methods such as WFS and the Spectral Division Method (SDM) [1, 13]. Spatially narrowband methods like the members of the Ambisonics family exhibit artifacts with slightly different properties. The extension of the presented results to narrowband methods will be outlined. The presented results on room modes are valid for all methods since room modes are only perceptually significant at lower frequencies [14, 15] where all methods exhibit a similar high accuracy [1].

This contribution focuses on the creation of artificial reverberation. It is not clear at this stage how the results can be transferred to reverberation recorded/measured with microphone arrays.

2. EARLY REFLECTIONS

2.1. Properties of the Spatial Aliasing Artifacts

The sound fields created by practical sound field synthesis systems exhibit a number of deviations from the prescribed virtual field [1]. These deviations are termed *artifacts* and the most important artifact in the present context is *spatial aliasing*. The term spatial aliasing is typically used in a very broad sense and often refers to all artifacts that arise due to the combination of spatial discretization of the secondary source contour and the radiation properties of the involved secondary sources (i.e. the loudspeakers). Note that spatial aliasing can theoretically be avoided by using a continuous distribution of secondary sources. A detailed treatment can be found in [1, Sec. 4.4.4] and [16].

We illustrate the most relevant results based on the example scenario depicted in Fig. 1: A virtual plane wave that is synthesized by a circular distribution of 56 monopole loudspeakers. The properties of other non-focused virtual sources are very similar. Focused virtual sound sources are a special case in which pre-echoes arise [17]. They are excluded from the present investigation.

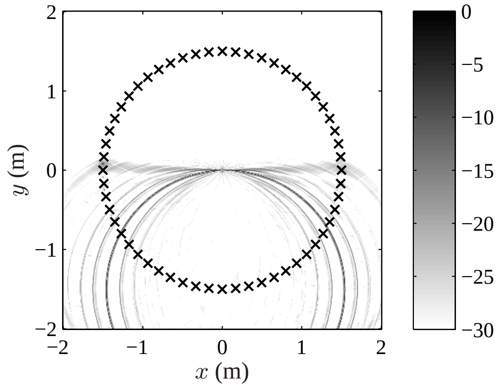
The two different basic options – spatially narrowband synthesis (27th order, Fig. 1(a)) and spatially fullband (infinite order, Fig. 1(b)) synthesis – are illustrated. It is evident that additional undesired wave fronts occur that follow the initial plane wave within a few ms (Fig. 2). A simple but useful model, especially for the spatially fullband example in Fig. 1(b), is the assumption that each active loudspeaker of the setup creates one additional spherical wave front that is emitted at the time instant at which the virtual plane wave passes the considered loudspeaker. As can be deduced from Fig. 2, the amplitudes of the additional wave fronts in Fig. 1(b) are only a few dB lower than the amplitude of the intended plane wave and are therefore perceptually relevant.

As illustrated in Fig. 1(c), the aliasing artifacts occur exclusively above a so-called aliasing frequency, which is approximately 1700 Hz in the current example[1]. Note that this behavior is very similar for spatially fullband and narrowband sound fields [1].

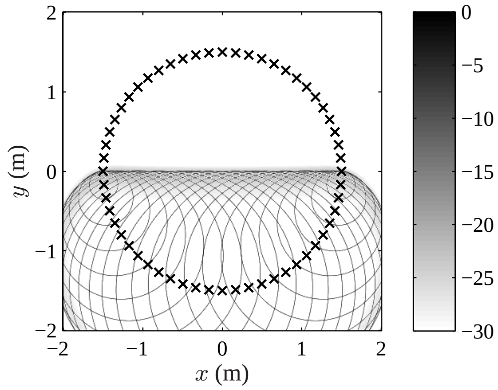
The timing and amplitude distribution of the aliasing artifacts is at least qualitatively similar to the timing and amplitudes of reflections in small reflective rooms. Fig. 3 shows some quantitative results. It is evident when comparing Fig. 3(a) and (c) that the artifacts arrive much denser than typical room reflections and in a time window that is much shorter. A situation in which a pattern evolves that is similar to the aliasing artifacts is when either the sound source or the receiver are located in a corner of the room. The proximity of the three walls that form the corner causes a very short delay between the direct sound and the first few reflections. Whether or not the two situations and spatial aliasing are similar from a perceptual point of view is not clear.

Another inconvenience arising from the densely spaced aliasing artifacts is the circumstance that the time interval between the direct sound – or the combination of direct sound and floor reflection – and the next following reflection is always very short in the artificial reverberation. Recording engineers often refer to this time interval as *pre-delay*. It can give important information about the size of the room and the location of the sound source. A large pre-delay suggests that the sound source is located at a significant distance from the closest wall. The room has therefore to be large. Manipulation of the pre-delay is a powerful audio mixing technique [18].

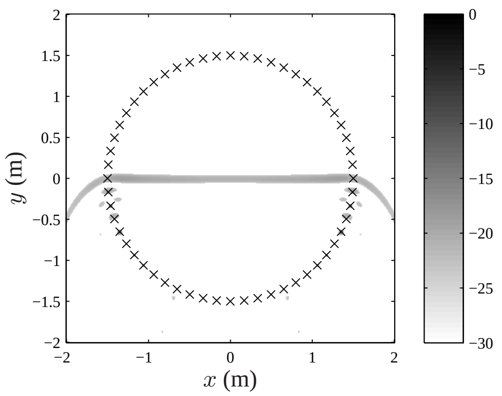
Despite certain differences, the working hypothesis in the present paper is that the human auditory system cannot distinguish between the aliasing artifacts and room reflections. This hypothesis bases mostly on the observations discussed above as well as on informal listening to setups like the one presented in [8], i.e. when the early room reflections are added as separate synthetic wave



(a) Spatially narrowband synthesis (e.g. Ambisonics).



(b) Spatially fullband synthesis (e.g. WFS, SDM).



(c) Sound field from Fig. 1(b) lowpass filtered with a minimum-phase FIR filter with a critical frequency of 1700 Hz.

Figure 1: Spatial impulse responses of a circular secondary source distribution in the horizontal plane when driven in order to synthesize a virtual plane wave propagating in positive y-direction. The absolute value of the time domain sound pressure is shown on in dB. (from [1, Fig. 4.19(c),(d)])

fronts (that exhibit their own aliasing artifacts). Informal listening shows that the reverberation in such a scenario tends to sound too dense. Note that each artificial reflection causes an entire set

[1] A minimum-phase filter is applied in Fig. 1(c) in order to obtain a compact support of the resulting wave front.

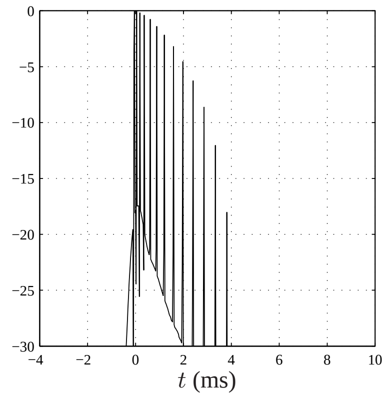


Figure 2: Time domain sound pressure of the field depicted in Fig. 1(b) at the coordinate origin on a logarithmic scale.

of wave fronts. A formal perceptual proof is not available at this point.

2.2. Adding the Low-frequency Content to the Aliased Wave Fronts

As noted above, a simplified interpretation of the aliasing artifacts is that an additional wave front arises from each loudspeaker. In the present context, we interpret these wave fronts as highpassed room reflections (recall Fig. 1(c)). The cutoff frequency (i.e. the spatial aliasing frequency) is typically between 1500 and 2000 Hz. It seems to be useful to artificially add the low frequency content to the highpassed reflections in order to make them natural. It should also be considered that room reflections experience diffraction at the boundaries of the reflecting surface at the very low end of the audible frequency range and a more modal behavior arises. It might therefore be preferable not to add the very low end to the (specular) reflection but treat it differently as discussed in Sec. 3. Note, however, that perception-based data are not available at this point.

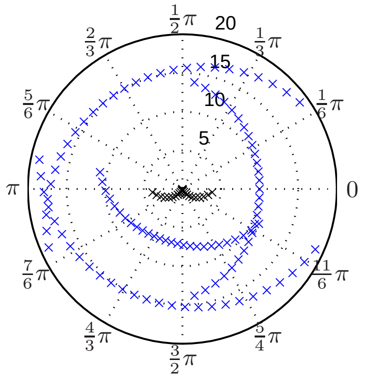
2.3. Adding Artificial Room Reflections

Adding artificial room reflections as separate synthetic wave fronts similar to the direct sound does not seem to be a favorable approach. Each added reflection will evoke a separate wave front pattern as illustrated by the blue marks in Fig. 3(a) so that only very few extra reflections lead to a very dense pattern of a high number of wave fronts. There are two obvious alternatives:

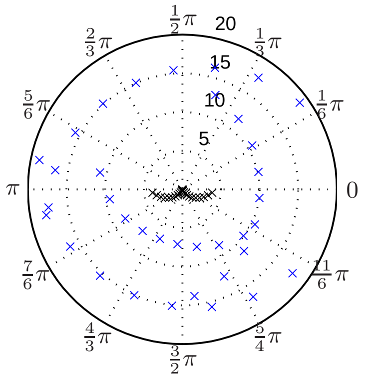
a) Use an individual loudspeaker for each reflection.

b) Synthesize the reflection using fewer loudspeakers with larger spacing than for the direct sound.

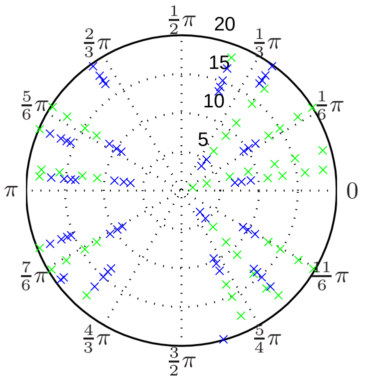
Ad a) This option is simple to implement and runs efficiently. However, the virtual room reflection that is produced by a single loudspeaker will exhibit an amplitude decay that is close to 6 dB for each doubling of the distance to the loudspeaker since the latter is typically very small and therefore acts like a monopole source. Larger sound sources – and therefore also reflections off large surfaces – exhibit a slower decay. This might not be an issue for small systems but with systems that have dimensions in the order of tens of meters the relative amplitudes of the artificial reflections change differently with the listening location than real reflections.



(a) Timing and incidence azimuths of the wave fronts of the sound field from Fig. 1(b) at the coordinate origin (black marks) and four virtual reflections (blue marks).



(b) Fig. 3(a) but with the virtual reflections synthesized by every third loudspeaker.



(c) Timing and incidence azimuths of the first couple of reflections calculated from a mirror image model [9] for a room of dimensions $5 \times 3.1 \times 2.33$ m. Green marks: Reflections that impinge from more than $30°$ off the horizontal plane.

Figure 3: Timing and incidence azimuths of the wave fronts/reflections. The radius represents the delay in ms of a given wave front/reflection relative to the first arriving wave front.

Ad b) Using fewer loudspeakers with larger spacing causes aliased wave fronts that are less dense than those of the direct sound. It is therefore possible to achieve a more balanced distribution of the wave fronts than when all loudspeakers are used as depicted in Fig. 3(b). Using more loudspeakers causes a slightly slower amplitude decay of the wave fronts than for a)[2].

Note that for both options a) and b) the incidence angles of the virtual reflections change somewhat differently with the listening location than those of real reflections. The perceptual significance of this circumstances is not clear.

2.4. Extension to Spatially Narrowband Methods

As mentioned in the Introduction, spatially narrowband methods for sound field synthesis such as the Ambisonics family of approaches exhibit spatial discretization artifacts that have somewhat different properties than those of the spatially fullband methods discussed so far. As evident from comparing Fig. 1(a) and (b), the additional wave fronts are fewer and are less homogeneously distributed over the listening area. Note that this is more of a quantitative than a qualitative difference. We assume that the considerations presented in the previous sections hold.

3. ROOM MODES

3.1. Physical Fundamentals

In order to illustrate the physical circumstances that lead to room modes, i.e., resonances of the room, we use the simplified model of a plane wave bouncing off an infinite rigid plane that extends normal to the propagation direction of the plane wave. We neglect phenomena like diffraction that occur at boundaries of finite extent. The reader interested in a detailed treatment is referred to [7].

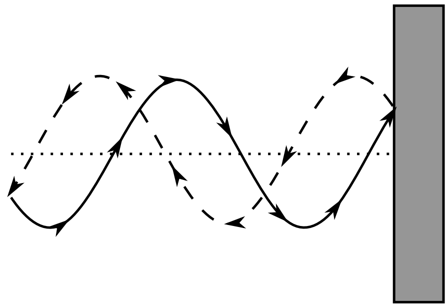
Rigid boundaries such as the walls of a room constitute Neumann boundary conditions for the sound waves inside the rooms as the particle velocity in the propagation direction of the wave vanishes at the boundary (the particles cannot move due to the boundary). As a consequence, the wave bounces back without a phase shift of the sound pressure. Refer to Fig. 4(a) for an illustration. The result is a field that consists of two plane waves of equal frequency and amplitude but with opposing propagation directions. Expressed in one-dimensional and in complex notation, this reads

$$p_{\text{incoming}} + p_{\text{reflected}} = \mathrm{e}^{-\mathrm{i}\frac{\omega}{c}x}\mathrm{e}^{\mathrm{i}\omega t} + \mathrm{e}^{\mathrm{i}\frac{\omega}{c}x}\mathrm{e}^{\mathrm{i}\omega t} = 2\cos\left(\frac{\omega}{c}x\right)\mathrm{e}^{\mathrm{i}\omega t} \tag{1}$$

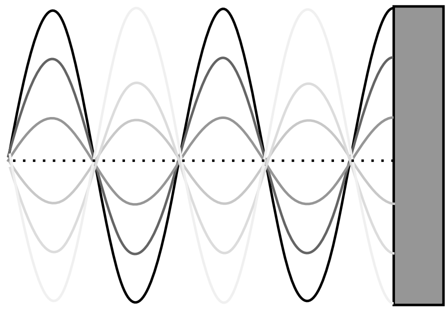
when assuming monochromatic waves of unit amplitude and choosing the time reference and coordinate system such that the two waves exhibit a relative phase shift of 0. The result is a standing wave of equal frequency like the component waves and with a pressure antinode at the boundary as illustrated in Fig. 4(b).

Note that this formation of standing waves occurs at all frequencies. We are usually not dealing with monochromatic waves but with waves that carry broadband and therefore time-varying content. In real rooms the waves bounce back and forth between the bounding surfaces. Standing waves occur only at particular frequencies in a room, namely at those frequencies for which the path length of the periodic path that a wave travels inside the room corresponds to an integer multiple of half the wavelength after the system has reached a steady state. The simplest case is a wave bouncing between two parallel walls of infinite extent. When the propagation direction of the wave is perpendicular to the walls and when the walls are perfectly reflective a persistent standing wave evolves at those frequencies specified above. Depending on the periodic path, different standing wave pattern evolve.

[2] Note that the slowest amplitude decay that e.g. an infinite linear loudspeaker array can produce is that of a cylindrical wave of 3 dB attenuation per doubling of the distance [1].



(a) One-dimensional wave impinging on a rigid boundary (solid line) and being reflected off (dashed line).



(b) Standing wave resulting from a superposition of the impinging and the reflected wave. Different shading of the wave refers to different time instants.

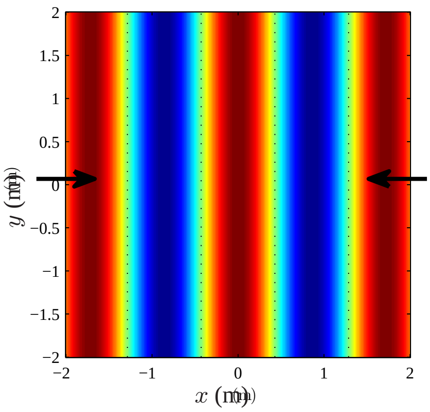
Figure 4: Illustration of the formation of a standing wave due to reflection off a rigid surface.

A significant amount of diffraction occurs in real rooms especially at low frequencies where the wave length is of similar order like the dimensions of the wall so that always a wave component that propagates perpendicular to a given wall arises. The amplitude and the Q-factor (and therefore the ringing duration) of the resonance depend on the acoustical properties of the boundaries, which are usually not perfectly rigid. Room modes occur all over the audible frequency bandwidth but only the low-frequency modes are perceptually relevant because of their sparsity [14, 15].

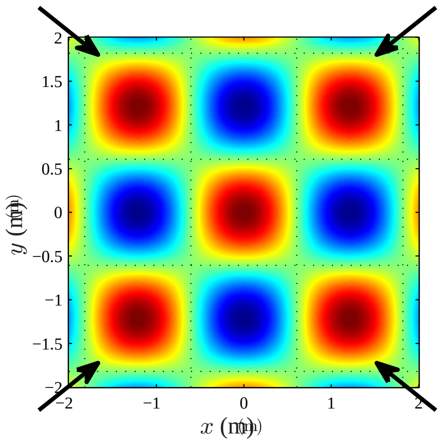
Fig. 5 illustrates the node/antinode patterns that evolve for the combination of different numbers of plane wave pairs so that different patterns can be realized. The standing waves exhibit their maximum amplitudes at the depicted time instant. Refer also to the animations at [19] that accompany this paper.

3.2. Room Modes in Sound Field Synthesis

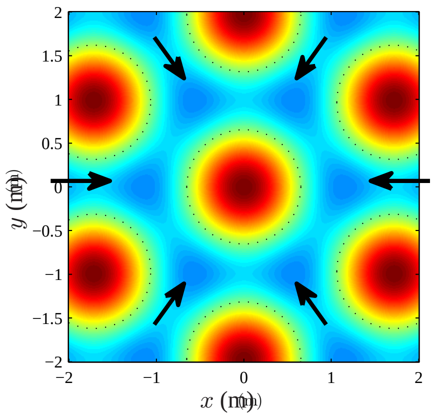
The parameters of modes in real rooms are complicated to determine because they depend heavily on the position of the source and many of the acoustical properties of the room. The interested reader is referred to [7]. It may be doubted that the human auditory system has a detailed expectation of plausible room modes so



(a) 1 pair of plane waves.



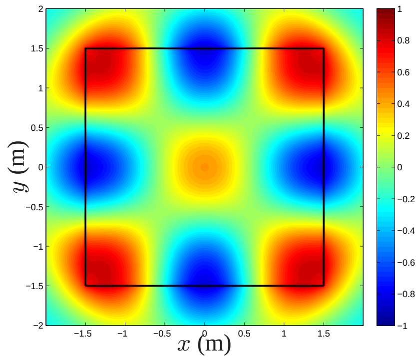
(b) 2 pairs of plane waves.



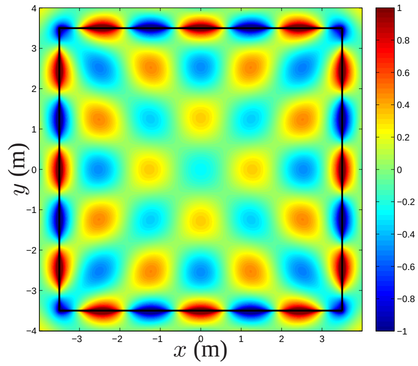
(c) 3 pairs of plane waves.

Figure 5: Cross-sections through the horizontal plane of the sound pressure evolving from different numbers of plane wave pairs of frequency $f = 200$ Hz. All sound fields are normalized to unity. The arrows illustrate the propagation directions of the plane wave components. The dotted lines mark the nodes.

that the more pragmatic approach of choosing the parameters like resonance frequency, amplitude, and bandwidth based on simple



(a) 3×3 m



(b) 7×7 m

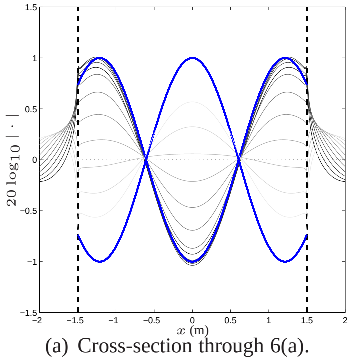
Figure 6: The standing wave from Fig. 5(b) synthesized by two quadratic sound field synthesis systems of different sizes. The black lines represent the secondary source contours. Note the different scalings of the axes.

statistical assumptions might be sufficient.

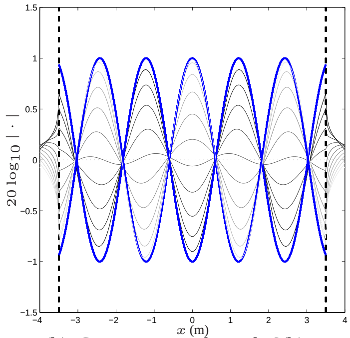
The creation of room modes based on pairs of plane waves as described in Sec. 3.1 is straightforward in three-dimensional sound field synthesis systems such as spherical arrangements of loudspeakers because plane waves with low-frequency content can be synthesized accurately. A set of narrow peak filters can be applied to the input signal of a given virtual source to efficiently create the narrowband input signals for the plane wave pairs.

The situation is more challenging in 2.5-dimensional – i.e. horizontal-only – synthesis. Here, synthetic plane waves exhibit an unavoidable amplitude decay of 3 dB for each doubling of the distance to the loudspeakers [1]. Short arrays exhibit an even faster amplitude decay because of the spatial truncation. Fig. 6 depicts the sound field from Fig. 5(b) synthesized by two loudspeaker systems of different sizes. Fig. 7 shows cross-sections through Fig. 6(a) and (b). Refer also to the animations at [19].

The simulations from Fig. 6 and 7 indicate that it is indeed possible to achieve standing waves in 2.5D synthesis. The deviation of the synthesized sound wave from the theoretic standing wave is small even for the mid-size array with 7 m edge length. It seems that the propagating components in the synthesized sound field are negligible. Future work has to investigate in what situations a considerable perceptual impairment arises.



(a) Cross-section through 6(a).



(b) Cross-section through 6(b).

Figure 7: Cross-section through Fig. 6(a) and (b) along the x-axis. Different gray shading represents different time instants. The blue lines represent the envelope of the prescribed (exact) standing wave calculated similarly to (1).

4. CONCLUSIONS

We presented guidelines for the creation of artificial early reflections and room modes in sound field synthesis. The most important aspect in the context of the creating of early reflections is the fact that practical sound field synthesis systems exhibit artifacts that are known as spatial aliasing that exhibit properties that are similar to room reflections. We presented two approaches for the design of reflection patterns that take the spatial aliasing artifacts into account.

We also suggested that the modal behavior of real rooms can be mimicked by synthesizing pairs of plane waves that propagate into opposing directions. Numerical simulations showed promising results even for 2.5D synthesis where the synthesized plane waves exhibit an undesired amplitude decay.

Acknowledgments

The author thanks Frank Schultz for valuable comments on the manuscript. The work presented in this paper is supported by grant AH 269/2-1 of Deutsche Forschungsgemeinschaft.

5. REFERENCES

[1] J. Ahrens, *Analytic Methods of Sound Field Synthesis*, Springer-Verlag, Berlin/Heidelberg, 2012.

[2] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *JASA*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[3] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new Ambisonic format," in *23rd International Conference of the AES*, Copenhagen, Denmark, May 2003.

[4] A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses," *JAES*, vol. 60, no. 11, pp. 887–898, Nov. 2012.

[5] J. Blauert and W. Lindemann, "Auditory spaciousness: Some further psychoacoustic analyses," *JASA*, vol. 80, no. 2, pp. 533–542, 1986.

[6] H. Kuttruff, *Room Acoustics*, Spon Press, London, fifth edition edition, 2009.

[7] F. Mechel, *Room Acoustical Fields*, Springer-Verlag, Berlin/Heidelberg, 2013.

[8] 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, 1994.

[9] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–948, Apr. 1979.

[10] J.-J. Sonke, "Variable acoustics by wave field synthesis," PhD thesis, Delft University of Technology, 2000.

[11] E. Hulsebos, "Auralization using wave field synthesis," PhD thesis, Delft University of Technology, 2004.

[12] F. Melchior, "Investigations on spatial sound design based on measured room impulses," PhD thesis, Delft University of Technology, 2011.

[13] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Trans. on Sp. and Audio Proc.*, vol. 18, no. 8, pp. 2038–2050, Nov. 2010.

[14] J. Fazenda, "Perception of room modes in critical listening spaces," PhD thesis, University of Salford, 2004.

[15] M. Karjalainen, P. Antsalol, A. Mäkivirta, and V. Välimäki, "Perception of temporal decay of low frequency room modes," in *116th Convention of the AES*, Berlin, Germany, May 2004.

[16] S. Spors and J. Ahrens, "Spatial aliasing artifacts of wave field synthesis for the reproduction of virtual point sources," in *126th Convention of the AES*, Munich, Germany, May 2009.

[17] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, "Physical and perceptual properties of focused sources in wave field synthesis," in *127th Convention of the AES*, New York, NY, Oct. 2009, p. paper 7914.

[18] R. Izhaki, *Mixing Audio - Concepts, Practices and Tools*, Focal Press, Oxford, 2007.

[19] J. Ahrens, "Animations," `http://www.soundfieldsynthesis.org/animations/eaa-2014`.