Properties of Large-scale Sound Field Synthesis

Jens Ahrens[1], and Hagen Wierstorf[2]

[1]*Quality and Usability Lab, University of Technology Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany*

[2]*Assessment of IP-based Applications, University of Technology Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany*

Correspondence should be addressed to Jens Ahrens (`jens.ahrens@tu-berlin.de`)

ABSTRACT

Sound field synthesis has been pursued as a promising approach for spatial audio reproduction for large listening areas. Research is typically performed on small and mid-size systems. An increasing number of systems of cinema size and larger exist, which have shown to exhibit properties that cannot be observed with smaller setups. In particular, practical limitations lead to artifacts whose perceptual saliency increases with array size. Depending on the situation, these artifacts are most prominent in time domain or in frequency domain. In this paper, we review the current state of knowledge on the properties of sound field synthesis using large-size loudspeaker arrays regarding both direct sound and reverberation.

1. INTRODUCTION

Sound field synthesis may be defined as the problem of driving a given ensemble of elementary sound sources such that the superposition of the sound fields emitted by the individual elementary sound sources produces a common sound field with given desired properties over an extended area [1]. Typically, a given listening volume or surface is surrounded by loudspeakers. Most common array geometries are rectangles, circles, and spheres. The approach of wave field synthesis (WFS) and a modern derivative of Ambisonics referred to as *near-field compensated higher order Ambisonics* or *Ambisonics with distance coding* are most commonly employed in order to compute the loudspeaker driving signals for rendering a given spatial audio scene. From a conceptual point of view, the difference between the two approaches is the circumstance that WFS solves the problem on the boundary of the listening area whereas Ambisonics solves the problem over the entire listening area. The equivalence of the two solutions can be shown using fundamental integral theorems of physics [1]. Significant differences between the two approaches arise in practical implementations as summarized below.

It can be shown theoretically that a given desired sound field can be synthesized perfectly under certain prerequisites. One requirement is the use of a continuous distribution of loudspeakers (*secondary sources*). This cannot be implemented with today's loudspeaker technology. Rather, the continuous distribution has to be approximated by a (larger) number of densely-spaced loudspeakers. A spacing of 10 cm–20 cm has been shown to be a good compromise between effort and result. A consequence of the fact that a discrete loudspeaker distribution instead of a continuous one is used is the circumstance that the synthesized sound field is identical to the desired one only below a given frequency, which lies typically between 1000 Hz and 2000 Hz. Above this so-called *spatial aliasing frequency*, the spatial structure of the synthesized sound field is distorted. The perceptual impairment due to these spatial distortions ranges from hardly noticeable to unacceptable depending on the particular scenario [2]. A significant part of the information that is relevant for spatial perception is contained in the frequency range that is synthesized correctly [3].

Unlike WFS, practical implementations of Ambisonics apply a so-called limitation of the spatial bandwidth. The consequence is that an area of very high accuracy evolves in the center of the loudspeaker setup with significant artifacts outside this area. In WFS, the physical artifacts are more evenly distributed over the entire listening area. In either case, the synthesized sound field extends over a significant area so that the potential of avoiding a stereophony-like *sweet spot* is still assumed by many authors. It has not been ultimately clarified if sound field synthesis can fulfill this expectation.

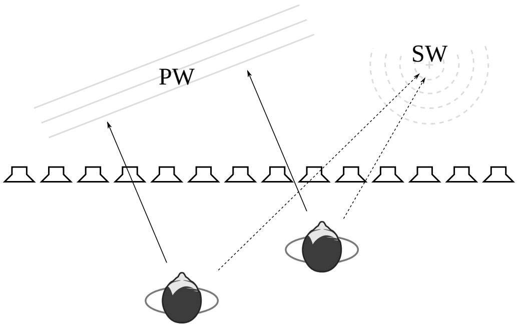


Fig. 1: Schematic of a scene composed of a plane wave object (PW) and a spherical wave/point source object (SW) synthesized by a linear array of loudspeakers (from [4]); the arrows indicate the auditory localization of the audio objects

Note that we focus on WFS in this work as large-size Ambisonics realizations with a spatial resolution that is comparable to that of WFS do not exist according to the authors' awareness. Numerical accuracy is one of the challenges in high-resolution Ambisonics implementations.

The audio scenes to be reproduced by sound field synthesis are typically described based on *audio objects* [5] rather than as loudspeaker signals as it has been the standard in cinema audio for a long time. An audio object is composed of an audio signal plus meta data. The audio signal could be, say, a speech signal that is radiated by a virtual sound source and the meta data would be the position of the sound source and its radiation characteristics. Refer to Fig. 1 for an illustration. This way, the representation of the audio scene is completely platform independent and operations like scaling of the scene or adapting the scene to circumvent limitations of a given reproduction system are straightforward. The process of computing the actual loudspeaker (or headphone) driving signals from such an abstract representation is termed *rendering*.

Systems used for research on sound field synthesis typically exhibit between 50 and 150 channels and dimensions in the order of several meters. Some of the commercial systems and a few experimental ones have more than 800 independent channels and dimensions of tens of meters [6]. Practical experimentation showed that such large systems can exhibit perceptual properties that are not evident with smaller setups.

In the present paper, we summarize the state of knowledge in this respect and illustrate the properties under discussion by means of numerical simulations. All simulations were performed using [7].

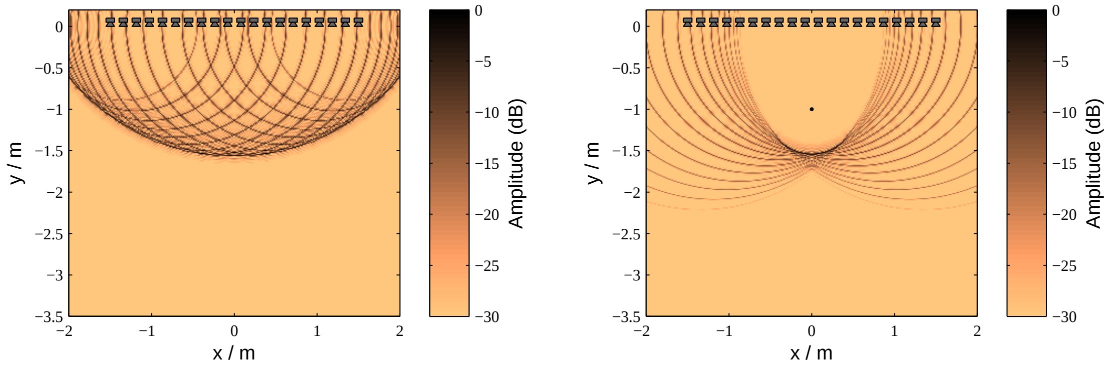
2. PROPERTIES IN TIME DOMAIN

The time-domain properties of synthesized sound fields differ substantially for the two different basic classes of virtual sound sources (non-focused and focused virtual sound sources). Both situations will be discussed in the following subsections.

2.1. Non-focused Sources

Non-focused sources are the most commonly employed source configuration, i.e. the virtual sound source is located outside of the listening area ("behind" the loudspeakers). We use a virtual point source in order to illustrate the properties of non-focused sources. Fig. 2(a) shows a time-domain snapshot of a virtual point source emitting an impulse. The source is located 1 m behind the loudspeaker array at $(0,1)$ m. The sound field exhibits a strong first wave front, which includes the desired wave front of the point source. After that, additional wave fronts appear containing frequency information above the spatial aliasing frequency. These components of the sound field are often referred to as *spatial aliasing* or *spatial aliasing artifacts*. Note that spatial aliasing also has some minor influence on the first occuring wave front. The additional wave fronts are equally distributed in the listening area. The strongest dependency on the listener position is in the direction of the *y*-axis, whereby less pronounced additional wave fronts occur at positions farther away from the loudspeaker array.

The number of additional wave fronts corresponds the number of employed loudspeakers and the length of the time interval between the first and the last wave front arriving at the listener depends on the size of the loudspeaker array. A linear proportionality is apparent. Fig. 3(a) shows the distribution of wave fronts in time and amplitude at two different listening positions for two loudspeaker array lengths. For the small loudspeaker array with a length of 3.1 m, the latest wave front arrives approximately 5 ms after the desired one for a listener positioned at $(1.5,-3)$ m. For a loudspeaker array length of 12.7 m, this time increases to over 25 ms. This circumstance will certainly lead to a difference in the perception as the additional wave fronts behave similar to reflections in a room [8]. One difference is that their time



(a) Point source at $(0,1)$ m

(b) Focused source at $(0,-1)$ m; the black mark indicates the position of the focus point

Fig. 2: Snapshots of synthesized sound fields in time domain; the secondary source spacing is 10 cm in all cases.

pattern is very regular and their distance in time is below 1 ms which is not the case for early reflections in a room. It is known from psychoacoustics experiments that two signals with a time distance below 1 ms are fused in their perception of location. This mechanism is known as *summing localization* [9]. Signal components arriving at a later time do not change the perceived location but contribute to the perception of the room. This latter phenomenon is known under the term *precedence effect* [10]. It includes also a signal dependent echo threshold, which is around 20 ms for broadband signals. If a reflection arrives after that threshold, it will be perceived as an additional signal (echo). This threshold is modulated by the difference in amplitude between the first wave front and the reflections. Due to the low amplitude of the additional reflections in WFS they will not become audible as distinct echoes even for very large loudspeaker arrays.

2.2. Focused Sources

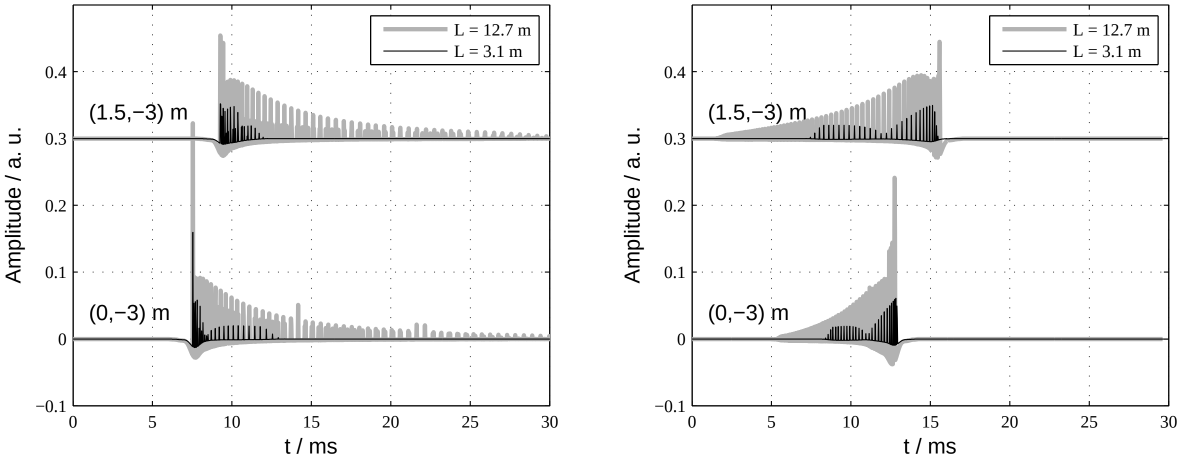
Focused sources are a feature of sound field synthesis that distinguishes it from other spatial audio presentation methods like amplitude panning. A focused source is a sound field that converges towards a focus point that is located inside the listening area. The sound field passes the focus point and diverges again and thereby mimics the sound field of a sound source at the location of the focus point. See Fig. 2(b) for an example.

The area between the focused source and the active loudspeakers can no longer be used as part of the listening area as localization cues are contradictory and do not correspond to the intended ones. However, the localization cues evoked by the diverging part of the sound field are similar to those of a point source placed at the focused point so that the perception of a sound source “in front of the loudspeakers” is achieved.

The spatial aliasing artifacts in focused sources exhibit a time domain behavior that is very different from that occurring with non-focused sources. The aliasing artifacts *precede* the desired wave front. This is depicted in Fig. 2(b) where a snapshot in time of a synthesized focused source radiating downwards is shown. Unlike with non-focused sources, the relative timing of the wave fronts depends on the listening position and the gap between adjacent wave fronts increases for listening at lateral positions.

The fact that the aliasing artifacts precede the desired wave front has a substantial impact on perception as such a wave front pattern is unnatural and neither a triggering of the precedence effect nor of summing localization can be observed. A formal evaluation presented in [11] showed that such wave front patterns can lead to strong coloration or unpleasant artifacts that accompany the focused source.

Most of the perceptual impairments of focused sources will become stronger with the size of the loudspeaker array. The advantage of having a larger viewing angle (effective listening area) and a better focusing at low frequencies [12] can hardly be exploited. If the time between the first undesired and the desired wave front exceeds 10 ms and the listener is located on the side of the listening area as it is the case in Fig. 3(b) the perceived



(a) Point source at $(-1,1)$ m

(b) Focused source at $(-1,-1)$ m

Fig. 3: Impulse response of two linear arrays of different length; the secondary source spacing is 10 cm in all cases.

auditory event begins to split into two [11]. One source is then perceived at the desired position and another high-passed version of it is located to the nearest edge of the loudspeaker array, where the first additional wave front arrives from. In addition, click-like artifacts can occur for long loudspeaker arrays.

The saliency of the perceptual impairments for focused sources and large loudspeaker arrays has led to the proposal of a reduction of the effective array length with larger loudspeaker arrays [11, 13]. In other words, it was proposed to use only a given section of a loudspeaker array for synthesizing a given focused source. Note that this reduces both the saliency of the artifacts as well as the size of the optimal listening area.

As the artifacts are most prominent for signals with strong transients, delicate combinations of array length, focused source position, and listener position should be used only with signals that do not exhibit such strong transients.

2.3. Other Time-Domain Properties

Another aspect related to the timing of the source signals is the employment of subwoofers in sound field synthesis. Loudspeakers designed for sound field synthesis have to be small since a small loudspeaker spacing is desired. As a consequence, such loudspeaker have a weak low-frequency response typically below 100 Hz or 200 Hz. While the information below these frequencies is not important for the presentation of spatial information, it is an important contributor to timbre and has definitely to be included. Though, the employment of subwoofers requires certain compromises since a reference point both for the amplitude as well as for the timing of the signals has to be defined as discussed below. Off this reference point, the balance of the amplitudes of the array loudspeakers and the subwoofer(s) as well as their timing relationship can be impaired.

Consider the case of a loudspeaker array that comprises one single subwoofer as depicted in Fig. 4. The amplitude of the subwoofer's signal has to be chosen such that an adequate timbre arises at the reference point whereby the distance attenuation of the virtual source's sound field – if apparent – has to be considered. The timing of the subwoofer's signal has to chosen such that the wave front synthesized by the loudspeakers of the array arrive at the reference point at the same time like the wave front emitted by the subwoofer. The timing will not be correct for listening positions off the reference point. For large arrays, timing discrepancies of several 10 ms can occur.

3. PROPERTIES IN FREQUENCY DOMAIN

3.1. Position-dependence of the Spatial Aliasing Frequency

As indicated in Fig. 2 and 3, spatial aliasing manifests

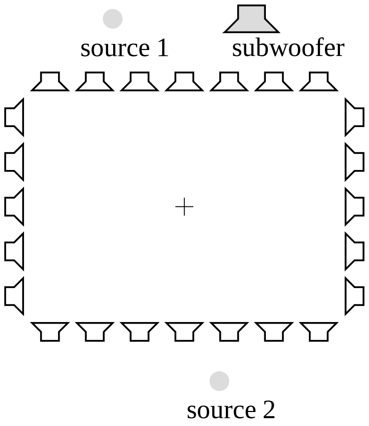


Fig. 4: Schematic illustration of a loudspeaker array with a subwoofer. The mark indicates the reference point for the timing.

itself as additional wave fronts that carry frequency content above the aliasing frequency. The fact that this high-frequency content is additional to the desired component of the synthesized sound field causes higher magnitudes of the transfer function of a system at those frequencies at which aliasing occurs. Note the steps that are apparent in Fig. 6 at a few kHz. Generally, this increase of high-frequency energy is compensated for by modifying the driving function above the aliasing frequency [14]. A closer look, however, reveals that the spatial aliasing frequency is position dependent to a certain extent.

Spatial aliasing causes additional wave fronts whereby these wave fronts do not necessarily exhibit considerable energy at all possible listening positions. This can be observed in Fig. 5, where the additional wave fronts due to spatial aliasing are apparent only close to the array at distances smaller than approximately half the array length. Note that there is a strong frequency-dependence of the number and the traveling directions of the aliasing components so that the situation looks somewhat different for other frequencies. It can be assumed for the setup depicted in Fig. 5 that aliasing is apparent anywhere in the listening area above approximately 6000 Hz.

Recall Fig. 6(a), which depicts the transfer function of an array that is identical to the one in Fig. 5. Note that Fig. 6(a) shows a point source instead of a virtual plane wave. The qualitative results are identical for both source types. It can be deduced from the step in the magnitude response that spatial aliasing kicks in at 2000 Hz for the close listening position and above 3000 Hz for the far position.

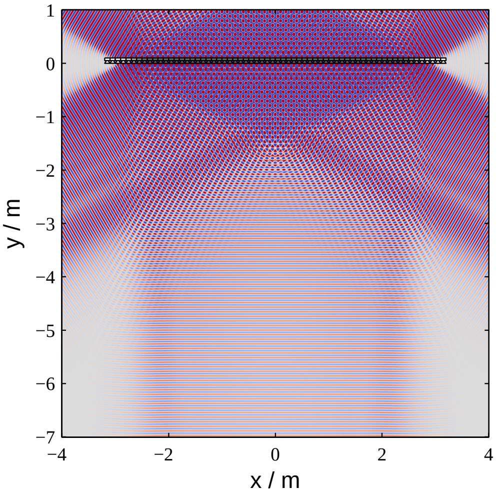
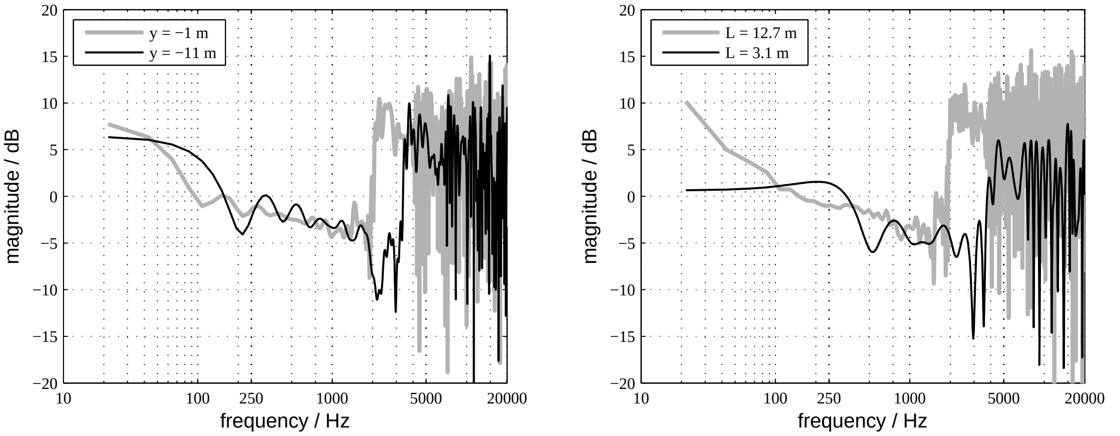


Fig. 5: Simulation of a monochromatic plane wave of $f = 4000$ Hz traveling in negative y-direction synthesized by a linear array of length $L = 6.3$ m with a loudspeaker spacing of 10 cm; notice the two aliasing components traveling southwest and southeast.

The situation is very similar for small and large arrays whereby it is such that the size of the region in which given aliasing artifacts are apparent or not scales linearly with array length. The differentiation between the two regions can be very relevant for rectangular arrays with significantly different edge lengths. It can occur that a given virtual source is synthesized only by the loudspeakers on one of the short edges and listeners can be located at significantly different distances to the active edge.

Note that the magnitude fluctuations apparent above 4000 Hz in Fig. 6 are not critical. Fig. 6 shows the magnitude spectrum of the sound pressure at one specific location. It is obvious from Fig. 2 that the sound field is composed of an entire set of wave fronts. The superposition of the wave fronts causes complicated interference patterns in the magnitude spectrum at a given location. However, the human auditory system is not a single pressure sensor. There are indications that the auditory system actually perceives that it is dealing with a set of wave fronts and interprets the field accordingly [2].

The general slopes apparent in Fig. 6 at frequencies be-



(a) Array of length $L = 6.3$ m; listener positions are along the y-axis

(b) Listener on the y-axis at $y = -3$ m; two different array lengths L

Fig. 6: Transfer functions of linear loudspeaker arrays of different lengths to different listener positions; the setup is similar to the one in Fig. 2; the transfer functions were normalized to have equal overall amplitude.

low 2000 Hz can be compensated for by an appropriate correction of the driving function [14]. This was waived here for transparency.

3.2. Nearfield vs. Farfield Radiation

Linear arrays of finite length exhibit what may be termed a nearfield and a farfield [14, 15]. At close distances from the array (no farther away than, say, the array length), the array radiates approximately like a line source. In other words, the amplitude decay of the radiated field is approximately proportional to $1/\sqrt{r}$ (i.e. 3 dB per each doubling of the distance) and the field exhibits a low-pass property with a slope of 3 dB/octave. This low-pass property is inherently compensated for by the loudspeaker driving function so that the radiated sound field effectively exhibits a flat frequency response. However, at sufficient distance the length of the array is negligible and it radiates approximately like a point source, i.e. with an amplitude decay that is proportional to $1/r$ (i.e. 6 dB per each doubling of the distance) and a flat frequency response.

The driving function is typically derived for what we termed nearfield above as the extent of the array is assumed to be infinite. This means that the spectral balance of the radiated field is good only in the nearfield but effectively exhibits a highpass property in the far-field with a slope of 3 dB/octave. The spatial extent of the two regions scales linearly with array size and there is a smooth transition between them. A noticeable effect is typically apparent for distances equal or larger that the array length compared to the ideal infinite setup [15].

Refer to Fig. 6(a) for a comparison of the transfer function of a given loudspeaker array at two different distances. Fig. 6(b) compares the transfer functions of two arrays of different lengths to a given listener position. Considerable deviations are apparent around 100 Hz in Fig. 6(a) and around 250 Hz and below 100 Hz in Fig. 6(b). Recall that the properties of the transfer functions in Fig. 6 at higher frequencies are discussed in Sec. 3.1.

It has been proposed to adapt the driving function to the effective length of the loudspeaker array under consideration in order to account for the varying extent of the nearfield [14]. This can, of course, only be optimized for a given listener distance (which may be tracked). Deviations will be apparent at other listener distances as the extent of the nearfield cannot be controlled.

There is an unproven potential that changing the listening distance by a given absolute amount results in weaker perceptual impairments for large arrays as the transition between nearfield and farfield occurs over a larger portion of space. It is also possible that this effect – if it exists – is masked by the position-dependence of the spatial aliasing frequency discussed in Sec. 3.1.

3.3. Other Frequency Domain Properties

A peculiarity arises with large arrays in that very pointed angles between the nominal orientation direction of a loudspeaker and the direction in which a listener is located relative to this orientation can occur. Most sound field synthesis approaches that are used in practice assume the loudspeakers to be omnidirectional, an assumption that holds for real-world loudspeakers only at lower frequencies [16]. It is unclear at this stage in how far this is audible as significant spatial aliasing occurs in the frequency range where loudspeakers exhibit a pronounced directivity. Approaches to sound field synthesis that can handle non-omnidirectional loudspeakers exist. An example is [17].

4. REVERBERATION

The presentation of artificial or recorded reverberation has been somewhat of a step-child in sound field synthesis ever since despite the practical importance of reverberation. Recently, a comprehensive concept was proposed that is currently under investigation [8].

4.1. Pre-delay

When a sound source and the receiver in a room are located at sufficient distance from the room boundaries then a considerable delay between the arrival of the source's direct sound/floor reflection and the first wall reflection arises. This is typically referred to as *pre-delay* in music production [18]. Modifying the pre-delay for a given sound source in a complex scene can have significant influence on the extent to which a source blends with the rest of the scene.

Recalling Fig. 3(a), we find that for large systems (gray line), the wave fronts that occur due to spatial aliasing fill the entire duration even for long pre-delays of, say, 15 ms. As a consequence, the spatial aliasing artifacts blend into the early reflections potentially without any perceptually relevant pre-delay occurring.

The impulse response of short loudspeaker arrays can vanish within a few ms as evident from the black lines in Fig. 3(a) so that long pre-delays can indeed be realized.

The situation is totally unclear for focused sources (Fig. 3(b)) from a perceptual point of view.

4.2. Early Reflections

An important aspect in the creation of reverberation using loudspeaker arrays is the (informal) observation that the spatial aliasing artifacts apparent in Fig. 3(a) indeed cause some sense of room impression. It is therefore to be expected that the human auditory system cannot reliably discriminate spatial aliasing and artificial room reflections. It was proposed in [8] to consider this circumstance in the design of the artificial reverberation and create the early reflections such that they evoke a plausible reflection/wave front pattern together with the spatial aliasing artifacts.

It has not been proven so far that this approach actually leads to a more convincing perception. In any case, the timing aspects discussed Sec. 2.3 hold also for the sythesized early reflections.

4.3. Diffuse Reverberation

It was shown in [19] that diffuse reverberation can be created by a set of plane wave carrying decorrelated signals. The user study considered one single listener position only. Similarly to the discussion Sec. 2.3, the relative timing of the individual plane waves changes significantly for two listening position that are at significant distance from each other as it can occur with large systems. It is unclear at this stage what the perceptual implications of these relative timing changes are and whether or not some sort of *sweet area* arises with respect to the reverberation. First results will be published in [20].

5. CONCLUSIONS

We presented an overview of the properties of large sound field synthesis systems. An important result is the fact that the length of the impulse response of a system to a given listener location depends on the size of the employed array. For short arrays, undesired wave fronts due to spatial aliasing arrive within a few milliseconds after the desired wave front for non-focused virtual sound sources or precede the desired wave front accordingly for focused sources. This interval during which the aliasing artifacts impinge on a listener position scales proportionally with array size. For large arrays this can lead to severe perceptual artifacts for the synthesis of focused sources. For non-focused sources the perceptual implications of this circumstance are largely unclear.

We also showed that the radiation properties of loudspeaker arrays evoke a nearfield and a farfield with different spectral properties.

Regarding the presentation of reverberation, we identified the relative timing of the components of the reverberation that varies with listener position as a potential challenge that requires further analysis. Also, there are limitations for large arrays regarding the realization of an effective pre-delay between direct sound and the first (virtual) wall reflection as this interval is filled with spatial aliasing artifacts.

ACKNOWLEDGMENTS

This work was partly funded by grant AH 269/2-1 of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).

6. REFERENCES

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

[2] H. Wierstorf. *Perceptual Assessment of Sound Field Synthesis*. PhD thesis, University of Technology Berlin, 2014. to appear.

[3] E. W. Start. *Direct sound enhancement by wave field synthesis*. PhD thesis, Delft University of Technology, 1997.

[4] G. Theile, H. Wittek, and M. Reisinger. Potential wave field synthesis applications in the multichannel stereophonic world. In *AES 24th International Conference*, Banff, Alberta, Canada, 2003.

[5] B. Shirley, R. Oldfield, F. Melchior, and J.-M. Batke. Platform independent audio. In O. Schreer, J.-F. Macq, O. A. Niamut, J. Ruiz-Hidalgo, B. Shirley, G. Thallinger, and G. Thomas, editors, *Media Production, Delivery and Interaction for Platform Independent Systems*. Wiley, Hoboken, 2014.

[6] D. de Vries. *Wave Field Synthesis*. AES Monograph, 2009.

[7] H. Wierstorf. The sound field synthesis toolbox. `https://github.com/sfstoolbox/sfs`, 2012. last accessed: Oct. 16 2014, commit 46962f7.

[8] J. Ahrens. Challenges in the creation of artificial reverberation for sound field synthesis: Early reflections and room modes. In *EAA Joint Symp. on Auralization and Ambisonics*, Berlin, Germany, 2014.

[9] J. Blauert. *Spatial Hearing*. The MIT Press, 1997.

[10] Ruth Y Litovsky, H. Steven Colburn, William A Yost, and S J Guzman. The precedence effect. *The Journal of the Acoustical Society of America*, 106(4):1633–54, 1999.

[11] H. Wierstorf, A. Raake, M. Geier, and S. Spors. Perception of focused sources in wave field synthesis. *Journal of the Audio Engineering Society*, 61(1/2):5–16, 2013.

[12] R. Oldfield. *The analysis and improvement of focused source reproduction with wave field synthesis*. PhD thesis, University of Salford, 2013.

[13] M.-H. Song, J.-W. Choi, and Y.-H. Kim. A selective array activation method for the generation of a focused source considering listening position. *The Journal of the Acoustical Society of America*, 131(2):EL156–62, 2012.

[14] S. Spors and J. Ahrens. Analysis and improvement of pre-equalization in 2.5-dimensional wave field synthesis. In *128th Convention of the AES*, London, UK, May 2010.

[15] F. Schultz and S. Spors. On the frequency response variation of sound field synthesis using linear arrays. In *DAGA*, Oldenburg, Germany, 2014.

[16] F. Fazi, V. Brunel, P. Nelson, L. Hrchens, and J. Seo. Measurement and Fourier-Bessel analysis of loudspeaker radiation patterns using a spherical array of microphones. In *124th Convention of the AES*, Amsterdam, The Netherlands, May 2008.

[17] J. Ahrens and S. Spors. An analytical approach to 2.5D sound field reproduction employing linear distributions of non-omnidirectional loudspeakers. In *IEEE ICASSP*, Dallas, Texas, USA, March 2010.

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

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

[20] J. 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.