Analysis of a Spatially Discrete Sound Field Synthesis Array in a Reflective Environment

Vera Erbes, Sascha Spors
Institute of Communications Engineering, University of Rostock, Rostock, Germany.

Stefan Weinzierl
Audio Communication Group, Technical University of Berlin, Berlin, Germany.

Summary
While the foundations of sound field synthesis demand an anechoic environment, this requirement cannot be met in real-world installations. Reflections at the walls of the listening room modify the synthesised sound field. Moreover, the synthesised sound field in a reflective environment, of e.g. a virtual point source, differs from the one of a real point source in the same environment. Filtered copies of the direct sound from each loudspeaker are mingled with spatial aliasing artefacts due to room reflections and a finite distance between loudspeakers, respectively. In a setup with typical loudspeaker spacings, the additional wave fronts that constitute spatial aliasing artefacts occur in a shorter timeframe than the early reflections of the listening room. This paper investigates how early reflections are filled in by these additional wave fronts produced by spatially discrete secondary source distributions. Different scenarios with real and virtual point sources in the free field and in a reflective environment are simulated. The reflective environment is simulated by an image source model. The spacing of the secondary sources is varied to generate different sequences of additional wave fronts. The resulting (room) impulse responses are analysed with respect to the structure and density of reflections and/or aliasing. Room acoustic measures like the reverberation time and early decay time of the system are considered. Finally, findings in the literature concerned with the perception of early reflections and diffuse reverberation allow for concluding on the perceptive impact of the listening room.

PACS no. 43.55.Br, 43.60.Fg

1. Introduction

Theoretical requirements of sound field synthesis (SFS) cannot always be fulfilled in practice. SFS theory assumes free field conditions, so that no energy is reflected back into the listening area. Real-world loudspeaker arrays typically have to be installed in rooms with more or less reflective boundaries. These reflections alter the synthesised sound field. Though the precedence effect is likely to ensure correct localisation in most cases as it does for real sources, other perceptual aspects of the sound field might be affected. Only the installation of a loudspeaker array in an anechoic chamber could remedy this completely.

Another assumption in the theory of sound field synthesis is that secondary sources are continuously distributed, which is not possible in practice. Instead, loudspeaker arrays represent a series of discrete sources sampling the secondary source contour. This leads to an erroneous synthesis for frequencies higher than the spatial aliasing frequency determined by the spacing between the loudspeakers. For impulse excitation of the virtual source this leads to a succession of pulses, one for each loudspeaker, with the interval between these pulses determined by the spacing between the loudspeakers. These undesired wave fronts are similar to room reflections [1], but occur in a much shorter timeframe.

While the perceptual consequences of discrete secondary source distributions have been explored, in particular with respect to localisation and colouration [2], the influence of the listening room has not been investigated thoroughly so far. Based on simulations, this paper examines how spatial aliasing fills in the time intervals between the early reflections of a room and leads to a more diffuse sound field in the later part of the room impulse response. Reverberation time and early decay time are calculated for different configurations and perceptual consequences are discussed.

2. Simulation method

Rooms have been simulated with an image source model and uniform reflection coefficients $\beta = -0.9$ for all boundaries [3]. A negative reflection coefficient is a simplified way to model that β can become negative for certain incidence angles, leading to impulse responses with positive and negative components [4]. To reduce adding up of reflections, oversampling has been used (with a sampling rate of 441,000 kHz). The order of the image source model according to [3] has been set to $N = 35$, leading to an exponential decay of the later part of the impulse responses to at least -65 dB. All rooms were of rectangular shape.
The SFS arrays in this paper are of linear and rectangular shape. As a closed-form expression, the driving signal of a virtual point source rendered with rectangular arrays is only available by Wave Field Synthesis (WFS). The driving signal used in this paper is [5]

$$d_{2,5\mathrm{D}}(\mathbf{x_0}, t) = \frac{1}{2\pi} \cdot \sqrt{2\pi|\mathbf{x}_{\mathrm{ref}} - \mathbf{x_0}|} \cdot w(\mathbf{x_0}) \times \cdot \frac{(\mathbf{x_0} - \mathbf{x}_{\mathrm{ps}})\mathbf{n}_{\mathbf{x_0}}}{|\mathbf{x_0} - \mathbf{x}_{\mathrm{ps}}|^{\frac{3}{2}}} \cdot a(t) \, * \, h_{2,5\mathrm{D}}(t) * \delta\left(t - \frac{|\mathbf{x_0} - \mathbf{x}_{\mathrm{ps}}|}{c}\right) \tag{1}$$

with $\mathbf{x_0}$, $\mathbf{x}_{\mathrm{ps}}$ and $\mathbf{x}_{\mathrm{ref}}$ denoting the positions of the secondary sources, the virtual point source and the reference point, respectively. $\mathbf{n}_{\mathbf{x_0}}$ is the normal vector of a secondary source at $\mathbf{x_0}$ directed inwards. $a(t)$ is the signal of the point source. $w(\mathbf{x_0})$ is the window function selecting the constructively contributing secondary sources and $h_{2,5\mathrm{D}}(t)$ is the inverse Fourier transform of the WFS pre-filter:

$$h_{2,5\mathrm{D}}(t) = \mathcal{F}^{-1}\left\{\sqrt{\mathrm{j}\frac{\omega}{c}}\right\}. \tag{2}$$

2.1. Simulated situations

Either real or virtual point sources are placed in free field or in a room. Five different cases have been investigated. The room a real point source is located in is termed virtual room (used in cases 1, 4 and 5). The room that contains the secondary sources is termed listening room (used in cases 2–5).

- Case 1: a point source in a virtual room
- Case 2: a virtual point source rendered by a linear array of secondary sources in free field
- Case 3: as case 2 but in a listening room
- Case 4: a rectangular array of secondary sources emulates case 1 (i.e. a virtual room) with virtual point sources for each horizontal reflection in free field
- Case 5: as case 4 but in a listening room.

Case 4 and 5 constitute a situation that is not feasible in practice as too many sources have to rendered by the array simultaneously. Nevertheless, they allow for investigation of a room-in-room situation especially in the time span of early reflections. While early reflections can be rendered by virtual point sources [6], diffuse reverberation is usually produced in practice by a limited number of plane waves from different directions [7, 8].

2.2. Variations in the simulation

Simulations have been varied with respect to the spacing of the secondary sources, the size of the virtual and of the listening room and the receiver position, each in three steps. The considered geometries are depicted in fig. 1. An overview of the simulated conditions can be found in table I.
Starting point is the arrangement of a typical sound field synthesis array at the Institute of Communcations Engineering, University of Rostock. 64 loudspeakers are arranged in a square of approx. 4 m length at 1.59 m height in a rectangular listening room of $5.75 \cdot 5 \cdot 3\,\mathrm{m}^3$ ($V = 86.25\,\mathrm{m}^3$). The position of the array inside the room (light grey area in fig. 1) is indicated in fig. 1 by the dashed line. Due to the construction of the array that allows for flexible placement of the loudspeakers, these are not equidistantly spaced (mean spacing is 23.4 cm). The real or virtual point source is placed outside the array but inside the room. For rendering of this virtual point source, only a line of loudspeakers is necessary (cases 2 and 3). This setup is termed the reference situation.
Loudspeaker spacing has been approx. doubled and halved to 12.7 and 56.4 cm, respectively. Placement inside the rooms is the same as for the original array. One larger and one smaller room are considered. The larger room has walls 0.5 m that are wider apart leading to an approx. doubled volume of $V = 162\,\mathrm{m}^3$. For the smaller room, floor and ceiling are closer by 0.5 m, walls by 0.3 m except the wall close to the location of the point source, which is only displaced by 0.2 m. The volume of the smaller room is $V = 46.2\,\mathrm{m}^3$ and thus approx. halved with respect to the middle-sized room.
Receiver positions have been chosen inside the loudspeaker array. One position is in the centre of the array which is also the WFS reference point. The other two positions have been chosen to be closer to the (virtual) point source by 1 m and more to the side by 0.8 cm, cf. fig. 1.

3. Room acoustical parameters

ISO 3382-1 [9] provides several room acoustical parameters serving as predictors for perceptual aspects in room acoustics. These have mostly been established in concert hall acoustics, and it is unclear to what extent they can be applied to the small rooms considered in the current paper [10, 11] and the special nature of synthesied sound fields.
In particular, the temporal separation in early and

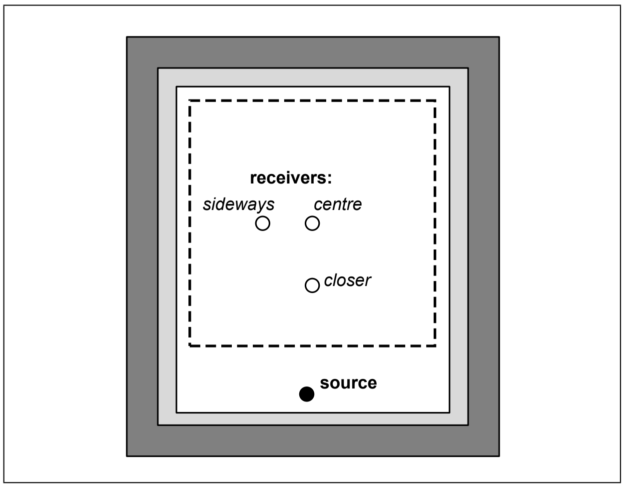

Figure 1. Simulated geometries as viewed from above. White and grey areas: different room sizes, dashed line: array contour, black dot: position of (virtual) point source, white circles: receiver positions.

late reflections at 50 and 80 ms for measures like clarity in music and definition in speech signals, respectively, are doubted to be adequate for small rooms [10]. Due to the differences in size, modal behaviour in the impulse response is shifted to a different frequency range.
Moreover, the very early and densely spaced pulses emerging from spatial aliasing differ considerably from the usual distribution of early reflections even in small rooms. A measure like the time gap between direct sound and the first reflection for prediction of perceptive attributes such as 'intimacy' [12] might lose its meaning in such situations.
The employed simulation model is not physically correct for low frequencies as it does not take wave-theoretic effects into acccount. In small rooms, the frequency range considered as low frequencies includes at least the two lower octave bands 125 and 250 Hz that are important for several room acoustical parameters [9]. Furthermore, the rooms in this paper are simulated with a reflection coefficient independent of frequency. For these reasons, spectral features of the sound field have not been investigated, although spectral differences might play a role in the employed simulations due to the interference of image sources and the WFS pre-filter.
Despite these restrictions, decay time porperties can be described by the Early Decay Time (EDT) and the reverberation time RT30. The just noticable difference (JND) for exponential decays is 4–5 % depending on signal type [13]. Concerning their perceptual meaning, the EDT has been shown to predict reverberance by several studies, e.g. [14, 15]. The reverberation time has often been found to be a predictor for perceived room size, e.g. [10], even if the latter might be influenced by other aspects, cf. e.g. [16, 17]. The applicability for the presented situations here will be discussed.

4. Results

The resulting normalised impulse responses for the reference situation are shown in fig. 2. Comparing cases 1, 2 and 3 in fig. 2 (a), one can see that the spatial aliasing fills the space between the early reflections evoked by a real point source. The relative amount of energy in the early and late part of the normalised impulse responses in cases 1 and 3 stays the same, though. This leads to almost exactly the same values for EDT and RT30 (first line of table I). The energetic equivalence becomes obvious when observing the EDCs in fig. 3, line a. As the EDT is a predictor for reverberance, it can be concluded that these two situations can be expected to sound equally reverberant. Case 3 does not exhibit the same sequence of pulses as case 1, as has already been observed in [18]. This might alter other perceptual impressions of the room.
In contrast, case 4 (fig. 2 (b) and fig. 3, line b), which renders only the horizontal image sources of case 1 as virtual point sources, exhibits an increased decay time with a higher EDT and RT30. Each virtual source is rendered by up to 32 secondary sources, leading to a high reflection density in the later part of the impulse response. Each synthesised reflection is spread over several pulses caused by spatial aliasing. The higher EDT indicates more reverberance than for the real point source in the room although only horizontal image sources are rendered. One might have expected this case to exhibit less reverberant energy than case 1 and 3. The reason for this behaviour lies in the driving function eq. (1). It renders virtual point sources with an amplitude increasing with distance to the reference point compared to the amplitude of a point source. Hence, the observations of the current paper stress the need for correct amplitude reproduction in WFS as has been proposed in [19].
Case 5 (fig. 2 (c) and fig. 3, line c) constitutes a room-in-room situation where the decay rates of both virtual and listening room matter. Consequently, EDT and RT30 are even more increased due to the large number of reflections und spatial aliasing with the EDT being even larger than RT30. This results in a non-exponential decay that is typical for room-in-room reproduction [20]. The evaluation of reverberation time by a regression line is not adequate for a non-exponential decay of this type, though, and might lead to overestimation of the reverberation time. That this case is the most reverberant has been confirmed by informal listening tests.
When applying RT30 as an indicator for perceived room size, it can be expected that case 1 and 3 with the same value for RT30 are perceived to be equal in size, while the more complex cases 4 and 5 might be

perceived to be larger. Particularly in case 4, RT30 suggests the perception of a room that is larger than the virtual room. In case 5, the room-in-room situation can be expected to create a sense of an even larger room than when virtual or listening room are present alone.

4.1. Variation of secondary source spacing

Considering decay rates, the variation of secondary source spacing results in only minor changes smaller than the JND for reverberation time since the energy is only distributed differently (cf. table I). Reverberance and perceived room size based on RT30 thus do not change compared to the reference situation. Nevertheless, the sequence of pulses is altered as can be seen in fig. 4 for case 3 in comparison with fig. 2 (a), leading to more densely spaced pulses for more densely spaced secondary sources and vice versa.

4.2. Variation of room size

If the virtual and the listening room size are made larger or smaller but are both of the same size, the energy measures EDT and RT30 for cases 1 and 3 are still the same (cf. table I). They only indicate that larger rooms lead to longer decay times as the reflections are spread over time. Cases 4 and 5 likewise show increased decay rates for larger rooms.
If virtual and listening room size differ, the decay rates of case 1 and 3 are not the same anymore, but follow the size of the virtual (case 1) and listening room (case 3), respectively. For the room-in-room situation (case 5), the listening room always increases the number of reflections compared to the same WFS simulation of the virtual room in free field (case 4).

4.3. Variation of receiver position

As the simple simulated room with a shoebox shape does not exhibit special geometrical properties, has equal reflection coefficients at all boundaries and all receiver positions are located in the diffuse field, it is not surprising that neither RT30 nor EDT vary much with receiver positions (cf. table I). All differences are smaller than the JND for reverberation time. Differences could have been expected for EDT, but less for RT30 which usually does not vary much with position [21].

5. Discussion

Although the chosen simulation models enable easy computation of room impulse responses for real and virtual sources in rooms and room-in-room situations, parts of the result have to be considered carefully.
Though temporal resolution has been chosen quite high in the simulations, the reflection density becomes so high in the later part of the impulse responses that several reflections cancel out due to the negative reflection coefficient. This is problematic when considering energetic measures. It leads to underestimation of the energy in the impulse responses especially in the late part and thus to shorter decay times. The effect becomes more pronounced for higher reflections densities, i.e. it matters most for case 5. Therefore, decay times should increase even more with the more complex cases 4 and 5 than is stated in table I. Comparisons still deliver valid conclusions, though not in terms of absolute values.

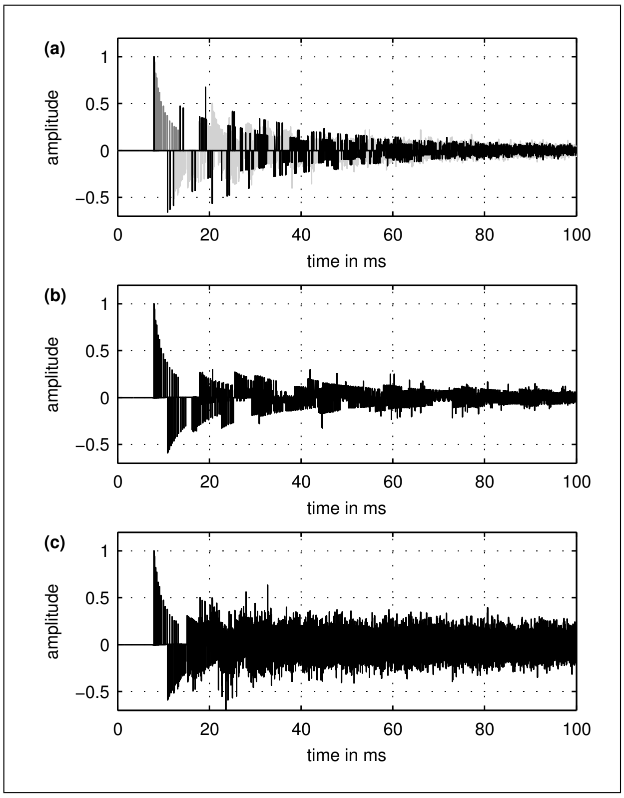

Figure 2. Normalised impulse responses for the reference situation. a: black: case 1, dark grey: case 2, light grey: case 3 (first pulses are concealed by case 2), b: case 4, c: case 5.

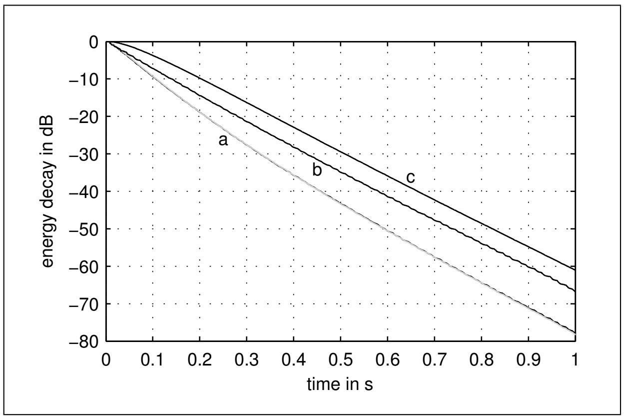

Figure 3. EDCs for the reference situation. a: black: case 1, light grey: case 3, b: case 4, c: case 5.

Table I. EDT and RT30 in s for the simulated conditions. Case 2 has been omitted as calculation of its decay curve is not meaningful. $\Delta\mathbf{x_0}$ is the secondary source spacing, V_{rec} and V_{list} denote the volume of virtual and listening room, respectively. Room geometry and receiver positions are depicted in fig. 1. Repetitive cells are left blank for readibility.

Condition:				**Case 1**		**Case 3**		**Case 4**		**Case 5**	
$\Delta\mathbf{x_0}$	V_{rec}	V_{list}	receiver	EDT	RT30	EDT	RT30	EDT	RT30	EDT	RT30
23.4 cm	86.25 m^3	86.25 m^3	centre	0.59	0.67	0.60	0.67	0.83	0.87	1.12	0.92
23.4 cm	86.25 m^3	86.25 m^3	closer	0.60	0.67	0.59	0.66	0.85	0.87	1.13	0.92
23.4 cm	86.25 m^3	86.25 m^3	sideways	0.59	0.67	0.59	0.67	0.82	0.86	1.13	0.91
23.4 cm	162 m^3	86.25 m^3	centre	0.74	0.80			0.96	1.03	1.24	1.05
23.4 cm	162 m^3	86.25 m^3	closer	0.75	0.80			0.99	1.03	1.25	1.05
23.4 cm	162 m^3	86.25 m^3	sideways	0.74	0.80			0.98	1.02	1.24	1.05
23.4 cm	46.2 m^3	86.25 m^3	centre	0.48	0.60			0.74	0.78	1.05	0.85
23.4 cm	46.2 m^3	86.25 m^3	closer	0.49	0.59			0.76	0.79	1.07	0.85
23.4 cm	46.2 m^3	86.25 m^3	sideways	0.48	0.60			0.73	0.77	1.05	0.85
23.4 cm	162 m^3	162 m^3	centre			0.74	0.80			1.35	1.09
23.4 cm	46.2 m^3	46.2 m^3	centre			0.49	0.59			0.97	0.82
56.4 cm	86.25 m^3	86.25 m^3	centre			0.59	0.67	0.82	0.87	1.13	0.92
12.7 cm	86.25 m^3	86.25 m^3	centre			0.60	0.67	0.83	0.87	1.13	0.92

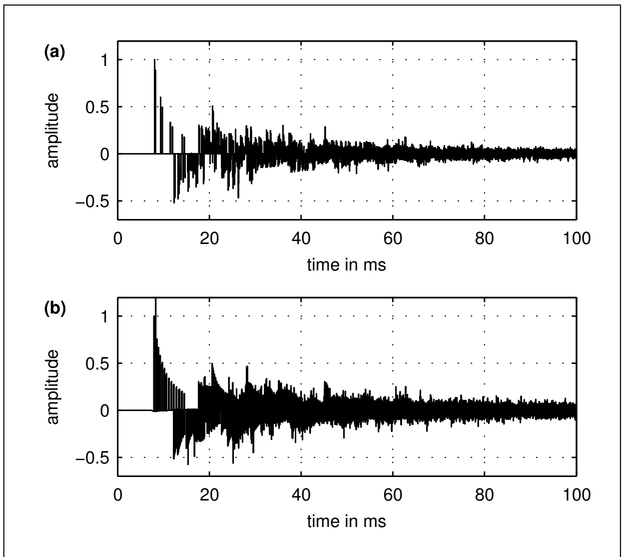

Figure 4. Normalised impulse responses for variation in loudspeaker spacing. a: $\Delta\mathbf{x_0} = 56.4$ cm, b: $\Delta\mathbf{x_0} = 12.7$ cm. Other settings are just as in the reference situation, cf. last two lines in table I.

Another drawback of the simulation model with image sources is the missing constitution of room modes and wave-theoretic effects which are an important part of the acoustics of small rooms, in particular for low frequencies. The current model covers only the specular part of wall reflections both for the virtual and the listening room.

The rendering of all reflections in the horizontal plane as virtual point sources of the WFS array is not a practically relevant approach as the number of sources is too high. But rendering early reflections with virtual point sources is a feasible approach [6]. Therefore, the conclusions concerning the early reflections are of more importance.

As has been discussed in section 3, there are limits to the application of traditional room acoustical parameters for the presented situations. Energetic measures fail to take into account the fine structure of reflections and spatial aliasing artefacts. When considering a room-in-room situation as in case 5, with a listening room that is smaller than the virtual room, the first reflections of the listening room arrive too early. Due to spatial aliasing and the mixing of the reflections of virtual and listening room in the later part of the impulse response, the sound field could be expected to become diffuse earlier as well, in a physical and perceptual sense. Another special property of WFS room simulation is apparent in case 4, where energy arrives at the listener only within the horizontal plane. The perception of this sound field in terms of reverberance and room size might differ from what calculated values for EDT and RT30 suggest.

6. CONCLUSIONS

Different cases of real and virtual point sources synthesised by WFS in free field and in rooms have been simulated with an image source model.

A real and a virtual point source synthesised with WFS in the same room are equivalent with respect to the energetic measures EDT and RT30. The emulation of the sound field of a point source in a virtual room leads to slower decay rates und thus a higher reverberance can be expected. This is remarkable, as the employed WFS array is only capable of 2.5D synthesis and generates only the energy fraction in the horizontal plane. The increased reverberant energy is caused by the behaviour of the WFS driving function that renders more distant point sources with a higher amplitude than the corresponding real source. Neither different receiver positions nor the variation

of secondary source spacing leads to differences in the investigated energetic measures, but rather generates different distributions of spatial aliasing pulses.
Future investigations should address other important percepts of room acoustics such as spatial and spectral features. To this end, the simulation method has to be refined.
Most importantly, the structure of impulse responses with spatial aliasing filling in the early reflections of a room differs considerably from typical room impulse responses. They bear no resemblance to a natural listening situation in the early part of the impulse response. This suggests that further research should lay emphasis on the distribution of these pulses and include listening tests since traditional room acoustical measures might not be adequate for predicting the perception of synthesised sound fields in rooms.

Acknowledgement

This work is part of the *Simulation and Evaluation of Acoustical Environments* (SEACEN) project funded by the German Research Foundation (DFG WE 4057/3-2).

References

[1] J. Ahrens: Challenges in the creation of artificial reverberation for sound field synthesis: early reflections and room modes. Proc. EAA Joint Symp. on Auralization and Ambisonics 2014, 1–6.

[2] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, F. Zotter: Spatial Sound With Loudspeakers and Its Perception: A Review of the Current State. Proc. of the IEEE 101 (2013) 1920–1938.

[3] J. B. Allen, D. A. Berkley: Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Am. 65 (1979) 943–950.

[4] E. A. Lehmann, A. M. Johansson: Prediction of energy decay in room impulse responses simulated with an image-source model. J. Acoust. Soc. Am. 124 (2008) 269–277.

[5] H. Wierstorf: Perceptual Assessment of Sound Field Synthesis. PhD thesis, Technical University of Berlin, 2014.

[6] D. de Vries, A. J. Reijnen, M. A. Schonewille: The Wave Field Synthesis Concept Applied to Generation of Reflections and Reverberation. Proc. 96th AES Conv. 1994, 3813.

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

[8] D. de Vries, E. M. Hulsebos: Auralization of room acoustics by wave field synthesis based on array measurements of impulse responses. Proc. 12th European Signal Proc. Conf. 2004, 1377–1380.

[9] ISO 3382-1: Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces. International Organization for Standardization, 2009.

[10] N. Kaplanis, S. Bech, S. H. Jensen, T. van Waterschoot: Perception of Reverberation in Small Rooms: A Literature Study. Proc. 55th AES Conf. 2014, P-3.

[11] F. E. Toole: Loudspeakers and Rooms for Sound Reproduction – A Scientific Review. J. Audio Eng. Soc. 54 (2006) 451–476.

[12] L. Beranek: Concert Halls and Opera Houses: Music, Acoustics, and Architecture. Springer, New York, 2004.

[13] H.-P. Seraphim: Untersuchungen über die Unterschiedsschwelle exponentiellen Abklingens von Rauschbandimpulsen. Acustica 8 (1958) 280–284.

[14] M. Barron: Interpretation of Early Decay Times in Concert Auditoria. Acustica 81 (1995) 320–331.

[15] G. A. Soulodre, J. S. Bradley: Subjective evaluation of new room acoustic measures. J. Acoust. Soc. Am. 98 (1995) 294–301.

[16] C. B. Pop, D. Cabrera: Auditory Room Size Perception for Real Rooms. Proc. Acoustics 2005, 115–121.

[17] M. Yadav, D. Cabrera, L. Miranda, W. L. Martens, D. Lee, R. Collins: Investigating auditory room size perception with autophonic stimuli. Proc. 135th AES Conv. 2013, 8934.

[18] T. Caulkins, E. Corteel, O. Warusfel: Wave field synthesis interaction with the listening environment, improvements in the reproduction of virtual sources situated inside the listening room. Proc. 6th Conf. on Digital Audio Effects 2003.

[19] F. Völk, H. Fastl: Wave Field Synthesis with Primary Source Correction: Theory, Simulation Results, and Comparison to Earlier Approaches. Proc. 133rd AES Conv. 2012, 8717.

[20] P. Vogel: Application of wave field synthesis in room acoustics. PhD thesis, Delft University of Technology, 1993.

[21] H. Kuttruff: Room Acoustics. Spon Press, Abingdon, 2009.