Assessment of the Perception of Synthesized Sound Fields with a Binaural Model

Hagen Wierstorf, Alexander Raake
Assessment of IP-based Applications, T-Labs, Technische Universität, Berlin, Germany

Sascha Spors
Quality and Usability Lab, T-Labs, Technische Universität, Berlin, Germany

Summary

Wave Field Synthesis (WFS) is a technique to synthesize a desired sound field in an extended area surrounded by a secondary source distribution. In practice these secondary sources are realized with loudspeakers and therefore a spatial sampling of the secondary sources occurs. The sampling may lead to aliasing artifacts in the sound field depending on the distance between the loudspeakers. In the time domain these artifacts will occur as additional wave fronts. If a linear loudspeaker array is used, its finite length will introduce further artifacts in the sound field. In this case the truncated array acts as a slit for the desired sound field and diffraction takes place, which leads to additional waves occurring from the edges of the loudspeaker array. The perception of these deviations from the desired sound field depends on the strength of the deviation and on the type of the desired sound field, e.g. if we have plane waves or a focused source located within the listener area. A test was conducted to rate the different perceptional dimensions of these artifacts. A binaural model after Lindemann (1986a) was used to predict the perception and to get insight into the mechanisms that may play a role in the perception of artifacts of synthesized sound fields.

PACS no. 43.66.Ba, 43.60.Sx

1. Introduction

Sound Field Synthesis (SFS) describes the ability to create a synthetic sound field within a defined and outspread listening area. This distinguishes it from other techniques such as stereophony, where the created sound field is only correct at one point, called the sweet spot. The principle of SFS lies in the Kirchhoff-Helmholtz-integral [1], hence the listening area has to be surrounded by secondary sources which are driven to create the desired sound field within the area. There exist different approaches to solve the underlying equations which lead to different techniques for SFS such as wave field synthesis (WFS) or higher-order Ambisonics (HOA). This study will limit its focus to WFS. The Kirchhoff-Helmholtz-integral assumes that the listening area is free of sinks and sources, nonetheless it is possible to synthesize virtual point sources within the listening area so called focused sources, with the restriction that this leads to a smaller listening area. In this study the sound field of a virtual point source and of a focused source are considered.

In practice, the secondary sources have to be spatially sampled due to the use of real loudspeakers as secondary sources. This leads to artifacts in the synthesized sound field which might be audible for the listener. At the moment it is not foreseeable to solve the problem of spatial sampling, therefore it is of great interest to know to what degree different artifacts of the synthesized sound field are audible. If this is known, the synthesis might be psychoacoustically tuned in a way to minimize audible artifacts. A few studies have already been conducted focusing on different aspects of the perception of synthesized sound fields. They have showed that, depending on the type of the synthesized sound field and the size of the loudspeaker array, wrong localisation, click artifacts [2] or coloration [3] are the most dominant unwanted perceptual effects.

In order to assess the perception of WFS in a more cost- and resource-efficient way than with listening tests, it will be of great interest to use an auditory model that can complement the subjective experiments. A first step into this direction is to apply existing binaural models, as it is apparent that binaural hearing plays a major role in the perception of sound fields. This study presents the successful use of a binaural model to predict wrong localisation in synthesized sound fields. In addition, we will discuss the limitations of current binaural models for the target application, and discuss what will be required in order to evaluate other aspects of the perception of synthesized sound fields.

2. Theory

2.1. Wave Field Synthesis

The theory of WFS for a linear loudspeaker array was initially derived from the Rayleigh integrals [4]. If the loudspeakers are located at the x-axis, they are able to synthesize a desired sound field in the x-y-half-plane with $y > 0$. The sound field P in the half-plane is then given by:

$$P(\mathbf{x}, \omega) = -\int_{-\infty}^{\infty} D(\mathbf{x}_0, \omega) G(\mathbf{x} - \mathbf{x}_0, \omega) dx_0 \ , \tag{1}$$

where $\mathbf{x} = (x, y)$ with $y > 0$, $\mathbf{x}_0 = (x_0, 0)$ denotes the position of the loudspeaker, $\omega = 2\pi f$ with frequency f, D is the driving signal of the loudspeakers and G is the 3D Greens function, which is a physical model of the point source used as the secondary source.

In this study we are interested in the synthesis of the sound field of a point source located behind the loudspeaker array and the sound field of a focused source located between the listener and the loudspeakers. For these kinds of virtual sources the driving signal D is given as [5]

$$D_{\mathrm{ps}}(\mathbf{x}_0, \omega) = \Psi(\omega) \frac{x_\mathrm{s} - y_\mathrm{s}}{|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|^{\frac{3}{2}}} e^{i\frac{\omega}{c}|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|} \ , \tag{2}$$

for a point source, and as [6]

$$D_{\mathrm{fs}}(\mathbf{x}_0, \omega) = \Psi(\omega) \frac{x_\mathrm{s} - y_\mathrm{s}}{|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|^{\frac{3}{2}}} e^{-i\frac{\omega}{c}|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|} \ , \tag{3}$$

for a focused source. $\mathbf{x}_\mathrm{s} = (x_\mathrm{s}, y_\mathrm{s})$ denotes the position of the virtual source, c the speed of sound and Ψ contains the spectrum of the desired virtual source and in addition amplitude and spectral correction terms. These corrections are necessary due to the use of point sources as secondary sources instead of line sources which are needed for a correct synthesis in a plane, but are not available in practice. In Figure 1, simulations of the sound fields for the two given driving functions are shown. In the case of a focused source the sound field converges towards the focal point at $(0, 1)$ m and diverges afterwards, which means the listening area is restricted to the area with $y > 1$ m.

If we transform the equations from above into the temporal domain we will get for the sound field

$$p(\mathbf{x}, t) = -\int_{-\infty}^{\infty} d(\mathbf{x}_0, t) g(\mathbf{x} - \mathbf{x}_0, t) dx_0 \ , \tag{4}$$

where g is again the three dimensional Greens function and d the driving signal. For the driving signals for the two desired virtual sources we get

$$d_{\mathrm{ps}}(\mathbf{x}_0, t) = \psi(t) * \frac{x_\mathrm{s} - y_\mathrm{s}}{|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|^{\frac{3}{2}}} \ \delta(t + \tfrac{|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|}{c}) \ , \tag{5}$$

$$d_{\mathrm{fs}}(\mathbf{x}_0, t) = \psi(t) * \frac{x_\mathrm{s} - y_\mathrm{s}}{|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|^{\frac{3}{2}}} \ \delta(t - \tfrac{|\mathbf{x}_0 - \mathbf{x}_\mathrm{s}|}{c}) \ , \tag{6}$$

where δ denotes the delta distribution and ψ the inverse Fourier transformation of Ψ. As can be seen, the driving signal for the focused source is a time reversal of the point source driving signal, which is a known property from the principle of acoustic focusing [7].

2.2. Discrete Loudspeakers

The sound field is synthesized by loudspeakers. In the calculations above the loudspeakers were handled as an infinitely long continuous distribution, which is not the case in reality. Hence, we have to handle the case of a real loudspeaker array, which is discrete and has a finite length. It has been shown that the use of a real loudspeaker array will lead to spatial sampling artifacts due to the discretization of the loudspeaker distribution and to truncation artifacts due to the finite length of the array. [8, 6]

Spatial sampling occurs above the spatial aliasing frequency $f_\mathrm{al} = \frac{2\Delta x_0}{c}$, where Δx_0 is the distance between two loudspeakers. In the sound field, the spatial aliasing will be present as additional unwanted contributions, because the contributions of the single loudspeaker will not cancel out each other. This can be seen in Figure 2, on the left side for a point source and on the right side for a focused source. Additional wave fronts exist besides the desired one. For the point source these additional wave fronts arrive at the listener position after the desired one. For the focused source the case is inverted due to the time reversal principle, so that the additional wave fronts arrive before the desired one at listener positions.

The truncation of the array leads to additional spherical waves originating from the edges of the array and interfering with the desired waves due to diffraction [9]. These can be reduced by applying a tapering window which drives the loudspeakers at the edges with a lower amplitude [4]. In addition to this, the listening area is smaller, and large amplitude differences occur at the side of the listening area due to diffraction minima and maxima, as can be seen in Figure 3. It shows the sound field synthesized by a loudspeaker array with a length of 0.75 m. The size of the focal point for the focused source is very large, which also is an effect of the truncation of the loudspeaker array and due to the diffraction limit for the focal point [10].

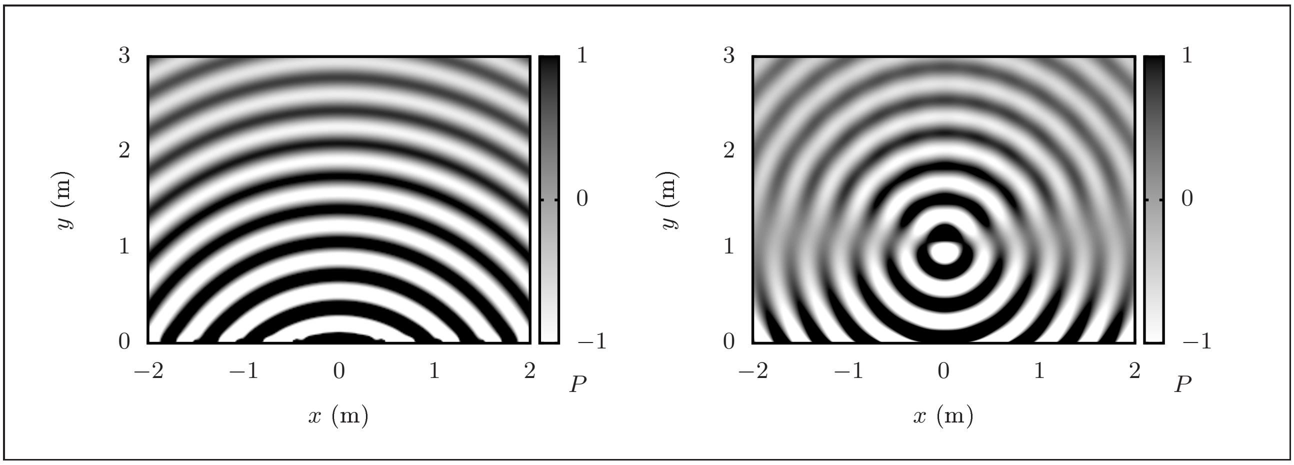

Figure 1. Simulation of the sound field $P(\mathbf{x},\omega)$ of a monochromatic virtual source with $f = 1000\,\mathrm{Hz}$. The sound field of a point source located at $\mathbf{x}_\mathrm{s} = (0,-1)\,\mathrm{m}$ (left) and of a focused source located at $\mathbf{x}_\mathrm{s} = (0,1)\,\mathrm{m}$ is shown.

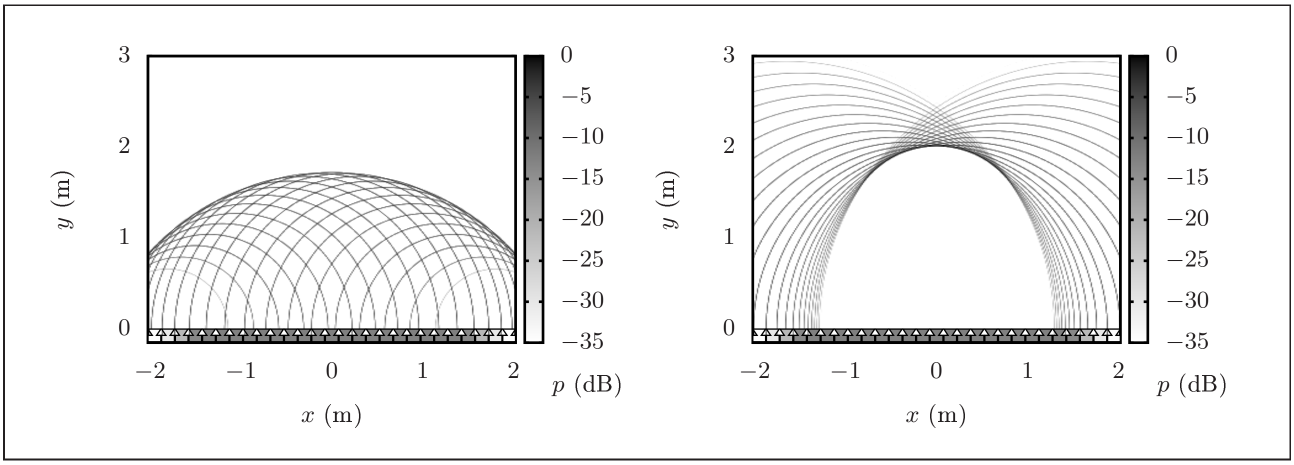

Figure 2. Simulation of the sound field $p(t,\omega)$ for a broadband virtual source. On the left, the sound field of a point source is shown that is located at $\mathbf{x}_\mathrm{s} = (0,-1)\,\mathrm{m}$ at a time $t = 7.9\,\mathrm{ms}$ after the impulse has startet at $\mathbf{x}_\mathrm{s}$. On the right, the sound field of a focused source is shown that is located at $\mathbf{x}_\mathrm{s} = (0,1)\,\mathrm{m}$ and $t = 2.9\,\mathrm{ms}$. The amplitude of the loudspeakers due to tapering is indicated by the color-intensity of speakers.

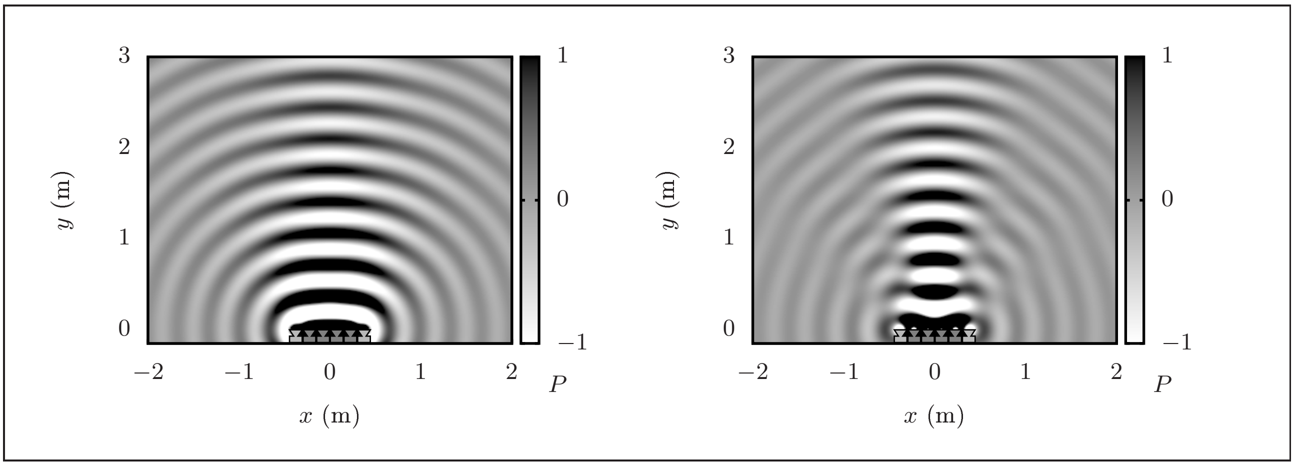

Figure 3. Simulation of the sound fields $P(\mathbf{x},\omega)$ of monochromatic virtual sources with $f = 1000\,\mathrm{Hz}$. The sound field of a point source located at $\mathbf{x}_\mathrm{s} = (0,-1)\,\mathrm{m}$ (left) and of a focused source located at $\mathbf{x}_\mathrm{s} = (0,1)\,\mathrm{m}$ is shown. The size of the loudspeaker array is $L = 0.75\,\mathrm{m}$. The amplitude of the loudspeakers due to tapering is indicated by the color-intensity of speakers.

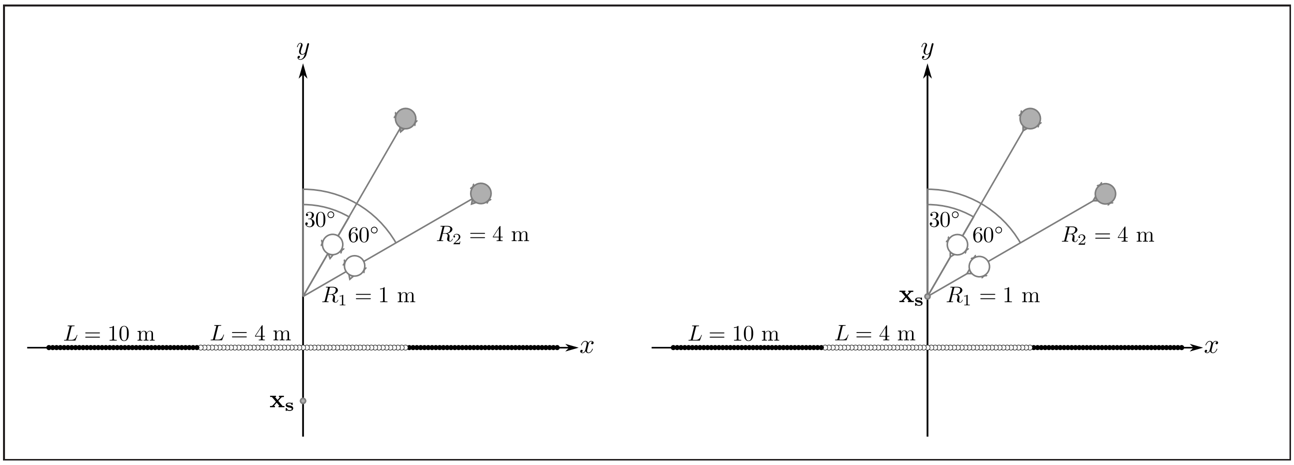

Figure 4. Geometry of the virtual WFS systems used in the experiment. The loudspeaker arrays were always located on the x-axis with their center at $x = 0$. A more detailed description is given in the text.

3. Experiment

It has been shown that the errors in the sound field due to spatial sampling and truncation of the loudspeaker array lead to different artifacts in the perceptual domain [3, 8, 6, 2].

In the previous section we have illustrated the fact that spatial aliasing leads to additional unwanted wave fronts in the sound field. For the perception of these extra wave fronts the precedence effect is of relevance [11, 12]. It describes the phenomenon that the first wave front arriving at the listener dominates the localization perception in a time frame of 1 ms – 40 ms after its arrival. In addition, the extra wave fronts are not heard as echoes in this time frame. This is why the precedence effect enables us to communicate in closed spaces. As a consequence we can assume, that in spite of the aliasing in the sound field, a virtual point source is perceived at the right location with no additional echoes, but with coloration due to the influence of the additional wave fronts, that affect the overall frequency spectrum. For focused sources, the additional wave fronts arrive at the listener position before the desired wave front. Geier et al. [2] and Wierstorf et al. [10] have shown in an experiment that in this case pre-echoes are audible for long arrays, and can be reduced by using shorter loudspeaker arrays. The localisation of a focused source can be disturbed by the localisation dominance of the precedence effect due to the fact that the first arriving wave front comes from a single loudspeaker position and not from the location of the focused source. In addition, for short loudspeaker arrays the truncation can lead to wrong binaural cues such as the interaural level difference (ILD).

The focus of the present study lies on the localization of the virtual sources. As mentioned above, due to the precedence effect the localization may depend on the first arriving wave front. One problem to be accounted for in SFS is the fact that we have not the classical precedence effect scenario, since instead of one well defined repetition, a bunch of repetitions are arriving with a distance in time of under 1 ms, all from different directions and with different amplitudes, depending on the loudspeaker they arrive from. For focused sources, the effect is dependent on the listener position, because the sampling artifacts are different at different positions [6]. In order to simplify the assessment of the localisation in SFS, we evaluate the performance of a binaural model to predict the perceived localization. A subjective test was done for the localization of focused sources in order to verify the model data. The test was part of a larger subjective test asking also for click artifacts, which was presented in [10]. After that, the localization of virtual sources has been modeled and will be presented in Section refsec:modelling.

3.1. Method

The method will only be presented briefly here, for a full description of the experiment refer to [10]. The test was conducted by a virtual WFS system realized by dynamic binaural resynthesis [13] and with headphone presentation. Binaural resynthesis gives the possibility to position different subjects in a consistent manner in the sound field and to switch instantaneously between different positions or virtual loudspeaker arrays. To create virtual loudspeaker arrays, a set of head-related impulse responses (HRIR), measured with the FABIAN dummy head [14], has been interpolated and summed up. The SoundScape Renderer [15] in combination with a head-tracker was used to realize the dynamic binaural presentation.

The test itself consisted of 17 different conditions resulting from a sample of four listener positions and five array lengths. See Figure 4 for a sketch of the used geometry. As array lengths, 0.3 m, 0.75 m, 1.8 m, 4 m, 10 m were used. Note that for an array length of 4 m only the two listener positions with $R = 1$ m,

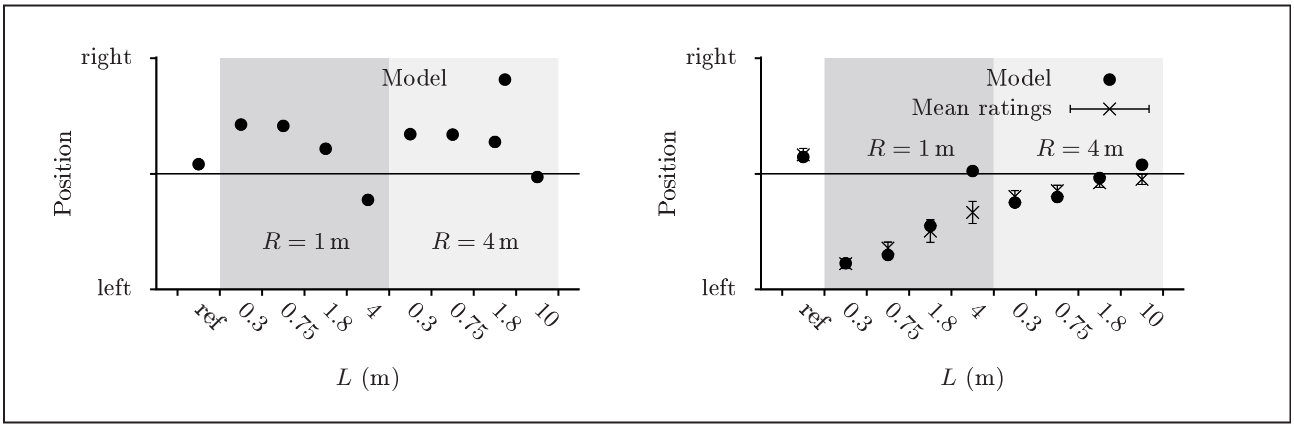

Figure 5. Results for the model dependent on the conditions for the point source (left) and the focused source (right). In addition the mean value and variance for the rating of the attribute pair *left* vs. *right* is shown for the focused source. The abscissa displays the different array lengths, the ordinate the judged or predicted position.

and for an array length of 10 m only the listener positions with $R = 4$ m have been tested. In addition to the applied SFS, a reference condition with a single loudspeaker located at the virtual source position was presented. As audio material a sentence uttered by a female speaker and piece of castanets were played.

Six subjects participated in the test. All of them had normal hearing levels and experience with such tests. In the test the subjects were presented a screen displaying on the top of the screen the attribute pair *left* vs. *right* with which the stimuli were to be judged, and below nine sliders, one for each of the eight different conditions with a fixed angle of 30° or 60°, and one for the reference condition. The subjects could switch between the different conditions instantaneously and as often as they wanted. They had to position the sliders according to the perceived lateralization of the stimuli. The test was run in two parts, one with speech, the other with castanets.

3.2. Results and Discussion

In Figure 5 (right), the results of the localization ratings are presented. The mean over all subjects, audio materials and the two angles have been calculated. It can be seen that the reference condition (arriving from the front of the listener) was rated to come slightly from the right side. All other conditions came from the left side, whereby shorter arrays and smaller radii lead to a rating further to the left.

The initial head orientation of the listener was always towards the location of the focused source. This means that the perceived lateralization of the focused source should have been near 0° for all conditions. On the other hand, the localisation dominance implies that the perceived lateralization should be dominated by the position of the loudspeaker which emanates the first wave front, which here is the speaker at the edge of the loudspeaker array. In this case, the lateralization of the focused source is expected to be to the left of the listener, which obviously is the case. But for shorter arrays the lateralization to the left is expected to be lesser due to the fact that the edge of the array moves to the right, from a listeners point of view. The result shows the opposite trend: the shorter the array, the more lateralization to the left.

As mentioned in Section 2.2 the truncation of the array leads to a diffraction of the sound field, and thus to errors in the binaural cues. The diffraction is the more pronounced, the shorter the array. Hence, for small array sizes, the result of the experiment may be accounted for by the diffraction. In order to test this hypotheses, a binaural model will be used in the next Section to predict the localisation based only on the two binaural cues interaural level difference (ILD) and interaural time difference (ITD).

4. Modelling

Localization prediction for an auditory event using a binaural model has been the scope of many studies [16]. For the purpose of this study, a binaural model after Lindemann [17] has been implemented in the Auditory Modelling Toolbox [18], and applied to the virtual sources. The binaural model examines the interaural time difference by calculating a running cross-correlation between the two ear signals. By incorporating a contralateral inhibition mechanism, also the ILD is accounted for, by shifting the peak of the cross-correlation. The same stimuli previously presented to test subjects in the listening test were used as input to the model. The same parameters of the model as described in the original paper by Lindemann were chosen. As a value for the lateralization of the source, the centroid of the cross-correlation output was calculated. In a first step this was done for all conditions of the focused source experiment, corresponding to the set-up sown on the right of Figure 4. The results are illustrated in Figure 5. The data from the model was scaled to have the same order of magnitude as the ratings for the *left* vs. *right* attribute pair.

As shown in the graph, the model is able to predict the lateralization for the reference conditions as well as for the three shortest arrays. On the other hand, for the two long arrays of 4 m and 10 m, the model is not able to predict the results. In the Lindemann model, the precedence effect is not included. It considers only ITD and ILD. Hence, the conclusion can be drawn that the lateralization of the focused source is dominated by the wrong binaural cues created due to the diffraction in case of the short arrays, and for larger arrays the lateralization is influenced by the precedence effect, so that the model fails in this case.

In a next step, the model was used to predict the lateralization for a point source for the listening positions as shown in the left of Figure 4. The predictions are depicted on the left side of Figure 5, this time without respective listening test results. Again, the head of the listener has always been oriented towards the position of the virtual source. The model predicts a lateralization of around 0° only for the reference condition, and for the 10 m-array condition. For 4 m, the perception is bounded to the left of the listener, and for shorter arrays to the right. For short arrays again wrong binaural cues are present in the sound field, and are likely to be the reason for values predicted by the model. For the 4 m array the predicted result cannot easily be explained at the moment.

5. Conclusions

The perceptual properties of SFS are still an open field of research. The artifacts of SFS due to the spatial sampling or the truncation of the used loudspeaker array cannot easily be avoided in practice, since a loudspeaker distance of 0.15 m already leads to spatial aliasing for frequencies above approximate 1000 Hz. To apply SFS, it is therefore important to know to what degree and what kind of artifacts a subject is able to perceive in a synthesized sound field, and which of these artifacts lead to an especially annoying perception. In order to reach this goal, additional subjective tests are needed. As a complement the usage of auditory models can provide first predictions for effects the models are able to address. In this study, a binaural model was used to predict the lateralization of virtual sources located in front of (focused source) and behind a loudspeaker array (virtual point source). It could be shown that the binaural model was able to predict localisation artifacts for focused sources synthesized by short loudspeaker arrays. These artifacts are due to the diffraction of the sound field for short arrays. For virtual point sources, the model also predicts localisation artifacts, which have to be verified in a future listening test. On the other hand, so far the model is not able to account for localisation dominance as part of the precedence effect. Future investigations will focus on the detection of further model limitations, and extensions to the models to finally serve for quality prediction for SFS.

Acknowledgement

This work was supported by DFG RA 2044/1-1.

References

[1] E. G. Williams: Fourier Acoustics. Academic Press, San Diego 1999.

[2] M. Geier et al.: Perception of focused sources in Wave Field Synthesis. Proc. 128th AES Conv. 2010.

[3] H. Wittek: Perceptual differences between wavefield synthesis and stereophony. PhD-thesis, University of Surrey 2007.

[4] A. J. Berkhout, D. de Vries and P. Vogel: Acoustic control by Wave Field Synthesis. JASA 93(5) (1993) 2764-2778.

[5] S. Spors, R. Rabenstein and J. Ahrens: The theory of Wave Field Synthesis revisited. Proc. 124th AES Conv. 2008.

[6] S. Spors, H. Wierstorf, M. Geier and J. Ahrens: Physical and perceptual properties of focused sources in Wave Field Synthesis. Proc. 127th AES Conv. 2009.

[7] S. Yon, M. Tanter and M. Fink: Sound focusing in rooms: the time-reversal approach. JASA 113(3) (2003) 1533-1543.

[8] S. Spors and J. Ahrens: Spatial aliasing artifacts of wave field synthesis for the reproduction of virtual point sources. Proc. 126th AES Conv. 2009.

[9] M. Born and E. Wolf: Principles of Optics. Cambridge University Press, New York, 1999.

[10] H. Wierstorf, M. Geier and S. Spors: Reducing artifacts of focused sources in Wave Field Synthesis. Proc. 129th AES Conv. 2010.

[11] H. Wallach, E. B. Newman and M. R. Rosenzweig: The precedence effect in sound localization. AJP **57** (1949) 315-336.

[12] J. Blauert: Spatial Hearing. The MIT Press, Cambridge, Massachusetts, 1997.

[13] A. Lindau, T. Hohn and S. Weinzierl: Binaural resynthesis for comparative studies of acoustical environments. Proc. 122th AES Conv. 2007.

[14] A. Lindau and S. Weinzierl: FABIAN – An instrument for the software-based measurement of binaural room impulse responses in multiple degrees of freedom. Proc. VDT Intern. Conv. 2006.

[15] M. Geier, J. Ahrens, and S. Spors: The SoundScape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods. Proc. 124th AES Conv. 2008.

[16] R. M. Stern and C. Trahiotis: Models of binaural interaction. In: Handbook of Perception and Cognition, Volume 6: Hearing. B. C. J. Moore (eds.). Academic Press, New York 1995.

[17] W. Lindemann: Extension of a binaural cross-correlation model by contralateral inihibition. I. Simulation of lateralization for stationary signals. JASA 80(6) (1986) 1608-1622.

[18] P. L. Søndergaard et al.: Towards a binaural modelling toolbox. Proc. EAA For. Acust. 2011.