



# Audio Engineering Society Convention Paper

Presented at the 129th Convention  
2010 November 4–7 San Francisco, CA

*The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42<sup>nd</sup> Street, New York, New York 10165-2520, USA; also see [www.aes.org](http://www.aes.org). All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

---

## Reducing Artifacts of Focused Sources in Wave Field Synthesis

Hagen Wierstorf, Matthias Geier and Sascha Spors

*Deutsche Telekom Laboratories, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany*

Correspondence should be addressed to Hagen Wierstorf ([Hagen.Wierstorf@telekom.de](mailto:Hagen.Wierstorf@telekom.de))

### ABSTRACT

Wave Field Synthesis provides the possibility to synthesize virtual sound sources located between the loudspeaker array and the listener. Such sources are known as focused sources. Previous studies have shown that the reproduction of focused sources is subject to audible artifacts. The strength of those artifacts heavily depends on the length of the loudspeaker array. This paper proposes a method to reduce artifacts in the reproduction of focused sources by using only a subset of the loudspeakers of the array. A listening test verifies the method and compares it to previous results.

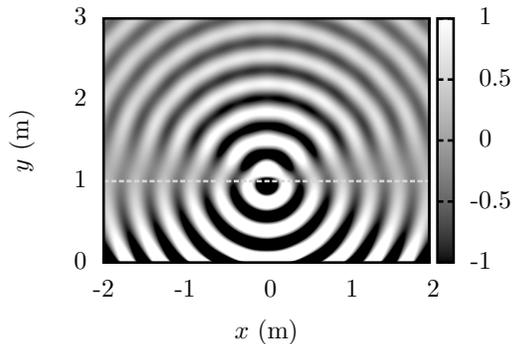
### 1. INTRODUCTION

*Wave Field Synthesis* (WFS) is one of the most prominent high-resolution sound field synthesis methods which are used and studied nowadays [1]. WFS offers the potential of creating the impression of a virtual point source located inside the listening area [2]. These virtual sources are termed *focused sources* due to the generation of an acoustic focus point at their position. Causality limits their position to be located between the loudspeakers and the listeners.

The theory of WFS assumes a spatially continuous distribution of acoustic sources (secondary sources)

around the listening area. In practical implementations the secondary source distribution is realized by a limited number of loudspeakers at discrete positions. Therefore a spatial sampling of the secondary sources occurs. For typical loudspeaker geometries and audio content this leads to spatial sampling artifacts that may become audible [3].

Previous studies have shown that these sampling artifacts are especially critical in the case of synthesizing focused sources [4, 5]. In this case a number of strong audible artifacts occur and reduce the number of possible applications of focused sources in real-life scenarios.



**Fig. 1:** Simulation of a wave field  $P(\mathbf{x}, \omega)$  for a monochromatic focused source located at  $\mathbf{x}_s = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . A continuous secondary source distribution is applied on the  $x$ -axis with a length of  $L \rightarrow \infty$ . The frequency of the source is  $f_s = 1000$  Hz. The amplitude of the wave field is clipped at  $|P| = 1$ . In the area below the grey dotted line the wave field is converging to the focus point. Above the line it is diverging from the focus point.

This paper introduces a method to synthesize focused sources with less audible artifacts by using only a subset of the secondary sources. In a listening test the proposed method is evaluated by using the main perceptual attributes that characterize the artifacts of focused sources. These attributes are extracted from a previous study [5]. In addition some results of the listening test are predicted by a binaural model.

The tests were conducted with a “virtual” WFS system realized by dynamic binaural resynthesis and presented to the participants by means of headphones in order to create reproducible test conditions.

## 2. THEORY AND PRELIMINARY WORK

The theory of WFS was initially derived from the *Rayleigh integrals* for a linear secondary monopole source distribution [6]. This distribution is capable of synthesizing a desired wave field in one of the half planes defined by the linear secondary source distribution. The wave field in the other half plane is a mirrored version of the desired one.

Without loss of generality a geometry can be chosen for which the secondary source distribution is

located on the  $x$ -axis of a Cartesian coordinate system. The reproduced wave field is then given by

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

where  $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$  with  $y > 0$  and  $\mathbf{x}_0 = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}$ . The functions  $D(\cdot)$  and  $G(\cdot)$  denote the secondary source driving signal and the wave field emitted by the secondary monopole sources, respectively. In WFS the driving function is given as

$$D(\mathbf{x}_0, \omega) = 2 \frac{\partial}{\partial y} S(\mathbf{x}, \omega) \Big|_{\mathbf{x}=\mathbf{x}_0}, \quad (2)$$

where  $S(\cdot)$  denotes the wave field of the desired virtual source.

In applications with cabinet loudspeakers as secondary sources the dimensional mismatch of a 3D secondary source for 2D reproduction has to be considered, too. This leads to a so called 2.5D driving functions which applies an amplitude correction to reduce this mismatch.

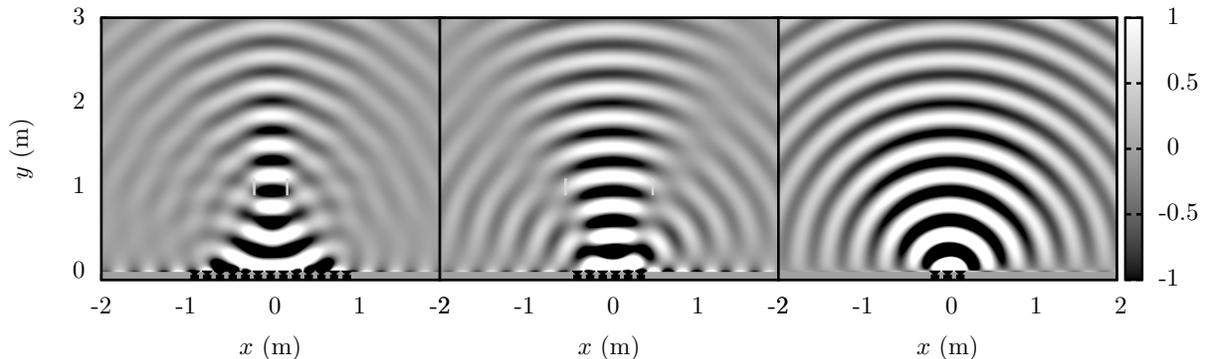
A reformulation of the theory based on the *Kirchhoff-Helmholtz integral* revealed that also arbitrary convex distributions can be employed [7, 8]. This study limits itself to linear arrays as these are mainly applied in real life scenarios at the moment. A detailed review of the theory of WFS can be found in the literature such as [6, 9].

### 2.1. Focused Sources

In model-based WFS spatial source models for the virtual sources are used to calculate the driving function. Typical source models such as plane or spherical waves are used which are driven with given input signals like speech or music. Spherical waves represent virtual monopole sources.

For the synthesis of a focused source, a synthesized wave field is desired which converges towards a focus point and diverges after passing the focus point. In a previous paper the 2.5D driving function for a focused source has been derived as [4]

$$D_{2.5D}(\mathbf{x}_0, \omega) = -g_0 \hat{S}_s(\omega) \sqrt{\frac{\omega}{2\pi j c}} \frac{y_0 - y_s}{|\mathbf{x}_0 - \mathbf{x}_s|^{\frac{3}{2}}} e^{j \frac{\omega}{c} |\mathbf{x}_0 - \mathbf{x}_s|}, \quad (3)$$



**Fig. 2:** Simulation of the wave field  $P(\mathbf{x}, \omega)$  of a monochromatic focused source positioned at  $\mathbf{x}_s = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . The used array lengths and positions are indicated by the loudspeakers. The array lengths are, from left to right: 1.8 m, 0.75 m, 0.3 m. The distance between two loudspeakers is  $\Delta x = 0.15$  m and the used frequency  $f_s = 1000$  Hz. The amplitude of the wave field is clipped at  $|P| = 1$ . The calculated width of the focus point  $\Delta_{f_s}$  due to diffraction is given in the left two figures by the distance between the two grey lines, respectively.

where  $g_0$  is a geometry dependent constant and  $\hat{S}_s(\omega)$  is the spectrum of the source signal. In Fig. 1 the wave field  $P(\mathbf{x}, \omega)$  for a monochromatic focused source located at  $\mathbf{x}_s = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  is simulated. The secondary source distribution is located at the  $x$ -axis. The wave field converges for  $0 < y < 1$  m towards the position of the focused source and diverges for  $y > 1$  m which defines the listening area for the given focused source position.

If the driving function (3) is transferred into the temporal domain, it is given as

$$d_{2.5D}(\mathbf{x}_0, t) = s(t) * h(t) * \frac{y_0 - y_s}{|\mathbf{x}_0 - \mathbf{x}_s|^{\frac{3}{2}}} \delta\left(t + \frac{|\mathbf{x}_0 - \mathbf{x}_s|}{c}\right), \quad (4)$$

where  $c$  is the speed of sound and  $h(t)$  denotes the inverse Fourier transformation

$$h(t) = \mathcal{F}^{-1} \left\{ \sqrt{\frac{\omega}{2\pi j c}} \right\}. \quad (5)$$

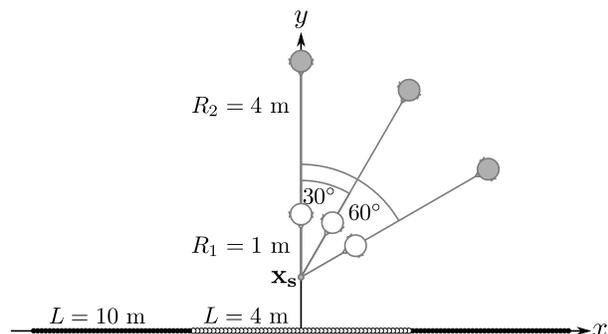
It is easy to see that this driving function can be implemented very efficiently by filtering the virtual source signal  $s(t)$  with the so called pre-equalization filter  $h(t)$  and applying a weighting and delaying of this pre-filtered signal for every loudspeaker. To obtain causality a pre-delay has to be applied.

## 2.2. Perception of Focused Sources

Theoretically, when an infinitely long continuous distribution of secondary sources is used, no other errors than an amplitude mismatch due to the 2.5D reproduction are expected in the perception of the wave field [4].

However, such a continuous distribution cannot be implemented in practice, because a finite number of loudspeakers has to be used. This results in a *spatial sampling* and *spatial truncation* of the secondary source distribution.

The spatial truncation of the array leads to two strong restrictions. On the one hand the listener area becomes smaller with a smaller array which can be seen in Fig. 2. The listening area can be approximated by the triangle that is spanned for  $y > y_s$  by the two lines coming from the edges of the loudspeaker array and crossing the focused source position. The other problem is that a smaller loudspeaker array has influence on the width of the focus point. The loudspeaker array can be seen as a single slit which causes a diffraction of the wave field going through it. This leads to a widening of the focus point depending on the wavelength  $\lambda = \frac{c}{f}$ . The width of the focus point at its position  $y_s$  can be defined as the distance between the first minima in



**Fig. 3:** Geometry of the *virtual* WFS system used in the experiment. The loudspeaker array is located on the  $x$ -axis with its center at  $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$  m. Two of the used array lengths  $L$  are indicated.  $\mathbf{x}_s = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  m denotes the position of the focused source. The position of the listener is given by the radius  $R_1 = 1$  m and  $R_2 = 4$  m and the angle  $\varphi \in [0^\circ, 30^\circ, 60^\circ]$ . The head orientation of the listener is always in direction of the focused source.

the diffraction pattern and is given by

$$\Delta_{fs} = 2|y_s - y_0| \tan \left( \sin^{-1} \frac{\lambda}{L} \right), \quad (6)$$

where  $\Delta_{fs}$  is the width of the focus point in meter,  $L$  the array length,  $y_s$  the focused source position and  $y_0$  the position of the loudspeaker array. As can be easily seen for wavelengths  $\lambda > L$  exists no minimum in the diffraction pattern and the loudspeaker array works as a point source. This is the case for a wave length of  $\lambda = 0.343$  m and an array length of 0.3 m as simulated in the right hand side of Fig. 2. For an array length of  $L = 0.75$  m the width of the focus point is  $\Delta_{fs} = 1.03$  m and for  $L = 1.8$  m it is  $\Delta_{fs} = 0.39$  m as can also be seen in the figure.

The spatial sampling leads to spatial aliasing which means that the wave field cannot be reproduced exactly for frequencies above the aliasing frequency  $f_{al}$  which depends on the listener position and the spacing of the loudspeakers. In the case of a virtual point source located behind the loudspeaker array spatial aliasing leads to additional wave fronts above the aliasing frequency arriving at the listener position from every single loudspeaker after the desired wave front. Due to the precedence effect [10] these additional wave fronts may be perceived as room impression and coloration. In the case of a focused source

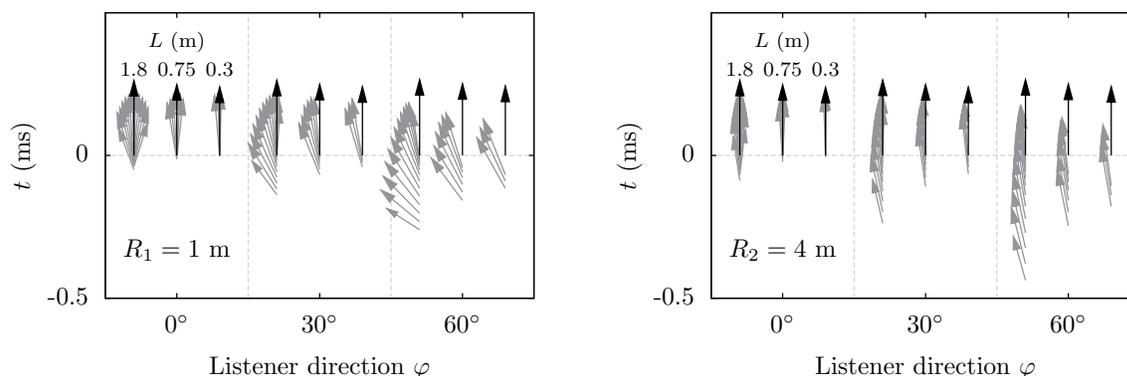
the loudspeakers that have the largest distance to the focused source point are driven first as can be seen in (4). This leads to additional wave fronts arriving at the listener position *before* the desired wave front of the focused source. It has been shown in detail ([4, 5]) that these additional wave fronts for focused sources may be perceived as strong click-artifacts, additional sources or a wrong localisation of the source compared to the case of a virtual point source located behind the array.

In a previous study [5] the *Repertory Grid Technique* (RGT) [11] was used to derive perceptual attributes that describe the artifacts of focused sources. The WFS arrays applied in this study were linear arrays of 4 m and 10 m. The listener positions of the test are shown in Fig. 3. In the RGT every subject creates her/his own attributes. In order to use a set of common attributes in this study, a clustering method was applied to identify common attributes among the subjects [12]. A group average clustering algorithm was used for every single subject to identify sets of perceptual attributes. Afterwards common attributes among the subjects were identified and given a common name. In the previous study speech and castanets have been used as audio materials and the attributes were derived independent for both signals. Therefore also the identification of common attributes among the subjects was done independently for both source materials. The two most common attribute pairs were pairs that referred to an attribute describing the amount of artifacts in the stimuli. This attribute pair was named *few artifacts* vs. *many artifacts*. Attribute pairs that referred to the position of the focused source were the second most common. This pair was named *left* vs. *right*. Both attribute pairs were used in the evaluation of the proposed method to reduce the artifacts in focused sources. This method will be presented in the next section.

### 2.3. Reduction of Artifacts

As mentioned in the last section, every single loudspeaker causes a pre-echo above the aliasing frequency for focused sources. The direction and time of arrival of the echoes is determined by the position of the loudspeakers.

The results from previous studies have also shown that the perception of artifacts is stronger for larger



**Fig. 4:** Direction, amplitude (dB) and time of appearance of pre-echoes for the different loudspeaker arrays at a radius of  $R_1 = 1$  m (left) and  $R_2 = 4$  m (right). For every listener direction  $\varphi$  the left entry is the 1.8 m array followed by the 0.75 m and 0.3 m one. For a sketch of the listener positions given by  $\varphi$ ,  $R_1$  and  $R_2$  see Fig. 3. The direction of every pre-echo is given by the arrow direction and its amplitude by the length of the arrow.

arrays and positions for which the pre-echoes arrive earlier. Therefore a method to reduce artifacts of focused sources is to reduce the time of arrival of the pre-echoes.

A straightforward technique to do this is to use only a subset of loudspeakers in the driving function of the focused sources. In this study sub-array lengths of  $L = 1.8$  m,  $L = 0.75$  m and  $L = 0.3$  m were chosen with 13, 6 and 3 active loudspeakers, respectively. The middle of the sub-array was always chosen as  $\binom{0}{0}$  in order to have a symmetric loudspeaker distribution around the  $x$ -position of the focused source. The low number of loudspeakers leads to arrival times of the pre-echoes under 1 ms for all considered positions (see Fig. 4).

However, reducing the pre-echo time comes not without its shortcomings. As can be seen for the listener positions of  $\varphi = 30^\circ$  and  $\varphi = 60^\circ$  in Fig. 4, the pre-echoes arrive mainly from the left hand side. This means the summing localization [13] may lead to a perception of the focused source from the left and not from the front as desired.

Also, the truncation of the array leads to strong restrictions as mentioned before. In Fig. 2 it can be seen that the focus point gets very large for a frequency of about  $f_s = 1000$  Hz. In this case it is possible that the listener perceives no focused source,

but a point source at the position of the loudspeaker array.

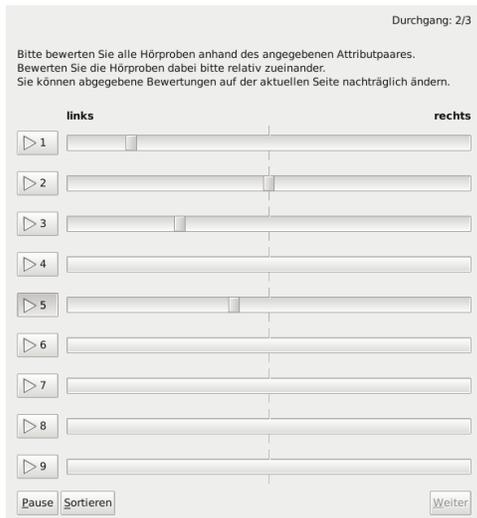
### 3. METHOD

A listening test was conducted to evaluate the use of sub-arrays to produce a focused source with reduced pre-echo artifacts. Therefore the subjects had to rate the perception of three short arrays in comparison to the arrays with a length of 4 m and 10 m used in the previous experiment [5] and a single loudspeaker reference. This was done for the six different listener positions shown in Fig. 3. For the rating three different pairs of German attributes were used: *wenig Störgeräusche* vs. *viele Störgeräusche* (few artifacts vs. many artifacts), *links* vs. *rechts* (left vs. right), *nah* vs. *fern* (close vs. far). For this study only the first two mentioned attribute pairs are considered.

#### 3.1. Participants

Six test subjects participated in the experiment. All of them were members of the Audio Group at the Quality and Usability Lab and had normal hearing levels.

The experiment was split into two sessions which were essentially the same except for different source material (speech, castanets) used in the stimuli.



**Fig. 5:** Screenshot of the rating GUI. At the top left and right, the attribute pair to rate is displayed. Below, there are sliders (one per condition) and for each slider there is a button to switch to the corresponding condition. A sort button gives the subjects the possibility to rearrange the presented sliders after their ratings.

### 3.2. Apparatus

The tests were conducted with a “virtual” WFS system realized by dynamic binaural re-synthesis and presented to the test subjects by means of headphones. See Fig. 3 for a sketch of the geometry of the virtual WFS arrays. Five linear loudspeaker arrays with a length of  $L = 10$  m,  $L = 4$  m,  $L = 1.8$  m,  $L = 0.75$  m,  $L = 0.3$  m, and a spacing of  $\Delta x = 0.15$  m between the loudspeakers were used. The transfer functions of the individual virtual loudspeakers were obtained by interpolating a database of anechoic *Head-Related Impulse Responses* (HRIRs) of the FABIAN mannequin [14] to the required directions and applying further weighting and delaying in order to account for the virtual loudspeakers’ distances.

For each possible head orientation from  $-180^\circ$  to  $180^\circ$  in  $1^\circ$  steps the driving function (5) of each virtual loudspeaker was convolved with that pair of HRIRs representing the given combination of loudspeaker and head orientation and the result was added for all loudspeaker positions. Each stimu-

lus was thus represented by a pair of impulse responses (left and right ear) which in turn represented the spatio-temporal transfer function of the loudspeaker system driven with the given configuration to the ears of the mannequin for a given head orientation [15]. This type of spatio-temporal transfer function is then typically referred to as *Binaural Room Transfer Function* or *Binaural Room Impulse Response* (BRIR) – when represented in time domain. The BRIRs were calculated for all possible head orientations. The headphone signal was then obtained by convolving a given input signal with the BRIRs representing the entire loudspeaker system as described above.

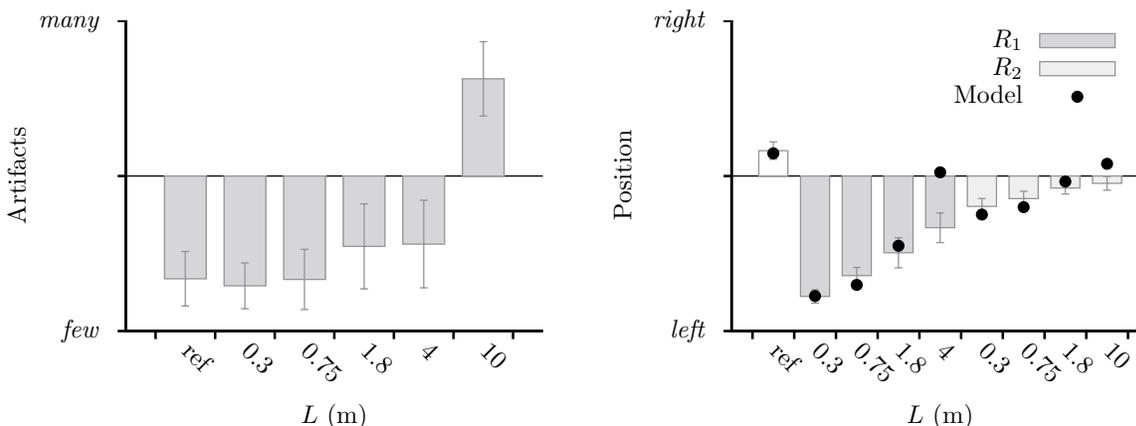
In order avoid biases in the subjects’ responses due to different levels of the stimuli, all BRIRs were normalized in amplitude based on the frontal direction.

For the three short arrays, six different listener positions on the right side ( $x \geq 0$ ) of the focused source position located on two half circles were used. The radii of the two circles were  $R_1 = 1$  m and  $R_2 = 4$  m. The 4 m array uses only the smaller radius and the 10 m array only the larger radius. Three different listener angles of  $\varphi = 0^\circ$ ,  $30^\circ$  and  $60^\circ$  were applied for both half circles. The different conditions are therefore named after their angle, radius and array length:  $\varphi_{LR}$ , for example  $0^\circ_{4m R_1}$  or  $60^\circ_{1.8m R_2}$ . The initial head orientation was always pointing towards the focused source, as shown in Fig. 3. This means for all conditions, the focused source was located directly in front of the listener.

As another condition, a reference stimulus (ref.) was created, which consisted of a single sound source straight in front of the listener. This was realized by directly using the corresponding HRIRs from the database. Stimuli examples containing the  $0^\circ$  head orientation of the BRIRs are available at [16].

As discussed in [4, 17] the pre-equalization filter  $h(t)$  has to be optimized separately for each listening position. This has been done in order to avoid systematic coloration by an improper choice of pre-equalization filter which was not part of the investigation.

As mentioned before, two different input signals were used – speech and castanets. The speech signal was chosen because it contains both periodic and aperiodic components and it is a very common and famil-



**Fig. 6:** Mean value and variance dependent on the condition for the ratings of the two attribute pairs *few artifacts vs. many artifacts* (left) and *left vs. right* (right). The mean is calculated about the different source materials (speech and castanets) and the different listener positions. For the artifacts attribute the condition  $0^\circ$  for the speech material was omitted. For the position attribute only the listener positions with  $\varphi = 30^\circ$  and  $60^\circ$  are considered and the mean was calculated separately for the two radii. For the position attribute also the results of a binaural model are plotted, see Section 5 for a detailed description of the used model.

iar type of signal. The castanets sample was chosen because it contains very strong transients which emphasize potential pre-echo artifacts.

The real-time convolution was performed using the *SoundScape Renderer* (SSR) [18, 19], an open-source software environment for spatial sound reproduction, running in *binaural room synthesis* (BRS) mode. The SSR convolves the input signal in real-time with that pair of impulse responses corresponding to the instantaneous head orientation of the test subject as captured by a Polhemus Fastrack tracking system. Due to the internal processing of the SSR the switching between different audio inputs leads to a smooth cross-fade with raised-cosine shaped ramps. AKG K601 headphones were used and a compensation of their transfer function was applied [20].

### 3.3. Procedure

After an introduction and a short training phase the participants started with the experiment containing speech or castanets as source material. In a second session the other source material was used. The subject was presented with a screen containing 9 sliders representing the following conditions: ref.,  $\varphi_{10m R_2}$ ,  $\varphi_{4m R_1}$ ,  $\varphi_{1.8m R_1}$ ,  $\varphi_{1.8m R_2}$ ,  $\varphi_{0.75m R_1}$ ,  $\varphi_{0.75m R_2}$ ,  $\varphi_{0.3m R_1}$ ,  $\varphi_{0.3m R_2}$ , where  $\varphi \in [0^\circ, 30^\circ, 60^\circ]$  was con-

stant for the given screen. At the top of the screen the attribute pair to rate was presented. A screenshot of the GUI can be seen in Fig. 5. After the subject had rated all conditions, the next attribute pair for the same conditions was presented. Thereby the order of the conditions attached to the slider and the appearance of the attribute pairs was randomized. This procedure was repeated three times, once for each angle  $\varphi$ . For an angle of  $\varphi = 0^\circ$  the attribute pair *left vs. right* was omitted.

## 4. RESULTS

For the purpose of this paper only the two attribute pairs *few artifacts vs. many artifacts* and *left vs. right* are considered. The left of Fig. 6 presents the mean of all subjects and both source materials (speech and castanets) for the rating of the *few artifacts vs. many artifacts* attribute pair. In addition, the means are calculated about all listener positions without the  $0^\circ$  position for the speech material. The figure hence presents the strength of artifacts only dependent on the array length. The  $0^\circ$  position for the speech material was removed as an outlier. At this position and with speech as source material there existed no or only few audible artifacts in the received signal. On the other hand there was coloration and four of the six subjects seemed to

have rated the coloration and not audible artifacts. It can be seen in the figure that the two shortest arrays resulted in as few artifacts as the reference condition. The 10 m array exposed strong artifacts as it was expected from the previous experiment. The 1.8 m array and the 4 m array caused few more artifacts than the reference condition. A one-way ANOVA shows that the mentioned three groups are different ( $p < 0.05$ ) from each other and not different within each group.

On the right hand side of Fig. 6 the results for the attribute pair *left* vs. *right* is presented. The means for the arrays were calculated for the  $30^\circ$  and  $60^\circ$  listener positions, but once for each radius. 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 two different source materials speech and castanets showed only significant differences for the 10 m array (only the  $30^\circ$  and  $60^\circ$  positions were regarded due to the outlier for the  $0^\circ$  speech condition).

## 5. DISCUSSION

As mentioned in section 2.2, the pre-echoes of focused sources lead to strong artifacts. The arrival time of the first pre-echo at the listener position can be reduced by using a shorter sub-array. This leads to a reduction of audible artifacts, as the results for the attribute pair *few artifacts* vs. *many artifacts* have shown. The two smallest sub-arrays with a length of 0.3 m and 0.75 m are rated to have the same amount of artifacts as the single loudspeaker reference.

All three loudspeaker arrays with a length of  $L < 2$  m have arrival times of the first pre-echo of under 1 ms. This means that they fall in a time window in which the summing localization is at work and no single echo should be audible. The artifacts audible for the array with  $L = 1.8$  m are therefore due to a comb-filter structure of the frequency spectrum of the signal. This structure resulted from the temporal delay and superposition procedure of the loudspeakers, see (4).

However, there are new problems due to a shorter array. The main problem is the localization of the

focused source. Fig. 6 shows a relation between the array length and the localization: the shorter the further left the focused source is perceived. This result means that the summing localization of the pre-echoes and the desired signal can not be the only reason for the wrong perception of the location. As can be seen in Fig. 4, the summing localization would propose the perceived location to be more to the left for the 1.8 m array than for the 0.3 m array. Therefore it is likely that the diffraction due to the short array length introduces wrong binaural cues like interaural time delay (ITD) and/or interaural level difference (ILD).

To verify this a binaural model after Lindemann [21, 22] was applied using the parameters from the original paper. This model analyses the ITD of a given signal with a cross-correlation in different frequency bands. Inherently it analyses also the ILD via its contralateral inhibition which shifts the resulting peak of the cross-correlation. The centroid of the mean cross-correlation (mean about frequency bands) was used as the model output for the perceived direction. The results were scaled to have the same order of magnitude as the rating results. The result is plotted in Fig. 6 together with the rating results. As the rating data, the model data are also means about the two listener directions  $30^\circ$  and  $60^\circ$ . The model results show a quite good congruence with the measured data. Only for the two large arrays clear deviations are visible. This is due to the fact that the first pre-echo time for those arrays is smaller than  $-1$  ms ( $-4$  ms for the 4 m array and  $-16$  ms for the 10 m array) and the perceived direction is dominated by the precedence effect, which is not accounted for in the binaural model. Also, a split image may be audible for the 10 m array – one source coming from the front and one from the left [5]. This indicates that the diffraction caused by small array sizes leads to the creation of wrong binaural cues at the used listener positions.

## 6. CONCLUSIONS/FURTHER WORK

In practice the perception of focused sources in WFS is not free from artifacts. The time-reversal technique used in the synthesis of focused sources causes the appearance of pre-echoes arriving from every single loudspeaker at the listener position before the desired focused source signal. It has been shown

that the number of pre-echoes and the time of arrival of the first pre-echo can be reduced by using a subset (sub-array) of the existing loudspeakers. In an experiment six subjects rated the perception of focused sources for different linear array lengths and different listener positions using the attribute pairs *few artifacts vs. many artifacts* and *left vs. right*. The results proof the reduction of artifacts by using fewer loudspeakers. On the other hand, the perception of a focused source as a small source located at a given position is limited with shorter sub-arrays. The diffraction causes a diffuse/wider focus point and the perceived location of the focused source is bound towards the side of the loudspeaker array for positions of the listener with  $x \neq x_s$ . This was also verified by a binaural model for the sub-arrays. To model the localization of focused sources for large arrays, the model still needs to be extended to account also for the precedence effect and split images due to very early arrival times of the pre-echoes.

In order to apply the proposed method in a real life scenario the localization problem has to be solved. In [23] the use of first order image sources has been proposed to enhance the localization of the focused source, which has been shown to enlarge the possible view angle of the listener for focused sources [24]. It has to be tested if it can be applied to avoid wrong localization due to additional binaural cues arriving from a diffraction by small sub-arrays.

## 7. REFERENCES

- [1] D. de Vries. *Wave Field Synthesis*. AES Monograph. AES, New York, 2009.
- [2] E. N. G. Verheijen. *Sound Reproduction by Wave Field Synthesis*. Ph.D. thesis, Delft University of Technology, 1997.
- [3] H. Wittek. *Perceptual differences between wave-field synthesis and stereophony*. Ph.D. thesis, University of Surrey, October 2007.
- [4] S. Spors, H. Wierstorf, M. Geier and J. Ahrens. Physical and perceptual properties of focused sources in Wave Field Synthesis. In *127<sup>th</sup> AES Convention*. October 2009.
- [5] M. Geier et al. Perception of focused sources in Wave Field Synthesis. In *128<sup>th</sup> AES Convention*. May 2010.
- [6] A. Berkhout, D. de Vries and P. Vogel. Acoustic control by Wave Field Synthesis. *JASA*, 93(5):2764–2778, May 1993.
- [7] E. W. Start. Application of curved arrays in Wave Field Synthesis. In *100<sup>th</sup> AES Convention*. May 1996.
- [8] J. Ahrens and S. Spors. On the secondary source type mismatch in Wave Field Synthesis employing circular distributions of loudspeakers. In *127<sup>th</sup> AES Convention*. October 2009.
- [9] S. Spors, R. Rabenstein and J. Ahrens. The theory of Wave Field Synthesis revisited. In *124<sup>th</sup> AES Convention*. May 2008.
- [10] H. Wallach, E. B. Newman and M. R. Rosenzweig. The precedence effect in sound localization. *American Journal of Psychology*, 57:315–336, 1949.
- [11] G. A. Kelly. *The Psychology of Personal Constructs*. Norton, New York, 1955.
- [12] J. Berg and F. Rumsey. Identification of quality attributes of spatial audio by repertory grid technique. *J. Audio Eng. Soc.*, 54(5), May 2006.
- [13] J. Blauert. *Spatial Hearing*. The MIT Press, Cambridge, Massachusetts, 1997.
- [14] A. Lindau and S. Weinzierl. FABIAN – An instrument for the software-based measurement of binaural room impulse responses in multiple degrees of freedom. In *24. Tonmeistertagung (VDT International Convention)*. November 2006.
- [15] M. Geier, J. Ahrens and S. Spors. Binaural monitoring of massive multichannel sound reproduction systems using model-based rendering. In *NAG/DAGA International Conference on Acoustics*. March 2009.
- [16] Audio examples. <http://audio.qu.tu-berlin.de/?p=392>.
- [17] S. Spors and J. Ahrens. Analysis and improvement of pre-equalization in 2.5-dimensional Wave Field Synthesis. In *128<sup>th</sup> AES Convention*. May 2010.

- [18] The SoundScape Renderer.  
<http://tu-berlin.de/?id=ssr>.
- [19] M. Geier, J. Ahrens and S. Spors. The SoundScape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods. In *124<sup>th</sup> AES Convention*. May 2008.
- [20] Z. Schärer and A. Lindau. Evaluation of equalization methods for binaural signals. In *126<sup>th</sup> AES Convention*. May 2009.
- [21] W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *JASA*, 80(6):1608–1622, December 1986.
- [22] Auditory Modeling Toolbox.  
<http://audio.qu.tu-berlin.de/?p=356>.
- [23] T. Caulkins, E. Corteel and O. Warusfel. Wave Field Synthesis interaction with the listening environment, improvements in the reproduction of virtual sources situated inside the listening room. In *6<sup>th</sup> DAFX conference*. London, September 2003.
- [24] R. Oldfield, I. Drumm and J. Hirst. The perception of focused sources in Wave Field Synthesis as a function of listener angle. In *128<sup>th</sup> AES Convention*. May 2010.