Perception of Focused Sources in Wave Field Synthesis

HAGEN WIERSTORF,[1] *AES Student Member*, **ALEXANDER RAAKE,**[1] *AES Member*, **MATTHIAS GEIER**[2],
(hagen.wierstorf@tu-berlin.de)
AND

SASCHA SPORS,[2] *AES Member*

[1]*Assessment of IP-based Applications, T-Labs, Technische Universität Berlin, Ernst-Reuter-Platz 7,*
10587 Berlin, Germany
[2]*Signal Theory and Digital Signal Processing, Institute of Communications Engineering, Universität Rostock,*
R.-Wagner-Str. 31, 18119 Rostock/Warnemünde, Germany

Wave Field Synthesis (WFS) allows virtual sound sources to be synthesized that are located between the loudspeaker array and the listener. Such sources are known as focused sources. Due to practical limitations related to real loudspeaker arrays, such as spatial sampling and truncation, there are different artifacts in the synthesized sound field of focused sources. In this paper we present a listening test to identify the perceptual dimensions that are associated with these artifacts. Two main dimensions were found, one describing the amount of perceptual artifacts and the other one describing the localization of the focused source. The influence of the array length on these two dimensions is evaluated further in a second listening test. A binaural model is used to model the perceived location of focused sources found in the second test and to analyze dominant localization cues.

0 INTRODUCTION

Wave Field Synthesis (WFS) [1] is one of the most prominent high-resolution sound field synthesis methods used and studied nowadays. Unlike traditional stereophonic techniques, it offers the potential of creating the impression of a virtual point source located inside the listening area—between the loudspeakers and the listeners [2]. These sources are known as *focused sources*, according to their strong relation to acoustic focusing [3].

The physical theory of WFS assumes a spatially continuous distribution of loudspeakers denoted as secondary sources. However, in practical implementations, the secondary source distribution will be realized by a limited number of loudspeakers placed at discrete positions. This implies a spatial sampling and truncation process that typically leads to spatial aliasing and truncation artifacts in the sound field [4], depending on the position of the virtual source and the position of the listener. For focused sources these artifacts are of special interest, as they may become clearly audible, especially for large loudspeaker arrays [5,6].

In this paper the physical properties of focused sources are studied (Section 1)—as well as their perceptual properties. It is shown that the synthesis of focused sources may be associated with a number of perceptually relevant artifacts, that will be increasingly audible the larger the WFS system, if no further action is taken. The perceptual properties are investigated by performing a formal listening test (Section 2) using the Repertory Grid Technique (RGT).

In a second listening test (Section 3), the two main perceptual dimensions, identified in the first listening test, were rated for loudspeaker arrays of different lengths. To further analyze the results, the results for the localization of focused sources are compared with the output of a binaural model (Section 4).

To create reproducible test conditions, all tests were conducted with a “virtual” WFS system realized by dynamic (head-tracked) binaural resynthesis, presented to the participants via headphones.

1 THEORY

The theory of WFS was initially derived from the *Rayleigh integrals* for a linear secondary source distribution [1]. With this source distribution it is possible to synthesize a desired two-dimensional sound field in one of the half planes defined by the linear secondary source distribution.

The sound 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. Then, the synthesized sound field is given by

$$P(\mathbf{x},\omega) = -\int_{-\infty}^{\infty} D_{\mathrm{2D}}(\mathbf{x}_0,\omega)\; G_{\mathrm{2D}}(\mathbf{x}-\mathbf{x}_0,\omega)\; dx_0\;, \qquad (1)$$

where $\mathbf{x} = (x, y)$ with $y > 0$, $\mathbf{x}_0 = (x_0, 0)$ and $\omega = 2\pi f$ with temporal frequency f. The functions D_{2D} and G_{2D} denote the secondary source driving signal and the sound field emitted by a secondary source, respectively. In WFS the driving function is given as

$$D_{\mathrm{2D}}(\mathbf{x}_0,\omega) = 2\frac{\partial}{\partial y} S(\mathbf{x},\omega)\,|_{\mathbf{x}=\mathbf{x}_0}\;, \qquad (2)$$

where $S(\mathbf{x},\omega)$ denotes the sound field of the desired virtual source.

The sound field $G_{\mathrm{2D}}(\mathbf{x}-\mathbf{x}_0,\omega)$ of a secondary source can be interpreted as the field of a line source intersecting the xy-plane at position $\mathbf{x}_0$. For practical applications only secondary sources with the field of a point source (G_{3D}) are available in most cases. Hence a dimensional mismatch of a three-dimensional secondary source for two-dimensional synthesis has to be considered. This leads to a so called two-and-a-half-dimensional driving function that applies an amplitude correction to reduce this mismatch. Using the far-field approximation $\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_0| \gg 1$ the following relationship between the field of a line source and the field of a point source can be derived [7]:

$$\underbrace{\frac{i}{4}\; H_0^{(1)}\!\left(\tfrac{\omega}{c}|\mathbf{x}-\mathbf{x}_0|\right)}_{G_{\mathrm{2D}}(\mathbf{x}-\mathbf{x}_0,\omega)} \approx \sqrt{2\pi\frac{ic}{\omega}|\mathbf{x}-\mathbf{x}_0|}\;\; \underbrace{\frac{1}{4\pi}\frac{e^{i\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|}}_{G_{\mathrm{3D}}(\mathbf{x}-\mathbf{x}_0,\omega)}\;, \qquad (3)$$

where $H_0^{(1)}$denotes the Hankel function of first kind and zeroth order.

This results in the so called 2.5D driving function, which is given with Eq. (3) as

$$D_{\mathrm{2.5D}}(\mathbf{x}_0,\omega) = \sqrt{\frac{ic}{\omega}}\;\underbrace{\sqrt{2\pi|\mathbf{x}_{\mathrm{ref}}-\mathbf{x}_0|}}_{g_0}\; D_{\mathrm{2D}}(\mathbf{x}_0,\omega)\;, \qquad (4)$$

where g_0 is chosen in such a way that it is a constant and does not depend on x. In this case the amplitude is correct at a line positioned at $|\mathbf{x}_{\mathrm{ref}} - \mathbf{x}_0| = y_{\mathrm{ref}}$ parallel to the loudspeaker array [2].

The synthesized sound field is given by

$$P(\mathbf{x},\omega) = -\int_{-\infty}^{\infty} D_{\mathrm{2.5D}}(\mathbf{x}_0,\omega)\; G_{\mathrm{3D}}(\mathbf{x}-\mathbf{x}_0,\omega)\; dx_0\;. \qquad (5)$$

A reformulation of the theory based on the *Kirchhoff-Helmholtz integral* revealed that also arbitrary convex distributions can be employed [8,9]. 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 [1,10].

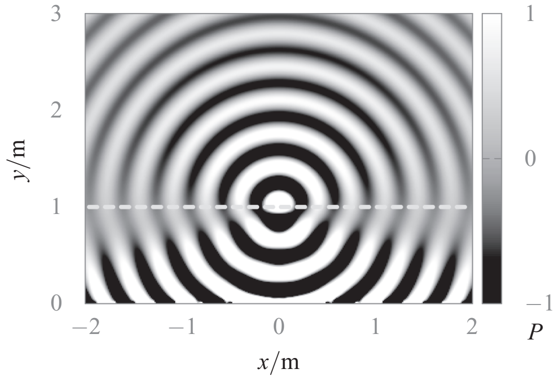

Fig. 1. Simulation of the sound field $P(\mathbf{x},\omega)$ for a monochromatic focused source with a frequency of $f = 1000$ Hz, located at $\mathbf{x}_{\mathrm{s}} = (0, 1)$ m. A continuous secondary source distribution with a length of $L \to \infty$ is placed on the x-axis. The amplitude of the sound field is clipped at $|P| = 1$. In the area below the gray dashed line, the sound field is converging to the focus. Above the line it is diverging from the focus.

1.1 Focused Sources

In WFS, sound fields can be described by using source models to calculate the driving function. For example, to synthesize the sound field of a human speaker positioned at point $\mathbf{x}_{\mathrm{s}}$, the model of a point source positioned at $\mathbf{x}_{\mathrm{s}}$ can be used. The point source is then driven by the speech signal of the human.

For the synthesis of a focused source, a sound field is desired that converges toward a focal point and diverges after passing it. This is known from the techniques of time-reversal focusing and can be reached by using a point sink as source model for the converging part of the focused source sound field [3]. In order to derive an efficient implementation of the driving function not a point sink, but a line sink with a spectral correction given by Eq. (4) is used as source model for $y < y_{\mathrm{s}}$ [5]

$$S(\mathbf{x},\omega) = S_{\mathrm{s}}(\omega)\sqrt{\frac{\omega}{ic}}\frac{i}{4}\; H_0^{(2)}\!\left(\tfrac{\omega}{c}|\mathbf{x}-\mathbf{x}_{\mathrm{s}}|\right), \qquad (6)$$

where $S_{\mathrm{s}}(\omega)$ denotes the frequency spectrum of the line sink, $H_0^{(2)}$the Hankel function of second kind and zeroth order, and $\mathbf{x}_{\mathrm{s}} = (x_{\mathrm{s}}, y_{\mathrm{s}})$ the position of the focused source. Using (2) and (4), this leads to the driving function

$$\begin{aligned} D_{\mathrm{2.5D}}(\mathbf{x}_0,\omega) = {} & -S_{\mathrm{s}}(\omega)\; g_0 \frac{i\omega}{2c}\frac{y_0 - y_{\mathrm{s}}}{|\mathbf{x}_0 - \mathbf{x}_{\mathrm{s}}|} \\ & \times H_1^{(2)}\!\left(\tfrac{\omega}{c}|\mathbf{x}_0-\mathbf{x}_{\mathrm{s}}|\right)\;, \end{aligned} \qquad (7)$$

where $H_1^{(2)}$denotes the Hankel function of second kind and first order. In Fig. 1 , the simulated sound field $P(\mathbf{x},\omega)$ for a monochromatic focused source located at $\mathbf{x}_{\mathrm{s}} = (0, 1)$ is shown. The sound field converges for $0 < y < 1$ m toward the position of the focused source and diverges for $y > 1$ m, which defines the listening area for the given focused source position. In addition, a phase jump occurs

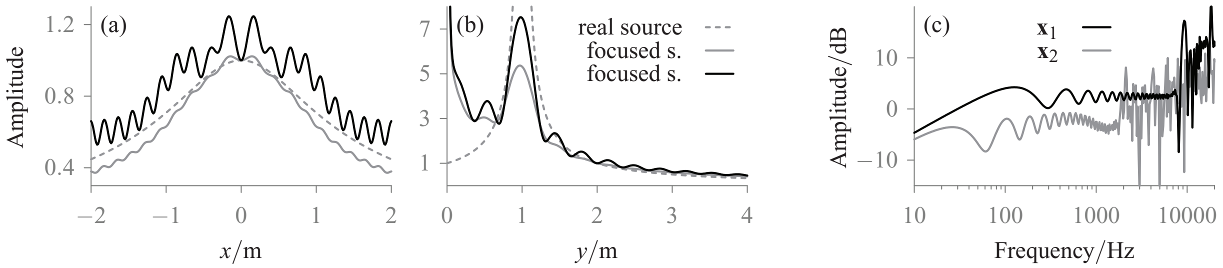

Fig. 2. Graph (a) and (b) show the amplitude distribution for a focused source positioned at $\mathbf{x}_s = (0, 1)$ m generated by the driving function Eq. (8) in black and by the driving function without large argument approximation Eq. (7) in gray. In addition the amplitude distribution for a real source located at the same position as the focused source is shown with the dashed line. (a) is parallel to the x-axis at $y = 2$ m and (b) parallel to the y-axis at $x = 0$ m. Graph (c) presents the frequency response of the focused source Eq. (8) at two listener positions $\mathbf{x}_1 = (0, 2)$ m and $\mathbf{x}_2 = (2, 2)$ m. The frequency of the monochromatic source in (a), (b) was $f = 1000$ Hz. Parameters: $L = 1000$ m, $\Delta x_0 = 0.15$ m, $y_{\text{ref}} = 2$ m.

at $y = y_s$, which is well known for focal points [11]. Due to the limitation of the used source model S to the area with $y < y_s$ the evanescent part of the focused source sound field is not identical for $y > y_s$ with that of a point source located at $\mathbf{x}_s$. In order to reproduce this part correctly, very high amplitudes in the region $y < y_s$ are needed because of the exponential decay of the evanescent waves along the y-axis. This can be shown by using the spectral division method to create the sound field [12].

Eq. (7) can be related to the traditional formulation of the driving function used in WFS [2, Eq. (2.30)] by replacing the Hankel function by its large-argument approximation [7]

$$D_{2.5\text{D}}(\mathbf{x}_0, \omega) \approx S_s(\omega)\; g_0 \sqrt{\frac{i\omega}{2\pi c}}\, \frac{y_0 - y_s}{|\mathbf{x}_0 - \mathbf{x}_s|^{\frac{3}{2}}}\; e^{-i\frac{\omega}{c}|\mathbf{x}_0 - \mathbf{x}_s|}\;, \tag{8}$$

where g_0 is explicitly given in [2].

When the driving function Eq. (8) is transformed into the time domain, it is given as

$$\begin{aligned} d_{2.5\text{D}}(\mathbf{x}_0, t) = s_s(t) * h(t) \;*\; & \frac{g_0}{2\pi}\, \frac{y_0 - y_s}{|\mathbf{x}_0 - \mathbf{x}_s|^{\frac{3}{2}}} \\ & \times \delta\left(t - \tfrac{|\mathbf{x}_0 - \mathbf{x}_s|}{c}\right), \end{aligned} \tag{9}$$

where c is the speed of sound, δ the delta function and $h(t)$ denotes the inverse Fourier transform

$$h(t) = \mathcal{F}^{-1}\left\{\sqrt{\frac{i\omega}{c}}\right\}\;. \tag{10}$$

It is easy to see that this driving function can be implemented very efficiently by filtering the virtual source signal $s_s(t)$ with the so-called pre-equalization filter $h(t)$ and weighting and delaying the pre-filtered signal for every secondary source appropriately.

In order to verify the influences of the applied large argument approximations, the amplitude distribution of the synthesized sound field can be studied. Fig. 2 shows the amplitude for a focused source positioned at $\mathbf{x}_s = (0, 1)$ m along two axes. The amplitude of a real source positioned at $\mathbf{x}_s$ is shown for reference as a gray dashed line. The black line shows the amplitude distribution for a focused source synthesized by the classical 2.5D driving function Eq. (8), the gray line for a focused source synthesized with the 2.5D driving function Eq. (7). It can be observed that for the focused source given by Eq. (7) the amplitude diverges from that of a real point source due to the 2.5D synthesis (see Fig. 2a). Interestingly, the amplitude distribution of the focused source synthesized by the classical driving function Eq. (8) has a more correct amplitude distribution for distances farther away from the focal point. But its additional large argument approximation reinforces the ripples of the amplitude distribution, that exist due to the 2.5D approximation.

1.2 Loudspeakers as Secondary Sources

Theoretically, when an infinitely long continuous secondary source distribution is used, no other errors than an amplitude mismatch due to the 2.5D synthesis are expected in the sound field [5].

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. In principle both can be described in terms of diffraction theory (see, e.g., [11]). Unfortunately, as a consequence of the dimensions of loudspeaker arrays and the large range of wave lengths in sound as compared to light, most of the assumptions made to solve diffraction problems in optics are not valid in acoustics. To present some of the basic properties for truncated and sampled secondary source distributions, simulations of the sound field are made and interpreted in terms of basic diffraction theory where possible.

1.2.1 Spatial Sampling

The spatial sampling that is equivalent to the diffraction by a grating only has consequences for frequencies greater than the aliasing frequency

$$f_{\text{al}} \geq \frac{c}{2\Delta x_0}\;, \tag{11}$$

where Δx_0 describes the distance between the secondary sources [5]. In general, the aliasing frequency is position dependent (cf., [8, Eq. 5.17]), but an analytical solution

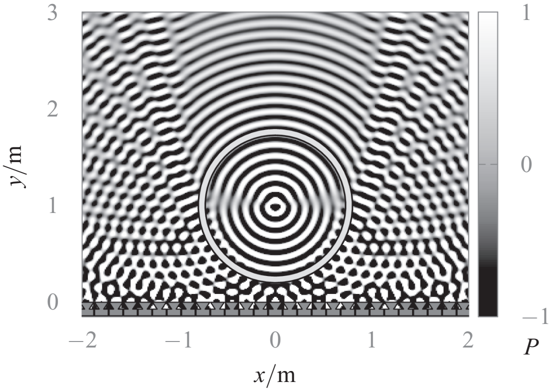

Fig. 3. Simulation of a sound field $P(\mathbf{x}, \omega)$ for a focused source with a frequency of $f = 3000$ Hz. A sampled secondary source distribution with a distance of $\Delta x_0 = 0.15$ m between the single sources was used. The amplitude of the sound field is clipped at $|P| = 1$. The gray circle indicates a region without aliasing given by Eq. (12). Parameters: $\mathbf{x}_\mathrm{s} = (0, 1)$ m, $L = 1000$ m, $y_\mathrm{ref} = 2$ m.

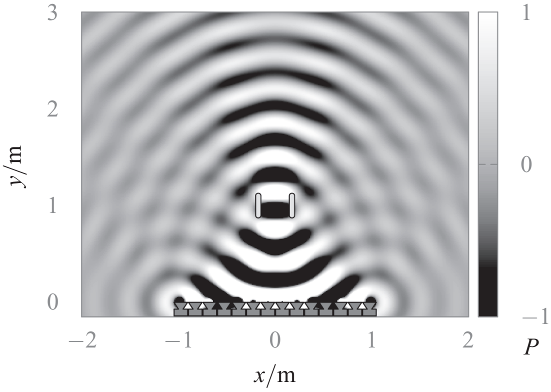

Fig. 5. Simulation of a sound field $P(\mathbf{x}, \omega)$ for a focused source with a frequency of $f = 1000$ Hz generated with a secondary source distribution of length $L = 2$ m. The amplitude of the sound field is clipped at $|P| = 1$. The two gray lines indicate the size of the focus after Eq. (13). Parameters: $\mathbf{x}_\mathrm{s} = (0, 1)$ m, $\Delta x_0 = 0.15$ m, $y_\mathrm{ref} = 2$ m.

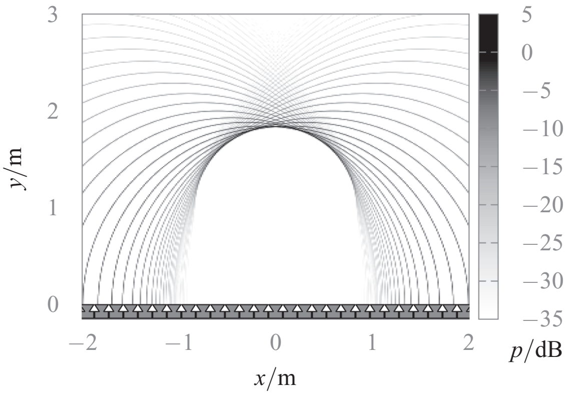

Fig. 4. Simulation of the sound field $p(\mathbf{x}, t)$ for a broadband focused source for $t = 2.4$ ms after the wave front has passed the focus. Parameters: $\mathbf{x}_\mathrm{s} = (0, 1)$ m, $L = 100$ m, $\Delta x_0 = 0.15$ m, $y_\mathrm{ref} = 2$ m.

for focused sources is not available at the moment. Fig. 3 shows the monochromatic sound field for a focused source with a frequency of 3000 Hz generated by a secondary source distribution with $\Delta x_0 = 0.15$ m. Clear interference artifacts are visible in the sound field, but there also is an area around the focus where no interference took place. This is a unique property of focused sources. The size of the area depends on the frequency f and becomes smaller with higher frequencies. It can be empirically described by a circle with a radius of (cf., [13])

$$r_\mathrm{al} = \frac{y_\mathrm{s} c}{f\,\Delta x_0}\,. \tag{12}$$

The area calculated using the parameters applicable to Fig. 3 is indicated by a gray circle.

Fig. 4 shows a snapshot of the sound field of a broadband focused source to examine more implications of the spatial sampling artifacts. Every single loudspeaker is sending a broadband signal according to Eq. (9). If no spatial aliasing occurs, the signals cancel each other out in the listening area, with the exception of the desired wave front of the focused source. In case of spatial aliasing and for frequencies above the aliasing frequency, the cancellation does not occur and a bunch of additional wave fronts reach a given listener position before the desired wave front characterizing the focused source. These additional wave fronts are very critical for the perception of focused sources, as we will see in Section 2. The additional wave fronts also add energy to the signal, which can be seen from the spectrum shown in Fig. 2c. The Figure depicts the frequency responses for two different listener positions $\mathbf{x}_1$ and $\mathbf{x}_2$, which are associated with different aliasing frequencies. Obviously, above the aliasing frequency the magnitude of the frequency response increases. This can be avoided by using the pre-equalization filter only until the aliasing frequency [5], which has the shortcoming of introducing position-dependency in the filter.

1.2.2 Truncation

The spatial truncation of the loudspeaker array leads to further restrictions. On the one hand, the listener area becomes smaller with a smaller array, which is shown in Fig. 5. The listening area can be approximated by the triangle that is spanned for $y > y_\mathrm{s}$ by the two lines coming from the edges of the loudspeaker array and crossing the position of the focused source. Another problem is that a smaller loudspeaker array introduces diffraction in the sound field. The loudspeaker array can be seen as a single slit that causes a diffraction of the sound field propagating through it. This can be described in a way equivalent to the phenomenon of edge waves as shown by Sommerfeld and Rubinowicz (see [11] for a summary). The edge waves are two additional spherical waves originating from the edges of the array, which can be softened by applying a tapering window [14]. The resulting diffraction pattern adds artifacts to the desired sound field. For example, the interaural level differences (ILD) will not be correct due to diffraction minima and maxima as shown in Fig. 9.

The diffraction also leads to a wavelength-dependent widening of the focus. The width of the focus at its position y_s can be defined as the distance between the first minima in the diffraction pattern and is given by

$$\Delta_s = 2|y_s - y_0| \tan\left(\sin^{-1}\tfrac{\lambda}{L}\right), \tag{13}$$

where Δ_s is the width of the focus, L the array length, y_s the y-position of the focused source and y_0 the y-position of the loudspeaker array. This formula is based on the assumption of Fraunhofer diffraction near the focus [11, 8.3 Eq. 34]. In Fig. 5, the calculated size of the focal point is indicated by the gray lines.

2 PERCEPTIONAL DIMENSIONS OF FOCUSED SOURCES

In the last section different artifacts in the synthesized sound field of a focused source were discussed. These artifacts raise the question whether and how they will affect the perception of focused sources. In this section a listening test is presented that investigates this perceptual impact.

It was shown in the previous section that the aliasing frequency f_{al} due to the spatial sampling introduced by the loudspeakers depends on the listening position. In addition, the diffraction due to truncation of the loudspeaker array depends on the size of the array. Hence, different array sizes and listener positions have to be considered in a respective listening test.

We have further shown that the time reversal technique used to create focused sources—in combination with spatial aliasing—leads to additional wave fronts arriving at the listener position from different directions and before the desired wave front. This is a situation that only occurs with such synthetic sound fields but not in case of everyday listening in natural environments. As a consequence, it can be expected that the description of the related perceptual effects requires multidimensional attributes in the perceptional domain. To address this issue, the Repertory Grid Technique (RGT) was used to identify perceptually relevant attributes [15,16]. With this method, in a first step each participant creates her/his own set of attributes and in a second step uses respective attribute scales for rating their perception. No attributes are provided by the experimenter, and, thus, the test subject has complete freedom in the choice of attributes.

A more detailed discussion of this first experiment was presented in [6].

2.1 Method

2.1.1 Stimuli

The tests were conducted with a "virtual" WFS system realized by dynamic binaural re-synthesis [17] using headphones. See Fig. 6 for a sketch of the geometry of the employed virtual WFS configurations. Two linear loudspeaker arrays with a length L of 4 m and 10 m and a loudspeaker spacing of $\Delta x_0 = 0.15$ m were synthesized. To handle truncation, a squared Hann tapering window with a length of $0.15L$ on both ends of the arrays was used. The impulse responses of the individual virtual loudspeakers of the array were obtained by interpolating and weighting a database of head-related impulse responses (HRIRs) of the FABIAN manikin [18] to the required positions and distances of the loudspeaker. Every virtual loudspeaker was then weighted and delayed according to the driving function Eq. (9) for a given focused source and listener position, summed up and filtered with the pre-equalization filter from Eq. (10). The result was a pair of HRIRs of the desired WFS array producing the given focused source for a given listener position. For the dynamic binaural re-synthesis these pairs of HRIRs had to be calculated for all possible head orientations of the listener ranging from −180° to 180° in 1° steps.

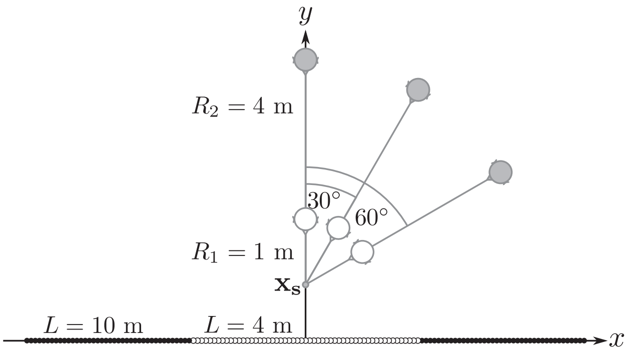

Fig. 6. Geometry of the experiment. Every listener was positioned at six different positions given by the head symbols. The focused source was always positioned at $\mathbf{x}_s = (0, 1)$ m, and the center of the used loudspeaker array was always positioned at $\mathbf{x} = (0, 0)$ m.

As discussed in Section 1.2, the aliasing frequency f_{al} depends on the listener position, therefore the WFS pre-equalization filter was calculated separately for each simulated listening position. Coloration introduced by an improper choice of the pre-equalization filter was not part of the investigation and should be avoided.

For both arrays, three different listener positions on a given circle around the focused source were used. The radius was $R_1 = 1$ m for the short array and $R_2 = 4$ m for the long array. Three different listener angles of $\varphi = 0°, 30°$, and $60°$ were applied for both array lengths (see Fig. 6). These six configurations will be referred to as $0^\circ_{4\text{m}}, 30^\circ_{4\text{m}}, 60^\circ_{4\text{m}}, 0^\circ_{10\text{m}}, 30^\circ_{10\text{m}}$, and $60^\circ_{10\text{m}}$. In all conditions, the focused source was located directly in front of the listener. A seventh reference condition ("ref.") was created, which consisted of a single sound source located at the position of the focused source. This was realized by directly using the corresponding HRIRs from the database.

As audio source signals, anechoic recordings of speech and of castanets were chosen. The speech signal was an 8 s sequence of three different sentences uttered by a female speaker. The castanets recording was 7 s long. The levels of the stimuli were normalized to the same loudness by informal listening by the authors for all conditions. The real-time convolution of these signals with the impulse responses for the WFS arrays was performed using the *SoundScape*

Renderer (SSR[1]) [19], an open-source software environment for spatial audio reproduction. The SSR performed a real-time convolution of the input signal with that pair of impulse responses corresponding to the instantaneous head orientation of the test subject as measured by a Polhemus Fastrak tracking system. In the SSR, switching between different audio signals is realized using a smooth cross-fade with raised-cosine shaped ramps. AKG K601 headphones were used, and the transfer functions of both earphones were compensated by appropriate filters [20]. Audio examples are available as supplementary material.[2]

2.1.2 Participants

In order to generate a large amount of meaningful attributes, test subjects with experience in analytically listening to audio recordings were recruited. The experiment was conducted with 12 Tonmeister students (3 female, 9 male, between 21 and 33 years old). The participants had between 5 years and 20 years of musical education, and all of them had experience with listening tests. They had normal hearing levels and were financially compensated for their effort.

2.1.3 Procedure

The participants received written instructions explaining their tasks in the two phases of the experiment.

The RGT procedure consisted of two parts, the *elicitation phase* and the *rating phase*. In the elicitation phase, groups of three conditions (*triads*) were presented to the test subject. The subjects were able to switch between them by pressing a corresponding button and could listen to each stimulus as long as they wanted. For each triad, the subject had to decide which two of the three stimuli were more similar and had to describe the characteristic that made them similar, and in which characteristic they were different from the third stimulus (which should be the opposite of the first property). If there were competing aspects, only the strongest one should be taken into account. One attribute pair per triad had to be specified, and two more could optionally be given if the test subject perceived several different properties. A screenshot of the used test GUI is shown in [6].

After a short training phase, every participant had to execute this procedure 12 times (using 12 different triads). Ten of the 12 triads resulted from a complete set of triads from the five conditions ref., $30^\circ_{4\mathrm{m}}$, $60^\circ_{4\mathrm{m}}$, $30^\circ_{10\mathrm{m}}$, and $60^\circ_{10\mathrm{m}}$. The two additional triads were (ref., $0^\circ_{4\mathrm{m}}$, $0^\circ_{10\mathrm{m}}$) and ($0^\circ_{4\mathrm{m}}$, $30^\circ_{4\mathrm{m}}$, $0^\circ_{10\mathrm{m}}$). These two have been chosen in order to consider the additional, very similar conditions together, to get attributes for the small differences between them. Complete triads for only five conditions have been chosen because of the time-consuming procedure (a complete set of triads for 7 conditions would have resulted in 35 triads).

The presented triads were the same for all participants, however, the order of the triads and the order of conditions within a triad was alternated over all participants based on a *Latin Square* design.

After the elicitation phase the participants took a break. During this time, the test supervisor removed repetitions of attribute pairs for constructing the attribute list used in the second RGT test phase.

For this rating phase in each trial one previously elicited attribute pair was displayed on top of the screen. Below, the seven conditions could be played back and had to be rated on corresponding sliders. The ratings were saved on a continuous scale ranging from -1.0 to 1.0. Once a rating was collected for all conditions, the test subject was able to switch to the next screen, a procedure repeated until all elicited attribute pairs were used. Before the actual test, a training phase had to be completed for two rating screens.

In the second session, which was in most cases done on another day, the elicitation and rating phase was repeated with the respective other source stimulus. Half of the subjects were presented with the speech sample in the first session and the castanets in the second session, and vice versa for the other half.

2.2 Results

One of the main results of the experiment were the elicited attribute pairs. They reflect the range of perceptual similarities and differences among the conditions. Their number was different between subjects, ranging from 6 to 17 pairs for individual subjects. The most prominent choices were artifacts (e.g., clean sound vs. chirpy, squeaky, unnatural sound) and localization (left vs. center). For the latter, it has to be noted that the focused source was always positioned straight in front of the listener. Attributes describing artifacts were provided by 10 of the 12 subjects for castanets and by 9 subjects for speech. Localization-related attributes were given by 7 subjects for castanets, and 5 subjects for speech. Other common attributes were related to coloration (original vs. filtered, balanced vs. unbalanced frequency response), distance (far vs. close) and reverberation (dry vs. reverberant). All elicited attributes were originally collected in German and were translated to English for this paper.

The ratings of the attributes can be used to identify the underlying dimensions that best describe the perception of focused sources. This was done using a principal component analysis (PCA) for individual subjects. For all subjects, two principal components could be identified as the main dimensions of the perceptual space. These dimensions can explain 90% of the variance for castanets and 97% for speech, respectively.

This also allows to determine the positions of the different conditions in the resulting perceptual space. Fig. 7 shows the PCA results for one individual subject for the speech and castanets, respectively. The PCA results for another subject can be found in [6]. The black dots represent the different conditions in this two-dimensional perceptual space. The gray lines show the arrangement of elicited attribute pairs in this space. From Fig. 7 it can be seen that for both castanets and speech the first principal component

[1] http://tu-berlin.de/?id=ssr

[2] http://audio.qu.tu-berlin.de/?p=625

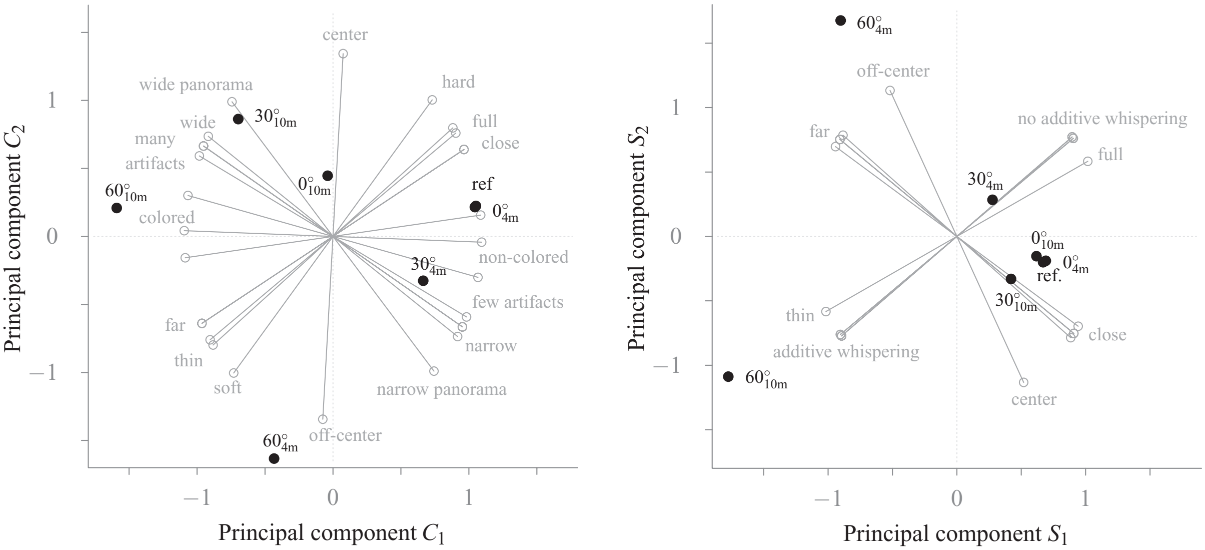

Fig. 7. Principal component analysis for castanets (left) and speech (right) for one single subject. The black points indicate the position of the conditions given in the two-dimensional space determined by the two given components for each stimulus type. The gray lines show the arrangement of the attribute pairs in these two dimensions.

C_1 resp. S_1 can be interpreted as a mixture of the amount of artifacts and the distance, and the second principal component C_2 resp. S_2 as the localization of the source. Considering individual conditions, it can be observed that the 10 m loudspeaker array was rated to produce artifacts in the perception of the focused source, while the artifact-related ratings for the 4 m array are more or less the same as for the reference condition. For the longer array, the amount of artifacts depends on the listener position, with the highest rating of artifacts at the lateral position $60^\circ_{10\text{m}}$. The perception of a wrong (off-center) direction is most distinct for the lateral positions of the shorter array, with the condition $60^\circ_{4\text{m}}$ as the most prominent case. Both lateral positions ($\phi = 60^\circ$) were perceived as more off-center than the other ones. Furthermore, it can be noted that the perceptual deviation from the reference condition occurs for more conditions for the castanets than for the speech stimuli.

2.3 Discussion

The results show that the amount of perceived artifacts depends on the length of the loudspeaker array and the position of the listener, being worse for a larger loudspeaker array and a more lateral position of the listener. This is due to the fact that for a larger loudspeaker array more additional wave fronts arrive before the desired one for the focused source. The perceived amount of artifacts further increases with the degree of lateral displacement of the listener relative to the focused source (see Fig. 6). The explanation for this finding can be illustrated using Fig. 8. Here, the direction of incidence of the desired (black arrow) and of the aliasing-related wave fronts for the focused sources are shown for the different listener and array configurations. Note that the arrows point into the direction of incidence from the listener perspective. The starting point of an arrow indicates the position in time of the wave front, and the

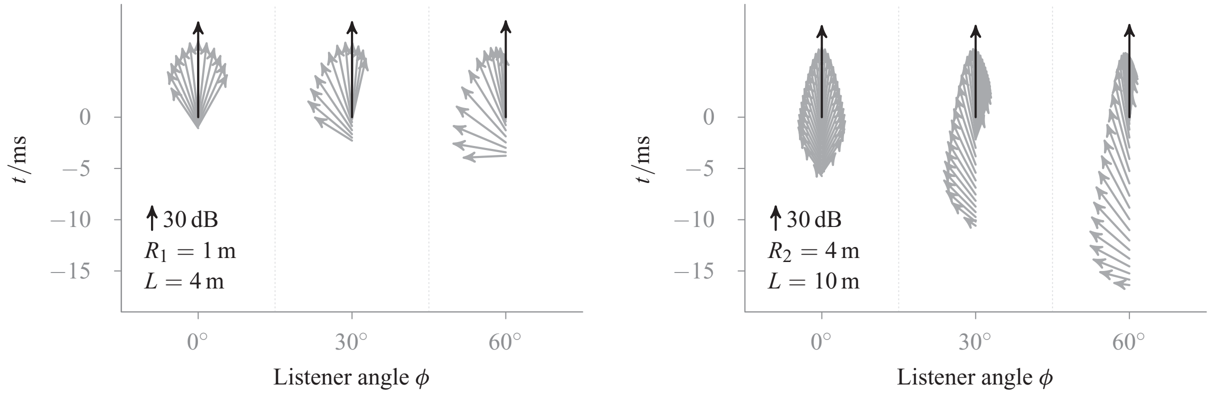

Fig. 8. Direction, amplitude and time of appearance of wave fronts for the 4 m loudspeaker array (left) and the 10 m array (right). The results are shown for different angles ϕ at a radius $R_1 = 1$ m (left) and $R_2 = 4$ m (right). The arrows are pointing toward the direction from which the wave fronts arrive. The time of appearance is given by the starting point of the arrow. Note that the (temporal) starting points lie closely together for listener positions close to the contributing loudspeakers of the array, and are further apart when the configuration involves larger distances from the loudspeakers. The length of the arrow is proportional to the amplitude of the wave front in dB. The length of the arrow in the legend corresponds to an amplitude of 30 dB. The black arrows indicate the desired wave fronts, the gray arrows aliasing-related wave fronts.

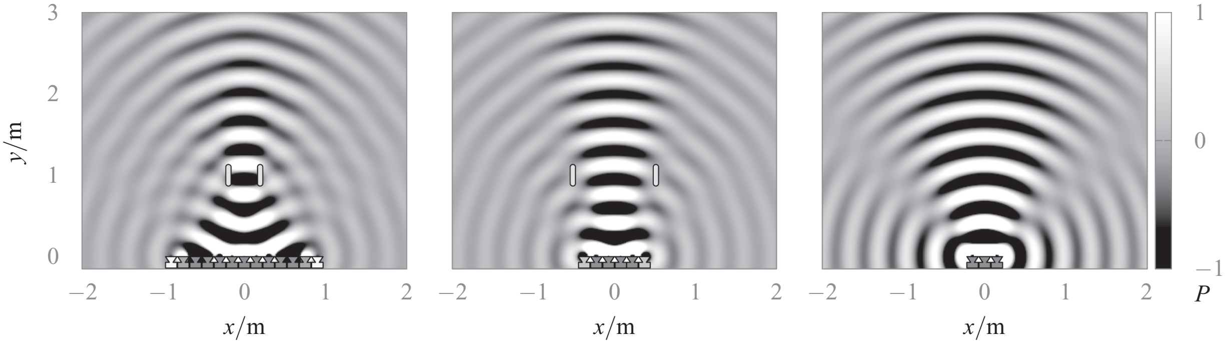

Fig. 9. Simulation of a sound field $P(\mathbf{x}, \omega)$ for a focused source synthesized with different array lengths, indicated by the loudspeaker symbols. The array lengths are, from left to right: 1.8 m, 0.75 m, 0.3 m. The amplitude of the sound fields is clipped at $|P| = 1$. The two parallel gray lines indicate the size of the focal point calculated using Eq. (13). A tapering window of 15% of the array length was applied at the end of the arrays, indicated by the different loudspeaker colors. Parameters: $\mathbf{x}_s = (0, 1)$ m, $f = 1000$ Hz, $\Delta x = 0.15$ m, $y_{\text{ref}} = 2$ m.

length of the arrow is proportional to its amplitude in dB. It is obvious that the larger the used loudspeaker array, the earlier the occurrence of additional wave fronts, and the higher their amplitude. This is due to the fact that every single loudspeaker adds a wave front. For a given array, the number of wave fronts will be the same regardless of the lateral listener position but the time of arrival of the first wave front will be earlier. This can be explained by the fact that the listener is positioned closer to one end of the loudspeaker array in this case. The loudspeakers at the ends of the array had to be driven as the first ones in order to create a focused source in the middle of the loudspeaker array, resulting in the significantly earlier incidence of the wave fronts from the loudspeakers close to the listener.

The results show a dependency of the perceived direction on the listener position and the array size. The condition $60^{\circ}_{4\text{m}}$ was perceived as most from the left. The perceived direction can be explained by the additional wave fronts, too. The conditions with $\phi = 0^{\circ}$ were perceived from the same direction for both array lengths as the reference condition in front of the listener. For these conditions, the additional wave fronts have no effect on the perceived direction, because they arrive at the listener position symmetrically from all directions (Fig. 8). For the lateral conditions, the first wave front will come mainly from the left side of the listener. Due to the precedence effect [21] this can lead to localization of the sound to the direction of the (first) wave front. For the 10 m array, the perceived direction is different from that of the shorter array. Most of the subjects localized the sound in the same direction as the reference. However, a few subjects indicated that they had heard more than one sound source—one high-frequency chirping source from the left and a cleaner source in front of them. This can be explained with the echo threshold related with the precedence effect, which means that further wave fronts that follow the first one with a lag larger than the echo threshold are perceived as an echo [22].

In order to verify this hypothesis, an experiment has been performed to examine the localization dominance for this kind of time-delayed wave front pattern [23]. Here, an approximated time of 8 ms between the first wave front and the desired one has been identified to be the threshold until which the perceived direction is dominated by the first wave front. This is in conformance with the results for the large array.

3 INFLUENCE OF ARRAY LENGTH ON THE PERCEPTION

As mentioned in Section 1.2, truncation of the loudspeaker array leads to two opposite effects. On the one hand, a smaller array leads to fewer additional wave fronts and reduces the perception of artifacts as shown in the last section. On the other hand, a smaller array leads to stronger diffraction of the sound field and therefore a smaller possible listening area as well as wrong binaural cues. Fig. 9 shows the wave fields for a focused source created at $\mathbf{x}_s = (0, 1)$ m using arrays of three different lengths, $L = 1.8$ m, $L = 0.75$ m, and $L = 0.3$ m, with the same fixed inter-loudspeaker distance of $\Delta x_0 = 0.15$ m as used previously. Hence, the arrays consist of 13, 6, and 3 loudspeakers. The figure illustrates that the focal point gets very large and even disappears for short arrays. This is indicated by the gray lines, which show the size of the focus as calculated using Eq. (13). For the short array with $L = 0.3$ m the equation is not defined for the given frequency of $f = 1000$ Hz, because $\lambda/L > 1$. In this case, no focal point exists, and the source is located near the position of the loudspeaker array, as can be seen in the rightmost graph of Fig. 9. In addition, the maxima and minima of the diffraction pattern introduce wrong interaural level differences (ILDs) at different listener positions. Note that these wrong binaural cues may deviate for a planar array due to the absence of the amplitude error of 2.5D WFS.

To verify if there is an array length for which the artifacts are not audible, and the wrong binaural cues are negligible as well, a listening test was conducted that included the three shorter array lengths as shown in Fig. 9 together with the two array lengths used in the first experiment. In the test two attribute pairs were rated by the subjects, one regarding the audible artifacts and one regarding the perceived position of the focused source. The middle of the array was again chosen at $\mathbf{x} = (0, 0)$ in order to have a symmetric loudspeaker distribution around the x-position of the focused source. A more detailed discussion of the experiment is presented in [24].

3.1 Method

3.1.1 Stimuli

The tests were conducted with a similar geometry and the same source materials as described in Section 2.1. The same listener positions as in Fig. 6 were used, now using the array sizes $L = 10$ m, 4 m, 1.8 m, 0.75 m, and 0.3 m. Again, the two different radius values $R_1 = 1$ m and $R_2 = 4$ m were used; only R_1 for the 4 m array, only R_2 m for the 10 m array, and both values for the three other array sizes. Altogether nine different conditions were created, again including the reference condition. Audio examples are available as supplementary material.[3]

3.1.2 Participants

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

3.1.3 Procedure

After an introduction and a short training phase with a violin piece as source material, one half of the participants started the first session presenting speech, the other half presenting castanets. In a second session, the speech and castanets source materials were switched between the groups. The subjects were presented with a screen containing nine sliders representing the nine different conditions. At the top of the screen, one of the two attribute pairs few artifacts vs. many artifacts and left vs. right were presented. After a subject had rated all conditions, the next attribute pair was presented for the same conditions. 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 all the array conditions assessed in case of each listening angle ϕ. For the listening angle of $\phi = 0°$, the attribute pair left vs. right was omitted.

3.2 Results

Fig. 10 presents the mean ratings over all subjects, all listener positions, and both source materials (speech and castanets) for the attribute pair few artifacts vs. many artifacts. Hence, the only independent variable is the strength of artifacts plotted on the x-axis. The 0° position for the speech material resulted as an outlier, and was not considered for the plot. At this position and with speech as source material, artifacts are only little audible. On the other hand, there is the coloration introduced by the spatial sampling and independent of the fact that focused sources were realized. An interview with the subjects revealed that four of them have rated this coloration rather than the targeted audible artifacts. It can be seen in the figure that the results for the different loudspeaker arrays build three different groups. The two shortest arrays resulted in as few artifacts as the reference condition. The 10 m array was found to lead to strong artifacts, as it was expected from the previous experiment. The amount of artifacts caused by the 1.8 m and the 4 m array are positioned between these two groups. A one-way ANOVA shows that the mentioned three groups are statistically different ($p < 0.05$) from each other and not different within each group.

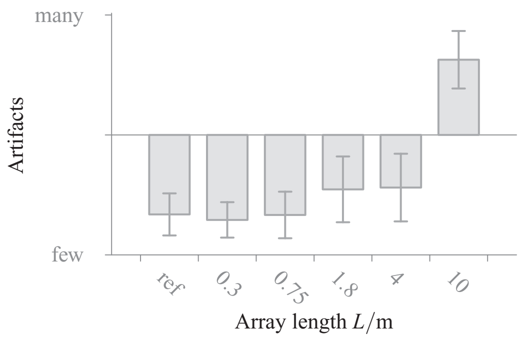

Fig. 10. Mean and variance for the rating of the attribute pair few artifacts vs. many artifacts plotted over the condition. The mean is calculated over all subjects, source materials and the different listener positions.

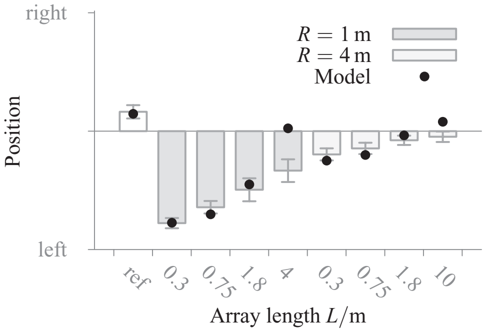

Fig. 11. Mean and variance for the rating of the attribute pair left vs. right plotted over the conditions. The mean is calculated over the different subjects, source materials, and the two listener angles 30° and 60°. The results are presented separately for the two radius values $R_1 = 1$ m and $R_2 = 4$ m. The black points are the results obtained using the Lindemann model, see Section 4.

In Fig. 11, the results for the attribute pair left vs. right are presented. The means for the arrays were calculated over the 30° and 60° conditions but once for each radius indicated by the two different shades of gray. It can be seen that the reference condition (arriving from straight ahead of the listener) was rated to come slightly from the right side. All other conditions came from the left side, where shorter arrays and smaller radii lead to a rating further to the left.

The two different source materials speech and castanets showed significant differences only for the 10 m array and the 30° and 60° positions, with more artifacts perceivable for the castanets stimuli.

3.3 Discussion

As mentioned in Section 2, the appearance of additional wave fronts due to spatial aliasing leads to strong artifacts

[3] http://audio.qu.tu-berlin.de/?p=625

for focused sources. The arrival time of the first wave front at the listener position can be reduced by using a shorter loudspeaker array. This leads to a reduction of audible artifacts, as shown by the results for the attribute pair few artifacts vs. many artifacts. The two smallest 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 wave front of below 5 ms. This means that they fall in a time window in which the precedence effect should work and no echo should be audible. The artifacts audible for the array with $L = 1.8$ m are therefore due to a comb-filter shaped ripple in the frequency spectrum of the signal, as a result of the temporal delay and superposition procedure of the loudspeakers, see (5) and (9).

However, there are other problems related with a shorter array. The main problem is the localization of the focused source. Fig. 11 shows a relation between array length and localization: the shorter the array, the further left the focused source is perceived. This result implies that the precedence effect cannot be the only reason for the wrong perception of the location. For a shorter array, too, the first wave front arrives from the loudspeaker at the edge of the array. This loudspeaker will be positioned less far to the left for a shorter array than for a longer array. Therefore, it is likely that the diffraction due to the short array length introduces wrong binaural cues, namely a wrong ILD.

4 MODELING THE PERCEPTION OF FOCUSED SOURCES

To verify the findings for localization perception, a binaural model according to Lindemann [25] was applied using the parameters from the original paper. The model is part of the auditory modeling toolbox.[4] This model analyzes binaural cues like the interaural time difference (ITD) and the interaural level difference (ILD). The ITD is calculated for a given signal by a cross-correlation ψ_n in different frequency bands n. The spacing of the frequency bands is 1 ERB. The model further analyzes the ILD via a contralateral inhibition mechanism, which leads to a shift of the resulting peak of the cross-correlation. This incorporates the ILD and ITD values into a single direction estimation, which has to be done manually in other binaural models—for example [27,28].

As a measure for the perceived direction the mean of the cross-correlation about the frequency bands $n = 5 \ldots 40$ was first calculated by

$$\psi = \frac{1}{36} \sum_{n=5}^{40} \psi_n \ . \tag{14}$$

Then the centroid d of ψ was used as the model output for the perceived direction

$$d = \frac{\int \tau \psi(\tau)\, d\tau}{\int \psi(\tau)\, d\tau} \ , \tag{15}$$

where τ is the time of the cross-correlation. The predicted localization was scaled to achieve the same order of magnitude as the rating results. The results are plotted in Fig. 11, together with the subject ratings. Like the rating data, the model data are also depicted as the mean over the two listener directions 30° and 60°. The model results show a quite good agreement with the test data. This indicates that the perceived localization is dominated by wrong binaural cues due to the diffraction artifacts for truncated arrays. Only for the two large arrays, clear deviations of the modeled results are visible. For these large arrays with $L = 4$ m and $L = 10$ m the time of arrival between the first and the last wave front is in the region of 3 ms to 15 ms, which suggests that the precedence effect plays a role in explaining the perceived direction. The model does not account for the precedence effect, which explains the deviation of its prediction from the subject data for the large arrays.

5 SUMMARY AND CONCLUSIONS

In practice, the creation of focused sources with WFS is not free from perceptual artifacts. The time-reversal technique used in the synthesis of focused sources causes the appearance of additional wave fronts arriving at the listener position from every single loudspeaker before the desired focused source signal. An experiment using the RGT method was carried out to identify attribute pairs that are able to describe the resulting perception of focused sources. The most dominant attribute pairs were those regarding audible artifacts, coloration, and the position of the focused source.

In a second experiment with different linear array lengths and different listener positions using only the attribute pairs few artifacts vs. many artifacts and left vs. right, it could be shown that artifacts could be reduced by using fewer loudspeakers. On the other hand, the perception of a focused source as a distinct source located at a given position is limited when using shorter arrays. The diffraction causes a wider focal point, and the localization of the focused source is disturbed. This was verified using a binaural model. The model results also indicated that the perceived localization for the small arrays is due to the wrong binaural cues introduced by the diffraction pattern. The results for the long arrays indicated that the precedence effect has to be considered for the perceived direction of focused sources created by these arrays.

This study shows that the usage of focused sources in WFS with typical linear loudspeaker arrays has to be handled with care. The appearance of additional wave fronts before the desired one introduces different artifacts in the perception of the focused source as compared to the synthesis of a point source located behind the loudspeaker array. In addition to these artifacts, the strong dependency of the spatial aliasing frequency on the position of the listener

[4] http://amtoolbox.sf.net [26]

for focused sources will introduce even more coloration in a real setup using a fixed pre-equalization filter for the whole listening area. This was not the case in this study, because the pre-equalization filter was chosen adaptive for the different listener positions in order to investigate only the effect of the additional wave fronts.

For further studies it could be interesting how the perception of focused sources behaves for WFS with multi-actuator panels [29]. These panels lead to a more chaotic distribution of the additional wave fronts that cause the perceptual artifacts for focused sources. Moreover it would be beneficial to investigate the perception of focused sources in other sound field synthesis techniques like near-field compensated higher order Ambisonics [30] or numerical methods—for example [31].

6 ACKNOWLEDGMENT

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

7 REFERENCES

[1] A. Berkhout, D. de Vries and P. Vogel, “Acoustic Control by Wave Field Synthesis,” *J. Acoust. Soc. Am.*, vol. 93, no.5, pp. 2764–2778 (1993).

[2] E. N. G. Verheijen, *Sound Reproduction by Wave Field Synthesis*, Ph.D. thesis, Delft University of Technology, 1997.

[3] S. Yon, M. Tanter and M. Fink, “Sound Focusing in Rooms: The Time-Reversal Approach,” *J. Acoust. Soc. Am.*, vol. 113, no. 3, pp. 1533–1543 (2003).

[4] H. Wittek, *Perceptual Differences between Wavefield Synthesis and Stereophony*, Ph.D. thesis, University of Surrey, 2007.

[5] S. Spors, H. Wierstorf, M. Geier and J. Ahrens, “Physical and Perceptual Properties of Focused Sources in Wave Field Synthesis,” presented at the *127th Convention* of the Audio Engineering Society (2009 Oct), convention paper 7914.

[6] M. Geier et al., “Perception of Focused Sources in Wave Field Synthesis,” presented at the *128th Convention* of the Audio Engineering Society (2010 May), convention paper 8069.

[7] E. G. Williams, *Fourier Acoustics* (Academic Press, San Diego, 1999).

[8] E. W. Stuart, “Application of Curved Arrays in Wave Field Synthesis,” presented at the *100th Convention* of the Audio Engineering Society (1996 May), convention paper 4143.

[9] J. Ahrens and S. Spors, “On the Secondary Source Type Mismatch in Wave Field Synthesis Employing Circular Distributions of Loudspeakers,” presented at the *127th Convention* of the Audio Engineering Society (2009 Oct), convention paper 7952.

[10] S. Spors, R. Rabenstein and J. Ahrens, “The Theory of Wave Field Synthesis Revisited,” presented at the *124th Convention* of the Audio Engineering Society (2008 May), convention paper 7358.

[11] M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, New York, 1999).

[12] S. Spors and J. Ahrens, “Reproduction of Focused Sources by the Spectral Division Method,” *4th International Symposium on Communications, Control and Signal Processing* (2010 Mar).

[13] S. Spors and J. Ahrens, “Efficient Range Extrapolation of Head-Related Impulse Responses by Wave Field Synthesis Techniques,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (2011).

[14] P. Vogel, *Application of Wave Field Synthesis in Room Acoustics*, Ph.D. thesis, University of Technology, 1993.

[15] G. A. Kelly, *The Psychology of Personal Constructs* (Norton, New York, 1955).

[16] J. Berg and F. Rumsey, “Spatial Attribute Identification and Scaling by Repertory Grid Technique and Other Methods,” presented at *16th AES International Conference*, Spatial Sound Reproduction (1999 Mar), conference paper 16-005.

[17] A. Lindau, T. Hohn and S. Weinzierl, “Binaural Resynthesis for Comparative Studies of Acoustical Environments,” presented at the *122th Convention* of the Audio Engineering Society (2007 May), convention paper 7032.

[18] A. Lindau and S. Weinzierl, “FABIAN—An Instrument for the Software-Based Measurement of Binaural Room Impulse Responses in Multiple Degrees of Freedom,” *24th VDT International Convention* (2006 Nov.).

[19] M. Geier, J. Ahrens and S. Spors, “The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods,” presented at the *124th Convention* of the Audio Engineering Society (2008 May), convention paper 7330.

[20] Z. Schärer and A. Lindau, “Evaluation of Equalization Methods for Binaural Signals,” presented at the *126th Convention* of the Audio Engineering Society (2009 May), convention paper 7721.

[21] H. Wallach, E. B. Newman and M. R. Rosenzweig, “The Precedence Effect in Sound Localization,” *Am. J. Psychol.*, vol. 57, pp. 315–336 (1949).

[22] J. Blauert, *Spatial Hearing* (The MIT Press, Cambridge, Massachusetts, 1997).

[23] H. Wierstorf, S. Spors, and A. Raake, “Die Rolle des Präzedenzeffektes bei der Wahrnehmung von Räumlichen Aliasingartefakten bei der Wellenfeldsynthese,” *DAGA German Conference on Acoustics* (2010 Mar).

[24] H. Wierstorf, M. Geier and S. Spors, “Reducing Artifacts of Focused Sources in Wave Field Synthesis,” presented at the *129th Convention* of the Audio Engineering Society (2010 Nov), convention paper 8245.

[25] W. Lindemann, “Extension of a Binaural Cross-Correlation Model by Contralateral Inhibition. I. Simulation of Lateralization for Stationary Signals,” *J. Acoust. Soc. Am.*, vol. 80, no.6, pp. 1608–1622 (1986).

[26] P. Søndergaard et al., “Towards a Binaural Modelling Toolbox,” *FORUM ACUSTICUM* (2011 June).

[27] E. Blanco-Martin, F. J. Casajús-Quirós, J. J. Gómez-Alfageme and L. I. Ortiz-Berenguer, “Objective Measurement of Sound Event Localization in Horizontal and Median Planes,” *J. Audio Eng. Soc.*, vol. 59, pp. 124–136 (2011 Mar.).

[28] M. Dietz, S. D. Ewert and V. Hohmann, "Auditory Model Based Direction Estimation of Concurrent Speakers from Binaural Signals," *Speech Communication*, vol. 53, no.5, pp. 592–605 (2011).

[29] B. Pueo, J. J. López, J. Escolano and L. Hörchens, "Multiactuator Panels for Wave Field Synthesis: Evolution and Present Developments," *J. Audio Eng. Soc.*, vol. 58, pp. 1045–1063 (2010 Dec.).

[30] J. Ahrens and S. Spors, "Spatial Encoding and Decoding of Focused Virtual Sound Sources," *Ambisonics Symposium*, Graz, Austria (2009 June).

[31] M. Kolundzija, C. Faller and M. Vetterli, "Reproducing Sound Fields Using Mimo Acoustic Channel Inversion," *J. Audio Eng. Soc.*, vol. 59, pp. 721–734 (2011 Oct.).

THE AUTHORS

Hagen Wierstorf

Matthias Geier

Alexander Raake

Sascha Spors

Hagen Wierstorf is a Ph.D. student in the Assessment of IP-Based Applications group at Technical University (TU), Berlin, Germany. He received his Diplom in Physics from Carl-von-Ossietzky-Universität Oldenburg, Germany in 2008. Since 2009 he is working at the Telekom Innovation Laboratories at the TU.

●

Matthias Geier is a Ph.D. student at the Institute of Communications Engineering at University of Rostock, Germany. He received his Diplom in electrical engineering/sound engineering at University of Technology and University of Music and Dramatic Arts in Graz, Austria in 2006. From 2008–2012 he was working at the Telekom Innovation Laboratories.

●

Alexander Raake is an Assistant Professor and heads the group for Assessment of IP-based Applications at Deutsche Telekom Labs, TU Berlin. From 2005 to 2009 he was a senior scientist at the Quality and Usability Lab of Deutsche Telekom Labs, TU Berlin. From 2004 to 2005, he was a Postdoctoral Researcher at LIMSI-CNRS in Orsay, France. From the electrical engineering and information technology faculty of the Ruhr-Universität Bochum, he obtained his doctoral degree (Dr.-Ing.) in January 2005, with a book on the speech quality of VoIP. After his graduation in 1997 he took up research at the Technical University in Lausanne (EPFL) on ferroelectric thin films. Before, he studied electrical engineering in Aachen (RWTH) and Paris (ENST/Télécom). Since 1999, he has been involved in the standardization activities of the International Telecommunication Union (ITU-T) on transmission performance of telephone networks and terminals, where he currently acts as a Co-Rapporteur for question Q.14/12 on monitoring models for audiovisual services.

●

Sascha Spors is full professor and heads the group signal theory and digital signal processing of the Institute of Communications Engineering, University of Rostock. From 2006 to 2012 he was senior research scientist at the Quality and Usability Lab of Deutsche Telekom Labs, TU Berlin, where he headed the audio technology group. From 2001 to 2005 he was a member of the research staff at the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Erlangen. He received the Dr.-Ing. degree with distinction from the University of Erlangen-Nuremberg, Erlangen, Germany, in 2006. Dr. Spors holds several patents and has authored or co-authored more than 150 papers in journals and conference proceedings. He is member of the IEEE Audio and Acoustic Signal Processing Technical Committee and chair of the AES Technical Committee on Spatial Audio.