Audio Engineering Society

Convention Paper

Presented at the 133rd Convention
2012 October 26–29 San Francisco, CA

Localization of a virtual point source within the listening area for Wave Field Synthesis

Hagen Wierstorf[1], Alexander Raake[1], Sascha Spors[2]

[1] *Telekom Innovations Laboratories, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany*

[2] *Institute of Communications Engineering, Universität Rostock, Richard-Wagner-Straße 31, 18119 Rostock, Germany*

Correspondence should be addressed to Hagen Wierstorf (hagen.wierstorf@tu-berlin.de)

ABSTRACT

One of the main advantages of Wave Field Synthesis (WFS) is the existence of an extended listening area contrary to the sweet spot in stereophony. At the moment there is only little literature available on the actual localization properties of WFS at different points in the listening area. One reason is the difficulty to place different subjects reliable at different positions. This study systematically investigates the localization performance for WFS at many positions within the listening area. To overcome the difficulty to place subjects, the different listening positions and loudspeaker arrays were simulated by dynamic binaural synthesis. In a pre-study it was verified that this method is suitable to investigate the localization performance in WFS.

1. INTRODUCTION

The task of sound reproduction is to achieve a convincing auditory scene in the mind of one or more listeners. The reproduced scene does not have to be a physical mirror image of a given sound scene to achieve this mission. Due to its physical limitations, stereophony has always tried to create a plausible and artistically convincing auditory scene. Other approaches started from a physical point of view and extended Snows idea of the acoustical curtain [1]. These are known as sound field synthesis techniques nowadays and Wave Field Synthesis (WFS) is one of its prominent representatives. Theoretical in WFS the sound field in a volume can be reproduced exactly given that the distance between adjacent loudspeakers on the boundary of the volume is infinitely small. In practice this doesn't hold, due to the finite distance between the loudspeakers. As a consequence, the generated sound field contains spatial sampling. The interesting question is how these sampling artifacts are perceived from a listener within the listening area.

This study focuses on the influence of the spatial sampling artifacts on the localization of a single virtual point source. For stereophony it is well known that the localization is disturbed outside of a small area, the sweet spot. If the listener is sitting outside of this area the localization is bounded to the nearest loudspeaker [2]. For WFS localization should be equally well throughout the listening area. Results of listening tests show that the localization of a single virtual point source is not or only slightly impaired for a loudspeaker spacing of less or equal to 22 cm [3, 4, 5]. All these tests were conducted with a linear loudspeaker array and a central listening position of the listener. Verheijen [6] varied the position of the source, which is equivalent to moving the listener around for a loudspeaker array which is much longer than the distance between the listener and the array. He used an array length of 2.53 m and the listener was positioned 3 m in front of the array. Hence the results can not directly be used to analyze the listening area. His study reported strong deviations for the localization of virtual point sources from real sources only for virtual sources which were placed beside the array. For a loudspeaker spacing of 0.11 m no difference to real sources were found and for a spacing of 0.22 m a slight increase of the root mean square (RMS) error (see [7] for a definition) of 0.5° was found. The present study focuses on the localization ability at different positions within the listening area for different loudspeaker spacings.

It is very difficult to realize a listening test at different positions for different loudspeaker array configurations. One would have to assure that all listeners are exactly positioned and they would have to move between the different positions during the test. To overcome these difficulties binaural synthesis of the WFS systems and different positions was applied. This is realized by applying HRTFs to synthesize the acoustic paths between the loudspeakers and the listener ears. If this is done for every possible head orientation of the listener, dynamic binaural synthesis can be realized. In this case the head orientation of a listener is measured by a head tracker and the HRTFs are switched accordingly. Results from the literature show that localization performance is not affected by the use of dynamic binaural synthesis in general. But the results can be influenced by the use of non-individualized HRTFs [8] and the used pointing method [9, 10]. Hence in a first experiment we investigated the accuracy of our method, which uses non-individualized HRTFs [11] and lets the subjects point there head towards the direction they perceive the auditory event [12]. This is assisted by a laser pointer mounted on the subjects head to indicate the look direction in order to overcome undershoots [13].

2. GENERAL METHOD

2.1. Listeners

Eleven adult listeners were recruited for both experiments (6 male, 5 female; aged 21–33 years, mean age 28.6 years). Four of them had prior experiences with psychoacoustic testing and wave field synthesis. The subjects were financially compensated for their effort.

2.2. Apparatus

Stimuli were digitally generated at a sampling rate of 44.1 kHz. Octave and the Sound Field Synthesis Toolbox [14] were used to compute binaural impulse responses for the binaural synthesis of WFS and single sources. For this purpose, HRTFs measured in an anechoic chamber [11] were combined to represent the binaural impulse response from the virtual source reproduced by WFS to the listeners ears. The SoundScape Renderer [15] was used to convolve the binaural impulse responses in real-time and to generate the loudspeaker signals. Pure Data (Pd) was used to play back the source signals and to control the experiment. The PC was equipped with a RME HDSP MADI card and the digital to analog conversion was done by a Cream Ware A16 converter. The listeners wore AKG K601 headphones and as loudspeakers 19 Fostex PM0.4 were arranged as a linear array with a spacing of 0.15 m between them. For the experiment only 11 of the loudspeakers were used, which are filled in Fig. 1. The head movement of the listener was tracked by a Polhemus Fastrak head tracker. The SoundScape Renderer was switching the HRTFs used for the dynamic binaural synthesis according to the actual orientation of the listener as provided by the head tracker. A laser pointer was mounted on the headphones. The listener was positioned within a acoustically damped listening room, 1.5 m in front of the loudspeaker array. An acoustical transparent curtain was placed in

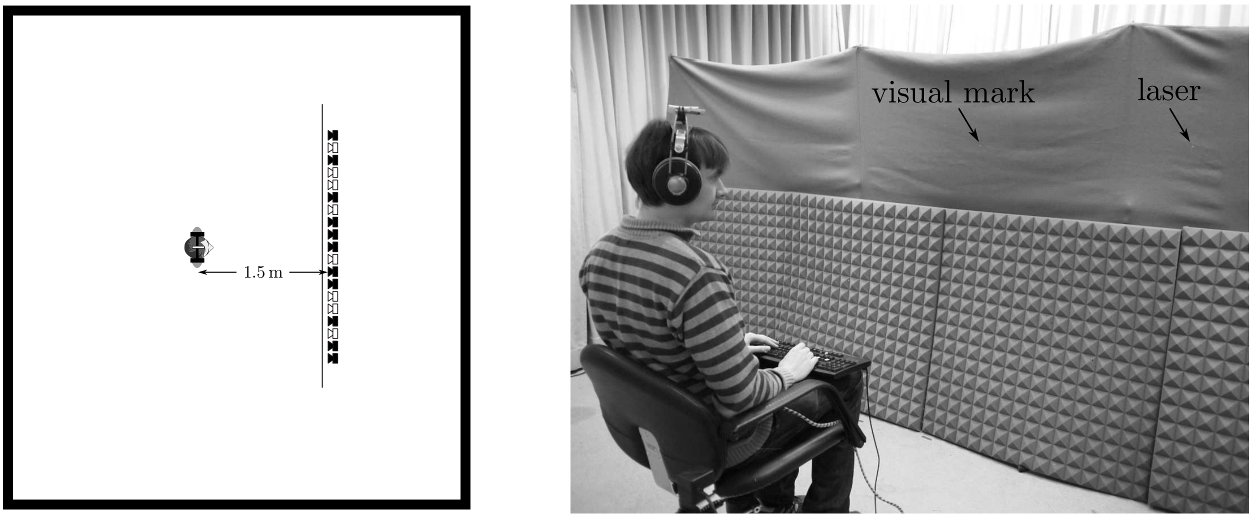

Fig. 1: Sketch of the experimental setup (left) and picture of a subject during the experiment (right). Only the filled loudspeakers were used in the first experiment. The light in the room was dimmed during all experiments.

between the loudspeakers and the listener. An illustration of the setup and a picture is shown in Fig. 1. The orientation and position of the subjects during the experiment was recorded with the head tracker.

2.3. Stimuli

As audio material, Gaussian white noise pulses with a duration of 700 ms and a pause of 300 ms between them were generated. A single pulses was windowed with a Hanning window of 20 ms length at the start and the end. The single pulses are independent white noise signals. The signal was furthermore bandbass filtered with a fourth order butterworth filter between 125 Hz and 20000 Hz. The signal with a total length of 100 s was looped in the experiment. For the headphone reproduction the stimuli were convolved with the corresponding HRTFs. It was assured that the stimuli had the same level for all conditions.

2.4. Procedures

The subjects sat on a heavy chair, wearing the headphones with the laser pointer and had a keyboard on their knees (see Fig. 1). They were instructed to point, with the laser pointer, into the horizontal direction where they perceived the auditory event by turning the head. If vertical deviations were perceived, these should be ignored. Once they were sure sure to point into the right direction, they were asked to hit the enter key. The subjects' head orientation is calculated as the mean over the following 10 values obtained from the head tracker, which corresponds to a time window of 90 ms. After the key press, the next trial started instantaneously, which implies that the subject started with the head orientation always from the last perceived position, and not from a fixed point. The subjects were instructed that they could turn their head freely if they were unsure about the direction of the sound.

At the beginning of every session, a calibration is carried out. First, the loudspeaker at 0° was active, and the subject had to look into the respective direction in order to calibrate the head tracker. In a second step, the subject was indicated to point towards a given visual mark on the curtain. The second step formed a connection between the head tracker orientation and the coordinate system in the room. After the calibration step, the illumination in the room was dimmed and the experiment started.

3. VERIFICATION OF THE LOCALIZATION METHOD

3.1. Experimental procedure

Three different acoustic conditions formed the experiment, *loudspeaker*, *room HRTFs*, and *anechoic*

	Loudspeaker	HRTF
unsigned error /°	2.4 ±0.59	2.0 ±0.56
standard deviation /°	2.2 ±0.15	3.8 ±0.30
time / s	3.5 ±0.65	5.5 ±1.72

Table 1: Mean values and 95% confidence intervals over all speaker positions and subjects.

HRTFs. For the first one, the noise pulses were played through one of the loudspeakers. For the other two conditions the sound was presented via headphones. The anechoic HRTFs consisted of the ones which were used also in the WFS localization test, and are described in [11]. The room HRTFs were recorded at the same position, the listener was placed during the experiment, see [16]. Three different acoustic conditions and eleven different loudspeakers positions lead to a total number of 33 experimental conditions. Every subject had to repeat all the 33 conditions six times. The first repetition of the 33 conditions constituted the training, thereafter a session with 66 trials and one with 99 trials were passed. In the sessions, the order of the acoustic conditions and speaker positions was randomized. The subjects required in average 15 minutes to complete the experiment without the training.

3.2. Results

Only the anechoic HRTFs are used in the WFS localization experiment conducted later. In order to validate the method, only the results dealing with these HRTFs in comparison to the real loudspeakers will be evaluated here. A detailed analysis has been presented in [16]. One of the eleven subjects showed a two to three times higher standard deviation than the other subjects and was removed before further data analysis. The localization ability for a given condition and subject is quantified by the unsigned error between the real azimuth and the perceived azimuth. First the mean and standard deviation of the error for every single subject and loudspeaker was calculated (each loudspeakers was presented five times to each subject). Then the unsigned error was calculated by building the mean about the absolute value of the difference between the real loudspeaker position and the mean position of the auditory event. The standard deviation was calculated by building the mean about the standard deviations of the single subjects.

The results are summarized in Tab. 1 for the acoustic conditions *loudspeaker* and *HRTF*. The unsigned error does not differ significantly between the two conditions. Only the standard deviation is 1.5° larger for the HRTF condition.

The time was measured the subjects needed to press the enter key after the presentation of the stimulus began. For the loudspeaker condition the subjects required, in average, 3.5 s seconds before pressing the enter key and 5.5 s for the HRTF condition.

3.3. Discussion

In agreement with the literature [12, 8, 10], the results indicate that subjects were able to localize auditory events at the same positions independently whether they were reproduced by real loudspeakers or via binaural synthesis and headphones. This holds also for the non-individualized HRTFs used here. The larger standard deviation indicates that the localization blur (cf. [2]) is slightly increased for the HRTFs. This is also supported by the longer response time of the subjects, which could indicate that it was more difficult to localize the source for the HRTF case.

Overall the results support that dynamic binaural resynthesis can be used for localization experiments in the given context.

4. LOCALIZATION OF VIRTUAL SOURCES IN WFS

4.1. Experimental procedure

All conditions were presented by dynamic binaural synthesis using headphones and head-tracking. A total of 48 conditions, resulting from 16 different listener positions and three different loudspeaker arrays were presented in the experiment. The loudspeaker array had a total length of 2.85 m. The number of loudspeakers was varied as 3, 8, and 15 active loudspeakers over the total length, respectively. The resulting distances between them was $\Delta x_0 = 1.43$ m, 0.41 m, and 0.19 m. The listener positions were located at two different distances 1.5 m and 2 m parallel to the loudspeaker array. At each of these distances, 8 listener positions were evaluated ranging from the center of the array at $X = 0$ m to the left side with $X = -1.75$ m in 0.25 m steps (see Fig. 3).

Δx_0 / m	unsigned error		max unsigned error	
	1.5 m	2.0 m	1.5 m	2.0 m
1.43	7.2°	6.0°	12.4°	10.4°
0.41	2.9°	2.4°	4.8°	3.9°
0.19	1.9°	1.7°	2.5°	2.4°
HRTF	2.0°		4.1°	
HRTF*	1.5°		2.0°	

Table 2: Mean and maximum of the unsigned error over all subjects and positions. The real source is represented by the results for the HRTFs shown in Tab. 1. The meaning of HRTF* is described in the text.

The 48 conditions were presented five times each to the subjects. The listening experiment was split into two sessions to avoid effects due to fatigue. One session for the listener positions with 1.5 m distance to the array and the other session for the distance of 2 m. Additionally, each session included ten times the presentation of a real loudspeaker at an azimuth of −5.7°. For the array with 8 speakers the array was rotated by 35°, and for the array with 15 speakers by 17.5° from the viewpoint of the listener. This was done to ensure an evenly distribution of the virtual source positions to the left/right of the listener.

4.2. WFS Implementation

The driving functions for WFS are computed by the Sound Field Synthesis Toolbox [14]. This requires an individual weighting and delaying of the virtual source signal for each loudspeaker, and a common pre-equalization filter [17, 6]. This filter should only be applied until the frequency prominent spatial sampling artifacts enter the synthesized sound field [18]. Therefore the cut off frequency of the pre-equalization filter was set differently for the three different loudspeaker configurations. The cut off frequencies were 120 Hz, 421 Hz, and 842 Hz ranging from the largest to the smallest loudspeaker spacing. To limit truncation artifacts due to the finite length of the arrays, a spatial (tapering) window was applied to the driving functions for the arrays with 8 and 15 speakers. In Fig. 3 the attenuated speakers are indicated by a color proportional to their attenuation. The reference point for the amplitude in 2.5D WFS was always identical with the listening position. The virtual source is a point source located at $(0, 1)$ m.

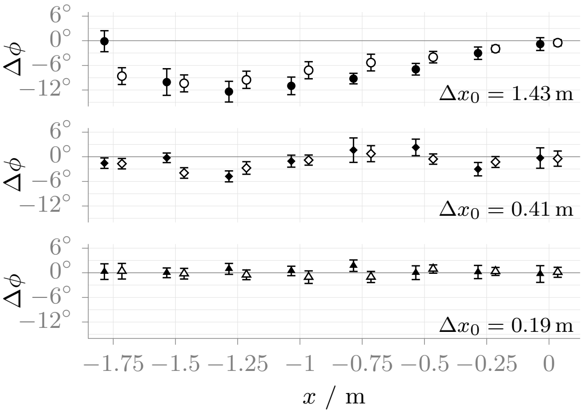

Fig. 2: Mean and 95% confidence interval of signed localization errors in WFS. Open symbols represents the distance of 2 m, closed symbols 1.5 m.

4.3. Data analysis

The offset of the pointing direction of the laser pointer was not equal for every subject. This was due to a variable placement of the headphones on the head of the listener and the laser pointer on the headphones. The offset was compensated by correcting the responses of the subjects using the results from the single speaker condition, for which the position was known. The results from the subject that was removed in the analysis of the first experiment are again not considered.

4.4. Results

The singed localization error was calculated for all responses of the subjects. It was calculated by calculating the mean perceived direction for every subject for every condition. Then the mean difference between these perceived directions and the real position of the virtual source was calculated. The mean and 95% confidence interval of the signed localization deviations in WFS are presented in Fig. 2. The results for the distance of 1.5 m between the listening positions and the loudspeaker array is presented with filled symbols, the distance of 2 m with open symbols. The results for the loudspeaker array with 15 speakers and a spacing of 0.19 m is shown at the bottom of the figure. For all 16 listening positions only minor deviations of the position of the auditory event from the position of the virtual source can

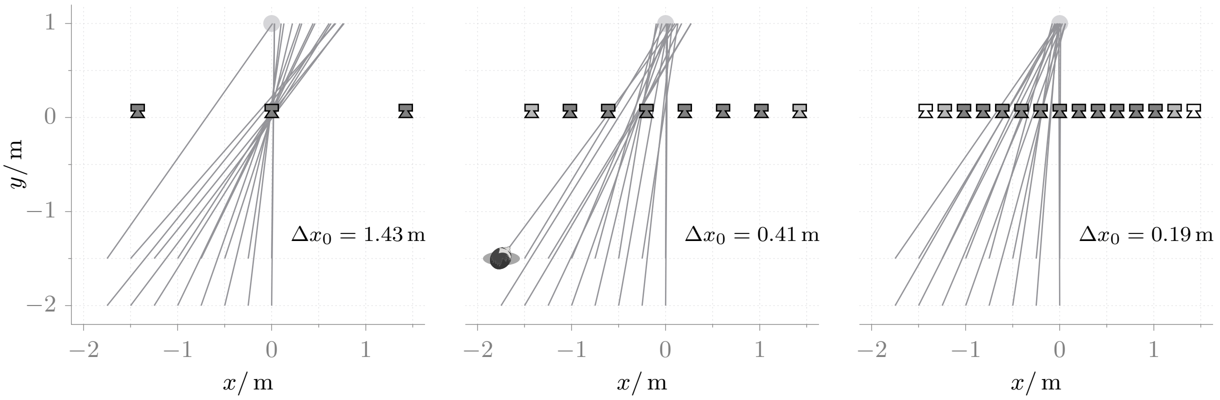

Fig. 3: Average direction the subjects were looking into from the evaluated 16 different listener positions. The results are shown for the three different loudspeaker distances. The gray point above the loudspeaker array indicates the desired virtual source position.

be observed. For a loudspeaker spacing of 0.41 m deviations can be observed at some listening positions, particularly for $(-1.5, -2)\,\mathrm{m}$, $(-1.25, -2)\,\mathrm{m}$, $(-1.25, -1.5)\,\mathrm{m}$, and $(-0.25, -1.5)\,\mathrm{m}$. For a loudspeaker spacing of 1.43 m, deviations are present for almost all positions besides the one at the center of the array and the one at $(-1.75, -1.5)\,\mathrm{m}$. In addition, this loudspeaker array has the largest deviation of −12° at the listening position $(-1.25, -1.5)\,\mathrm{m}$.

Another way of presenting the data is to draw a line starting from the listening position towards the mean direction the auditory event was perceived. This is done for all three loudspeaker arrays in Fig. 3. The position of the virtual source is indicated by the grey circle at $(0, 1)\,\mathrm{m}$. Again, this figure shows that the position of the auditory event corresponds to that of the virtual source for the loudspeaker array with 15 speakers and has the largest deviations for the array with 3 speakers. In addition, a tendency towards the position of a real speaker can be observed. This is the case for almost all listening positions for $\Delta x_0 = 1.43\,\mathrm{m}$, and for the positions around $X = -1.25\,\mathrm{m}$ for $\Delta x_0 = 0.41\,\mathrm{m}$.

In Tab. 2 the mean unsigned error together with its maximum for the listening positions parallel to the array is presented for the three loudspeaker arrays together with the result for a real source from Section 3.2. For the HRTFs the deviations are higher for a loudspeaker to the side. In order to have a fair comparison, the values for the HRTFs were recalculated considering only positions within an angle of $|\phi| < 35°$ which was the maximum mean result for the virtual source positions in WFS. The recalculated values are shown in the Table as HRTF*. Again the larger the spacing between the loudspeakers the larger the localization deviation. A slight dependency on the distance can be observed as well, with a less pronounced deviation for the larger distance. It can be seen that the results for the array with a loudspeaker spacing of 0.19 m is comparable to the real source case (HRTFs).

4.5. Discussion

For sound localization in the horizontal plane, humans exploit primarily selected differences between the two ears [2]. These are the interaural time differences (ITDs) and interaural level differences (ILDs), whereby the ITDs in the low frequency region up to 1 kHz dominate the localization [19]. Hence, one can assume that the localization for WFS works correctly if the aliasing frequency is greater than 1 kHz. This is supported by the results for the loudspeaker spacing of 0.19 m, were the aliasing frequency is around 842 Hz. To investigate the influence of spatial aliasing on the ITDs and localization in more detail, the ITDs were calculated for the different listening positions. The ear signals were fed into a binaural model after [20, 21], where the signals are filtered by a gammatone filterbank and the ITD is calculated independently for every frequency channel. The result is presented in Fig. 4 together with the ITD values for a point source located at the virtual source position (light gray) and a point source

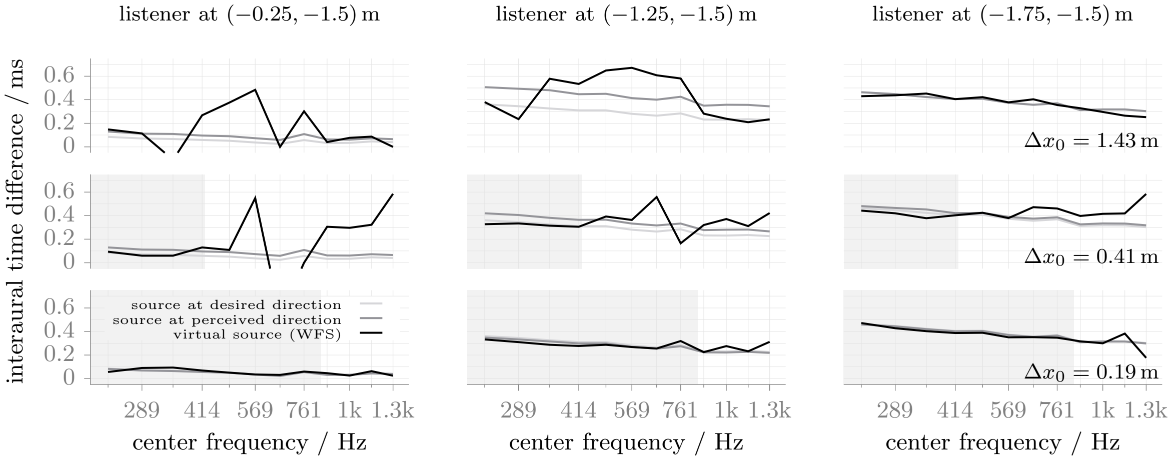

Fig. 4: Interaural time differences (ITDs) between the left and right ear signals of the listener at three different positions for all three loudspeaker arrays at center frequencies between 236 Hz to 1296 Hz. Black lines indicate the WFS signals, gray a point source placed at the position of the auditory event for WFS, and light grey a point source placed at the physical position of the virtual source. The light gray areas indicate the frequency region for which no spatial sampling artifacts are present.

located at the position of the auditory event (dark gray). The frequency region for which no aliasing occurred is indicated by the gray background. In the center column, the results are presented for the position with the greatest deviation between the virtual source and the auditory event location for the loudspeaker array with only three speakers. In this case it can be observed that the ITD deviates in a large frequency range towards the direction of the perceived auditory event. For the loudspeaker array with $\Delta x_0 = 0.41\,\mathrm{m}$ this is only the case for frequencies above the aliasing frequency of 421 Hz. For the array with 15 speakers the ITD is correct for most of the frequencies, and only slightly affected above the aliasing frequency. The right column presents the only position besides the central listening position, where the array with 3 speakers showed no localization deviations. This can be explained by the calculated ITDs, which is synthesized correctly in this case. The left column shows a position near the center of the array, where only slight localization deviations occurred. Nonetheless major deviations of the ITD can be observed for the two arrays with the fewest speakers. One explanation for the correct position of the auditory event may be the fact that the deviations occur in both directions and may cancel each other.

5. CONCLUSION

The ability to localize a virtual point source synthesized by WFS at different positions within the listening area was investigated. In order to be able to seamlessly switch between the loudspeaker array configuration or the position of the subjects during the experiment, dynamic binaural synthesis was applied to simulate the ear signals via headphones. In a prior experiment it was verified that this method together with the applied pointing procedure was reliable and accurate. In the main experiment the localization of a virtual point source in WFS for 16 different listener positions and three different loudspeaker spacings was investigated. The results show that for a loudspeaker spacing of around 20 cm the localization error is below 2° and no sweet-spot can be observed within the listening area. If the loudspeaker spacing is increased up to 41 cm, the localization error increases also. Its mean is still below 3°, but there are differences between the individual positions up to 5°. If the spacing is further increased to 1.43 m with an array consisting of only three speakers, the localization results show the same behavior as for a stereophonic setup. A sweet-spot can be observed at the center position and a localization of the auditory event towards the nearest loudspeaker at the other positions.

The results lead to the conclusion that localization in WFS works well in the entire listening area also for relatively large loudspeaker spacings up to 40 cm. More critical for the application of WFS seems to be coloration of the virtual point source for larger loudspeaker spacings. We will investigate on this in future experiments.

Only virtual point sources were investigated in this study, the synthesis of extended sources or ambience is still an open question in WFS. The localization properties of focused sources have already been investigated [22]. Comparing the results presented there with the ones shown in this article allows the conclusion that localization is more critical for focused source for the same array configuration.

This study has proven that dynamic binaural synthesis can be used to evaluate the localization within an extended listening area for sound field synthesis methods. It enables a systematic comparison between different loudspeaker spacings, array geometries and listener positions. Furthermore different sound field synthesis methods can be compared, like for instance WFS and Higher Order Ambisonics.

ACKNOWLEDGMENTS

This work was supported by the research grant RA 2044/1-1 of Deutsche Forschungsgemeinschaft (DFG).

6. REFERENCES

[1] Snow WB. Basic Principles of Stereophonic Sound. *Journal of the Society of Motion Picture and Television Engineers*, 61:567–587, 1953.

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

[3] Vogel P. *Application of Wave Field Synthesis in Room Acoustics*. Ph.D. thesis, Technische Universiteit Delft, 1993.

[4] Start E. *Direct Sound Enhancement by Wave Field Synthesis*. Ph.D. thesis, Technische Universiteit Delft, 1997.

[5] Wittek H. *Perceptual differences between wavefield synthesis and stereophony*. Ph.D. thesis, University of Surrey, 2007.

[6] Verheijen E. *Sound Reproduction by Wave Field Synthesis*. Ph.D. thesis, Technische Universiteit Delft, 1997.

[7] Hartmann WM. Localization of sound in rooms. *The Journal of the Acoustical Society of America*, 74(5):1380–1391, 1983.

[8] Bronkhorst AW. Localization of real and virtual sound sources. *The Journal of the Acoustical Society of America*, 98(5):2542–2553, 1995.

[9] Seeber BU. A New Method for Localization Studies. *Acta Acustica*, 83:1–5, 1997.

[10] Majdak P, et al. The Accuracy of Localizing Virtual Sound Sources: Effects of Pointing Method and Visual Environment. In *124th Audio Engineering Society Convention*. 2008.

[11] Wierstorf H, et al. A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances. In *130th Audio Engineering Society Convention*. 2011.

[12] Makous JC and Middlebrooks JC. Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America*, 87(5):2188–200, 1990.

[13] Lewald J, et al. Sound localization with eccentric head position. *Behavioural Brain Research*, 108(2):105–25, 2000.

[14] Wierstorf H and Spors S. Sound Field Synthesis Toolbox. In *132nd Audio Engineering Society Convention*. 2012.

[15] Geier M, et al. The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods. In *124th Audio Engineering Society Convention*. 2008.

[16] Wierstorf H, et al. Perception and evaluation of sound fields. In *59th Open Seminar on Acoustics*. 2012.

[17] Spors S, et al. The theory of Wave Field Synthesis revisited. In *124th Audio Engineering Society Convention*. 2008.

[18] Spors S and Ahrens J. Analysis and Improvement of Pre-equalization in 2.5-dimensional Wave Field Synthesis. In *128th Audio Engineering Society Convention*. 2010.

[19] Wightman FL and Kistler DJ. The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3):1648–61, 1992.

[20] Dietz M, et al. Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Communication*, 53(5):592–605, 2011.

[21] Søndergaard PL, et al. Towards a binaural modelling toolbox. In *FORUM ACUSTICUM*, pp. 2081–2086. 2011.

[22] Wierstorf H, et al. Reducing Artifacts of Focused Sources in Wave Field Synthesis. In *129th Audio Engineering Society Convention*. 2010.