Localization in Wave Field Synthesis and higher order Ambisonics at different positions within the listening area

Hagen Wierstorf[1], Alexander Raake[1], Sascha Spors[2]
Email: hagen.wierstorf@tu-berlin.de
[1] *Assessment of IP-based Applications, Technische Universität Berlin, 10587 Berlin, Deutschland*
[2] *Institute of Communications Engineering, Universität Rostock, 18119 Rostock/Warnemünde, Deutschland*

Introduction

Sound field synthesis methods try to synthesize a desired sound field within an extended listening area. If this task can be complied, the perception of a listener placed within the listening area will not differ from his/her auditory perception within a real sound field. Due to the practical limitation in the number of loudspeakers than can be applied for sound field synthesis methods, the produced sound field is in most cases only indistinguishably from the real sound field in a small area and/or under a certain frequency.

This study investigates the ability to correctly localize sources in a synthesized sound field for two sound field synthesis methods. These methods are Wave Field Synthesis [1] and near-field compensated higher order Ambisonics [2]. In order to investigate the localization by the listeners in the whole listening area, the area was sampled at 16 different positions. To further investigate on the number of loudspeakers used, three different loudspeaker arrays with different spacings of the single loudspeakers were used. The different listener positions and loudspeaker arrays were presented to the subjects via dynamic binaural synthesis [3]. This has the advantage of seamlessly switching between the different conditions and reliably placing of the different listeners at exactly the same places. In a former test it was shown that dynamic binaural synthesis has no influence on the localization performance [5].

In the next chapter a brief introduction into Wave Field Synthesis and near-field compensated higher order Ambisonics will be given, with a highlight on the differences between the two methods. Afterwards a localization experiment will be introduced and its results will be presented and discussed.

Wave Field Synthesis and near-field compensated higher order Ambisonics

In sound field synthesis methods one is looking for solutions to the following equation.

$$P(\mathbf{x},\omega) = \sum_{\mathbf{x}_n \in \partial V} G(\mathbf{x}|\mathbf{x}_n,\omega) D(\mathbf{x}_n,\omega) \,, \qquad (1)$$

where $G(\mathbf{x}|\mathbf{x}_n,\omega)$ denotes the sound field that originates from a single loudspeaker placed at $\mathbf{x}_n$ on the boundary ∂V of a volume V, and $D(\mathbf{x}_n,\omega)$ is the signal that is fed into the loudspeaker. $P(\mathbf{x},\omega)$ is the desired sound field we targeting to create in V. The equation has to be solved with respect to $D(\mathbf{x}_n,\omega)$.

For special geometries like a circle or a sphere this can be done directly and results in a solution known as near-field compensated higher order Ambisonics, if we assume a point source like characteristic for a single loudspeaker. In this case the sound field is represented by circular or spherical harmonics. If we sample the boundary with M loudspeakers, only sound fields represented by harmonics up to an order of $N = \frac{M-1}{2}$ can be synthesized correctly for a circular array.

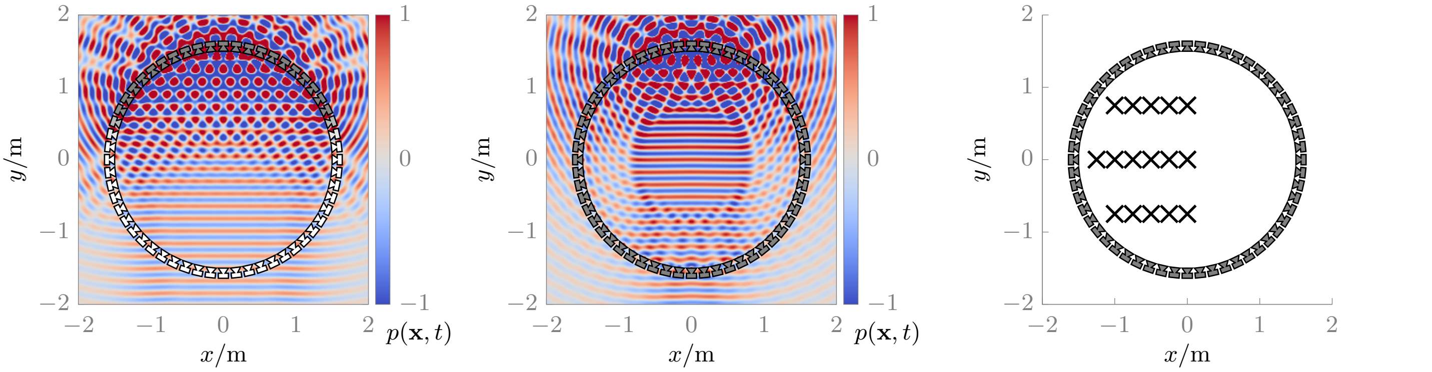
In Wave Field Synthesis the equation is solved by a high frequency approximation that solves the problem for small linear array elements. Applied to a circular array this leads to a selection of an active sub-array for a given source. This is illustrated in Fig. 1a, where the sound field of a plane wave coming from above is synthesized as the desired sound field. The plane wave has a frequency of 2 kHz. In addition to the difference in the active loudspeakers there is also a difference in the artifacts that are present in the synthesized sound field with discrete loudspeakers. The loudspeaker array has a spacing of 0.17 m between its loudspeakers which enables only a correct reproduction of frequencies up to 1 kHz. For Wave Field Synthesis theses artifacts vanish for positions farer away from the array, whereas for near-field coming higher order Ambisonics there is always a artefact-free region in the center of the array – compare Fig. 1a and Fig. 1b.

In the next section a localization experiment is presented that investigates the influence of the different properties of Wave Field Synthesis and near-field compensated higher order Ambisonics on the perception of the direction of a desired source.

Method

For both sound field synthesis methods the same circular loudspeaker array with a diameter of 3 m was used. The array consisted of 56, 28, or 14 loudspeakers, corresponding to spacings of 0.17 m, 0.34 m, and 0.67 m between the loudspeakers. Within the loudspeaker array 16 different listening positions were chosen as shown in Fig. 1c. All the positions were only in the left half of the listening area due to the symmetry of the problem.

Both sound field synthesis methods synthesized the field of a point source located at $(0, 2.5)$ m and a plane wave traveling into the direction $(0, -1)$. For Wave Field Syn-



(a) Wave Field Synthesis **(b)** near-field compensated higher order Ambisonics **(c)** listener positions

Figure 1: Sound fields and listener positions. The sound fields are plotted for a monochromatic plane wave $p(\mathbf{x}, t)$ traveling into the direction $(0, -1)$ with a frequency of $f = 2000\,\mathrm{Hz}$.

thesis the upper frequency of the pre-equalization filter was set to the aliasing frequency which approximately is given as $f_{\mathrm{aliasing}} = \frac{c}{2\Delta x_n}$. For near-field compensated higher order Ambisonics the order was limited by the number of available loudspeakers.

All stimuli were created via dynamic binaural synthesis. In this case all loudspeakers of the different loudspeaker arrays were simulated with a set of anechoic head-related transfer functions [4]. The signals were calculated with the Sound Field Synthesis-Toolbox[1] [5]. The stimuli were presented using headphones (AKG K601). A head tracker (Polhemus Fastrak) measured the head orientation of the listener in order to switch to the appropriate set of head-related transfer functions for the direction the listener is looking to. The impulse response from the virtual source to the left/right ear was convolved with a white noise pulse signal (700 ms noise with a 20 ms on- and off-ramp, 300 ms pause and a total length of 100 s) by a convolution engine (SoundScape Renderer).

12 self reported normal hearing listeners, mostly students, participated in the experiment. The experiment was conducted within an acoustically damped room. The listeners were placed on a heavy chair and were looking towards a curtain at 1.5 m distance – compare Fig. 2. The listener pointed with a laser pointer mounted on the headphones into the direction she/he perceived the stimuli. In order to do this, the listener turned their heads until the location of the laser light was corresponding to the position of the auditory event. The listeners were instructed to rate only the horizontal component of the auditory event, due to the fact that non-individual head-related transfer functions were used and the dynamic binaural synthesis was only working in the horizontal domain.

The experiment was carried out in 12 runs. One run consisted of a given sound field synthesis method, a given virtual source, and a given number of loudspeakers. Within a run the 16 listener positions were randomized and there were 5 repetitions per listener position, leading to a total number of 960 answers per listener, lasting around 90 minutes that were distributed to two days.

Figure 2: Picture of a subject during the experiment. The light was dimmed during the experiment and is only present here to take the picture.

For the runs with near-field compensated higher order Ambisonics only three listeners participated in the experiment, and the results are preliminary at the moment.

Results and Discussion

The results of the localization experiment are presented in Fig. 3. The upper row shows the average perceived direction for Wave Field Synthesis and the bottom row the average direction for near-field compensated higher order Ambisonics. At every listening position an arrow is indicating the direction of the auditory event and a gray line extends the array in order to allow a comparison with the position of the virtual source. The color of the arrows indicates the deviation from the desired direction. For the point source the desired position of the source is illustrated by the gray circle, positioned at $(0, 2.5)\,\mathrm{m}$.

[1] http://github.com/sfstoolbox/sfs, revision aa52df7b6

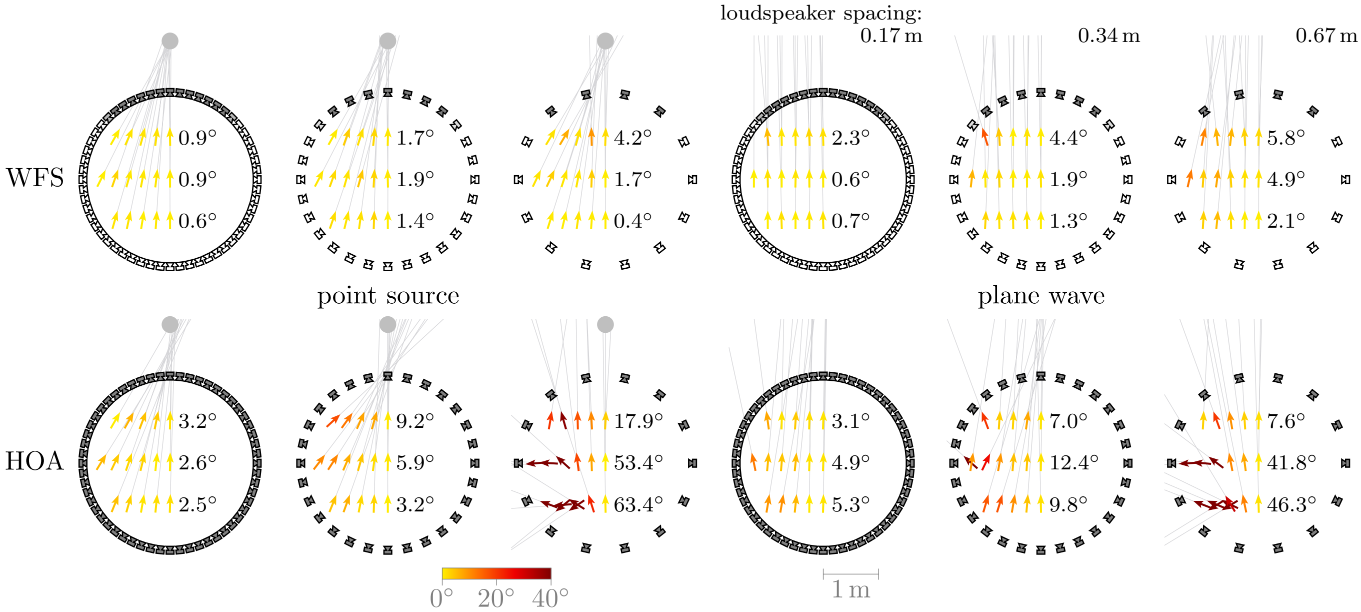


Figure 3: Localization results for Wave Field Synthesis and near-field compensated higher order Ambisonics. On every employed listening position a arrow is indicating the average result for this particular position. The arrow is pointing into the direction the subjects localized the stimuli. The color coding of the arrow indicates the deviation of this position from the desired one. For the point source condition the gray circle symbolizes the desired position of the virtual point source.

For the plane wave the desired direction is straight to the top of the graph. The numbers besides the three rows of arrows are the average deviation from the desired direction for the positions in this row, corresponding to y-positions of 0.75 m, 0 m, or −0.75 m.

For Wave Field Synthesis the localization is only slightly dependent on the listening position. Especially positions near to the active loudspeakers have a slightly larger deviation, as it is the case for all positions with $y = 0.75$ m. For a point source the overall deviation averaged over all listener positions is 0.8° with a mean confidence interval of 1.4° for a loudspeaker spacing of 0.17 m. The localization performance is not distinguishable from the localization performance for a real source. For the maximum loudspeaker spacing of 0.67 m the overall deviation increases to still only 2.1° with a mean confidence interval of 1.8°. The results are slightly worse for a plane wave as virtual source. In this case the performance is 1.2° for a spacing of 0.17 m and 4.3° for a spacing of 0.67 m. The confidence interval is comparable to the point source case. The slightly worse results for a plane wave can be explained by the larger deviations for positions at the side of the array. At these position the localization tends to be bound to the edge of the active loudspeakers, which is not the case for the synthesized point source.

For near-field compensated higher order Ambisonics the overall localization performance is lower than that of Wave Field Synthesis. For 56 loudspeakers the deviation is 2.8° for a point source and 4.4° for a plane wave. For the plane wave, positions to the side of the array seem to be even more critical than for Wave Field Synthesis. If we lower the number of used loudspeaker and as a consequence the Ambisonics order the results get even worse for the side positions, reaching deviations of more than 90°. In this case the low amount of spatial information and the usage of all loudspeakers to synthesize the sound field could even lead to the perception of more than one auditory event for side positions, as reported by the subjects and as indicated by the results. For example at some positions no reasonable average could be calculated over the listeners ratings, because everyone was looking towards the direction of another loudspeaker. In this case more than one arrow is plotted in the figure, showing grouped individual results. For example, if two arrows are plotted, one shows an individual result and the other one the average over the other two listeners which had no major deviation between their results.

To sum up the results, Wave Field Synthesis leads to a reasonable localization accuracy in the whole listening area. This holds also for large loudspeaker spacings. Only in the very near field of the loudspeakers (< 0.3 m) larger deviations occur. Near-field compensated higher order Ambisonics shows a stronger dependency on the number of loudspeakers and on the position of the listener. For large loudspeaker spacings only a region at the center of the listening area shows still good localization performance, whereas at positions out of the center large deviations or more than one auditory event could appear.

Conclusion

The localization accuracy in Wave Field Synthesis and near-field compensated higher order Ambisonics was investigated. Three different circular loudspeaker arrays were applied for both methods, with different spacings between the loudspeakers. For sound field synthesis methods like Wave Field Synthesis and near-field compensated higher order Ambisonics it is of interest how the perception for the listener scales in the whole listening area. To investigate this 16 different listening positions were applied in a way, that a great amount of the listening area was covered. In order to make it possible to directly compare different listening positions and different loudspeaker arrays, all loudspeaker setups and listening positions were simulated via dynamic binaural synthesis.

The results indicate that the localization for Wave Field Synthesis is not impaired in comparison to a free field situation with a real source for loudspeaker spacings of 0.2 m or smaller. Doubling the spacing introduces only small deviations of the desired directions and even for the largest spacing of 0.67 m the performance deteriorates only to an average of around 4°.

For near-field compensated higher order Ambisonics the localization accuracy is not as evenly distributed in the listening area as for Wave Field Synthesis. It shows good localization performance in the center of the array, but degradations nearer to the loudspeakers. That becomes especially worse for large loudspeaker spacings and the low number of employed loudspeakers in that case. Here, listeners can hear more than one auditory event.

Overall the results emphasize good localization abilities for listeners for Wave Field Synthesis in the whole listening area, even for a small number of loudspeakers. In contrast to that results, near-field compensated higher order Ambisonics showed deteriorated localization for off-center listening positions for a low number of loudspeakers. In that case it will be a better choice to apply Ambisonics variants that apply also a sort of loudspeaker selection like Wave Field Synthesis, for example like the methods presented in [8].

Besides correct localization, sound field synthesis methods should also reproduce sound without perceptible coloration. That is especially an issue for Wave Field Synthesis [7] and still a matter of investigation.

Acknowledgement

The authors which to thank Vincent Kuscher, who helped with the experiments. This work was supported by the grant DFG RA 2044/1-1.

References

[1] Berkhout, A.J., de Vries, D., Vogel, P.: Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America* 93(5): 2764-78, 1993.

[2] Daniel, J.: Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. Université Paris 6, 2001.

[3] Völk, F., Konradl, J., Fastl, H.: Simulation of wave field synthesis. In *Proceedings Acoustics 08* 6419-6424, 2008.

[4] Wierstorf H., Geier, M., Raake, A., Spors S.: A free database of head-related impulse response measurements in the horizontal plane with multiple distances. In *Proceedings 130th Convention of the Audio Engineering Society*, 2011.

[5] Wierstorf H., Spors S.: Sound field synthesis toolbox. In *Proceedings 132nd Convention of the Audio Engineering Society*, 2012.

[6] Wierstorf, H., Spors, S., Raake, A.: Perception and evaluation of sound fields. In *Proceedings 59th Open Seminar on Acoustics*, 2012.

[7] Wittek H.: Perceptual differences between wavefield synthesis and stereophony. University of Surrey, 2007.

[8] Zotter, F., Frank, M., Pomberger, H.: Comparison of energy-preserving and all-round Ambisonic decoders. In *Proceedings of the German Annual Conference on Acoustics (DAGA)*, 2013.