# Perception and evaluation of sound fields

Hagen Wierstorf, Sascha Spors, Alexander Raake

*T-Labs, Technische Universität Berlin, 10587 Berlin, Deutschland, Email: hagen.wierstorf@tu-berlin.de*

### Summary

Sound field synthesis techniques claim to recreate a desired sound field within an extended listening area. In order to investigate the perceptual properties of the synthesized sound field the listener has to be placed at different positions. In practice that can be quite difficult with real loudspeakers. Another possibility to perform listening tests is to present the field via binaural synthesis. This study investigates whether binaural synthesis is perceptually transparent for the purpose of localization studies for sound field synthesis. A localization test is performed comparing real loudspeakers to two different binaural synthesis configurations using non-individual head-related transfer functions (HRTFs), once with and once without reflections. The results show only slight differences between real speakers and HRTFs-based synthesis, resulting in a one degree greater localization blur for the HRTFs without reflections than for the other two cases.

## 1. Introduction

In the last years several sound reproduction techniques beside the classical stereophony emerged. One prominent class are sound field synthesis (SFS) methods. They apply a bunch of loudspeakers to synthesize a given sound field within an extended listening area. The best known representatives are Wave Field Synthesis (WFS) [1] and Higher Order Ambisonics (HOA) [3].

SFS methods assume that the loudspeaker distribution at the surface is continuous, which is of course not given in reality. In practice a spatial sampling of the surface occurs. This sampling leads to artifacts in the synthesized sound field above the so called aliasing frequency. For a loudspeaker distance of 0.15 m this frequency is around 1 kHz for WFS. Hence, one has to consider that SFS methods are not strictly physically motivated, but rely on the possible masking of the artifacts in the sound field, which means on psychoacoustics.

One of the advantages of SFS methods over stereophony, the actual synthesis in a given volume, is also a critical point for psychoacoustic experiments. It is not suitable to place a subject at one point within the listening area. In contrast, the perception of the whole sound field is under consideration. Even more complicated is a systematic study of the dependency of the perception on the aliasing frequency. For this purpose, the number and distance of the loudspeakers has to be varied considerably, which is often not possible with a real setup. In practice, this has lead to only a few psychoacoustics experiment regarding the perception of a sound field created by SFS methods. In addition, most of these studies have investigated only one or two listener positions and a small range of different loudspeaker distances (for WFS see e.g. [14, 11]).

In this study, we propose to overcome these shortcomings by using dynamic binaural synthesis to simulate different positions within the listening area and different loudspeaker arrays. In a first step we restrict this study to localization experiments. Dynamic binaural synthesis simulates a loudspeaker by applying the respective head-related transfer functions (HRTFs) to a desired audio material and playing it back to a listener via headphones. Simultaneously, the orientation of the head of the listener is tracked and the HRTFs are switched accordingly to the head position of the listener. If this dynamic part is applied, results from the literature have shown that the localization performance for a virtual source is more or less equal to the case of a real loudspeaker. The localization error for real sources, that is the difference between the direction of a real loudspeaker and direction of the auditory event is between 2°–5° [10, 2, 5, 8]. If individual HRTFs of the subjects were used, no difference between localization of real and virtual speakers were found. For non-individual HRTFs deviations around 1° were found [10]. One cause for the varying results for the localization performance in the literature is the fact that localization experiments are critical regarding the used pointing method. Due to the fact, that the localization erorr can be as small as 1° the error of the method has to be smaller than 1°, which can not be achieved with all methods [10, 7].

This study applies a method similar to the one used in [8]. Here, the subject has to point with her head towards the direction of the auditory event, while the sound event is present. This has the advantage that the subject is directly facing the source, a region in which the minimum audible angle is smallest [9]. If the subject is only pointing with its nose in the direction of the source, an estimation error of the sources at the side will occur, due to an interaction with the motor system. This can be overcome by adding a visual pointer, which indicates to the subject where her nose is pointing [6].

The aim of this study is on the one hand to test

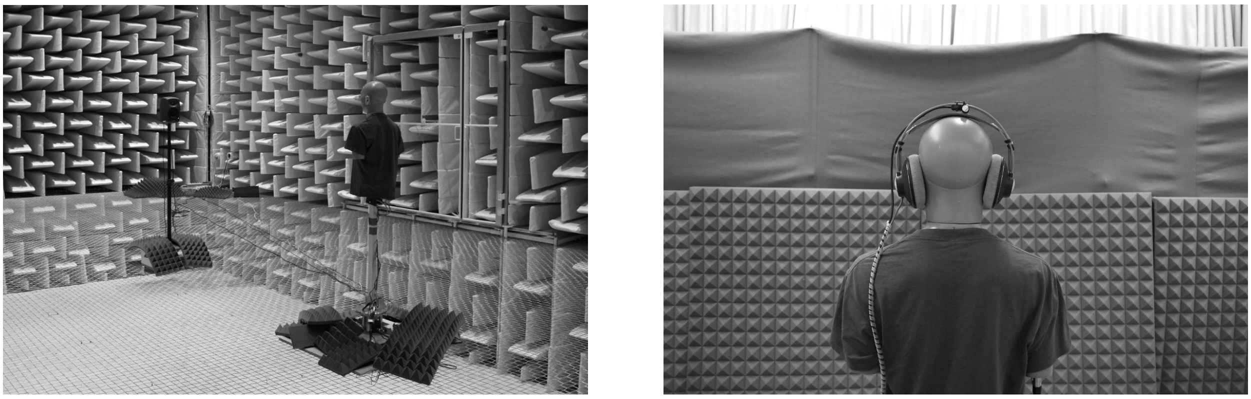

**Figure 1:** Measurements with the artificial head in the anechoic chamber (left) and in the listening room, which housed the experiment (right). Note, that the dummy head was wearing the same headphones as the subjects and was also suited at the same position.

the resolution of our pointing method. Another task is to investigate if HRTFs measured with a dummy head could be used, and possible loudspeaker arrays could be constructed by interpolation from a HRTF set measured for only one speaker. Hence this study will compare the localization of a single loudspeaker with its simulation via dynamic binaural synthesis using HRTFs recorded with an artificial head. One HRTF set was recorded at the same position as the subjects were placed within the experiment and for every used loudspeaker. The other one is recorded in an anechoic chamber with a single loudspeaker, as described in [12]. If the localization of a virtual source reproduced with an anechoic HRTF set is not affected or only to a known small degree, the method can be applied to investigate the localization in sound fields generated by different SFS methods.

## 2. Method

### 2.1. Head-related transfer functions

This study employs two different sets of head-related transfer functions (HRTFs). The first one was measured in an anechoic chamber with a resolution of 1° in the horizontal plane and a distance of 3 m between loudspeaker (Genelec 8030A) and artificial head (KEMAR, type 45BA). It is freely available and its measurement is described in [12]. In this case, the different loudspeaker positions were realized by inter- and extrapolating the measured HRTF set.

The second HRTF set was measured using the same equipment. But this time the artificial head was placed within the listening room, in which the localization experiment took place. It was positioned exactly at the same position the subjects were seated in during the experiment. As loudspeakers, all 19 speakers from the experiment were used (see Sec. 2.3). To mimic the same configuration as for real loudspeaker listening, the artificial head was wearing headphones during the measurements (compare Fig. 1). Again, a resolution of 1° was chosen and the measurement was done for angles ranging from $-90°$ to $90°$ of the head orientation of the dummy head. Both HRTF sets are available at [15].

### 2.2. 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.

### 2.3. Apparatus

Stimuli were digitally generated at a sampling rate of 44.1 kHz. A PC with Octave was used to generate impulse responses for the binaural synthesis. To this aim the time signals of HRTFs measured in the reproduction room at the listener position and HRTFs measured in an anechoic chamber [12] were given as input signals for convolution with the stimuli. The SoundScape Renderer [4] and puredata [16] finally played the headphone and loudspeaker signals. The PC was equipped with a RME HDSP MADI card and the digital to analog conversion was done by CreamWare A16 converters. The listeners wore AKG K601 headphones and as loudspeakers 19 Fostex PM0.4 were arranged as a linear array applying a distance of 0.15 m between them. For the experiment only 11 of this loudspeakers were used, the chosen ones are marked in black in Fig. 2. The head movements of the listener were tracked by a Fastrak Polhemus head tracker, which passed its signal to puredata. The SoundScape Renderer was then switching the HRTFs for the dynamic binaural synthesis, according to the orientation of the listener given by the head tracker data. A small laser pointer was mounted onto the headphones. The listener was positioned within an acoustically damped listening room, 1.5 m in front of the loudspeaker array, with an acoustical transparent curtain in between. A sketch of the setup and a picture is shown in Fig. 2. The orientation and position of the subjects during the experiment was recorded with the

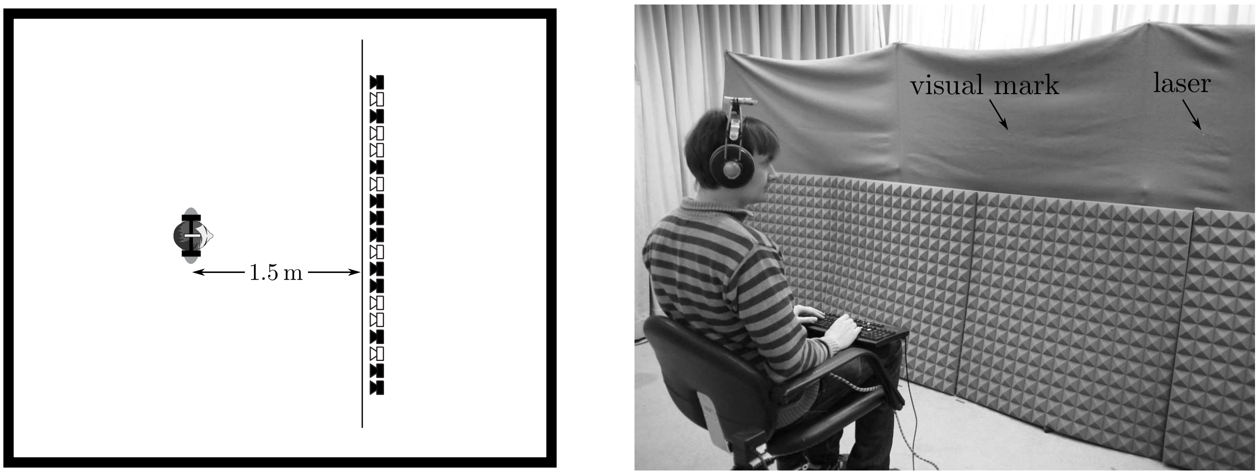

**Figure 2:** Sketch of the apparatus in the listening room (left) and a subject during the experiment (right). Only the blacked marked loudspeakers were used in the experiment. Note, that the room was dark during the experiment.

same head tracker.

2.4. Stimuli

As audio material, Gaussian white noise pulses with a duration of 700 ms and a pause of 300 m between them were applied. The single pulses were windowed with a Hanning window of 20 ms length at the start and the end. The signal was bandbass filtered with a fourth order butterworth filter between 125 Hz and 20000 Hz. The signal with a length of 100 s was stored and played back in a loop in the experiment. The single pulses of this signal were independent white noise signals. For the headphone reproduction the noise file was convolved with the time signal of the corresponding HRTF.

2.5. Procedure

The subjects sat on a heavy chair, wearing headphones with the mounted laser pointer and had a keyboard on their knees (see Fig. 2). They were instructed to point, with the laser pointer, into the horizontal direction where they perceive the auditory event by turning the head. The vertical direction was to be ignored. If they were 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 of 90 ms. After the key press, the next trial started instantaneously, which implies that the subject started with the localization always from the last position, and not from a fixed point. The subjects were instructed that they could turn their head if they were unsure about the direction of the sound.

There were three conditions in the experiment, *loudspeaker*, *room HRTF*, *anechoic HRTF*. For the first one, the noise pulses were just played through one of the loudspeakers. For the other two conditions the sound was played via headphones. There were three different conditions and eleven different loudspeakers, which leads to a number of 33 trials. Every subject had to pass all the 33 trials six times. The first 33 trials were for training, thereafter a session with 66 trials and one with 99 trials were passed. In the sessions, the order of the conditions and speakers was randomized. The subjects needed, average, 15 minutes to complete the experiment not counting the training.

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 room. After the calibration step, the room was darkened and the experiment started.

2.6. Data analysis

A subject was able not only to turn her head, but also to move the head in a translatory way. For the two conditions employing headphone reproduction, this had no influence on the results for the perceived direction, because the dynamic binaural synthesis compensated only for the angle of the head, not its absolute position. Hence, the virtual source was moving with the subject in case of translational movements. For the loudspeaker condition, this is no longer true, and the perceived angle between a single loudspeaker and the head of the subject is changing with possible head movements. To calculate the direction of the auditory event, the data was compensated for these head movements, which were acquired by the head tracker as well, by the following formula.

$$\phi' = \tan^{-1}\left([(1.5 - y)\tan\phi - x]/1.5\right) \qquad (1)$$

Here $\phi$ is the measured head orientation, $x, y$ are the measured coordinates of the head tracker, assuming that the origin of the coordinate system is at the center of the chair, and $\phi'$ is the final value

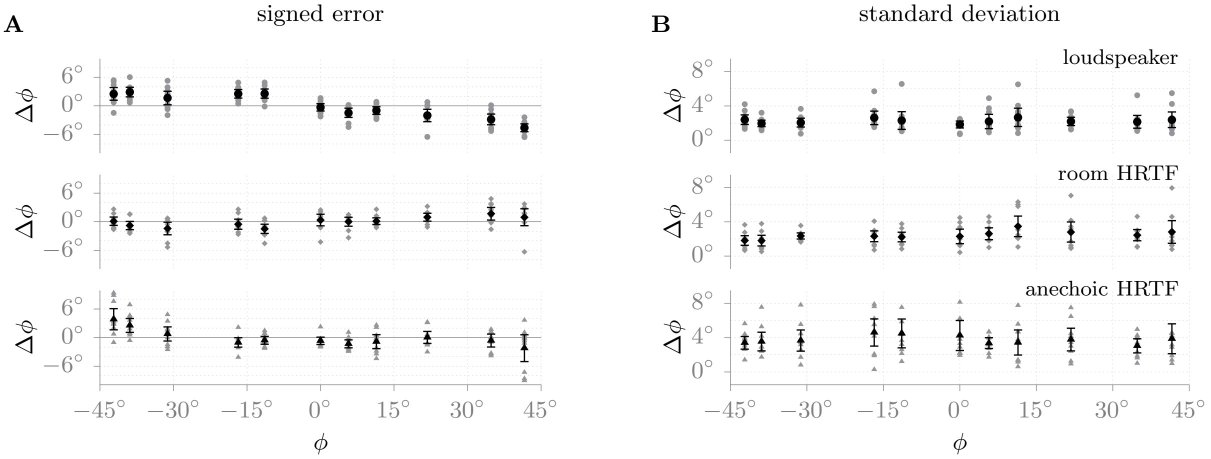

**Figure 3:** The mean over all subjects together with the 95% confidence interval is shown. In grey, the individual subjects' results are presented. In graph A, the signed error of the localization of the eleven speakers is shown. In graph B, the mean standard deviation for the localization task is depicted. The top row represents the condition with the real loudspeakers, the middle row the room HRTFs, and the bottom row the anechoic HRTFs.

for the direction of the auditory event. In an additional step, the measured orientation of the head had to be connected with the orientation of the subject within the room. This step is needed, because the orientation of the head tracker is not an absolute value and was chosen anew in every session. Furthermore, the laser pointer had to be switched on again before every session, which could affect its position on the headphone slightly, and finally the position of the headphones on the head of the listener was not the same for every subject. In practice, this was solved by compensating the measured head orientation data with the position of a tiny visual mark on the curtain. Its position in the current head tracker orientation coordinate system was measured in a calibration step.

The main problem which had to be solved for the two HRTF conditions was relative zero point of the head tracker. To compensate for this, we assumed that the localization was symmetrical to the left and the right. Then the zero point of the localization data is the mean value of all measured directions of the auditory event in one session. In order to avoid an overreaching of the HRTF conditions this step was also done for the loudspeaker condition.

After the data calibration, the results from both sessions were pooled for every subject and the mean and standard deviation were calculated. The mean over all subjects together with the confidence interval was then calculated using these data.

## 3. Results

The data analysis has shown a standard deviation for one subject, that was twice as high as that of the other subjects. Hence, this subject was excluded from the results. Fig. 2.6 shows the mean over the subjects, together with the 95% confidence interval. In the top row, the results for the loudspeaker condition are presented, in the middle row the results for the room HRTFs and at the bottom the one for the anechoic HRTF condition. The signed error is calculated by subtracting the real position of the given loudspeaker from the mean localization values of the subjects. The individual results of the subjects are given by the grey symbols. It can be seen that the signed error never exceeds 5° for all conditions and speakers. For the loudspeaker condition and the anechoic HRTFs, a slight underestimation of the speakers at the sides can be observed. This effect is not present for the room HRTFs. In addition, the mean of the standard deviations of the subjects was calculated. The loudspeaker condition and the room HRTFs show very similar results around 2.3°. For the anechoic HRTF condition the standard deviation is slightly higher at 3.8° as the mean over all speakers.

In Tab. 1 the mean values are presented. In addition, the mean values of the unsigned errors over the speakers are presented. The unsigned error is calculated by using the absolute value of the difference between the real speaker and the position of the auditory event. The loudspeaker condition shows a mean value of 2.4°, the room HRTFs of 1.5° and the anechoic HRTFs of 2.0°.

The head tracker was not only used to save the answers of the subjects, in addition the whole movements of the subjects were saved. This way it was possible to take a look at the response times and movement patterns of the subject and to check whether they were different for the different conditions. Figure 4 shows some selected results from one subject. One trial for every condition has been selected. At a time of 0 s, the subject starts at its last point, where she has answered with the enter key, and then moves her head to the next condition. It is obvious that the time span was different for the con-

|  | Loudspeaker | room HRTF | anechoic HRTF |
|---|---|---|---|
| unsigned error /° | 2.4 ±0.59 | 1.5 ±0.26 | 2.0 ±0.56 |
| standard deviation /° | 2.2 ±0.15 | 2.4 ±0.28 | 3.8 ±0.30 |
| time / s | 3.5 ±0.65 | 3.7 ±0.55 | 5.5 ±1.72 |
| turning points | 1.6 ±0.36 | 1.8 ±0.23 | 3.2 ±0.92 |

**Table 1:** Mean values about all speaker positions and subjects together with the confidence interval.

ditions, with ranging from 3.5 s for the loudspeaker condition to 5.5 s for the anechoic HRTF condition, on average (see Tab. 1). Another possible measure is the number of turning points a subject required while moving her head to the direction of the auditory event. The number of turning points was calculated by a differentiation. Only those points were counted that differ in their position from the last one by more than 1°. For example, for the loudspeaker condition in Fig. 4 0 turning points, for the room HRTF 2, and for the anechoic HRTF 2. The number of turning points is correlated with the processed time, as can be seen from the mean values in Tab. 1.

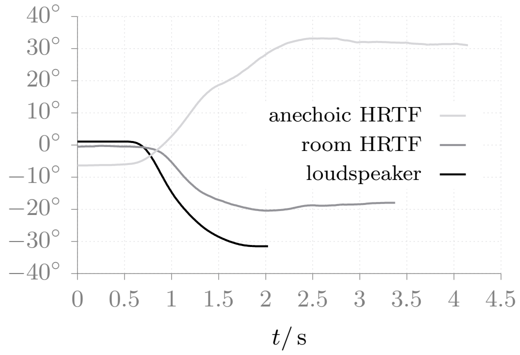

**Figure 4:** Head movements of a single subject between two source localizations (pressing of the enter key).

## 4. Discussion

The results are in agreement with results from the literature. Accordingly, a localization task could be done with a similar accuracy for virtual sources presented via headphones using dynamic binaural synthesis as with real sources.

The accuracy for virtual sources are higher if the used HRTF are measured at the place of the subjects in the experiment, including room reflection. A surprising result is the fact that the localization error was higher for the real sources than for the virtual sources in the room HRTF condition. At this point it is not possible to verify if this is a real effect, or due to the slightly different method of measuring the localization in both cases.

The standard deviation as an indicator of the localization blur is the same for the real speakers and the room HRTFs, but larger for the anechoic HRTFs. This indicates that room reflections could be a helping factor in localization of virtual sources.

Further differences between the real and virtual sources were found for the response times and head movements of the subjects. For real sources, the subjects tend to directly look to the source, for virtual sources a slightly to-and-from movement of the head is present.

The results show that the method of dynamic binaural synthesis is very promising for localization tests in the context of sound field synthesis. The localization performance is only slightly degraded for anechoic HRTFs. On the other side, the method allows to compare instantaneously different position within a sound field. In addition, it was shown that the anechoic HRTF of a single loudspeaker can be inter- and extrapolated (in order to simulate different loudspeaker arrays) without degrading the results. In a current test, we use the method to evaluate the localization for WFS at several positions in the listening area.

## Acknowledgments

The authors of this work wish to thank Lars-Erik Riechert who helped with the experiments, and Pjotr Majdak for fruitful discussions on localization with non-individual HRTFs. This work was funded by DFG RA 2044/1-1.

## References

[1] Berkhout, A. J. et al. (1993). Acoustic control by wave field synthesis, J Acoust Soc Am, 2764-78

[2] Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. J Acoust Soc Am, 98(5), 2542-2553.

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

[4] Geier, M., Ahrens, J., Spors, S. (2008). The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods. *124th AES Conv.*

[5] Hess, W. (2004). Influence of head-tracking on spatial perception. *117th AES Conv.*

[6] Lewald, J., Dörrscheidt, G. J., Ehrenstein, W. H. (2000). Sound localization with eccentric head position. Behav Brain Res, 108(2), 105-25.

[7] Majdak, P., Laback, B., Goupell, M., Mihocic, M. (2008). The Accuracy of Localizing Virtual Sound Sources: Effects of Pointing Method and Visual Environment. *124$^{th}$ AES Conv.*

[8] Makous, J. C., Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. J Acoust Soc Am, 87(5), 2188-200.

[9] Mills, A. W. (1958). On the minimum audible angle. J Acoust Soc Am, 30(4), 237-246.

[10] Seeber, B. U. (2003). Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode. Technische Universität München.

[11] Start, E. (1997). Direct Sound Enhancement by Wave Field Synthesis. Technische Universiteit Delft.

[12] Wierstorf, H. et al. (2011). A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances. *130$^{th}$ AES Conv.*

[13] Wierstorf, H., Spors, S. (2012). Sound Field Synthesis Toolbox. *132$^{nd}$ AES Conv.*

[14] Wittek H. (2007). Perceptual differences between wavefield synthesis and stereophony, University of Surrey.

[15] `https://dev.qu.tu-berlin.de/projects/measurements/wiki/`

[16] `http://puredata.info/`