Audio Engineering Society

Convention Paper

Presented at the 128th Convention
2010 May 22–25 London, UK

Perceptual Assessment of Delay Accuracy and Loudspeaker Misplacement in Wave Field Synthesis

Jens Ahrens, Matthias Geier, and Sascha Spors

Quality and Usability Lab, Deutsche Telekom Laboratories, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany

Correspondence should be addressed to Jens Ahrens (jens.ahrens@telekom.de)

ABSTRACT

The implementation of simple virtual source models like plane and spherical waves in wave field synthesis employs delays which are applied to the input signals. We present a formal experiment evaluating the perceptual consequences of different accuracies of these delays. Closely related to the question of delay accuracy is the accuracy of the loudspeaker positioning. The second part of the presented experiment investigates the perceptual consequences of improperly placed loudspeakers. Dynamic binaural room impulse response based simulations of a real loudspeaker array are employed and a static audio scene setup is considered.

1. INTRODUCTION

Wave field synthesis (WFS) is an approach to the physical synthesis of sound fields over an extended receiver area by means of arrays of secondary sources (i.e. loudspeakers). Although dynamic scenarios including moving sources with Doppler effect [1, 2] are possible, we restrict the presented investigation to static scenarios.

The implementation of simple virtual source models like plane and spherical waves in WFS employs delays which are applied to the input signals. These delays can take values which are not equal to integer multiples of the sampling interval on a time discrete system but require the application fractional delays. For practical implementations the application of delays equal to integer multiples of the sampling interval are desired since these delays are computationally significantly more efficient than fractional delays. One aspect of the experiment presented in this paper is the investigation of the question whether such a quantization of the delays in practical implementations leads to a perceptual impairment.

Closely related to the question of delay accuracy is the accuracy of the loudspeaker positioning. The second aspect of the presented experiment is the investigation of the perceptual consequences of both random and systematic misplacement of the loudspeakers. An instrumentalized study of the latter subject based on simulations of sound fields can be found in [3].

The loudspeaker array employed in the experiments presented in this paper is the circular 56-channel array with a nominal radius of 1.495m installed at the Usability Laboratory at Deutsche Telekom Laboratories. In order to assure equal conditions for all subjects and in order to be able to seamlessly switch between different listening positions, dynamic binaural re-synthesis was performed based on measurements of the binaural room impulse responses (BRIRs) of each individual loudspeaker for different head orientations and listening positions.

2. WAVE FIELD SYNTHESIS

The theory of Wave Field Synthesis (WFS) was initially derived from the Rayleigh integrals which require the employed secondary source distributions to be linear in the two-dimensional case or to be planar in the three-dimensional case. A reformulation of the theory based on the Kirchhoff-Helmholtz integral revealed that also arbitrary convex distributions can be employed with only low error [4, 5]. As already mentioned in Sec. 1, a circular array was used in the presented experiment. Refer to Sec. 3 for details. We refer the reader to the literature such as [6, 7] for a detailed review of the theory of WFS. This section summarizes selected practical aspects.

The scenario considered is the synthesis of a virtual plane wave with constant propagation direction. The driving signal $d(\mathbf{x}_0, t)$ for a loudspeaker at position $\mathbf{x}_0$ in order to reproduce such a virtual plane wave can be calculated in time domain via [7]

$$d(\mathbf{x}_0, t) = w(\mathbf{x}_0) \cdot A(\mathbf{x}_0) \cdot [\, f(t) *_t s(t - \Delta t)\,] \ , \quad (1)$$

whereby $f(t)$ denotes the impulse impulse response of a filter with transfer function $\sqrt{\frac{1}{i\omega}}$, the asterisk $*_t$ denotes convolution with respect to time, and $s(t)$ denotes the input signal. The window $w(\mathbf{x}_0)$ represents the fact that not all loudspeakers of the circular array contribute to the reproduced sound field. It performs a *selection of active secondary sources* and equals 1 if the position $\mathbf{x}_0$ is *illuminated* by the virtual plane wave, and 0 if it is not [5, 7, 8].

The calculation of the driving signal for a loudspeaker at a given position in the illuminated area involves thus

- a filtering operation (represented by $f(t)$). This operation is equal for all loudspeakers and is therefore performed on the input signal directly. Note that this filtering is only applied below the spatial aliasing frequency [9].
- a weighting of the input signal. The weight $A(\mathbf{x}_0)$ is individual for each loudspeaker.
- a delaying of the input signal by Δt. The delay is also individual for each loudspeaker.

The filtering and weighting operation can be performed at the highest accuracy of the underlying system without compromising the computational efficiency.

Delays which are equal to integer multiples of the sampling interval on time discrete systems can be implemented efficiently using standard delay lines. For delays of other values, so-called fractional delays [10] have to be employed which are computationally costly.

The calculation of the driving signal (1) is typically implemented as a *driving function* which may be represented by an impulse response with which the input signal is convolved.

3. PREPARATION OF STIMULI

Stimuli were present to subjects via headphones in order to assure equal conditions for all subjects and in order to be able to seamlessly switch between different listening positions. To achieve a headphone simulation of a loudspeaker system which is a close as possible to a real loudspeaker system, the binaural room impulse responses (BRIRs) of the loudspeaker system installed at the usability laboratory of Deutsche Telekom Laboratories were measured using the FABIAN mannequin [11]. The loudspeaker system is a circular arrangement of a nominal radius of 1.495m and composed of 56 equiangularly spaced loudspeakers. Refer to Fig. 1 and 2.

Fig. 1: A section of the loudspeaker array which was re-synthesized via headphones. The system is circular with a nominal radius of 1.495m and is composed of 56 equiangularly spaced loudspeakers.

The BRIRs were measured from each of the loudspeakers to each ear of the mannequin for 161 head orientations (i.e. -80° to 80° with 1° resolution) for three listening positions (*center*, *side*, and *front*) resulting in 27048 pairs of impulse responses. The listening positions were chosen as indicated in Fig. 2 whereby the reference orientation of the mannequin was in positive y-direction. In order to exclude the influence of the reproduction room the impulse responses were carefully windowed in time domain such that only the direct sound from the loudspeakers is used. Throughout the experiment a temporal sampling rate of 44.1kHz was used. Refer to Sec. 4 for a description of the hardware and software setup.

The virtual loudspeakers system was driven in order to synthesize a virtual plane wave sound field with propagation direction in negative y-direction (refer to Fig. 2). From the listeners perspective, the plane wave was thus impinging “from the front”. The parameters for the different stimuli are outlined in Sec. 3.1 to 3.3 and are summarized in table 1. Sample stimuli can be downloaded from [12].

For each head orientation, the driving function of each loudspeaker (refer to Sec. 2) was convolved with that pair of BRIRs representing the given head orientation and the result was added for all loudspeakers. Each stimulus was thus represented by a pair of impulse responses (left and right ear) which in turn represent the spatio-temporal transfer function of the loudspeaker system driven with the given configuration to the ears of the mannequin for a given head orientation [13]. This spatio-temporal transfer function was then calculated for all possible head orientations. The headphone signal was then obtained by convolving a given input signal with the BRIRs representing the entire loudspeaker system as described above.

Crucial for the presented experiment is the accuracy of the loudspeaker placement. The loudspeaker rig was manufactured by a company using state-of-the-art exhibition stand construction methods. Manual measurements with a laser distance meter did not reveal deviations larger than the measurement accuracy of this measurement procedure.

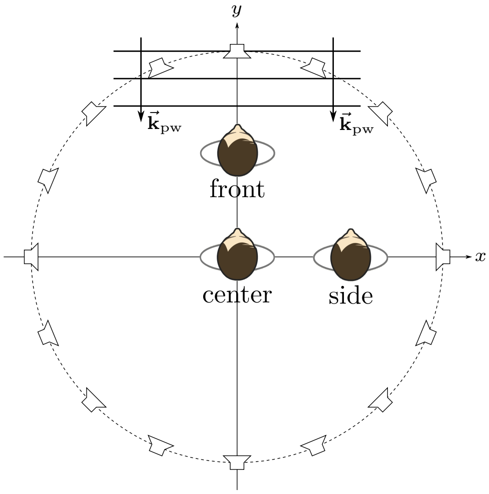

y
$\vec{\mathbf{k}}_{\text{pw}}$
$\vec{\mathbf{k}}_{\text{pw}}$
front
x
center
side

Fig. 2: Schematic of the simulated setup. At the non-central listening positions the center of the head is at half-way between center and loudspeaker contour.

3.1. Delay Accuracy

Two different basic types of delays were employed for testing the perceptual consequences of different delay accuracy: 1) fractional delays using Lagrange interpolation [10], and 2) delays equal to integer multiples of the time-domain sampling interval. The fractional delays were implemented as convolution with impulse responses representing the respective delay. The toolkit provided at [14] was used. The integer delays were implemented using standard delay lines.

The different accuracies tested are listed in table 1. In the table e.g. 'f10' refers to a fractional delay of 10th order, and 'i2' refers to a delay which is quantized in steps equal to two times the sampling interval. Note that the experiment was carried out at 44.1kHz sampling frequency so that the sampling interval equals 1/44100s ($\approx 23\mu$s).

The 'f10'-delay served as reference throughout the entire experiment. Informal pre-testing suggested that further increasing the accuracy of the fractional delay does not lead to audible changes.

The stimuli were aligned in time in order to compensate for the unavoidable pre-delay introduced due to the Lagrange interpolation to allow seamless switching between the stimuli.

3.2. Radius Accuracy

As mentioned above, the estimated radius of the circular loudspeaker array is 1.495m. The radius was deduced from the diameter of the system which was measured with a laser distance meter between the assumed location of the acoustical centers of the tweeters of two opposing loudspeakers. The measurement was confirmed at multiple locations.

Note that the assumed location of the acoustical center of the low/mid-range drivers is on a radius of approximately 1.53m. Preliminary investigations suggested that misplacement of the loudspeakers primarily affects high-frequency content. Therefore, it was decided to use the radius on which the tweeters are positioned as reference. We term the latter radius *nominal radius*.

The radii used in the calculation of the driving signals are listed in table 1.

3.3. Random Loudspeaker Misplacement

In order to simulate a random misplacement of individual loudspeakers, random delays/antizipations were added to the individual loudspeaker signals. These delays/antizipations are uniformly distributed within a given range such that they represent a radial misplacement of the loudspeakers with respect to the listening position. Table 1 lists the ranges of simulated radial displacement in meters which were employed in the test. A value of e.g. 0.002m means that delays/antizipations were applied on the loudspeaker signals which correspond to a radial displacement in the range of ±0.002m.

4. HARDWARE/SOFTWARE SETUP

Audio processing and graphical user interface (GUI) were running on a single computer which was located outside the room in which the test was performed. The GUI was implemented in Python.

Realtime auralization of the stimuli was performed using the SoundScape Renderer (SSR) [15, 16], an open-source real-time spatial audio framework, running in *binaural room scanning* mode. The BRIR sets off all stimuli were loaded into memory and an input port was created for each BRIR set. The GUI indicated the desired audio file and stimulus condition as text messages sent via TCP/IP to Pure Data(Pd [17]) which in turn replayed the audio to that input of the SSR which corresponded to the desired test condition. Due to the internal processing in the SSR this switching of the audio file between different inputs leads to a smooth cross-fade with raised-cosine shaped ramps.

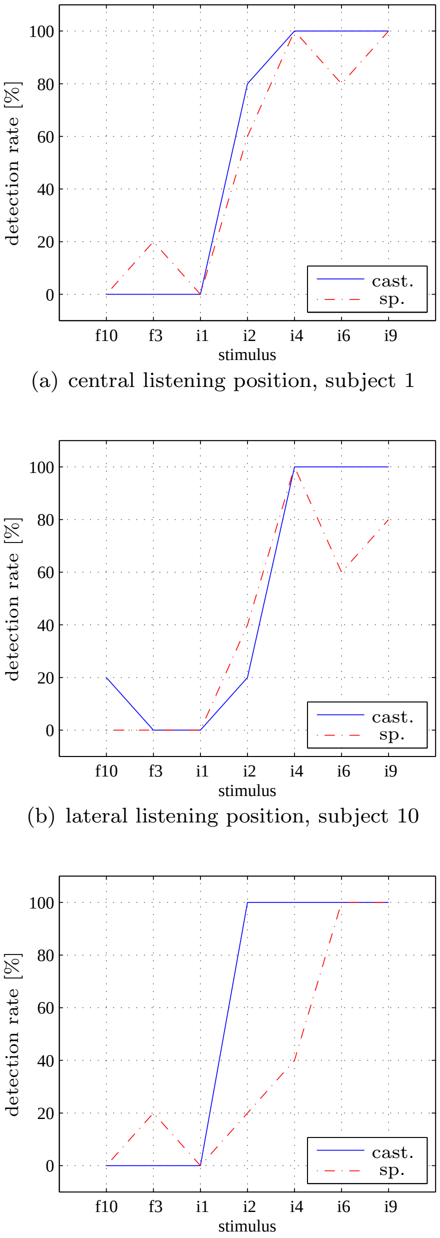
The SSR then convolved the input signal in realtime with that pair of impulse responses corresponding to the instantaneous head orientation of the test subject as indicated by a Polhemus Fastrack tracking system.

AKG K601 headphones were used with a compensation of the transfer function applied [18]. The experiment was carried out with the subject positioned inside the real loudspeaker array at a potential real listening position in order to support the headphone simulation with the appropriate visual impression and acoustical environment.

5. TEST PROCEDURE

delay types	radii (m)	rand. displacement (m)
'f10'	1.1	**0.000**
'f3'	1.3	0.002
'i1'	1.45	0.005
'i2'	1.485	0.01
'i4'	**1.495**	0.02
'i6'	1.505	0.03
'i9'	1.55	
	1.8	
	2.0	

Table 1: Summary of parameters employed in the calculation of the driving signals. The reference conditions are written in boldface.

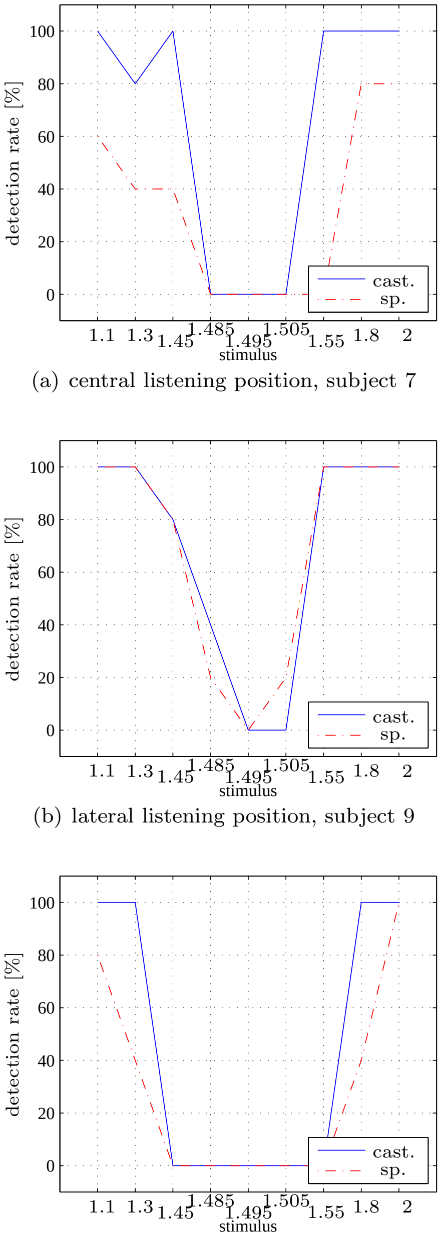


(a) central listening position, subject 1

(b) lateral listening position, subject 10

(c) frontal listening position, subject 5

Fig. 3: Representative individual difference detection rates for the delay accuracy condition. The horizontal axes use arbitrary scaling.



(a) central listening position, subject 7

(b) lateral listening position, subject 9

(c) frontal listening position, subject 10

Fig. 4: Representative individual difference detection rates for the radius accuracy condition. The horizontal axes use arbitrary scaling.

The test was designed as a pairwise comparison of a given stimulus and the according reference whereby it was not indicated which of the two stimuli in a pair was the reference. For each stimulus, the reference stimulus at the corresponding simulated listening position and with highest implemented delay accuracy was used. Stimulus pairs were presented in random order. Each possible pair of stimuli was repeated 5 times throughout the test. The subjects' task was to indicate whether or not they hear a difference between the two stimuli of a given pair or not.

10 subjects (both male and female, expert and non-expert listeners; all reported not to be aware of any hearing impairment) participated in the test which was performed in two session for each subject. Female speech and castanet samples each of approximately 7 seconds duration were used as input signals. In each of the sessions exclusively one type of input signal was used. Each session was composed of a training of 20 stimuli pairs followed by 3 runs of approximately 100 stimuli each. Duration of each session was between 35 minutes and 45 minutes.

In each of the 3 runs in one session only one parameter was tested (either radius or delay accuracy, or random displacement). After each run, subjects were asked to describe in free text what kind of differences they detected in those cases where they did detect differences.

The GUI was operated by subjects via a keyboard. The space bar (operated by the left hand) was used in order to switch between stimuli, and two different keys operated by the right hand were used to indicate either *I hear a difference* or *I do not hear a difference*. The keyboard operation was considered convenient and efficient. Once an answer was given by the subject the next pair of stimuli was immediately presented.

6. RESULTS

Representative individual results are presented in Fig. 3-5 and results accumulated over all subjects and input signals in Fig. 6. It can be deduced from that subjects were very reliable in detecting the reference stimulus, i.e. when reference stimulus was compared to itself, no difference was perceived with very few exceptions. Accordingly, for the stimuli the parameters of which departed strongest from the reference stimulus difference detection rates are nearly 100%. Typically, a smooth transition in the difference detection rate between the reference stimulus and those stimuli with strongest modified parameters occurs. Generally, no obvious differences between the performances of expert and non-expert listeners were detected.

The results of the different conditions are analyzed in detail in the following sections.

6.1. Detection of Differences

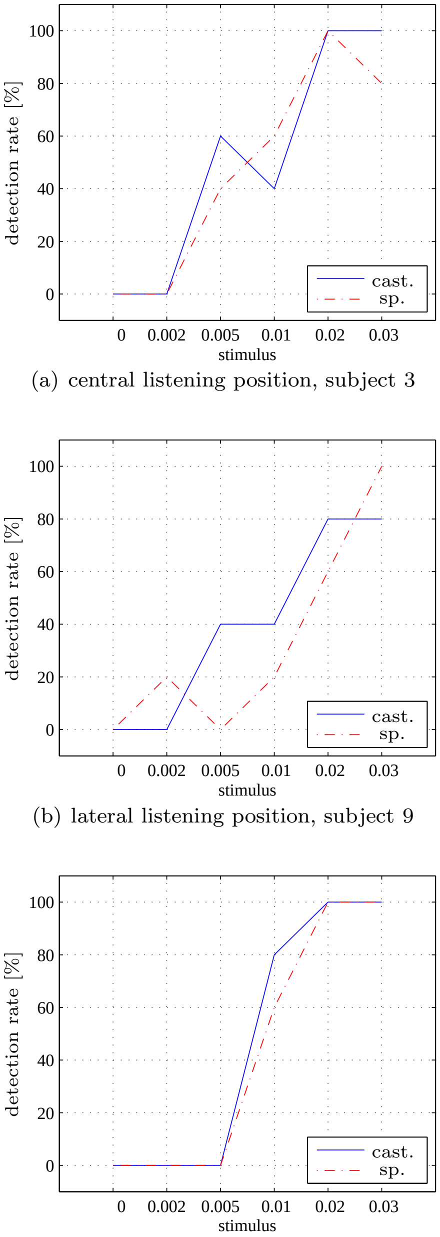
6.1.1. Delay Accuracy

Representative individual difference detection rates for varying delay accuracy are depicted in Fig. 3. Observations are summarized below.

- No obvious difference in the detection rates between central, lateral, and frontal listening position (Fig. 3(a), 3(b), and 3(c) respectively) can be observed.
- No obvious difference between the input signals can be observed whereby detection rates are occasionally slightly lower for the speech signal than for castanets.
- The difference detection rates for varying accuracies of the involved delays are typically between 0% and 20% for the reference compared to itself as well as the 'f3' and 'i1' stimuli. This means that out of the 5 presentations of one stimulus pair a difference was perceived at maximally one single presentation.
- For all other conditions (i.e. 'i2', 'i4', 'i6', 'i9'), detection rates are generally high.
- The 'i1' condition represents thus the lower bound of the delay accuracy which is indistinguishable from highest accuracy.
- Fig. 6(a) depicts the detection rates accumulated over all subjects, listening, positions, and input signals in order to indicate what has to be expected when neither the listening position nor the input signal is known. Observations are similar to those of the individual results described above.

6.1.2. Radius Accuracy

Representative individual difference detection rates for varying radius accuracy are depicted in Fig. 4. Observations are summarized below.

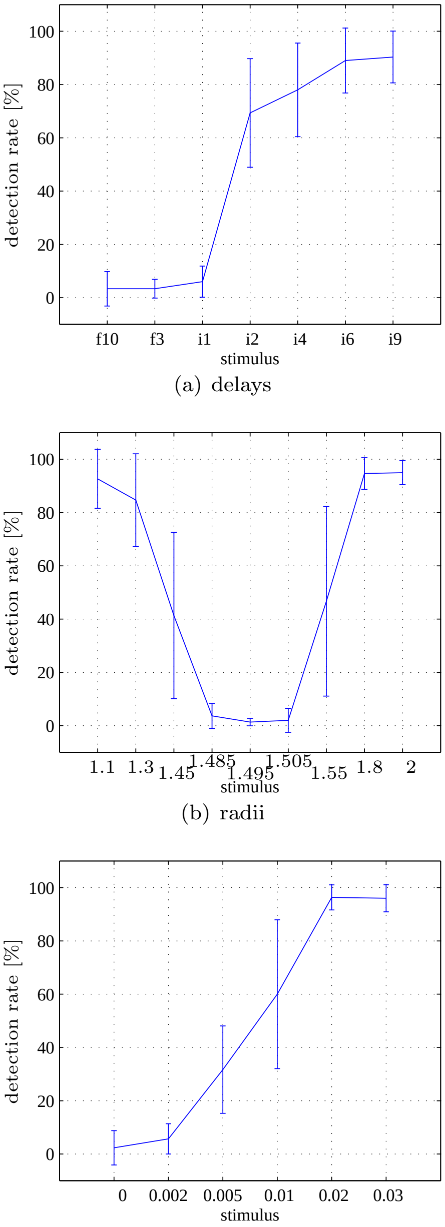


(a) central listening position, subject 3

(b) lateral listening position, subject 9

(c) frontal listening position, subject 5

Fig. 5: Representative individual difference detection rates for the random displacement condition. The horizontal axes use arbitrary scaling.



(a) delays

(b) radii

(c) random displacement

Fig. 6: Difference detection rates of all subjects accumulated over all listening positions and input signals. The errorbars indicate the standard deviation for the detection rates of the individual subjects for the individual input signals. The horizontal axes use arbitrary scaling.

- The frontal listening position (Fig. 4(c)) appears to be least critical, the lateral listening position (Fig. 4(b)) appears to be most critical.
- Detection rates are generally lower for the speech signal than for castanets.
- For radii which deviate by more than 20cm from the reference radius, detection rates are generally very high.
- The highest undetectable radius inaccuracy for current setups is thus between 1cm and 5cm depending on the listening position.
- The accumulated results depicted in Fig. 6(b) confirm above described observations.
- The symmetry of the detection rates with respect to the 1.495 m condition (especially in the accumulated results in Fig. 6(b)) indicates that the choice of using the 1.495 m condition as reference was reasonable.

6.1.3. Random Displacement

Representative individual difference detection rates for a simulated random displacement of the individual loudspeakers are depicted in Fig. 5. Observations are summarized below.

- The slope of detection rates is typically steeper for the frontal listening position (Fig. 5(c)) than for the central and lateral listening positions (Fig. 5(a) and 5(b)).
- No obvious difference between the input signals can be observed whereby detection rates are occasionally slightly lower for the speech signal than for castanets.
- The difference detection rates for random displacements in the ranges of larger than 0.01m (i.e. 1cm) are typically high.
- The accumulated results depicted Fig. 6(c) confirm above described observations.

6.2. Comments of Subjects

As mentioned in Sec. 5, subjects were asked after each run to described freely what differences they detected. The answers were rather fuzzy and somewhat inconsistent. While some subjects mentioned primarily timbral attributes for a given condition, other mention primarily spatial attributes. In general, answers were composed of one or several of the following attributes:

- timbre
- distance of the virtual source
- amount of reverberation
- apparent size of the virtual source

The only condition which lead to rather consistent answers was the radius-accuracy condition in combination with the lateral listening position. In this case, subjects frequently reported changes in the position of the virtual source with respect to the direction. Analysis of the resulting sound fields when an incorrect radius is assumed in the calculation of the driving functions shows that this circumstance can lead to an incorrect curvature of the wave front which obviously affects localization.

Although not essentially represented in the difference detection rates, the subjects reported that the detection task was perceived to be significantly more difficult for the speech input signal than for the castanets.

7. CONCLUSIONS

We have presented a formal experiment based on a dynamic binaural re-synthesis of a real loudspeaker system in order to assess the perceptual consequences of varying delay accuracy in the calculation of the driving signals and misplacement of the loudspeakers.

We have shown that a quantization of the delays to be applied to the input signals in steps of one time sampling interval at 44.1kHz sampling frequency is perceptually indistinguishable from higher accuracy. Lower accuracy in turn leads to timbral coloration as well as differences in spatial attributes like perceived size of the virtual source and its distance. It is a remarkable circumstance that the sampling interval at a sampling frequency which is just high enough to perfectly represent a continuous signal over the entire audible frequency range represents also that quantization of the delay which is just inaudible for static scenarios.

We showed that a systematic loudspeaker misplacement like incorrect dimensions of the loudspeaker setup in the calculation of the loudspeaker driving signals affects localization of the virtual source in terms of a position bias and apparent source width for some listening positions. Also timbral coloration occurs. We assume that other systematic deviations of involved parameters like an incorrectly estimated speed of sounds leads to similar perceptual impairments. Additionally, the results suggest that the position of the acoustical centers of the tweeters shall be used as position of the loudspeaker in the calculation of the driving signals.

Random loudspeaker misplacement leads also to trimbral coloration and changes in spatial parameters like perceived distance and size of the virtual source. The required accuracy in the loudspeaker position is in the range of a few centimeters.

In general, no substantial dependence of the perception on the input signal has been detected.

Note that the presented results represent an overcritical assessment: 1) They were retrieved under laboratory conditions and 2) an A/B-comparison with a reference stimulus was performed which allows to detect even the slightest perceptual differences between stimuli. We assume that a practical situation is significantly less critical since the influence of the reproduction room is included and a comparison to a reference is not possible.

Finally, note that the presented results only hold for static scenarios. Dynamic scenarios (i.e. moving sources) exhibit peculiar properties with respect to a number of practically relevant parameters [1, 2]. In particular, the delays applied on the input signals vary over time so that it is likely that delay accuracy is more crucial than with static scenarios. A formal evaluation is in preparation.

ACKNOWLEDGEMENTS

The thank Alexander Lindau of TU Berlin for performing the BRIR measurements.

8. REFERENCES

[1] A. Franck, A. Gräfe, T. Korn, and M. Strauß. Reproduction of moving virtual sound sources by wave field synthesis: An analysis of artifacts. In *32nd Int. Conference of the AES*, Hillerød, Denmark, Sept. 2007.

[2] J. Ahrens and S. Spors. Reproduction of moving virtual sound sources with special attention to the Doppler effect. In *124th Conv. of the AES*, Amsterdam, The Netherlands, May 17–20 2008.

[3] M. Strauß and M. Munderloh. Influence of loudspeaker displacement on the reproduction quality of wave field synthesis systems. In *19th Int. Congress on Acoustics (ICA)*, Madrid, Spain, Sept. 2–7 2007.

[4] E. W. Start. Application of curved arrays in wave field synthesis. In *100th Convention of the AES*, Copenhagen, Denmark, May 11–14 1996.

[5] J. Ahrens and S. Spors. On the secondary source type mismatch in wave field synthesis employing circular distributions of loudspeakers. In *127th Convention of the AES*, New York, NY, Oct. 9–12 2009.

[6] A.J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *JASA*, 93(5):2764–2778, May 1993.

[7] S. Spors, R. Rabenstein, and J. Ahrens. The theory of wave field synthesis revisited. In *124th Convention of the AES*, Amsterdam, The Netherlands, May 17–20 2008.

[8] F. Fazi, P. Nelson, and R. Potthast. Analogies and differences between 3 methods for sound field reproduction. In *Ambisonics Symposium*, Graz, Austria, June 25–27 2009.

[9] S. Spors and J. Ahrens. Analysis and improvement of pre-equalization in 2.5-dimensional wave field synthesis. In *128th Convention of the AES*, London, UK, May 22–25 2010.

[10] T. I. Laakso, V. Vlimki, M. Karjalainen, and U. K. Laine. Splitting the unit delay. *IEEE Signal Proc. Magazine*, Jan. 1996.

[11] A. Lindau and S. Weinzierl. FABIAN - An instrument for the software-based measurement of binaural room impulse responses in multiple

degrees of freedom. In *Proc. 24. Tonmeistertagung (VDT International Convention)*, Leipzig, Germany, Nov. 16–19 2006.

[12] Quality and Usability Lab. Spatial Audio Research - Audio research at Quality and Usability Lab. http://audio.qu.tu-berlin.de.

[13] M. Geier, J. Ahrens, and S. Spors. Binaural monitoring of massive multichannel sound reproduction systems using model-based rendering. In *NAG/DAGA International Conference on Acoustics*, Rotterdam, The Netherlands, March 2009.

[14] T. I. Laakso, V. Vlimki, M. Karjalainen, and U. K. Laine. Splitting the unit delay - tools for fractional delay filter design. http://www.acoustics.hut.fi/software/fdtools, Retrieved Dec. 09.

[15] M. Geier, S. Spors, and J. Ahrens. The SoundScape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods. In *124th Conv. of the AES*, Amsterdam, The Netherlands, May 17–20 2008.

[16] The SoundScape Renderer team. The SoundScape Renderer. http://www.tu-berlin.de/?id=ssr.

[17] PD-community. Pure Data. http://puredata.info.

[18] Z. Schärer and A. Lindau. Evaluation of equalization methods for binaural signals. In *126th Convention of the AES*, Munich, Germany, May 7–10 2009.