Perceptional comparison of wave field synthesis implemented with integer and fractional delays

Jens Ahrens, Matthias Geier, and Sascha Spors
Deutsche Telekom Laboratories, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany

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 loudspeakers. 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 [1]. For practical implementations the application of delays equal to integer multiples of the sampling interval is desired since these delays are computationally significantly more efficient than fractional delays. The experiment presented in this paper investigates the question whether such a quantization of the delays in practical implementations leads to a perceptional impairment.

Preparation of Stimuli

The theory of WFS was initially derived from the Rayleigh integrals and has later been extended based on the Kirchhoff-Helmholtz integral. The calculation of the according driving signal for a loudspeaker at a given position involves a filtering operation and a weighting and delaying of the input signal. We refer the reader to the literature such as [2] for a detailed review of the theory of WFS.

Stimuli were present to subjects via headphones. Head-tracking was applied. The binaural room impulse responses of a real loudspeaker system were measured using the FABIAN mannequin. The loudspeaker system is a circular arrangement of a nominal radius of 1.495m and composed of 56 equiangularly spaced loudspeakers. Measured listening positions were center as well as half-way between center and loudspeakers to the side and to the front. The virtual loudspeakers system was driven in order to synthesize a virtual plane wave sound field impinging “from the front” from the listeners perspective. Sample stimuli can be downloaded from [3].

Two different basic types of delays were employed for testing the perceptual consequences of different delay accuracy: 1) fractional delays using Lagrange interpolation [1], and 2) delays equal to integer multiples of the time-domain sampling interval. The different accuracies tested are 'f10', 'f3', 'i1', 'i2', 'i4', 'i6', 'i9' whereby 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). Realtime auralization of the stimuli was performed using the SoundScape Renderer (SSR) [4], an opensource real-time spatial audio framework.

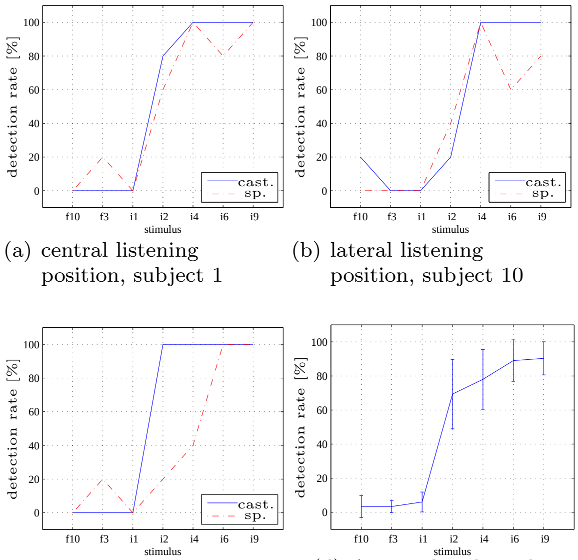
Test Procedure

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. The subjects task was to indicate either *I hear a difference* or *I do not hear a difference* via dedicated buttons. 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.

10 subjects (both male and female, aged 25-37, expert and non-expert listeners, no one reported being aware of any hearing impairment) participated in the test. Female speech and castanets were used as input signals.

Results

Representative individual results are presented in Fig. 1(a)-(c) and results accumulated over all sub-



(a) central listening position, subject 1

(b) lateral listening position, subject 10

(c) frontal listening position, subject 5

(d) Accumulated results.

Figure 1: Representative individual and overall difference detection rates. The horizontal axes use arbitrary scaling.

jects and input signals in Fig. 1(d). The errorbars in Fig. 1(d) indicate the standard deviation of the detection rates w.r.t. the individual results. It can be deduced from the figures 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 with lowest delay accuracies detection rates are nearly 100%. Typically, a smooth transition in the difference detection rate between the reference stimulus and those stimuli with lowest accuracy occurs. Generally, no obvious differences between the performances of expert and non-expert listeners were detected. Further observations are summarized below.

- No obvious difference in the detection rates between different listening positions 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.
- For conditions '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.

Comments of Subjects

Subjects were asked after each run to describe freely what differences they detected. The answers were rather fuzzy and somewhat inconsistent. While some subjects mentioned primarily timbral attributes others mention primarily spatial attributes. In general, answers were composed of one or several of the attributes timbre, distance of the virtual source, amount of reverberation, and apparent size of the virtual source.
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.

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 in WFS. It was shown that delays quantized in integer multiples of the sampling interval are perceptionally indistinguishable from delays of higher accuracies. Note that the presented results only hold for static scenarios. Dynamic scenarios deserve special attention since the duration of the delays varies over time. An according experiment in preparation.

References

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

[2] 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.

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

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