Binaural Monitoring of Massive Multichannel Sound Reproduction Systems Using Model-Based Rendering

M. Geier, J. Ahrens and S. Spors

Deutsche Telekom Laboratories, Technische Universität Berlin, Germany, E-Mail: Matthias.Geier@telekom.de

Introduction

Nowadays, several massive multichannel sound reproduction systems are available in research institutes and entertainment venues. They utilize a large number of output channels (up to several hundred channels) and provide an extended listening area. The most widespread reproduction methods in this context are Wave Field Synthesis and Higher-Order Ambisonics.

The process of producing content for such systems is often not satisfying, because time for on-site set-up and fine-tuning of the performance is limited and expensive.

This paper investigates how binaural monitoring can be used to make the process easier and cheaper by doing most of the production work using headphones. Only final adjustments have to be done on the target system.

Data-based vs. Model-based

In data-based rendering, soundfields are recorded with microphone arrays. For reproduction, the measured signals are extrapolated according to the given loudspeaker positions. The limitations of this extrapolation procedure are not known in detail. Additionally, this method is computationally expensive, needs to handle large amounts of measurement data and a realtime implementation is very hard to achieve.

In model-based rendering, on the other hand, *sound objects* are rendered according to a scene description. The same scene can then be reproduced with different rendering methods. One example of a storage format for such a scene is the *Audio Scene Description Format* (ASDF) [1].

For model-based rendering separate source signals are needed, which are typically dry recordings. If desired, room information can be added separately, either by a room-acoustical model or by measured room impulse responses.

Several software applications for model-based rendering are available, in the context of this paper the *SoundScape Renderer* [2] was used.

Different Methods

There are two different methods for using binaural monitoring. The first one is to simulate a given loudspeaker system, authentically reproducing its strengths and weaknesses. This includes reproduction of artifacts like coloration, spatial aliasing and array truncation effects [3, 4, 5].

Simulation of the loudspeaker system can be done by using Binaural Room Impulse Response (BRIR) measurements of each loudspeaker. The BRIRs have to be measured separately for each combination of listener position and loudspeaker. To be able to use head-tracking, all BRIRs have to be measured for all possible head orientations, usually in angular increments of one to five degree. This leads to several hundreds or thousands of BRIRs which have to be handled by the reproduction system. Depending on the reverberation time of the reproduction room and therefore the length of the BRIRs, the rendering process can be computationally very demanding. To reduce the computational complexity, the loudspeaker system can be measured without room information, i.e. in an anechoic room, resulting in much shorter impulse responses. However, the measurement effort may be the same (or even more) and the result will sound less realistic.

Possible applications of this first approach are subjective system evaluation and binaural documentation/archiving of spatial reproduction scenarios.

The other alternative to employ binaural rendering is the reproduction of the original virtual audio scene disregarding the actual loudspeaker-based reproduction system. For this task, free-field Head Related Impulse Responses (HRIRs) can be used, which needs much less measurements, makes the rendering process less complex than in the BRIR-case and allows arbitrary virtual listener positions. However, room information is not included and the artifacts caused by the loudspeaker system are not reproduced. Both room information and rendering artifacts can be added separately, if desired.

HRIRs are typically measured for only one distance and the signal is attenuated or amplified to simulate different source-listener distances. For more accuracy, HRIR measurements can be done for several distances [6].

Suitable applications for the latter approach are scene authoring and live monitoring (e.g. in theater or DJ performances).

BRIR Measurements

The measurements used in the context of this paper were done for a circular array with a diameter of three meters consisting of 56 loudspeakers installed in a room with a volume of about $50\,\mathrm{m}^3$. The room was acoustically treated to get a reverberation time (T_{60}) of 0.1 seconds at 1000 Hz. The exitation signal was a colored sinus sweep with a 20 dB emphasis for frequencies below 100 Hz and a band-limitation from 50 Hz to 20 kHz. Its length was chosen to be about 1.5 seconds to achieve a satisfying signal to noise ratio. The measurements were done for head orientations in a range of ±80° in increments of one degree resulting in $161 \times 56 = 9016$ impulse responses (for each ear). The whole procedure, which takes about 8 hours, was repeated for several listener positions, however, for the test described in this paper, only the measurements for one listener position in the center of the loudspeaker array were used. The BRIRs were measured with the FABIAN system which was developed at TU Berlin [7].

Tests

The objective of the informal tests conducted for this paper was to get an overview over the qualitative differences between the reproduction and monitoring methods which are presented below. The different methods were presented to a handful of expert listeners who then described their perception of sound color and several spatial attributes with regard to the difference or similarity of the stimuli.

Only Wave Field Synthesis (WFS) was used as loudspeaker based reproduction method. For all binaural setups, head tracking was used. This improves localisation and externalisation of virtual sources [8]. The binaural signals were equalized to minimize the influence of the headphones [9].

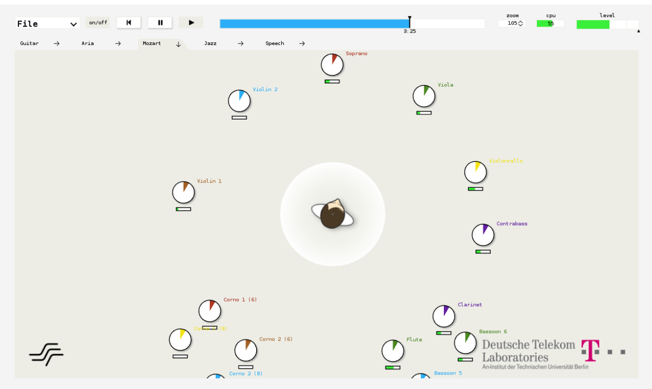

Figure 1: Screenshot of the *SoundScape Renderer* playing a scene consisting of anechoic recordings of a Mozart aria from [10].

Three different audio scenes have been utilized, featuring speech (three virtual sound sources), mallet percussion (22 sources) and music for chamber orchestra (14 sources from [10], see figure 1). Only static scenes with fixed source positions were used. All experiments were done in the same room, therefore the same visual cues were present for all experiments.

The different setups which were compared in the tests are presented in the following sections.

Setup 1, "pure" binaural rendering

In this setup, the virtual audio scene is directly rendered to a binaural signal without taking into account the original loudspeaker reproduction system. The dry source signals are convolved with generic free-field HRIRs (see figure 2). This was done with the *SoundScape Renderer* operating in "binaural" mode. No additional room information was added to the signal, no WFS artifacts were simulated.

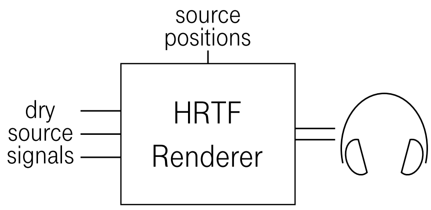

Figure 2: "pure" binaural rendering

Setup 2, "real" WFS

This setup was the reference condition. The *SoundScape Renderer* was used (in "WFS" mode) to generate the loudspeaker signals from the source signals (see figure 3). A circular array of 56 loudspeakers with a diameter of three meters was used.

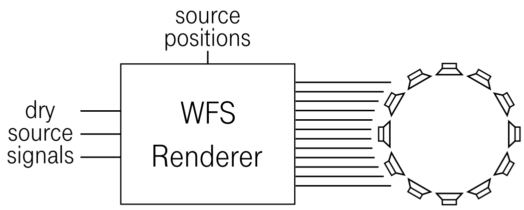

Figure 3: WFS rendering

Setup 3, "virtual" WFS

In this setup, the loudspeaker signals are convolved with the measured (BRIRs) of the loudspeakers.

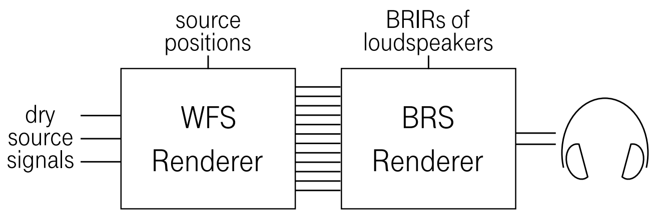

Figure 4: "virtual" WFS rendering – direct approach

To implement "virtual" WFS directly (as shown in figure 4) would be inefficient because two separate rendering stages are needed. First, 56 loudspeaker signals would need to be generated with a WFS renderer and then, all of the loudspeaker channels would have to be convolved with their respective BRIRs. This can be done much more efficiently by applying (for each loudspeaker) the WFS driving functions (depending on the source position) onto the measured BRIRs of the loudspeaker (depending on the head orientation) [11]. The modified BRIRs can then be summed up over all loudspeakers resulting in only one set of BRIRs for each virtual source position. For convenience, the combination of the BRIRs and the WFS driving functions was done off-line using MATLAB© functions from [12]. Using the combined BRIRs, the binaural signal was generated with the *SoundScape Renderer* operating in "BRS" (*Binaural Room Scanning*) mode as shown in figure 5.

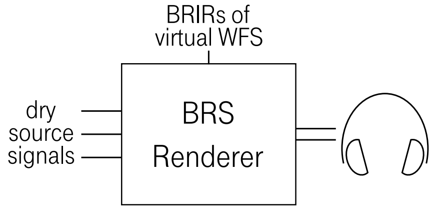

Figure 5: "virtual" WFS rendering – combined approach

Results and Conclusions

The tests have shown that "virtual" WFS using BRIRs of the original loudspeaker system comes very close to "real" WFS in both sound color and spatial impression. However, it requires much effort for the measurement of the BRIRs which have to be done for each loudspeaker setup and for each desired listener position and orientation individually.

"Pure" binaural rendering of the original scene is very useful for monitoring because of its simplicity and flexibility but it exhibits a different sound quality which leaves much room for improvement.

It was observed that headphone equalization of the binaural signals is essential for spatial impression and natural sound color [9]. Naturally, headphones have a limited response in the low frequency range, this could be enhanced by using a subwoofer. If binaural monitoring is used for scene authoring, this limitation has to be taken into account and the low frequencies have to be adjusted in the reproduction room.

The localisation and externalisation of sound sources is very convincing in the "virtual" WFS setup but not quite as good in "pure" binaural rendering. This is most probably due to the lack of room information in the free-field HRIRs. It has to be investigated in future studies if virtual room acoustics can improve that and lead to a spatial impression closer to "virtual" WFS and ultimately "real" WFS.

Further Work

The findings of this paper have to be backed up with formal listening tests.

Dynamic scenes have to be investigated. These are scenes where sources are moving automatically or where sources can be manipulated in realtime by the test subjects. This is a crucial point since the artifacts produced by loudspeaker systems can become very prominent for moving sources [13].

An efficient implementation of "virtual" WFS can be achieved by an on-line combination of loudspeaker-BRIRs and the corresponding WFS driving functions.

The binaural rendering with free-field HRIRs can be extended with a room acoustic simulation. Its impact on the spatial perception has to be evaluated.

Furthermore, it should be evaluated how much influence the presence or absence of visual cues (e.g. the loudspeaker array) has on the perceived spatial impression.

Acknowledgements

We thank Alexander Lindau from Technische Universität Berlin for the BRIR measurements.

References

[1] M. Geier, J. Ahrens and S. Spors. ASDF: Ein XML Format zur Beschreibung von virtuellen 3D-Audioszenen. In *34. Jahrestagung für Akustik (DAGA)*. Dresden, Germany, March 2008.

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

[3] J. Ahrens and S. Spors. Alterations of the temporal spectrum in high-resolution sound field reproduction of different spatial bandwidths. In *126th Convention of the AES*. Munich, Germany, May 2009.

[4] S. Spors and J. Ahrens. A comparison of Wave Field Synthesis and Higher-Order Ambisonics with respect to physical properties and spatial sampling. In *125th Convention of the AES*. San Francisco, CA, USA, October 2008.

[5] S. Spors, R. Rabenstein and J. Ahrens. The theory of Wave Field Synthesis revisited. In *124th Convention of the AES*. Amsterdam, Netherlands, May 2008.

[6] T. Lentz et al. Precise near-to-head acoustics with binaural synthesis. *Journal of Virtual Reality and Broadcasting*, 3(2), 2006.

[7] A. Lindau and S. Weinzierl. FABIAN – An instrument for software-based measurement of binaural room impulse responses in multiple degrees of freedom. In *24. Tonmeistertagung*. Leipzig, Germany, 2006.

[8] P. Minnaar et al. The importance of head movements for Binaural Room Synthesis. In *Proc. ICAD*. Espoo, Finland, 2001.

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

[10] T. Lokki, J. Pätynen and V. Pulkki. *Anechoic Recordings of Symphonic Music*. Helsinki University of Technology, 2008.
`http://auralization.tkk.fi/node/4`.

[11] H. Wittek. *Perceptual Differences Between Wavefield Synthesis and Stereophony*. Ph.D. thesis, Department of Music and Sound Recording, School of Arts, Communication and Humanities, University of Surrey, 2007.

[12] M. Frank. *Perzeptiver Vergleich von Schallfeldreproduktionsverfahren unterschiedlicher räumlicher Bandbreite*. Master's thesis, Institute of Electronic Music and Acoustics, Graz, Austria, 2009. Work in progress.

[13] J. Ahrens and S. Spors. Reproduction of moving virtual sound sources with special attention to the doppler effect. In *124th Convention of the AES*. Amsterdam, Netherlands, May 2008.