Audio Engineering Society
Convention Paper
Presented at the 117th Convention
2004 October 28–31 San Francisco, CA, USA

Head-Tracked Auralization of Acoustical Simulation

Christoph Moldrzyk[1], Wolfgang Ahnert[2], Stefan Feistel[2], Tobias Lentz[3], and Stefan Weinzierl[4]

[1] Technical University Berlin, FSP-PV/PRZ, Sekr. MA 073, Str. d. 17. Juni 136, 10623 Berlin, Germany
aural@prz.tu-berlin.de

[2] ADA Acoustic Design Ahnert, Arkonastr. 45-49, 13189 Berlin, Germany
wahnert@ada-acousticdesign.de, ease@auralisation.de

[3] RWTH Aachen University, Institute of Technical Acoustics, Neustr. 50, 52066 Aachen, Germany
tle@akustik.rwth-aachen.de

[4] Technical University Berlin, Institute of Communication Research, Sekr. EN-8, Einsteinufer 17, 10587 Berlin, Germany
stefan.weinzierl@tu-berlin.de

ABSTRACT

A desirable feature of modern acoustical simulation programs is the easy, fast and reliable auralization of prediction results. To be considered as a serious tool, the auralization results should be equivalent to human perception in reality. In this paper we consider a new auralization technique, based on a head-tracked headphone system with high spatial resolution and real-time convolution. We discuss the measurement of directional head-related transfer functions, the calculation of directional binaural impulse responses and the realisation as a real-time convolution software. A listening test was performed, comparing reality, measurement and prediction results for an example room.

1. INTRODUCTION

1.1. Auralization

For a reliable auralization of simulation the sound of the simulated room must be close to reality. It is not only important to have a high quality sound-system and a correct spatial representation of the complex sound field. Also the listener should have the opportunity to interact with the scene and to behave naturally. As an example, head rotation causes a variation of the sound, the direction of the source and the directions of the reflections. The system should react to these changes. In a first approach, presented here, the system only provides head rotation in the horizontal plane, because this causes the largest changes in the Room Impulse Response (RIR) and because the perceived direction of the source will be highly dependent on alterations of the Interaural Time Difference (ITD) and Interaural Level Difference (ILD).

For an auralization, different approaches of reproduction systems are possible. For a variety of resasons [1], binaural representation technology should give the most reliable (comparable to reality) reproduction results. This can be done by crosstalk-cancelled loudspeakers (transaural) or by headphones (binaural). The system presented here is using head-tracked headphones for reproduction due to practical (application) reasons: without any room-acoustical treatment highly immersive acoustical environments are possible. Due to head-tracking, low latency and high angular resolution, the system also avoids the effects of in-head-localization and front/back ambiguity.

1.2. Spatial representation (binaural approach)

For visual stimuli, the brain compares pictures from both eyes to determine the objects' placement in a scene. This generates a three-dimensional cognitive representation of the environment. In straight analogy, stimuli present at the eardrums are compared to determine Gestalt and direction of a sound event.

The sound pressure at the eardrum can be represented in the time domain by the Head-Related Impulse Response (HRIR) and in the frequency domain by the Head-Related Transfer Function (HRTF). These transfer functions can be measured individually with small in-ear microphones in the auditory canal of a human or of an artificial head [2].

The Binaural Impulse Response (BIR) of a room can be considered as the superposition of many HRIRs. The direct sound incidence and every reflection contribute to the complete impulse response with a specific time alignment and level. The impulse response can be understood as a fingerprint of the room, but it is only valid for one position.

To allow for interaction, a set of impulse responses (filters) has to be generated, either measured in a real room or simulated with appropriate software. For playback those binaural impulse responses (one impulse response for every angular direction) have to be loaded for every sound source. When the user turns his head, the head tracker selects a new set of impulse responses thus providing a stable localization of the source.

2. THE AURALIZATION SYSTEM

2.1. Requirements for a dynamic system

Important factors for the implementation of a dynamic system are the angular range and the maximum angular velocity. The actual angular direction and rotating speed enter the convolution procedure as varying parameters. The maximum range of head rotation in the horizontal plane without rotation of the shoulder is ±90°, usually the rotation is ±75°.

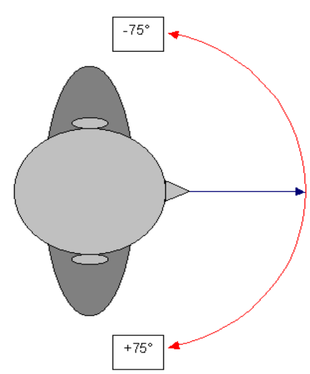

Figure 1: Horizontal head movement

The angular velocity of the head movement can increase up to 1200°/s, the normal speed is between 10°/s and 100°/s. A slow rotation of the head has ~50°/s, which is one degree every 20ms, for 100°/s it is one degree every

10ms. This sets the requirement for the performance of all subsystems, such as the head-tracker, the sound device and the processor. As a platform a PC-system should be used to avoid additional DSP-based hardware.

2.2. Latency problem

A system for interactive auralization requires a very low latency. The latency of the system is defined as the time difference between the action of the user and the change of sound at the user's ears. It is the sum of the individual latencies of the various parts such as head-tracker, convolution, filter and the sound output device. The drawback of MS Windows sound reproduction systems is a latency of usually 2048 samples or more, which comes to more than 46ms at 44.1 kHz. If a filter is changed immediately after the sound device switches to a new block it will last for the playing time of this block (e.g. 46ms). ASIO (Audio Streaming In Out) as a new audio standard allows the latency of the audio output to be reduced down to 64 samples or ~1.5ms respectively.

2.3. Convolution

Zero latency filtering is possible using FIR filters but also poses the highest requirements on processing power. RIRs with 130000 filter taps (3s) or more cannot be processed in real-time on a PC-system. Block convolution is a method to reduce the requirement to a minimum, but it also causes the latency to increase in proportion to the filter length. The only way to minimize the latency of the convolution is a special treatment of filter blocks. A further advantage is that the ASIO device can also operate on those block sizes. Hence, the optimal convolution adds a time delay of only one block to the latency of the system.

2.4. HRTF Measurement

HRTF measurements were taken for a dummy head with torso, moulded from a real person [3][4]. 52.000 transfer functions have been measured, 720 horizontal positions in ½°-steps, every 5° vertical. The measurement procedure is shown in figure 3, showing the dummy head sitting on a rotatable chair in the anechoic chamber of the Institute of Technical Acoustics RWTH Aachen.

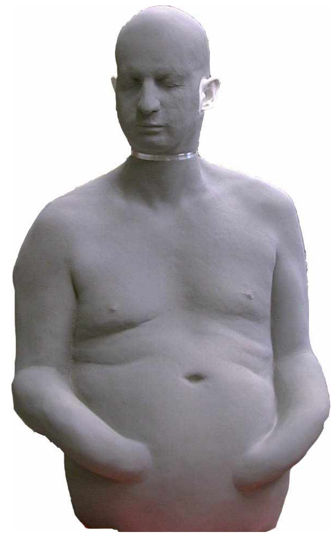

Figure 2: Rotatable dummy head

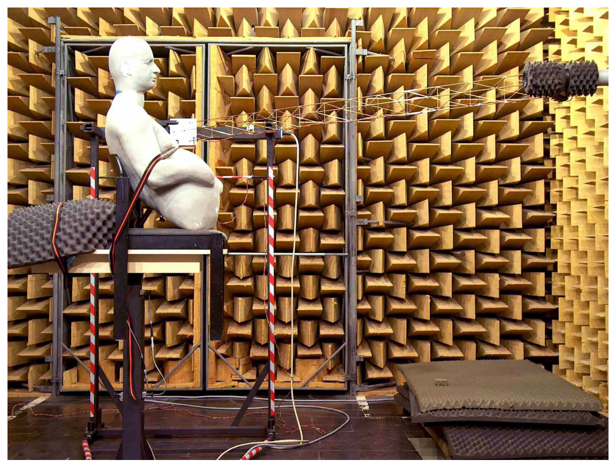

Figure 3: Transfer function measurement

2.5. Headphone compensation

An important criterion when selecting the headphone for a binaural auralization system is the impedance of the transfer path headphone-eardrum compared to air. Of all existing headphones a circumaural electrostatic headphone was shown to have the lowest impedance, closest to natural hearing [5]. This makes it easier to compensate for the transfer function headphone-eardrum. In this context, it is not sufficient to compensate for the transfer function from headphone to microphone at the blocked entrance of the ear canal, as done by the commonly used Inverse Headphone Transfer Function (IPTF). Since the influence of the ear canal itself is difficult to assess technically, we think that the compensation of the transfer function headphone-eardrum can only be obtained in satisfying quality by comparing the listening impression in reality with the results of the auralization of measured BIRs, although this is in contrast to the common procedure of how the headphone transfer function is compensated.

2.6. Head-Tracking

To track head movements several devices have been tested: an ultrasound tracker, a gyrotracking device, and an electromagnetic tracking system. The receiver was mounted on top of the headphones, the transmitter was located in front of the head or above the head.

Though nominally having update rates fast enough, fast head movements free of perceptible artifacts were possible only with the gyrotracking device and the electromagnetic tracking system. The ultrasound tracker adds about 35ms processing time to the overall latency. Assuming not more than 10ms for the remaining parts of the system, the resulting total system latency is still lower than 45ms. This indicates, that a system latency of less than 45ms is desirable to avoid perception artifacts. These results are different from those by Mackensen [6], who has specified a maximal latency time of 85ms. The difference could be attributed to listeners being more sensitive to latency problems for signals with fast slew rates and longer decay times as a result of the reverberation time in our test room of ca. 2.5s. Similar effects of signal duration on latency tolerance were reported by Brungart [7].

3. LISTENING TESTS

3.1. Listening room

As listening room the Audimax of Technical University Berlin was used. The Audimax is a multipurpose rehearsal room with a volume of ~10.000m^3, 1280 seats and a reverberation time RT_{60} of 2.1s (high frequencies) to 2.8s (low frequencies). Built of static elements made of concrete, the sidewalls are made of brick with Helmholtz-resonators for low frequency absorption. The ceiling is made of stepped light elements and reflectors.

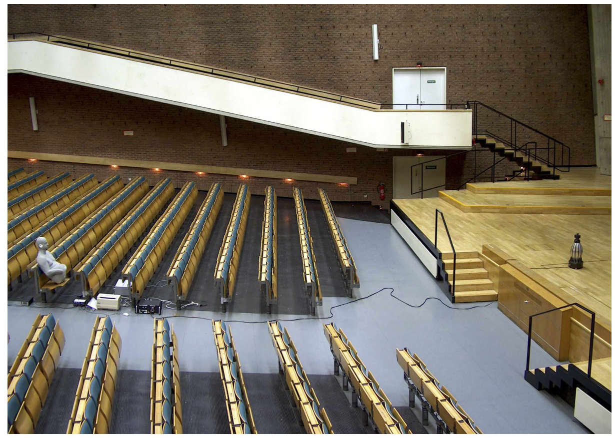

Figure 4: Listening room, Audimax of Technical University Berlin with dummy head and source

3.2. Listening test

Listening tests have been performed to compare listening in reality with the auralization of measurements and with the auralization of room simulation. The mentioned listening room was simulated with AURA, the room acoustical analysis module of EASE [8], incorporating source data (measured in 5° steps) and receiver data (measured in 1/2° horizontal and 5° vertical intervals). In addition, a full measurement of the source in the room has been done to obtain the equivalent set of BIRs in reality. The source was a MBL 101 B, an omnidirectional loudspeaker (figure 5). The distance between source and receiver was ~10m.

Measured and calculated RIRs were convolved with dry sound in real-time. The head positions of the listening subjects were tracked with a head-tracker. The stimuli were presented by an electrostatic headphone, STAX Lambda Pro New, with the solid state amplifier device Monitor Broadcast. 50 Subjects repeatedly listened to three different stimuli (speech, cello, orchestra) and answered a questionnaire with 16 semantic differentials.

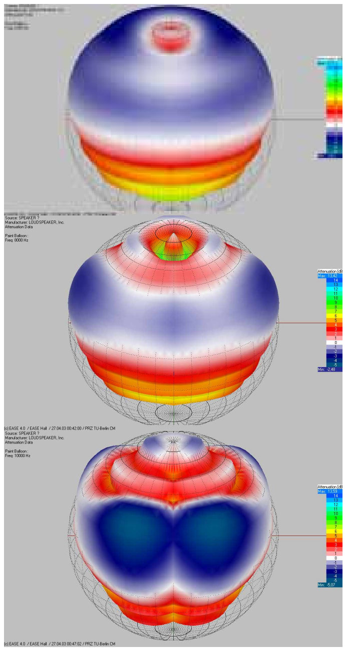

Figure 5: Directivity of MBL 101 B at 6.3/8/10kHz (top down)

3.3. Listening test results

The comparative listening experiments had the following preliminary results:

- Listening impression of reality and the auralization of measurements were quite close: test subjects were not able to hear any differences in 40% of all stimuli.
- Only 4% (2 of 50 subjects) immediately recognised differences. Differences applied to colorations of the sound rather than to spatial artifacts.
- Comparison of simulation results with auralization of measurements yielded a good agreement, comparable to the differences perceived for measurement and reality.

4. CONCLUSION

We have described an approach to create a PC-based binaural auralization system with listening impressions close to reality. A viable solution was presented utilizing a head-tracking system to allow for the consideration of fast head movements in the horizontal plane. Taking into account the characteristics of the human head, the HRTFs of a rotatable dummy head were measured with a high angular resolution and applied to the RIR. A realistic reproduction of the room was reached using a low-latency convolution software to allow for switching the combined filters in real-time synchronously to the head movements. Listening tests showed good agreement of the auralization of measured and simulated RIRs with the listening impressions in reality. Acoustical demonstrations of architectural design results, of sound reinforcement systems, perception experiments and VR environments, even historical reconstructions [9] are potential applications.

5. ACKNOWLEDGEMENTS

This work was supported by FSP-PV/PRZ of Technical University Berlin, ITA Aachen and ADA Berlin. We thank A. Goertz for fruitful discussions and technical help.

6. REFERENCES

[1] Moldrzyk, C.: “Akustischer Entwurf“, Dissertation, TU Berlin, in progress

[2] Møller, H.: “Fundamentals of binaural technology", Applied Acoustics, Vol. 36, No. 3/4, 1992, pp. 171-218

[3] Moldrzyk, C.: “An innovative dummy head for the verification of an acoustical design methodology for architects”, Proc. Tonmeister Convention 2002, Hannover

[4] Moldrzyk, C.: “Design Tool Auralisation”, Proc. CFA/DAGA 2004, Strasbourg

[5] Møller, H., Hammershøi, D., Jensen, C. B., Sørensen, M. F.: “Transfer Characteristics of Headphones Measured on Human Ears“, JAES, Vol. 43, No. 4, pp. 203-217; April 1995

[6] Mackensen, P.: “Auditive Localization. Head movements, an additional cue in Localization”, Dissertation, TU Berlin 2004

[7] Brungart, D. S. et al.: “The Interaction between Head-Tracker, Latency, Source Duration, and Response Time in the Localization of Virtual Sound Sources”, Proc. ICAD, Sydney 2004

[8] AURA module of acoustic software EASE4.1, see www.ada-acousticdesign.de/german/ease/index_ease.html

[9] Weinzierl, S.: “Beethovens Konzerträume. Akustik und symphonische Aufführungspraxis an der Schwelle zum modernen Konzertwesen“, Frankfurt a. M. 2002