ROOM SIMULATION FOR BINAURAL SOUND REPRODUCTION USING MEASURED SPATIOTEMPORAL IMPULSE RESPONSES

Cornelia Falch, Markus Noisternig, Stefan Warum, Robert Höldrich

Institute of Electronic Music and Acoustics
University of Music and Dramatic Arts, Graz, Austria
http://iem.at

ABSTRACT

In binaural sound reproduction systems the incorporation of room simulation is important to improve sound source localisation capabilities. Thus, the localisation error can be decreased, while equivalently an enhanced externality (out of head localisation) is achieved. Previously proposed works are based on simple geometrical approaches for room simulation.

In this paper an alternative method using measured room impulse responses (RIRs) is presented. Therefore, it is possible to obtain a convincing acoustical image of an existing room. The RIRs are measured using a circular microphone array to capture both temporal and spatial information of the desired room.

1. INTRODUCTION

A review of literature states that incorporating room simulation in binaural sound reproduction systems is important to improve localisation capabilities as well as out of head localisation [1].

In previously proposed systems [2] room simulation is divided into two stages of computation. First, image sources of low order are calculated using a simple geometrical approach. To cover the direction of the image source signals, representing the early reflections, they are encoded into the binaural system according to their position in virtual space. Therefore, the proposed system uses a virtual ambisonic approach for sound source spatialisation. Second, late reverberation is taken into account by computationally efficient algorithms introduced in [3]. Dattoro recommends the implementation of reverberators relying on all-pass circuits embedded within very large globally recursive networks. Because of the fact that late reverberation signals are highly decorrelated, low order ambisonics is sufficient for encoding. The acoustic properties of the reflecting walls are taken into account by simple infinite impulse response (IIR) filters of low order.

Now, the present study deals with the opportunity to capture the acoustic properties of “real rooms” in binaural sound reproduction systems. The main idea is to measure not only the temporal but also the spatial behaviour of the room impulse responses using a microphone array. By convolving a sound source signal with these spatial room impulse responses (SRIRs) and encoding the convolved signals to the binaural system according to the direction of the several SRIRs it is possible to simulate the measured room.

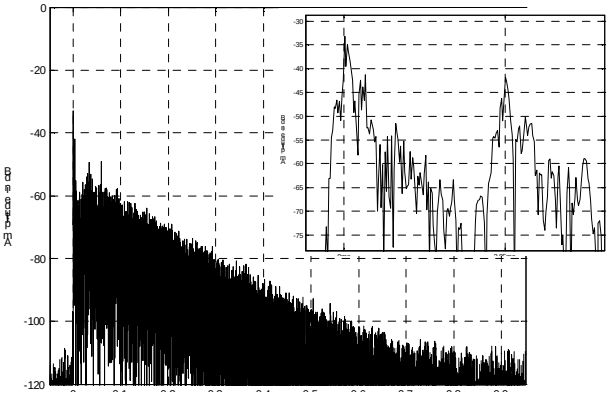


Figure 1: *Impulse response.*

Theoretically, the system is able to process multiple, time variant sound sources, though for the ease of illustration a single, time invariant sound source serves as input signal. As a second simplification, the model is restricted to the 2 dimensional case. The next chapter gives a detailed description of the theory and structure, followed by an explanation of the measurement of the spatial room impulse responses (SRIRs). Chapter 4 deals with measuring the reference signal, the output of the proposed work will be compared to in an objective as well as a subjective evaluation preocedure. The paper concludes with a brief outlook on future work, especially the modification required to extend the system's ability to handle general multiple, time variant sound sources.

2. THEORY

Time variant binaural sound source reproduction using ambisonics [2], [4], [5] results in a fixed set of head related impulse response (HRIR) filters. The decoder depends only on the arrangement of virtual loudspeakers and the order of the ambisonic system.

Here, the fundamental idea was to capture the acoustic properties of the IEM-CUBE for binaural sound reproduction systems. The IEM-CUBE is a small concert room incorporating a periphonic sound recording system based on ambisonics [6]. Therefore, SRIR measurements have been carried out for the horizontal plane using a circular microphone array as explained in chapter 3. Direct signal and first floor reflection are extracted from the several measured impulse responses, one of them is depicted in figure 1. The small plot shows an enlargement of the

first few milliseconds of the impulse response to visualise the direct signal and the first floor reflection that is delayed by 2.85ms. Considering the first floor reflection as a separate sound source provides the opportunity to partly compensate for the inability of the planar microphone array to resolve a three dimensional sound field.

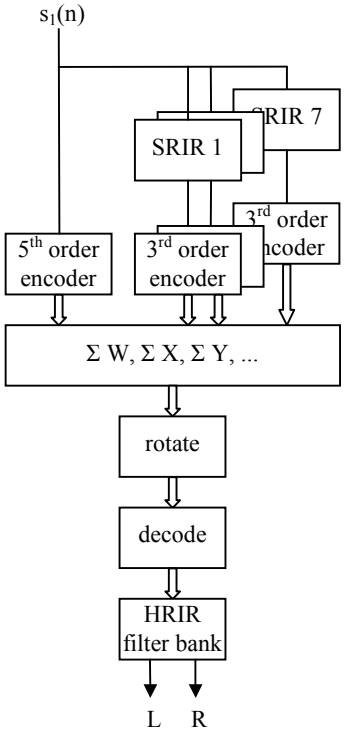


Figure 2: *System for one static sound source.*

A block diagram of the entire system is illustrated in figure 2 for one single static sound source. Both, direct signal and first floor reflection are directly encoded to the ambisonic domain using a 5th order ambisonic encoder. In contrast, the sound source is convolved with the associated SRIRs to enable sound source spatialisation via room simulation and the result is encoded applying a 3rd order system.

Within the ambisonic domain all three signal parts are channel-wise superimposed and decoded to virtual loudspeakers. Due to the virtual ambisonic approach, the listener is able to move through the virtual room. Finally, the decoded signals are convolved with the HRIRs to create the left and right headphone signals. Therefore, it is possible to create static sound sources situated in an image of the original room.

Humans are able to improve their localisation capabilities by small unconscious head movements [7]. To be able to utilise this significant characteristic in the technical set-up, great importance is attached to the incorporation of head tracking [8]. A simple rotation matrix is implemented in the ambisonic domain for this purpose.

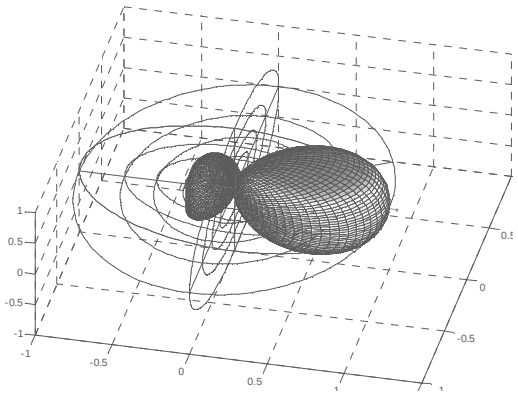
Figure 3: *The microphone on the rotating table .*

3. MICROPHONE ARRAY

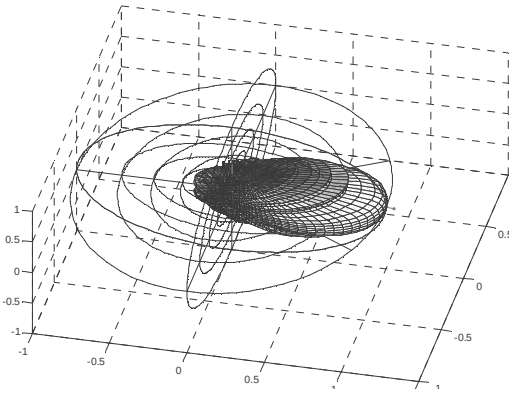
The circular microphone array employed for the impulse response measurements in this paper is illustrated in figure 3. A single microphone is mounted on a slowly rotating turntable via a rod of 0.5 meter length. The minimum angular resolution of the apparatus is 2.5°, which corresponds to 144 different microphone positions. The output signals are merged into seven channels by a simple delay and sum beamformer. Figures 4(a) and (b) depict the beamformer patterns for 250Hz and 5kHz in the horizontal plane. (The elliptical lines indicate the 0/-3/-6/-10dB amplitude surface.) The set is able to capture a frequency range approximately between these two frequencies.

According to Zielinski [9] the apparently low upper frequency limit of the beamformer (5kHz) is nevertheless sufficient for a satisfying reproduction of the room acoustics. Additionally, Poletti states in [10] that for a 3rd order ambisonic system the reproduction error of the system is negligible within this bandwidth, too. Figure 4(c) shows the dependence of the main lobe with respect to the elevation angle.

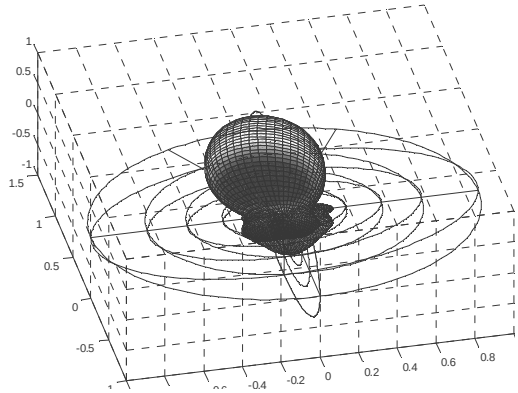
For a given loudspeaker direction, the SRIRs are measured using the method introduced by Farina in [11]. The basic principle is to use logarithmic sweeps as excitation signals. In contrast to linear frequency sweeps it is thus possible to separate the impulse responses (IRs) for each harmonic distortion order. These unwanted artefacts occur due to the non linear property of the desired room. Thus, the final result of the whole measurement procedure is a sequence of IRs clearly separated along the time axis. Furthermore, unwanted distortions can be windowed out easily. To guarantee for a sufficient time delay between consecutive IRs, the logarithmic sweep has to be very long.



(a): *Horizontal beam, f=250Hz.*



(b): *Horizontal beam, f=5kHz.*



(c): *Vertical beam, f=500Hz.*

Figure 4: *Patterns of the D&S beamformer.*

Additionally, the time delay must also exceed the reverberation time of the room, so that subsequent IRs will not produce aliasing problems.

Figure 5: *The dummy head "SOURCE".*

4. EVALUATION OF THE METHOD

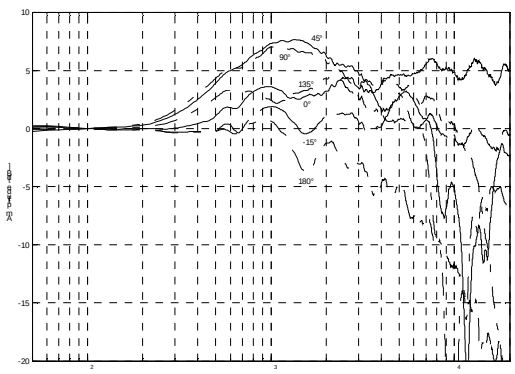
To investigate the performance of the proposed system, its binaural output is compared to an equivalent signal acquired through a second recording procedure. The measurement of the binaural reference signal is carried out with a special designed artificial head recording system, termed the "SOURCE". Unlike many other dummy heads which mimic the external geometry of the pinna, earcanal, head, shoulders and torso, the dummy head SOURCE is the physical realisation of a mathematical model based on the finite element method, see figure 5. The interaural level and time differences for the correct reproduction of an incident sound wave denote the design criteria. Psychoacoustic tests prove advantages of the SOURCE compared to other dummy heads concerning sound coloration for varying angles of sound incidence, yielding a more neutral acoustical image, [12], [13]. Head related transfer functions (HRTFs) of the dummy head SOURCE are depicted in figure 6(a) for various azimuth angles and (b) for various elevation angles.

The objective evaluation part is based on general energy considerations and error signal estimation. Extensive listening tests will gain information about the subjective behaviour of the system. However, both procedures are not yet accomplished, thus at this time no final quality statement can be presented.

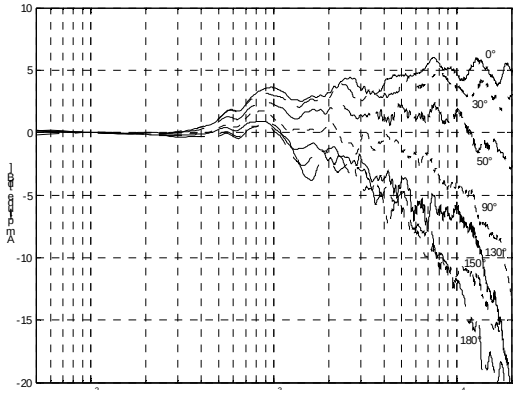
5. FUTURE WORK

Using the abovementioned approach only static sound sources may be reproduced. To overcome this limitation, some modifications have to be applied on the system. Figure 7 presents a block diagram of the proposed variant. Additionally, SRIR measurements were carried out for every single loudspeaker position of the IEM-CUBE.

For time variant sound source reproduction, the direct signals are encoded to the ambisonic domain due to their actual position. Thus, the time variant property is inherent in the encoder. In order to filter the signal with the appropriate SRIRs, a decoder is employed that allocates the ambisonic signals to N virtual loudspeaker positions. They have to be arranged at exactly the same positions as the actual loudspeakers of the IEM-CUBE.



(a): *Different azimuth angles.*



(b): *Different elevation angles.*

Figure 6: *HRTFs of the dummy head "SOURCE".*

Within this virtual domain the loudspeaker signals are convolved with their dedicated SRIRs resulting in N times L signals, where L denotes the number of impulse responses used to cover the entire horizontal plane. However, signals assigned to the same SRIR are superimposed reducing the total number of signals to L. A low order encoder transforms them back into Ambisonic domain. Together with the encoded direct signal we thus have obtained full information on the time varying sound field. Additional signal manipulation (rotation) will follow before the decoding stage. The last step includes filtering the remaining signals with the appropriate HRIRs to obtain the binaural reproduction of time variant sound sources.

Although we still deal with 2D considerations the computational effort of the system is enormous, due to the necessary N times L filter of about 700ms length. To overcome this problem, the SRIRs can be further split up into more than one section, eg. two, one of length 70ms and the other one of length 630ms. The first part contains substantial spatial information, thus the horizontal plane is covered by seven beamformer patterns and fed into a 3rd order ambisonic encoder as before. For the remaining portion a subdivision into three sections is sufficient, equally the encoder is reduced to a 1st order

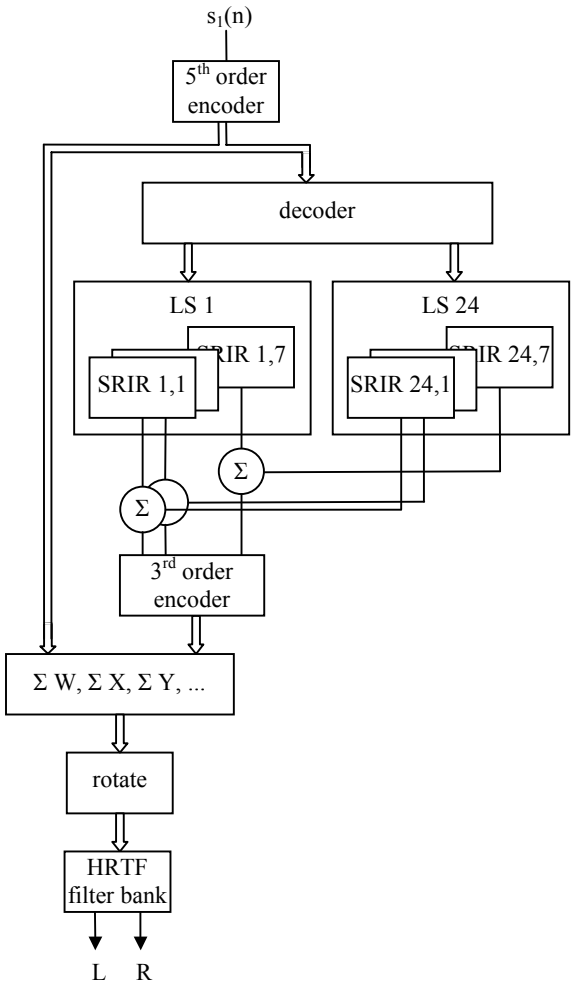


Figure 7: *System for time variant sound sources.*

system. A schematic representation of this idea is given in figure 8, where "long/medium/short filter" denotes the filter length and "low/higher/high order" the order of the ambisonic encoder. The previously explained decoding, filtering and encoding steps denote matrix manipulations and result in a [LxL] matrix, the elements of which express the filtered ambisonic channels. This is true due to chosing the same number of SRIRs and associated ambisonic channels.

6. CONCLUSION

The proposed system allows to capture the acoustical properties of real rooms for binaural sound reproduction. Therefore, the temporal as well as the spatial behaviour of the room impulse responses are measured using a circular microphone array. The implementation of the virtual Ambisonic approach offers the possibility to deal with time variant sound sources. Furthermore, one suggestion is presented to decrease the enormous computational complexity that originates from the multiple filtering with the long impulse responses.

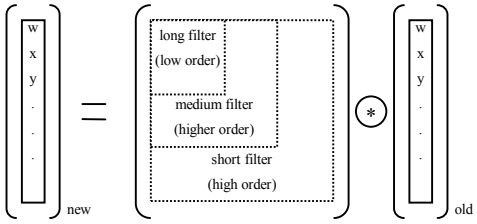


Figure 8: *Scheme of the matrix decomposition.*

7. ACKNOWLEDGEMENTS

The authors would like to thank Mr. Martin Pflüger at AVL List GmbH, Graz, for providing technical data on the dummy head SOURCE.
This study has been performed in cooperation with AKG Acoustics, Vienna.

8. REFERENCES

[1] F. L. Wightman and D. J. Kistler, *Headphone stimulation of free field listening I: stimulus synthesis*, J. Acoust. Soc. Am., vol. 85, pp. 858-867, 1989.

[2] M. Noisternig, A. Sontacchi, T. Musil and R. Höldrich, *A 3D Ambisonic based Sound Reproduction System*, Proc. of the 24th Conf. of the Audio. Eng. Soc., Banff, Canada, July 2003.

[3] J. Dattorro, *Effect Design: Part 1: Reverberator and Other Filters*, J. Audio Eng. Soc., vol. 45, no. 9, pp. 660-684, September 1997.

[4] M. A. Gerzon, *Ambisonic in multichannel broadcasting and video*, J. Audio Eng. Soc., vol. 33, pp. 859-871, 1985.

[5] M. Poletti, *The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems*, J. Audio Eng. Soc., vol. 44, no. 11, pp. 1155-1182, November 1996.

[6] J. Zmölnig, W. Ritsch and A. Sontacchi, *Der IEM-CUBE – ein periphones Reproduktionssystem*, 22nd Tonmeistertagung, Hannover, Germany, November 2002.

[7] J. Blauert, *Spatial Hearing*, 2nd edition, MIT press, Cambridge, MA, 1997.

[8] D. R. Begault and E. M. Wenzel, *Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized HRTFs on the Spatial Perception of a Virtual Speech Source*, J. Audio Eng. Soc., vol. 49, no. 10, October 2001.

[9] S. K. Zielinski, F. Rumsey and S. Bech, *Effects of Bandwidth Limitation on Audio Quality in Consumer Multichannel Audiovisual Delivery Systems*, J. Audio Eng. Soc., vol. 51, no. 6, pp. 475-501, June 2003.

[10] M. Poletti, *The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems*, J. Audio Eng. Soc., vol. 44, no. 11, pp. , November 1996.

[11] A. Farina, *Simultanous measurement of impulse response and distortion with a swept-sine technique*, preprint of 110th AES Convention, Paris, February 2000.

[12] F. Graf, M. Pflüger, P. Röpke and G. Graber, *Aufnahmesysteme für psychoakustische Analysen – Vergleich Kunstkopf vs. Alternativkonzepte*, DAGA, Oldenburg, March 2000.

[13] M. Pflüger, F. Brandl, Biermeyer, *SOURCE – A Stereophonic System for Engine and Vehicle Sound Recordings*, SAE Conference, Travese City, May 2003.