A 3D Real Time Rendering Engine for Binaural Sound Reproduction

Markus Noisternig, Thomas Musil, Alois Sontacchi, Robert Höldrich

Institute of Electronic Music and Acoustics
University of Music and Dramatic Arts Graz
Inffeldgasse 10/3
A-8010 Graz, AUSTRIA
noisternig@iem.at

ABSTRACT

A method of computationally efficient 3D sound reproduction via headphones is presented using a virtual Ambisonic approach. Previous studies have shown that incorporating head tracking as well as room simulation is important to improve sound source localization capabilities. The simulation of virtual acoustic space requires to filter the stimuli with head related transfer functions (HRTFs). In time-varying systems this yields the problem of high quality interpolation between different HRTFs. The proposed model states that encoding signals into Ambisonic domain results in time-invariant HRTF filters.

The proposed system is implemented on a usual Notebook using Pure Data (PD) a graphically based open source real time computer music software.

1. INTRODUCTION

The following paper deals with the theory and practice of 3D sound reproduction using headphones.

A review of literature on creating virtual environments using loudspeaker states that methods based on physical reconstruction of the acoustical field, like Ambisonic and the holophonic approach offer a good localization performance over an extensive listening area. The main advantage of the general Ambisonic approach [1] is the high computational efficiency. Using generalized Ambisonic the reconstruction of the sound field is accurate only over a small listening area. However, the proposed system is related to a binaural synthesis method based on decoding Ambisonic to virtual loudspeakers avoiding the problems caused by a small listening area.

Sound source spatialization in virtual acoustic environments using headphones requires the filtering of the sound streams with head related transfer functions (HRTFs). The HRTFs capture both, the frequency and time domain aspects of the listening cues to a sound position. The measurement of HRTFs has been researched extensively by Wightman and Kistler [2]. In the proposed system generic HRTFs using the KEMAR [3] as well as the CIPIC [4] database have been used. Wenzel *et al.* [5] state that the use of nonindividualized transfer functions leads to a degradation of localization accuracy, increasing the following errors:

- *localization error*, referring to the deviation of the perceived to the synthesized direction of a virtual sound source
- *localization blur*, that describes the "width" of the perceived stimulus
- *externalization error*, also termed as "inside-the-head localization"
- *cone of confusion*, which refers to localization errors caused by contours of constant interaural time difference (ITD) and interaural level difference (ILD) resulting in front back confusions

Regarding hearing in natural sound fields humans are able to improve sound source localization using small head movements. Begault and Wenzel [6] have shown the importance of incorporating head tracking as well as reverberation in binaural sound reproduction systems to improve localization capabilities. This yields the problem of high-quality time-variant interpolation between different HRTFs. Using Ambisonic this problem can be solved as shown in the next section.

2. THEORY

The following section gives a brief introduction into Ambisonic theory. Furthermore a 3D binaural sound reproduction system is developed, incorporating head tracking as well as room simulation.

2.1. The Ambisonic Approach

Ambisonic is a technique for spatial audio reproduction introduced in the early seventies by Gerzon [1]. Further details of Ambisonic are published in [7, 8, 9, 10].

The basic idea of the generalized Ambisonic approach is the expansion of a wave field into spherical harmonics, assuming that the original wave field is a plane wave. However this assumption is not compulsory because any acoustical field can be expressed as a superposition of plane waves. It is claimed in [7, 9] that Ambisonic systems are asymptotically holographic. Holographic theory states that the Kirchhoff-Helmholtz Integral relates the pressure inside a source free volume of space to the pressure and velocity on the boundary at the surface. Deriving the Ambisonic coding/decoding equations from the Kirchhoff-Helmholtz theory it can be shown that the original wave field may be reconstructed exactly by arranging infinitely many loudspeakers on a closed contour assuming plane wave signals. Using N loudspeakers arranged on a sphere Ambisonic can synthesize a good approximation of the original acoustical field over a finite area. Poletti has shown in [9] that higher order Ambisonic systems are increasingly accurate. Beyond, some indication of the upper frequency limit of the system is given as well.

The decoding stage will only depend on the actual loudspeaker layout. By reproducing a 2D field using a finite number of loudspeakers the so called angular sinc functions (Asincs) arise. The Asinc functions describe the Ambisonic decoding process, more precisely the panning of the Ambisonic signals to the several loudspeakers. Furthermore sound source localization may be confused by signals coming from loudspeakers far away from the intended virtual source position. Weighting the amplitudes of higher order Ambisonic channels, representing higher order spherical harmonics, yields a reduction of the sidelobes as well as a broadening of the main lobe. Therefore the confusing far away speaker signals may be attenuated considering just noticeable difference (JND) thresholds to decrease the localization error. Though as a result of the wider main lobe the localization blur increases. Hence, "windowing" of the Ambisonic channels can be used to improve the capabilities of the decoder to minimize the error of synthesis [10].

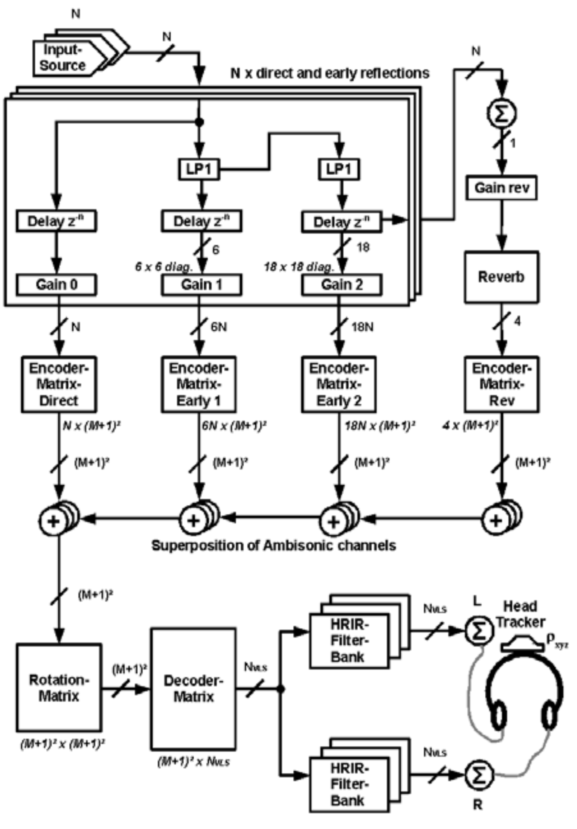

Figure 1: *Block diagram of the 3D binaural sound reproduction system incorporating head tracking and room simulation*

2.2. Binaural Sound Reproduction

In time-varying binaural sound reproduction systems, as mentioned above, the problem of high quality interpolation between different HRTFs occur. This interpolation yields artifacts decreasing the localization performance of the system. Therefore the source signals are encoded into Ambisonic domain dependent to their position in the virtual acoustic space (figure 1). Due to the variable distance of the sound sources in virtual space signals are delayed relative to the listeners position first. Using 3D Ambisonic of M^{th} order leads to $(M+1)^2$ Ambisonic channels. The number of Ambisonic channels is independent of the number of virtual sound sources to encode. This is an important fact for incorporating room simulation as shown later in section 2.3. Henceforth head rotation is taken into account with simple time-variant rotation matrices in the Ambisonic domain using three degrees of freedom. Head rotation is identified using a head tracking device mounted on the headphones. Then the Ambisonic channels are decoded to virtual loudspeakers. The binaural signals are created by filtering these loudspeaker signals with their appropriate HRTFs and superimpose them yielding left and right ear headphone signals.

Due to the fact that the decoding matrix is defined by the virtual loudspeaker setup it is important to distribute them as uniformly as possible over the spheres surface. Otherwise, ill conditioning or even singularities in the decoder matrix may occur.

Moreover it is possible to reduce the number of Ambisonic channels by using a mixed order Ambisonic setup. Humans are able to localize sound sources in horizontal plane more precisely than in vertical directions [11]. As a consequence the Ambisonic order for encoding elevation may be reduced, not affecting the encoding in the horizontal plane. Using this approach the computational efficiency of the system may be increased.

Consider the fact that filtering with HRTFs is a highly computational task using shorter HRTFs yields an increase of efficiency. Error analysis of a 2D system states, that HRTFs less than 128 points lead to a satisfactory localization performance (figure 2 and figure 3), as shown in [12]. Furthermore the use of individualized HRTFs will increase localization capabilities. Further investigations on the influence of different HRTFs to the overall system localization performance have not been carried out during this work.

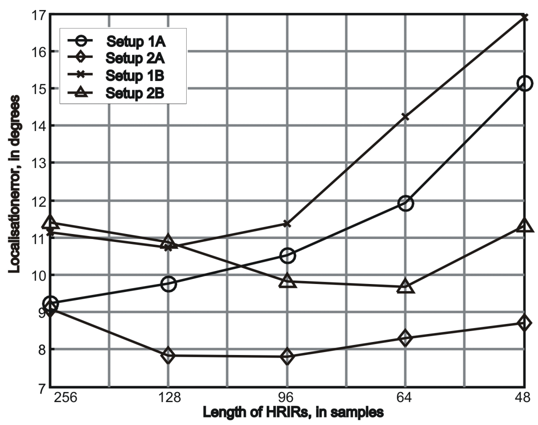

Figure 2: *Localization error in degrees for binaural sound reproduction systems using different HRTFs (1,2) and different interpolation techniques (A,B)*

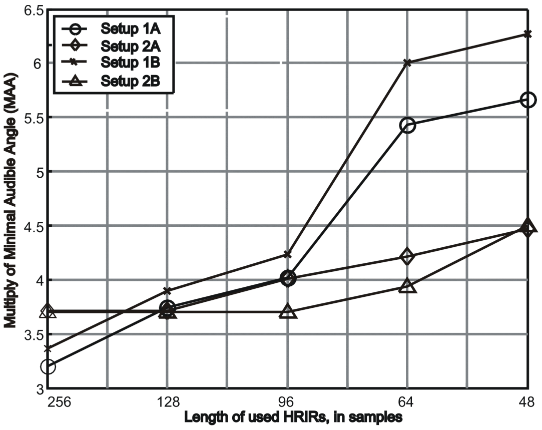

Figure 3: *Localization blur in multiples of the minimum audible angle (MAA) [11] for binaural sound reproduction systems using different HRTFs (1,2) and different interpolation techniques (A,B)*

2.3. Room Simulation

As mentioned above to improve localization accuracy and perceived externality of virtual sound sources it is important to incorporate room simulation.

In this section we focus on sound reverberation as a natural phenomenon occurring when sound waves propagate in an enclosed space. We are considering a rectangular room containing omni-directional virtual point sources. Room simulation is divided into two stages of computation:

- early reflections of first and second order
- diffuse sound field

The early reflections are taken into account using a simple geometrical acoustic approach calculating image sources. Every image source is filtered with a first order IIR lowpass filter to consider the acoustic properties of the reflecting walls. Then the signals of the image sources are delayed and attenuated according to their distance to the listener. Now the image sources are encoded to Ambisonic dependent to their position in the virtual acoustic space. To increase the computational efficiency for encoding image sources the virtual room is divided into several subspaces. Image sources situated in same subspaces are bundled. Henceforth the bundled signals are encoded according to the direction dedicated to their respective subspace.

Another approach to improve computational performance is to encode higher order early reflections with Ambisonic of lower order. Because of the fact that higher order reflections become more and more diffuse the loss of localization accuracy may be accepted to enhance computational efficiency.

Late reverberation creates an ambient space in the perception of the listener. Dattorro [13] states that the most efficient implementations of reverberators rely on all-pass circuits embedded within very large globally recursive networks (figure 4). The first step is to delay the input signal to handle the time where late reverberation starts. Then the signal is low pass filtered to consider the coloration due to the absorption of high frequency signal components at the enclosing walls. Furthermore the first set of input diffusers quickly decorrelate the incoming sound to prepare it for the next stage. The second diffuser stages are arranged to feed back globally on themselves, to loop the decorrelated sound indefinitely. By multiplying a gain factor <1.0 in the feedback paths it is possible to control the decay time. The lowpass filter in the feedback paths incorporates the texture of materials at the enclosing walls as before.

Finally the reverberation signals are encoded into Ambisonic domain. Because of the fact that reverberation signals do not affect localization accuracy low order Ambisonic is sufficient for encoding.

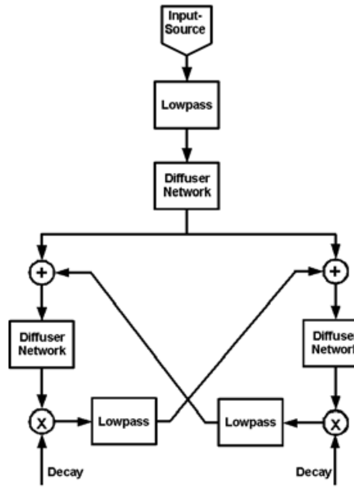

Figure 4: *Recursive reverberation network*

Incorporating this reverberation network yields a simple way to control parameters of the reverberated sound like

- diffusion of the input and decay signals (decorrelation)
- reverberation pre delay
- reverberation gain
- decay rate
- cut off frequency of the high-frequency damping.

The effects of the surrounding environment can be used to change the perceived distance of the virtual sound source.

3. IMPLEMENTATION

The proposed system is implemented on a standard notebook using Pure Data (PD). PD is a graphically based open source real time computer music software by Miller Puckette[1].

First a 2D system with 4th order Ambisonic was implemented in PD as well as on a digital signal processor (DSP) running a PC as a host system. The 2D system has been optimized using the results of an objective mathematical model as described in [12]. Furthermore the objective mathematical model as well as the localization accuracy of the optimized system have been evaluated using listening tests [14].

The next step was the implementation of a 3D system using 4th order Ambisonic on a usual notebook running a 1.6 MHz CPU. Because of the fact that for a 2 channel system with room simulation the required CPU performance goes up to 2GHz the

1 http://crca.ucsd.edu/~msp/software.html

computational efficiency of the overall system has been optimized as described above. With the rapid increasing of CPU power it will become possible to run the binaural application as a background task for computer music software.

4. CONCLUSIONS

In this paper the relations between Ambisonic and time-variant binaural sound reproduction systems have been discussed. We have shown that the use of the virtual Ambisonic approach for binaural synthesis in 3D systems containing multiple input sources yields an enormous benefit regarding to the computational efficiency. As a consequence it is possible to implement a full 3D system incorporating room simulation as well as head tracking on a usual notebook.

The main advantages of using Ambisonic in time varying binaural systems are as follows:

- Rendering the sound field using Ambisonic in time-variant binaural sound reproduction systems yields time-invariant HRTFs without the need of interpolation between them.
- The number of HRTFs is independent of the number of virtual sound sources to encode. This is quite important because incorporating room simulation by calculating early reflections of first and second order yields an enormous increase of virtual sound sources.
- Ambisonic provides a decoupling of the encoder and decoder. Hence, the awareness of the playback configuration can be limited to the decoder while only the universal multichannel format is implemented in the encoding stage.
- Head rotation may be taken into account with simple time-variant rotation matrices using a head tracker mounted on the headphones.

As future research, a comprehensive localization error analysis of the proposed system would be interesting using the objective mathematical model of localization as well as listening tests.

5. REFERENCES

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

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

[3] Gardner, W. G. and Martin, K. D., „HRTF Measurement of a KEMAR“, in *J. Acoust. Soc. Am.*, vol. 97, pp. 3907-3908, 1995

[4] Algazi, V. R., Duda, R. O., Thompson, D. M. and Avendano, C., “The CIPIC HRTF Database”, in *Proc. IEEE Workshop on Applications of Sig. Proc. to Audio and Electroacoustics*, pp. 99-102, NY, 2001 October

[5] Wenzel, E. M., Arruda, M., Kistler, D. J. and Wightman, F. L., “Localization using nonindividualized head-related transfer-functions”, in *J. Acoust. Soc. Am.*, vol. 94, pp. 111-123, 1993

[6] Begault, D. R. and Wenzel, E. M., “Direct Comparison of the Impact of Head Tracking, Reverberation and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Sound Source”, in *J. Audio Eng. Soc.*, vol. 49, no. 10, 2001 October

[7] Nicol, R. and Emerit M., “3D Sound Reproduction over an Extensive Listening Area: A Hybrid Method Derived from Holophony and Ambisonics”, in *Proc. AES 16th Int. Conf.*, pp. 436-453, 1999

[8] Jot, J. M., Larcher, V. and Pernaux J.-M., “A Comparative Study of 3D Audio Encoding and Rendering Techniques”, in *Proc. AES 16th Int. Conf.*, pp. 281-300, 1999

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

[10] Poletti, M., ”A Unified Theory of Horizontal Holographic Sound Systems”, in *J. Audio Eng. Soc.*, vol. 48, no. 12, 2000 December

[11] Blauert, J., “Spatial Hearing”, 2nd ed., MIT Press, Cambridge, MA, 1997

[12] Sontacchi, A., Noisternig, M., Majdak, P. and Höldrich, R., “An Objective Model of Localisation in Binaural Sound Reproduction Systems”, in *Proc. AES 21st Int. Conf.*, St. Petersburg, Russia, 2001 June

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

[14] Sontacchi, A., Majdak, P., Noisternig, M. and Höldrich, R., “Subjective Validation of Perception Properties in Binaural Sound Reproduction Systems”, in *Proc. AES 21st Int. Conf.*, St. Petersburg, Russia, 2001 June