A 3D AMBISONIC BASED BINAURAL SOUND REPRODUCTION SYSTEM

MARKUS NOISTERNIG, ALOIS SONTACCHI, THOMAS MUSIL, AND ROBERT HÖLDRICH

Institute of Electronic Music and Acoustics,
University of Music and Dramatic Arts, Graz, Austria
noisternig@iem.at, alois.sontacchi@kug.ac.at,
musil@iem.at, robert.hoeldrich@kug.ac.at

A computationally efficient 3D real time rendering engine for binaural sound reproduction via headphones is presented. Binaural sound reproduction requires to filter the virtual sound source signals with head related transfer functions (HRTFs). To improve humans localization capabilities head tracking as well as room simulation have to be incorporated. This yields the problem of high-quality, time-varying interpolation between different HRTFs. To overcome this problem a virtual ambisonic approach is used that results in a bank of time-invariant HRTF filter.

INTRODUCTION

A review of perceptual literature states that humans are able to locate the position of a sound source with remarkable accuracy using a variety of acoustic cues [1]. Real-world signals are acoustically filtered by the pinna, head and torso of the listener. Referring to the duplex theory of sound source localization the main cues for horizontal perception are the interaural time difference (ITD) and the interaural level difference (ILD) caused by wave propagation time differences and the shadowing effects as mentioned above. In vertical directions monaural (spectral) cues affect the perceived elevation of a sound source. Well, the head related transfer functions (HRTFs) capture both, the frequency domain and time domain aspects of the listening cues to a sound position. The measurements of HRTFs have been researched extensively by Wightman and Kistler [2].

In binaural sound reproduction systems, the spatialization of virtual sound sources requires to filter the signals with HRTFs appropriate to their desired position in virtual space. Wenzel et al. state in [3] that the use of nonindividualized HRTFs yields a degrading localization accuracy, externalization errors (inside-the-head localization) and reversal errors. In the proposed system generic HRTFs using the KEMAR [4] as well as the CIPIC database [5] have been used.

Regarding hearing in natural sound fields humans are able to improve sound source localization capabilities due to small unconscious head movements. Begault and Wenzel [6] have shown the importance of incorporating head tracking as well as room simulation in binaural sound reproduction systems. Thus, incorporating multiple moving sound sources and head tracking yields the problem of high-quality, time-varying interpolation between different HRTFs. The proposed system overcomes this problem by using a virtual ambisonic approach that results in a bank of time-invariant HRTF filters. The following section gives a brief introduction into ambisonic theory. In section two a binaural sound system is derived from the generalized ambisonic approach and optimization criteria are discussed as well.

1 AMBISONIC THEORY

1.1 The ambisonic approach

Ambisonic is a sound reproduction technique involving a limited number of playback channels while allowing reproduction of a full 3D virtual acoustic space with several moving sound sources. This sound reproduction technique was originally introduced by Gerzon [7]. Further details of ambisonic are published in [8-12].

In [8], [9] it is shown that ambisonic is asymptotically holographic. The holographic theory states that any sound field can be expressed as a superposition of plane waves. 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. Therefore, it is possible to reproduce the original sound field by an infinite number of loudspeakers arranged on a closed contour. The loudspeaker signals are assumed to be plane waves. By using a finite number of loudspeakers arranged on a sphere, a good approximation of the original sound field may be synthesized over a finite area (sweet spot). A lower limit of the number of needed loudspeakers is given by the number of transmit channels which is defined by the ambisonic order. Consequently, it can be shown that higher order ambisonic systems are increasingly accurate.

Deriving ambisonic from the homogenous wave equation

$$\Delta p(t,\mathbf{r}) - \frac{1}{c^2}\frac{\partial}{\partial t^2} p(t,\mathbf{r}) = 0 \quad (1)$$

where $p(t,\mathbf{r})$ is the sound pressure at position **r** and c is the speed of sound, yields the so called matching conditions [11]

$$s \cdot Y^{\sigma}_{m,\eta}(\Phi,\Theta) = \sum_{n=1}^{N} p_n \cdot Y^{\sigma}_{m,\eta}(\varphi_n,\vartheta_n) \quad (2)$$

The left side of (2) represents the ambisonic encoding equation which can be written in vector notation

$$\mathbf{B}_{\Phi,\Theta} = \mathbf{Y}_{\Phi,\Theta}.s \quad (3)$$

where $\mathbf{B}_{\Phi,\Theta}$ is the ambisonic channel vector, s is the pressure of the original sound wave coming from direction (Φ,Θ) and $Y^{\sigma}_{m,\eta}(\Phi,\Theta)$ describes the corresponding spherical harmonics. On the right hand side of (2) p_n represents the signal of the n^{th} loudspeaker at direction (φ_n,ϑ_n). The spherical harmonics $Y^{\sigma}_{m,\eta}(\varphi_n,\vartheta_n)$ are defined as follows

$$Y^{\sigma}_{m,\eta}(r) = \begin{cases} A_{m,\eta}\, P^{\eta}_{m}(\cos\Theta)\cos(m\Phi) & \text{for } \sigma = 1 \\ A_{m,\eta}\, P^{\eta}_{m}(\cos\Theta)\sin(m\Phi) & \text{for } \sigma = -1 \end{cases} \quad (4)$$

where $P^{\eta}_{m}(\cos\Theta)$ are the Legendre Polynomials.
By defining

$$\mathbf{p} = [p_1, p_2, \ldots p_N]^T \quad (5)$$

as the loudspeaker signal vector and

$$\mathbf{B} = [Y^{1}_{0,0}(\Phi,\Theta), Y^{1}_{1,0}(\Phi,\Theta), \ldots Y^{-1}_{M,M}(\Phi,\Theta)]^T \cdot s \quad (6)$$

as the ambisonic signal vector, equation (2) can be written as

$$\mathbf{B} = \mathbf{C} \cdot \mathbf{p} \quad (7)$$

The matrix **C** contains the spherical harmonics for encoding the several loudspeaker signals into ambisonic. Now, the decoder can be calculated from the encoder as follows

$$\mathbf{D} = \mathrm{pinv}(\mathbf{C}) = \mathbf{C}^T \cdot (\mathbf{C} \cdot \mathbf{C}^T)^{-1} \quad (8)$$

Hence, the awareness of the playback configuration can be limited to the decoder while only the universal multi-channel format, defined by the ambisonic order, is implemented in the encoding stage. So, ambisonic provides a decoupling of the encoder and decoder. The decoder is only defined by the loudspeaker arrangement. Furthermore, the number of loudspeakers is independent of the number of virtual sound sources to encode. The minimum number N of required loudspeakers for a 3D system is limited by the number L of transmit channels, defined by the ambisonic order M, as follows

$$N \geq L = (M+1)^2 \quad (9)$$

1.2 Weighting of higher order ambisonic channels

The limitation to a finite number of ambisonic channels yields a degradation of the localization accuracy. As mentioned above, the main idea of the ambisonic approach is a decomposition of the incident plane wave into spherical harmonics that results in a series expansion. Considering a two dimensional system, the ambisonic equations may be derived using the two dimensional Fourier Transform [9]. Consequently the limitation to a finite number of ambisonic channels equals a multiplication of the series by a rectangular window. Now performing the Fourier Transform yields the so called angular sinc functions (asincs). The asinc functions describe the panning of the ambisonic signals to the several loudspeakers. Therefore, the localization may be confused by out of phase signals coming from loudspeakers far away from the intended virtual source position (figure 1). Weighting the amplitudes of higher order ambisonic channels, representing higher order spherical harmonics, yields a reduction of the side lobes amplitude as well as a broadening of the main lobe of the asinc functions. Though, as a result of the wider main lobe the localization blur increases. Now, several windowing techniques may be used to optimize the localization capabilities by considering just noticeable difference (JND) thresholds.

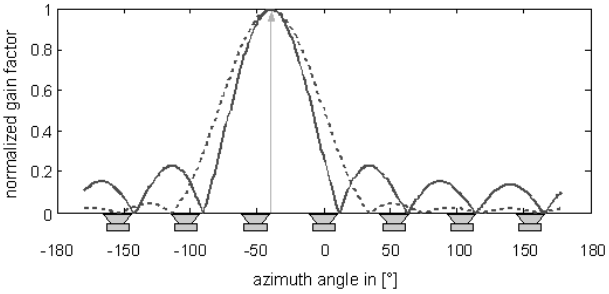

Figure 1: Absolute value of the angular sinc functions (asincs) with (dotted line) and without weighting (solid line) of the ambisonic channels, for an intended virtual source position of - 40° (gray arrow)

2 BINAURAL SOUND REPRODUCTION

2.1 The virtual ambisonic approach

To overcome the problem of high-quality, time-varying interpolation between different HRTFs in time-variant binaural sound reproduction systems, a virtual

ambisonic approach is used (figure 2). This approach is based on the idea to decode ambisonic to virtual loudspeakers. Then the binaural signals are created by convolving the virtual loudspeaker signals with HRTFs appropriate to their position in space. Now, the filtered signals are superimposed to create the left and right ear headphone signals.

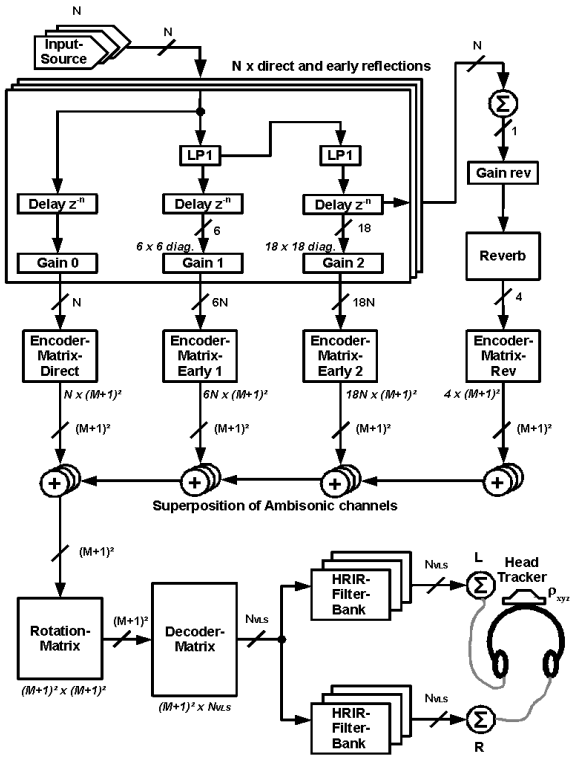

Figure 2: Block diagram of a 3D binaural sound reproduction system incorporating head tracking and room simulation

As is shown in figure 2, the sound source signals are encoded into ambisonic domain depending on their position in virtual space. To cover the different distances relative to the listener, the signals are delayed first. The number of ambisonic channels depends only on the ambisonic order and is therefore independent from the number of sound sources to encode. This fact is quite important for incorporating room simulation as is shown later. Now, head tracking is taken into account by simple rotation matrices in the ambisonic domain. Therefore, a head tracking device, based on the gyroscope principle, is mounted on the headphones to identify the actual head rotation.

As mentioned above, using the virtual ambisonic approach in time-varying binaural systems results in a bank of time-invariant HRTF filters. The decoding process is defined by the loudspeaker arrangement only. Therefore, to avoid ill conditioning or even singularities in the decoder matrix, it is important to distribute the loudspeakers as uniformly as possible over the spheres surface.

2.2 Room simulation

The incorporation of room simulation is an important fact to improve source localization capabilities as well as out-of-head localization in binaural systems. Room simulation is divided into two stages of computation:

- the early reflections
- the reverberant sound field

A simple geometrical approach assuming a rectangular room is used to calculate image sources of first and second order. Furthermore, the sound sources are assumed as omni-directional point sources. The acoustic properties of the reflecting walls are taken into account by low-pass filtering the image source signals using first order infinite impulse response (IIR) filters. Now, the image source signals are delayed due to their relative distance to the listener. To cover the direction of the early reflections, the signals are encoded into ambisonic domain according to their position in virtual space. The disadvantage of this approach is the huge increase of virtual sound sources to encode.

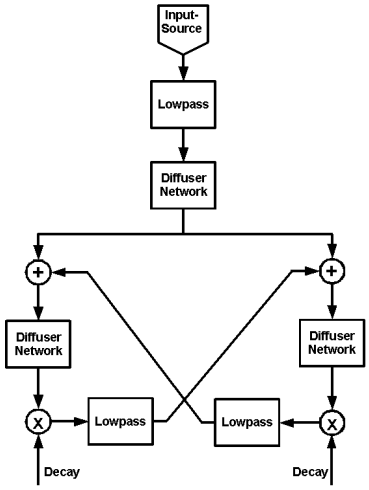

Figure 3: Recursive reverberation network

Late reverberation is taken into account by algorithms proposed by Dattoro [13]. Dattoro states that the most efficient implementations of reverberators rely on all-pass circuits embedded within very large globally recursive networks (figure 3). To consider the coloration due to the absorption of the enclosing walls, the signal is low-pass filtered first. Furthermore, the time where late reverberation starts is taken into account by simple delays. The overall structure of the reverberator is divided into two levels of computation. First, the input diffuser quickly decorrelates the input signals to prepare them for the next stage of computation. In the next stage the decorrelated sound signals are looped infinitely by globally feedback paths. Now, attenuating the feedback

path signals by multiplication with a gain factor less than one, offers a parameter to control the decay time. The texture of materials at the reflecting walls are taken into account by low-pass filters as before.
Finally the reverberation signals are encoded into ambisonic domain. Because of the fact that late reverberation signals are highly decorrelated, low order ambisonic is sufficient for encoding. This yields a reduction of the computational cost.
The proposed algorithm yields a simple way to control several parameters of reverberation like decorrelation, reverberation pre delay, reverberation gain, decay rate and cut-off frequency to take the wall characteristics into account.

2.3 Optimization

For implementation it is necessary to increase the computational efficiency of the proposed system. The following improvements have been carried out:
The effect of shortening the HRTF filters have been studied extensively for a two dimensional binaural system by carrying out listening tests [14] as well as by objective validation of the localization accuracy using simulation results [15]. Hence, the results have shown, that shorten the HRTF filters up to 128 taps does not heavily affect the localization accuracy.
Because of the fact, that humans localization accuracy is much more better in horizontal than in vertical directions, it is possible to encode vertical directions with ambisonic of lower order. This approach is termed mixed order ambisonic. Furthermore, higher order reflections for room simulation become more and more diffuse. So, they are encoded with lower ambisonic order as well.
Another approach to reduce the computational cost of room simulation is to divide the virtual acoustic space into subspaces. Now, image source signals situated in the same subspace are regarded as coming from the same direction. Henceforth, the bundled signals are encoded into ambisonic domain according to the direction of the associated subspace.
For increasing the localization accuracy by incorporating head tracking, rotation around the z-axis is quite sufficient.
Moreover, filtering the virtual loudspeaker signals with their appropriate HRTFs is done in the frequency domain. So, the signals are superimposed to create the left and right ear signals in the frequency domain as well. Therefore, the inverse Fourier transform has do be performed only for two channels.

3 IMPLEMENTATION

The proposed system is implemented on a usual notebook using pure data (pd), a graphically based open source real time computer music software by Miller Puckette [16].
A 2D system was implemented on a digital signal processing platform (DSP) first, running a PC as a host system.
A full 3D system incorporating ambisonic of 4th order has been optimized using the results as mentioned in [14, 15]. Reducing the computational cost of the algorithm made it possible to implement the system on a usual notebook, running a 1.6 GHz CPU.
Furthermore, the system was implemented on a DSP PCI platform to create a user interface for blind persons [17]. To ease the development of user applications for the system, an application programming interface (API) is intended.

4 CONCLUSIONS

In this paper the advantages of using a virtual ambisonic approach in binaural sound reproduction systems has been discussed. The main advantages of the virtual ambisonic approach are as follows.

- In time-varying binaural systems using the virtual ambisonic approach results in a bank of time-invariant HRTF filters.
- For multiple virtual sound sources the number of HRTF filters is independent of the number of sources to encode. Therefore, room simulation using image sources is possible.
- Head rotation is taken into account by simple rotation matrices in ambisonic domain.
- The awareness of the playback configuration can be limited to the decoder while only the universal multi-channel format is implemented in the encoding stage.

Therefore, convincing the advantages of the virtual ambisonic approach, several optimization criteria have been discussed to reduce the computational cost. With the rapid increase of CPU power it will become possible to run multi-channel binaural sound reproduction as a background task for applications.

5 ACKNOWLEDGEMENT

This work was partially supported by AKG-Acoustics GmbH, and the author wishes to thank Stefan Leitner and Martin Opitz for inspiring discussions.

6 REFERENCES

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

[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] 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)

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

[5] 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)

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

[8] 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)

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

[10] 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)

[11] Daniel J., Rault J.-B. and Polack J.-D., "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions", in *Proc. 105th Conv. Audio Eng. Soc.*, preprint 4795 (1998)

[12] 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)

[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., 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)

[15] 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)

[16] Pure data (pd):
http://crca.ucsd.edu/~msp/software.html
http://pd.iem.at

[17] Frauenberger, Ch., and Noisternig, M., "3D Audio Interface for the Blind", submitted to the *Int. Conf. on Auditory Displays*, Boston, MA, USA, (2003, July 6-9)