VECIMS 2003 - International Symposium on
Virtual Environments, Human-Computer Interfaces, and Measurement Systems
Lugano, Switzerland, 27-29 July 2003

3D Binaural Sound Reproduction using a Virtual Ambisonic Approach

Markus Noisternig, Thomas Musil, Alois Sontacchi, Robert Höldrich
Institute of Electronic Music and Acoustics
University of Music and Dramatic Arts
Inffeldgasse 10/3, 8010 Graz, Austria
Phone: +43-316-389-3170, Fax: +43-316-389-3171, Email: noisternig@iem.at

Abstract – Convincing binaural sound reproduction via headphones requires to filter the virtual sound source signals with head related transfer functions (HRTFs). Furthermore, humans are able to improve their localization capabilities by small unconscious head movements. Therefore it is important to incorporate head-tracking. This yields the problem of high-quality, time-varying interpolation between different HRTFs. A further improvement of humans localization accuracy can be done by considering room simulation yielding a huge amount of virtual sound sources.

To increase the computational efficiency of the proposed system a virtual Ambisonic approach is used, that results in a bank of time-invariant HRTF filter independent of the number of sources to encode.

I. INTRODUCTION

The following paper deals with the theory and practice of fully 3D binaural sound reproduction systems using a virtual Ambisonic approach.

Sound source spatialization in virtual acoustic environments requires to filter the source signals with head related transfer functions (HRTFs) to create the left and right ear signals. The HRTFs capture the diffraction of a sound wave for a certain angle of incidence caused by the torso, head, shoulders and pinnae of the listener. Consequently they show a great person-to-person variability. Wenzel et al. state in [1] that the use of nonindividualized transfer functions yields a degradation of humans localization accuracy. In the proposed system generic HRTFs have been incorporated using the KEMAR [2] as well as the CIPIC database [3].

In real-world environments humans are able to improve their sound source localization accuracy by performing small unconscious head movements [4]. To benefit from this phenomenon in binaural sound reproduction systems, head tracking has to be taken into account. Furthermore, the use of generic HRTFs decreases the perceived externalization of the virtual sound source, also termed as out-of-head-localization. Begault and Wenzel state in [5] that the incorporation of head tracking and room simulation improves the localization accuracy as well as the perceived externalization.

Therefore, the implementation of head tracking and moving virtual sound sources 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. Ambisonic is a sound reproduction technique involving a limited number of playback channels, while allowing reproduction of a full three dimensional acoustic space with several moving virtual sources.

The following section gives a brief introduction into Ambisonic theory. In section III a binaural sound reproduction system is developed using the Ambisonic approach, incorporating head tracking as well as room simulation.

II. THE AMBISONIC APPROACH

Ambisonic is a technique for spatial audio reproduction introduced in the early seventies by Gerzon [6]. Further details of Ambisonic theory are published in [7] – [11].

The holographic theory of wave field reproduction 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. Recently it has been shown that the Ambisonic formulation is asymptotically holographic assuming plane wave loudspeaker signals [7], [8]. Deriving the Ambisonic coding and decoding equations from the Kirchhoff - Helmholtz integral it can be shown, that the original wave field may be reconstructed exactly by arranging infinitely many loudspeakers on a closed contour. Using a finite number of N loudspeakers arranged on a sphere a good approximation of the original sound field may be synthesized over a finite area (sweet spot). In [9] it is shown, that higher order Ambisonic systems are increasingly accurate.

The decomposition of the incoming wave field into spherical harmonics can be shown deriving Ambisonic from the homogenous wave equation

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

where $p(t,\mathbf{r})$ is the sound pressure at the position $\mathbf{r}$ and c is the speed of sound. Solving the wave equation for the incoming sound wave as well as for the plane waves of the several loudspeakers yields the so called matching conditions [10]

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

The left side of (2) represents the Ambisonic encoding equation,

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

where $\mathbf{B}_{\Phi,\Theta}$ represents the Ambisonic channels in vector notation, s is the pressure of the original sound wave coming from direction (Φ,Θ) and $Y^{\sigma}_{m,\eta}$ describes the spherical harmonics. On the right hand side of (2) p_n is the signal of the n^{th} loudspeaker at direction (φ_n,ϑ_n). The spherical harmonics $Y^{\sigma}_{m,\eta}$ can be calculated 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^{η}_{m} represents Legendre polynomials. Now, using vector notation (2) may be written as

$$\mathbf{B} = \mathbf{C} \cdot \mathbf{p} \quad (5)$$

where

$$\mathbf{p} = [p_1, p_2, \ldots p_N]^T \quad (6)$$

is the loudspeaker signal vector and

$$\mathbf{B} = [Y^{1}_{0,0}(\Phi,\Theta), Y^{1}_{1,0}(\Phi,\Theta), \ldots Y^{-1}_{M,M}(\Phi,\Theta)]^T \cdot s \quad (7)$$

represents the Ambisonic channels. The Matrix **C** contains the decomposition of the several loudspeaker signals using spherical harmonics.

Now it is possible to calculate the decoding from the encoding equations as follows

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

As can be seen, the decoding stage will only depend on the actual loudspeaker arrangement. Consider the reproduction of a 2D field using a finite number of loudspeakers and deriving the horizontal holographic approach from the two dimensional spatial Fourier transform it is shown in [9], [11], that the decoding process may be described by the so called angular sinc functions (asincs). Consider the asinc-functions it can be seen that the auditory localization may be confused by signals coming from loudspeakers far away from the intended position of the virtual sound source. Like in conventional signal processing techniques using the Fourier transform, windowing can be used to attenuate the sidelobes of the asinc functions. Weighting the amplitudes of higher order Ambisonic channels is used to attenuate far away speaker signals considering just noticeable difference thresholds (JND) to increase localization accuracy. A broadening of the main lobe will occur as well yielding a broadening of the perceived sound source increasing the localization blur.

III. DEVELOPMENT OF A BINAURAL SYSTEM

Incorporating head tracking and moving virtual sound sources in binaural sound reproduction systems, the conventional approach of spatialization by convolving the source signals with HRTFs yields the problem of time-varying interpolation between different HRTFs. This interpolation results in artifacts decreasing the localization performance of the system. As mentioned above, to overcome the problem of time-varying interpolation between different HRTFs a virtual Ambisonic approach is used, as described in the following section.

A. The virtual Ambisonic approach

The virtual Ambisonic approach is based on the idea to decode Ambisonic to virtual loudspeakers. Then, after decoding the binaural signals are created by convolving the virtual loudspeaker signals with HRTFs appropriate to the loudspeaker position in space. Now the filtered signals are superimposed to create the left and right ear headphone signals (figure 1).

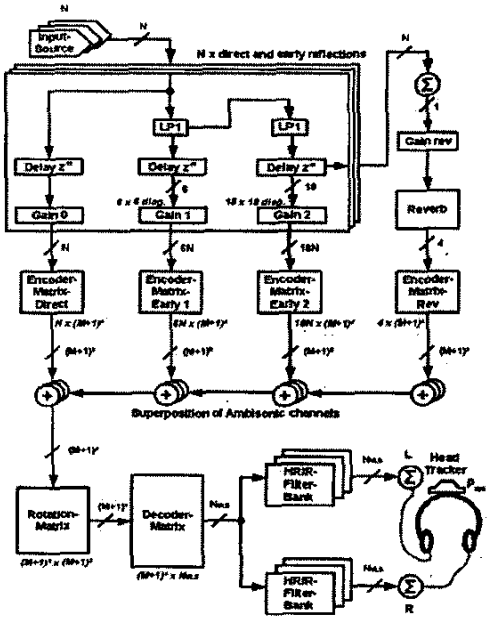

Figure 1. Block diagram of the 3D binaural sound reproduction system incorporating head tracking as well as room simulation.

A review of literature states that the Ambisonic approach offers a good localization performance only over a small listening area (sweet spot). However, binaural sound reproduction does not suffer from this effect, because due to the headphones the listeners position is always in the sweet spot area.

Moreover, the virtual sound source signals are encoded into Ambisonic domain dependent to their position in virtual space. To embrace their variable distances, the sound source signals are delayed relative to the listeners position first. Furthermore, the high frequency damping due to the sound waves propagation path is taken into account by simple IIR low-pass filter. Using fully 3D Ambisonic of M^{th} order yields $(M+1)^2$ transmit channels independent of the number of sources to encode. As shown later, this fact is important for room simulation. Henceforth, head rotation is taken into account by simple rotation matrices in the Ambisonic domain. Head rotation is identified by the use of a head tracking device mounted on the headphones.

As mentioned above, the decoder is defined solely by the position of the virtual loudspeakers. To avoid ill conditioning or singularities in the decoder matrix, it is important to distribute the virtual loudspeakers as uniformly as possible over the spheres surface.

Filtering with HRTFs is a highly computational task. Listening tests and error analysis of a 2D system have shown that shorten HRTFs up to 128 points yields a satisfactory localization accuracy [13], [14] (figure 2). Furthermore the use of individualized HRTFs will increase localization capabilities. Further investigations on the influence of different HRTFs to the overall systems localization performance have not been carried out during this work.

B. Room Simulation

To improve the perceived externalization of a virtual sound source the incorporation of room simulation is an important fact for binaural sound reproduction. In this section we focus on the simulation of sound reverberation as a natural phenomenon occurring when sound waves propagate in an enclosed space. The calculation of the room simulation is divided into two stages. First, the early reflections of first and second order are calculated. Then a model for efficient simulation of the diffuse sound field is introduced.

The early reflections are taken into account using a simple geometrical acoustic approach by calculating image sources. We are considering a rectangular room containing omni-directional virtual point sources. To consider the acoustic properties of the reflecting walls, the image source signals are filtered with a low order IIR low-pass filter. According the different distance of the image sources to the listening position, the image source signals are delayed and attenuated as well. Then every image source is encoded to Ambisonic dependent to their position in the virtual acoustic space.

Due to the fact, that higher order early reflections become more and more diffuse they are encoded with lower order Ambisonic. The loss of localization accuracy may be accepted to increase the computational efficiency.

Another approach to decrease computational cost for encoding the image sources is to divide the virtual space into several subspaces or regions of influence. Now image sources situated in same subspaces are bundled and encoded to Ambisonic domain according to the direction dedicated to their respective subspace.

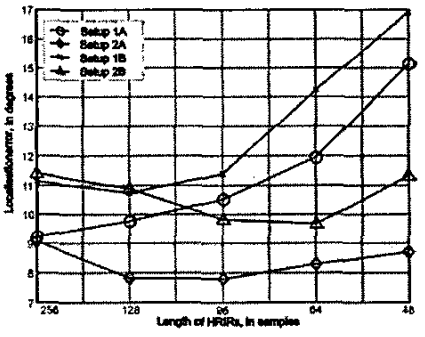

Figure 2. Localization error in degrees for binaural sound reproduction systems using different HRTFs (1,2) and different interpolation techniques (A,B)

Late reverberation creates an ambient space in the perception of the listener. A computational efficient calculation is to implement reverberators with all-pass circuits embedded within very large globally recursive networks (figure 3). To handle the start time of late reverberation the input signal is delayed. Then the signal is low-pass filtered to consider the coloration due to the absorption of high frequency signal components at the enclosing walls.

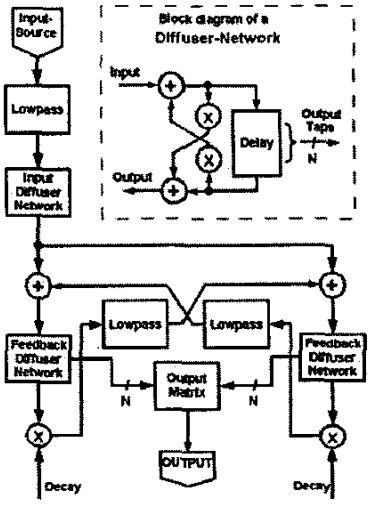

Figure 3. Recursive reverberation network

The first set of input diffusers is used to quickly decorrelate the incoming signal to prepare it for the next stage of computation. To loop the decorrelated sound indefinitely the second diffuser stages are arranged to feed back globally on themselves. By multiplying a gain factor <1.0 in the feedback paths it is possible to control the decay time. The low-pass 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.

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

C. Optimization

To increase the computationally efficiency of the overall system, the following improvements have been carried out:

Using shorter HRTF-filter increases the computational efficiency. Further studies and listening tests have shown [13], [14] that a filter length less then 128 taps yields a satisfactory localization performance.

Moreover, the use of a mixed order Ambisonic setup reduces the computational cost of the encoding as well as of the decoding stage. Because of the fact that humans localization accuracy is much more better in horizontal than in vertical directions, it is possible to encode elevation cues with Ambisonic of lower order.

The possibilities to reduce the computational cost of room simulations using lower order Ambisonic for encoding higher order reflections as well as bundling the image source signals according to their direction in virtual space are described in the previous section.

IV. IMPLEMENTATION

First a 2D system using 4^{th} order Ambisonic was implemented on a digital signal processor (DSP) running a PC as a host system. The system has been optimized using the mathematical model described in [13]. Furthermore the mathematical model as well as the systems localization accuracy has been evaluated by listening tests, using the 2D DSP platform to render the binaural signals in real-time [14].

The incorporation of the abovementioned optimizations made it possible to implement a fully 3D system with Ambisonic of 4^{th} order on a usual notebook running a 1.6GHz CPU. The proposed system was programmed using Pure Data (PD). PD is a graphically based open source real time computer music software by Miller Puckette [15].

V. CONCLUSIONS

In this paper the advantages of using the virtual Ambisonic approach to carry out binaural sound reproduction in real-time has been discussed. For multiple moving sound sources this approach brings an enormous benefit in increasing the computational efficiency of the system. The main advantages are

- 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 to encode.
- 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 multi channel 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.

With the rapid increase of CPU power it will become possible to run multi channel binaural sound reproduction systems as background tasks for applications like plug-ins for computer music software and for example interactive acoustical interfaces for blind persons.

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.

VI. ACKNOWLEDGEMENT

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

VII. REFERENCES

[1] 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

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

[3] 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

[4] Blauert, J., "Spatial Hearing", 2nd ed., MIT Press, Cambridge, MA, 1997
[5] 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
[6] Gerzon, M. A., "Ambisonic in multichannel broadcasting and video", in *J. Audio Eng. Soc.*, vol. 33, pp. 859-871, 1985
[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] Poletti, M., "A Unified Theory of Horizontal Holographic Sound Systems", in *J. Audio Eng. Soc.*, vol. 48, no. 12, 2000 December
[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] 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
[11] 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
[12] Dattorro, J., "Effect Design: Part 1: Reverberator and Other Filters", in *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 660-684, 1997 September
[13] 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
[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
[15] http://crca.ucsd.edu/~msp/software.html