A LIBRARY FOR REALTIME 3D BINAURAL SOUND REPRODUCTION IN PURE DATA (PD)

Thomas Musil, Markus Noisternig, Robert Höldrich

Institute of Electronic Music and Acoustics
University of Music and Dramatic Arts, Graz, Austria
{musil, noisternig, hoeldrich}@iem.at

ABSTRACT

This paper presents a library for programming 3D binaural sound reproduction systems using Pure Data (pd), an open source computer music programming language.

The theory that forms the basis of the proposed library is an improved virtual Ambisonics approach. Using this approach provides computationally efficient implementation of multiple moving sound sources, room simulation, head tracking and time varying listener positions in virtual space. Furthermore, simple GUI objects for controlling parameters are implemented to provide easy to use environment.

1. INTRODUCTION

The motivation to create a library for binaural sound reproduction using Miller Puckette's open source computer music programming language Pure Data (pd) [1], was to create a simple application programmers interface (API) for research projects as well as artistic productions. Previous works on the improvement of binaural systems [2] followed by reseach projects in different fields of virtual acoustics, psychoacoustics and multimedia, e.g. auditory interfaces for users with visual disabilities [3], [4], virtual environments [5], and realtime audio rendering systems for internet based applications, have shown the needs for an easy to use and computationally efficient object library.

The following paper deals with the implementation of fully 3D binaural sound reproduction systems using an improved virtual Ambisonics approach.

Concerning the creation of virtual acoustic environments using headphones, sound source spatialization requires to filter the sound streams with head related impulse responses (HRIRs). Real-world signals are acoustically filtered by the pinna, head and torso of the listener. The filtering can be described by HRIRs which capture both, the frequency and time domain listening cues to a sound source position. In the proposed system non-individualized HRIRs using the KEMAR [6], as well as the CIPIC [7], databases have been used. The use of non-individualized HRIR filter yields a degradation of localization accuracy [8],[9]. The synthesized sound source is usually localized incorrect either in distance or direction. Beyond, it is very common that the use of non-individualized HRIRs yields front back confusions which means, that sound sources synthesized in the front hemisphere are perceived in the back hemisphere. Therefore, the library provides fully interchangeability of HRIRs during runtime.

Regarding hearing in natural sound fields humans are able to improve sound source localization using small head movements. In [10] the importance of incorporating head tracking to binaural systems is shown. To overcome the problem of high-quality time-variant interpolation between different HRIRs a virtual Ambisonics approach is used.

Binaural sound reproduction systems often suffer from externalization errors, also termed inside-the-head-localization. Objects for room simulation, based on image source models for early reflections of first and second order as well as recursive reverberation networks for late reverberation, help to improve externalization. Therefore, the library provides objects for room simulation.

The next chapter gives a brief introduction into the virtual Ambisonics approach and the improvements that have been made during recent works. A detailed description of the overall algorithm is given in [2]. In the following sections an overview of library objects and control structures is given. The last chapter also shows in detail how to create room reverberation using the proposed library.

2. THEORY

As mentioned above, the proposed library uses a virtual Ambisonics approach. Ambisonic is a rendering technique for spatial audio reproduction, developed independently by several researchers in the early seventies [11] - [14]. However, the proposed binaural system is based on decoding Ambisonic channels to *virtual* loudspeakers, convolving the loudspeaker signals with HRIRs related to their position in the virtual room and superimposing related signals to create the left and right ear signals [15],[16].

2.1. Ambisonics Theory

Deriving the Ambisonic encoding and decoding equations from the Kirchoff-Helmholtz integral, it can be shown, that the original wavefield may be reconstructed exactly by arranging infinitely many loudspeakers on a closed contour. Using a finite number of N loudspeakers on a close contour a good approximation of the original sound field may be synthesized over a finite area (sweet spot). Further reading, e.g. [17], shows that higher order Ambisonic systems become increasingly accurate.

The decomposition of the incoming wave field into spherical harmonics can be shown [18]

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

where *p(t,***r***)* denotes the sound pressure at position **r** and c is the speed of sound. Solving the wave equation for the incoming sound wave as well as for the waves of the several loudspeakers, under the assumption of plane waves, yields the so called matching conditions [19]

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

The left side of equation (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 $Y_{m,\eta}^{\sigma}$ 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). $Y_{m,\eta}^{\sigma}$ may be calculated as follows

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

where P_m^{η} is the Legendre polynomial of degree m and $A_{m,\eta}$ are complex constants in general, whose values are determined by satisfying boundary conditions [18].

Using vector notation (2) may be written as

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

where

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

denotes the vector with the N several loudspeaker signals.

$$\mathbf{C} = \begin{bmatrix} Y_{0,0}^{1}(\Phi_1,\Theta_1) & Y_{0,0}^{1}(\Phi_2,\Theta_2) & \cdots & Y_{0,0}^{1}(\Phi_N,\Theta_N) \\ Y_{1,0}^{1}(\Phi_1,\Theta_1) & Y_{1,0}^{1}(\Phi_2,\Theta_2) & \cdots & Y_{1,0}^{1}(\Phi_N,\Theta_N) \\ \vdots & \vdots & \ddots & \vdots \\ Y_{M,M}^{-1}(\Phi_1,\Theta_1) & Y_{M,M}^{-1}(\Phi_2,\Theta_2) & \cdots & Y_{M,M}^{-1}(\Phi_N,\Theta_N) \end{bmatrix} \quad (7)$$

denotes the Matrix containing spherical harmonics $Y_{m,\eta}^{\sigma}$ of up to degree M to derive the several loudspeaker signals and

$$\mathbf{B} = [Y_{0,0}^{1}(\Phi,\Theta), Y_{1,0}^{1}(\Phi,\Theta), \dots Y_{M,M}^{-1}(\Phi,\Theta)]^T \cdot s \quad (8)$$

represents the Ambisonic channels.

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

$$\mathbf{D} = pinv(\mathbf{C}) = \mathbf{C}^T \cdot (\mathbf{C} \cdot \mathbf{C}^T)^{-1} \quad (9)$$

As equation (8) shows, the decoding stage only depends on the actual loudspeaker positions. Therefore, an optimal loudspeaker setup for 3D Ambisonic systems of 4^{th} order, that is aligned as uniformly as possible along the enclosures surface. Furthermore, the Ambisonics representation of a sound field is independent of the number and positions of sound sources creating the entire sound field.

2.2. Binaural Sound Reproduction using Ambisonics

The virtual Ambisonics approach is based on the idea to decode Ambisonics to virtual loudspeakers, and convolve the loudspeaker signals with HRIRs to create the left and right ear signals (Figure 1). As is shown in detail in [2], the virtual Ambisonics approach is very efficient to create binaural sound reproduction systems. For multiple moving sound sources this approach brings an enormous benefit in increasing the computational efficiency of the system. The main advantages are

- Rendering time varying sound fields using Ambisonics yields to a set of time-invariant HRIR filters without the need of interpolation
- The number of required HRIRs is independent of the number of virtual sound sources to encode. Regarding to room simulation, the enormous increase of virtual sound sources depends only to the efficiency of the encoding stage.
- Ambisonics 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 head tracking devices.

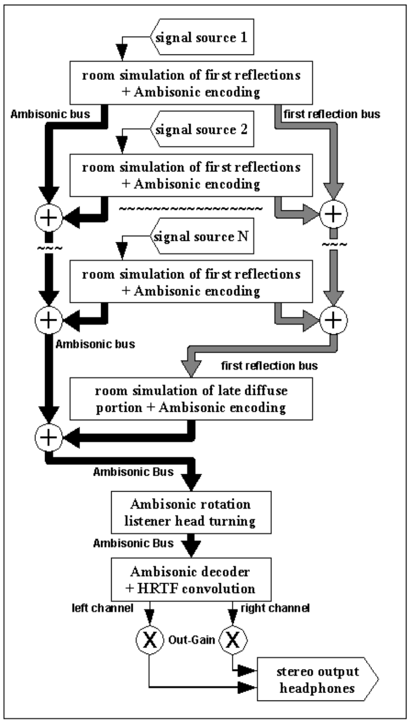

Figure 1: *Block diagram of a binaural system*

2.3. Room Simulation

Room simulation is used to improve the perceived externalization of a virtual sound source rendered by binaural systems. The calculation of the room simulation is divided into two stages [2]. In the first stage, early reflections of first and second order are calculated using a simple geometrical acoustic approach. Secondly, late reverberation is implemented by reverberators embedded in feedback delayed network structures.

To increase computational efficiency early reflections and late reverberation are encoded into Ambisonics domain regarding to the localization accuracy needed.

2.4. Optimization

To increase the computational efficiency of the overall system, the following improvements have been carried out [2]:

- Shorten length of HRIR filter due to results of listening tests
- Improved localization accuracy by windowing Ambisonic channels
- Improved encoders by using mixed order Ambisonics

The following chapter introduces the librariy objects and gives detailed descriptions to their functionality.

3. THE LIBRARY

The library described in the following sections provides objects for the use with Pure Data (pd). Pure Data is a graphically based open source software for real time audio and multimedia programming [1]. Furthermore, the library is available in precompiled versions for Linux and Windows XP.
As common practice in pd, to ensure computationally efficient implementation the library consists of objects for signal manipulation (signal domain, marked by ~) as well as of objects for controlling parameters (message domain).

3.1. iem_matrix

iem_matrix provides objects for matrix operations in signal domain. As mentioned above, matrix functions are needed to encode signals into Ambisonics domain, rotate the Ambisonics field and decode Ambisonic channels to virtual loudspeakers. Furthermore, in multichannel environments gain control may be simplified by the use of matrix operations. Therefore, *iem_matrix* is a collection of basis functions.

Matrix elements may be changed and updated via structured lists in message domain using library objects described in the following sections.

3.2. iem_ambi

iem_ambi provides objects for encoding, decoding and rotating Ambisonics in message domain. Each object sends parameter lists to signal manipulation objects (e.g. *iem_matrix* objects). The following objects are included in the library:

3.2.1. ambi_enc

This object provides parameter lists for matrix objects to encode a sound source into Ambisonics domain dependent on azimuth angle, angle of elevation and the order of the Ambisonic system.

3.2.2. ambi_rotate

This object provides parameter lists at the output for rotating the entire sound field in Ambisonics domain using 3 degrees of freedom. Therefore, the object *ambi_rotate* is fundamental to incorporate head tracking to increase localization accuracy.

3.2.3. ambi_decode

This object provides parameter lists at the output for decoding Ambisonic channels to loudspeaker setups. As input parameters this object needs lists with loudspeaker positions and the order of the Ambisonics system to decode.
In real world sound enhancement systems often just the upper hemisphere is used for arranging loudspeakers. Therefore the user has to take care to avoid singularities in the decoding matrix and has to compensate rippling of the virtual sound sources energy to smooth perceived movements.

3.3. iem_bin_ambi

iem_bin_ambi provides objects for decoding Ambisonic channels to a virtual loudspeaker setup. Furthermore, to increase computational efficiency the following optimizations have been carried out:

3.3.1. Reduced HRIR set

By assuming symmetry of the listeners head (that means symmetry of HRIRs) and establishing a left-right symmetrical loudspeaker setup, the number of required HRIR filter may be reduced by fifty percent.

3.3.2. Frequency domain filter

As former listening tests have shown, HRIR filter less than 128 taps decrease the localization accuracy heavily. Therefore, assuming longer filter, frequency domain filtering improves computational efficiency compared to time domain implementations.
Assuming real valued input signals and taking spectral redundancy into account, further improvements on computational efficiency may be obtained by using real FFT algorithms.

3.3.3. Reduced IFFT

By combining the decoder with the subsequent HRIR filter considering symmetries, a reduced filter set may be calculated as mentioned above. Now, by bringing these filters into frequency domain, one may reduce the number of required IFFTs to the number of two, one for the left and one for the right ear signal.

Taking a look to the Ambisonics decoder illustrates that the decoding matrix depends only on the loudspeaker setup. Irregular setups may cause matrix singularities. Preferably, the loudspeaker setup is aligned as uniformly as possible along a spheres surface. We found in our work that the optimal setup for Ambisonics of 4^{th}

order results in a **truncated icosahedron**, better known as the shape of a football.
Furthermore, considering the symmetry of the upper and lower hemisphere we found that totally symmetrical setups yield to singularities in the decoding matrix.

3.4. iem_roomsim

iem_roomsim provides objects for the calculation of early reflections of first and second order (Figure 2). A simple geometrical acoustic approach is used to calculate image sources considering a rectangular room containing omnidirectional virtual point sources. The output provides parameter lists containing delay times, damping factors according to distance and angles of arrival according to listeners position. The input of the object expects a parameter list containing listeners position, sound source position and the dimensions of the room.

Figure 2 shows a typical block diagram of *iem_matrix* objects to realize room simulation with first and second order early reflections by the use of the library *iem_roomsim*. The input signal is divided into direct and reflection signal paths where the initial filter take air and wall absorption into account. Each path is delayed due to the position of the signal source in the virtual room using a simple geometrical approach. Now signals are collected considering a bus structure for further processing. Then, transmission loss is taken into account by multiplication with damping factors. Finally all signals are encoded into Ambisonics domain dependent on their direction of arrival (DOA) referring to the listeners position in the virtual room.

3.5. iem_reverberation (subpatch)

This subpatch provides a computationally efficient calculation of late reverberation with globally recursive householder matrix structures based on former works of J.-M. Jot and A. Chainge [20], and M. Puckette [1]. Referring to Figure 3, the signals of the first reflection input bus (cp. Figure 2) are spreaded using a split matrix S acting as an input diffuser stage. Furthermore, the basic reverb algorithms from [20] are extended by equalizers using multiband shelving filter (LMH) to cover the acoustic properties of the reflecting walls. A dense sounding late reverberation signal is provided using compaction algorithms via matrix manipulation structures. Now, the signal is lowpass filtered (LP2) to consider the damping through the wave propagation path. Finally, after gain adjustment stage (G) the several reverberation signals are encoded to Ambisonics domain using sparsely spreaded spots in the virtual room.

3.6. iemgui

iemgui provides objects for creating graphical user interfaces. Using the Pure Data external Gem [21] allows to create and control 3D scenes within the environment of Pure Data, based on SGI's openGL standard. OpenGL instruction sets are mostly implemented in todays graphic chips, thus freeing CPU ressources for critical audio processing and controlling tasks.

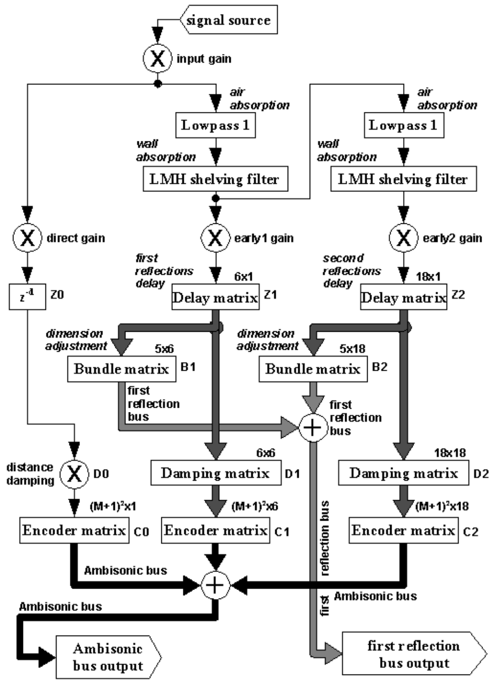

Figure 2: *Block diagram using iem_matrix objects to realize first and second order early reflections*

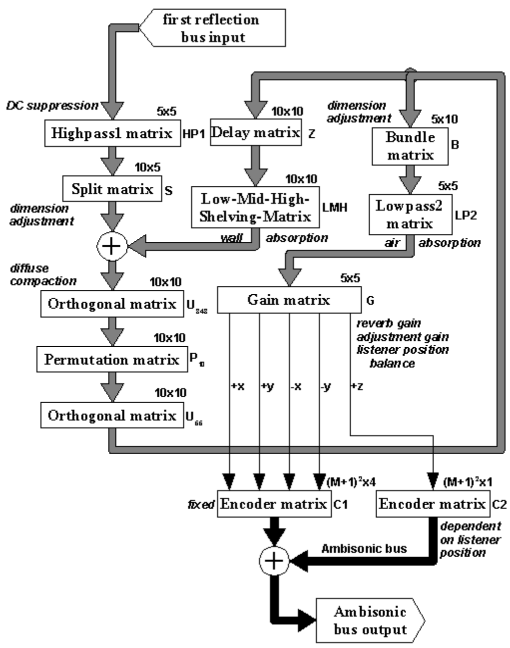

Figure 3: *Block diagram of the subpatch for late reverberation.*

.

4. CONCLUSIONS

In this paper a library for programming binaural systems using the open source programming language Pure Data has been presented, which is based on the virtual Ambisonics approach.

The advantage of the virtual Ambisonics approach is the computational efficiency for the use in 3D real time audio rendering applications incorporating head tracking, several moving sound sources, time varying listener position as well as room simulation.

Referring to the VARESE project by using pd's GEM library it is also possible to implement 3D real time graphics rendering.

The proposed library was evaluated during the realization of several research projects as well as artistic productions in the field of multimedia and contemporary music.

The proposed library will be available under GPL online under
http://iem.at/ projekte/dsp/bin_ambi

5. REFERENCES

[1] M. S. Puckette, Realtime Audio Programming Language *Pure Data (pd)*, http://crca.ucsd.edu/~msp/software.html, March 2005

[2] M. Noisternig, T. Musil, A. Sontacchi, and R. Höldrich, "3D Binaural Sound Reproduction using a Virtual Ambisonic Approach", in *Proc. Int. Symp. on Virtual Environments, Human-Computer Interfaces and Measurement Systems (VECIMS)*, Lugano, Switzerland, July 2003

[3] C. Frauenberger, and M. Noisternig, „3D Audio Interfaces for the Blind", in *Proc. Int. Conf. on Auditory Displays (ICAD)*, July 2003

[4] C. Frauenberger, V. Putz, and R. Höldrich, "Spatial Auditory Displays: A study on the use of virtual audio environments as interfaces for users with visual disabilities", in *Proc. Int. Conf. on Digital Audio Effects (DAFs'04), Naples, Italy, October 2004*

[5] V. Zouhar, R. Lorenz, T. Musil, J. M. Zmölnig, and R. Höldrich, „Hearing Varèse's Poème électronique inside a virtual Philips Pavilion", in *Proc. Int. Conf. on Auditory Displays*, Limerick, Ireland, July 2005

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

[7] V. R. Algazi, R. O. Duda, D. M. Thomson, and C. Avendano, "The CIPIC HRTF Database", in *Proc. IEEE Workshop on Applications of Sig. Proc. to Audio and Electroacoustics*, pp. 99-102, New York, October 2001

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

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

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

[11] D. H. Cooper , and T. Shiga, "Discrete-Matrix Multichannel Stereo", in *J. Audio Eng. Soc.*, vol. 20, pp. 346-360, June 1972

[12] M. A. Gerzon, "Multi-System Ambisonic Decoder, Part 1: Basic Design Philosophy", in *Wireless World*, vol. 83. no. 1499, pp. 43-47, July 1977

[13] M. A. Gerzon, "Multi-System Ambisonic Decoder, Part 2: Main Decoder Circuits", in *Wireless World*, vol. 83. no. 1500, pp. 69-73, August 1977

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

[15] D. G. Malham, "3-D sound for virtual reality using ambisonic techniques", in *Proc. 3rd Annual Conf. on Virtual Reality*, London, 1993

[16] C. Travis, "A virtual reality perspective on headphone audio", in *Proc. of the Audio Eng. Soc. U. K. Conference "Audio for New Media"*, 1996

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

[18] L. J. Ziomek, "Fundamentals of Acoustic Field Theory and Space-Time Signal Processing", *CRC Press Inc., Florida*, pp. 269, 1995

[19] J. S. Bamford, "An Analysis of Ambisonic Sound Systems of First and Second Order", in *Thesis at the University of Waterloo, Ontario, Canada*, 1995

[20] J.-M. Jot, and A. Chaigne, "Digital delay networks for designing artifical reverberators", in *Proc.of the 90th Conv. of the Audio Eng. Soc.*, Februar 1991

[21] J. M. Zmölnig, "GEM", *http://gem.iem.at,* April 2005