HEARING VARÈSE'S POÈME ÉLECTRONIQUE INSIDE A VIRTUAL PHILIPS PAVILION

Vit Zouhar, Rainer Lorenz†, Thomas Musil‡, Johannes Zmölnig‡, Robert Höldrich‡*

* Department of Music, Palacký University, Olomouc, Czech Republic.
vit.zouhar@upol.cz
† Institut für neue Musik und Medien, Hochschule für Musik, Karlsruhe, Germany.
lorenz@hfm-karlsruhe.de
‡ Institute of Electronic Music and Acoustics, University of Music and Dramatic Arts, Graz, Austria.
{musil,zmoelnig,hoeldrich}@iem.at

ABSTRACT

Topic of this paper is an interactive sound reproduction system to create virtual sonic environments to visual spaces, called *Virtual Audio Reproduction Engine for Spatial Environments* (**VARESE**). Using *VARESE*, Edgard Varèse's *Poème électronique* was recreated within a simulated *Philips Pavilion* (a construction originally designed by Le Corbusier and Iannis Xenakis for the 1958 World Fair in Brussels, then dismantled after its closing). The system draws on binaural sound reproduction principles including spatialization techniques based on the Ambisonics theory. Using headphones and a headtracker, listeners can enjoy a preset reproduction of the *Poème électronique* from their individual standpoint as they freely move through the virtual architectural space. While *VARESE* was developed specifically for its use in reconstructing *Poemè électronique*, it is flexible enough to function as a standard interpreter of sounds and visual objects, enabling users to design their own spatializations. The system runs on a standard laptop PC with a modern graphics card, using Miller Puckette's computer music software Pure Data (*pd*), plus an extended graphic interface (running equally on Linux and MS Windows operating systems and featuring a maximum of five sound sources using Ambisonics).

1. HISTORICAL BACKGROUND AND CONCEPTION

From 1956 until 1958, the star architect Le Corbusier, composers Iannis Xenakis and Edgard Varèse and film director Philippe Agostini co-authored a *Gesamtkunstwerk* commissioned by Philips for the World Fair EXPO 1958 in Brussels. An important part of the work was a spatialized reproduction of the electroacoustic work by Varèse and multiple projections of a film by Agostini. Marc Tribe describes the construction:

"The audio component was to be a demonstration of the effects of stereophony, reverberation, and echo. Sounds were meant to appear to move in space around the audience. Varèse was finally able to realize the movement of sounds through space from different directions. The audio tape consisted of three tracks, to give the illusion of three simultaneous sound sources that could be moving around the space. There were 350 speakers in 20 amplifier combinations (six amplifiers assigned to track one, eight assigned to track two, and six assigned to track three). A fifteen track control tape sent signals to the projectors and amplifiers. A series of "sound routes" were conceived so that the sounds could appear to move through the space from different directions. These were realized by designing an amplifier that would "span" a group of speakers, iterating through them five speakers at a time... This was one of the most elaborate site-specific projects ever created. The sound was written for the space and vice versa."[1]

Eight minutes in duration, its performance is followed by an electroacoustic composition: *Concrete PH* by Xenakis, 'PH' being an abbreviation for "paraboloid hyperboloid".

Due to being not a permanent construction, the *Philips Pavilion* was dismantled in January 1959; therefore the physical space for an authentic experience of the multimedia work ceased to exist. Only a two-channel version of Varèse's *Poème électronique* had been recorded and was subsequently published. Today, few drawings, sketches and other materials remain that illustrate - with limited degree of exactitude - principles of the original spatialization [1][2].

1.1. Reconstructions

Although there were attempts to rebuild a copy of the original pavilion in Eindhoven, not one project was realized in the end [3]. Various other performance reconstruction projects were undertaken over the last twenty years. The common aim of these projects was to recreate the original sound spatialization as closely as possible and to combine this with the film and colour-projections of the 1958 performances. The first such project was presented by the Asko Ensemble and Bart Lootsma on 11th of February 1984, in Eindhoven, Netherlands. Taking place in the Grote Zaal Auditorium Technische Hogeschool, with 72 loudspeakers situated on large panels, this performance matched the setup in the *Philips Pavilion*. Sound Paths, as the performance was called, was controlled by computer via Midi data.

Another reconstruction was realized on the Musica Sczienzia festival, on the 1st and 2nd of June in 1999, at Giardini della Filarmonica in Rome. It resulted from a collaboration between the architect and musician Valerio Casali and the artistic-scientific team of the Centro Ricerche Musicali. This production also featured a film projection. Kees Tazelaar (from the Studio of Sonology in The Hague) created a reconstruction at the Musica Viva 2004 festival in Lisboa. Tazelaar's setup included 34 sound sources. Tazelaar used digitizations of all original production tapes, which had been made by de Bruin and Varèse in Eindhoven.

1.2. Virtual reproduction

Presently, our team at the Institute of Electronic Music and Acoustics Graz has been experimenting with the 3D simulation of Varèse's *Poème électronique* since 2002. It was a logical next step in the application of spatialized and binaural systems developed at the IEM (Institute of Electronic Music and Acoustics) during the past few years[4]. The essential sources in use are the four digitized tapes of the original production tapes collection (3 mono, 1 stereo). We have kindly gained access to these sound materials (which were made by Tazelaar) with the permission of Konrad Boehmer, head of the Institute of Sonology at the Royal Conservatory in The Hague. Other source materials were the figure published by de Bruyn[5], a sketch of the sound paths by Xenakis[1] and various additional letters, diagrams and papers.[6]

The first spatialization of *Poème électronique* in the *IEM CUBE* concert space at the IEM Graz took place in 2003. The *IEM CUBE*'s reproducing sound equipment can generate an unlimited number of virtual sound sources. It meant that approximations of the original sound paths could be positioned into space, involving Midi and the software Samplitude and *pd* (see Fig.1).

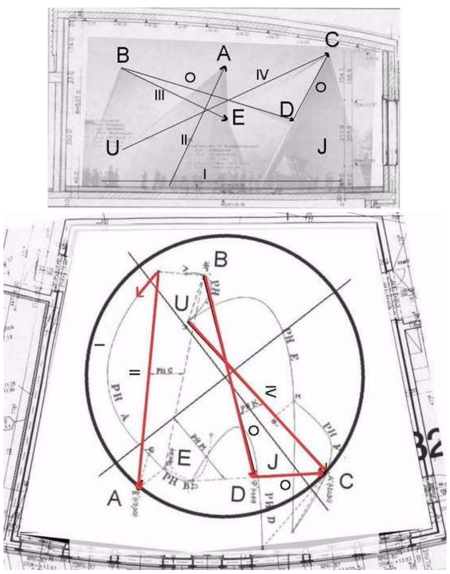

Figure 1: Layout and vertical section of the sound-paths in the *IEM CUBE*

From this, the conception of the system *VARESE* developed. In 2004, the graphics model of the *Philips Pavilion* and the audio-engine were implemented, leading to the current version as demonstrated in this paper. Now, a listener or interpreter can modify existing and record new paths of sound sources via *pd*, using a graphical interface.

2. VISUALIZATION OF THE PHILIPS-PAVILION

The basis for the visualization of the *Philips Pavilion* at the '58 World-exposition in Brussels is a virtual reconstruction of the building that was made in a 3D-animation program. Because of incomplete original source material schematic drawings and photographs which had appeared in several publications served as templates for the reconstruction. Additionally, research in various archives supplemented these materials, providing additional photos, diagrams, and other documents. The result is a computeraided 3-D simulation of the building, upon which the present realtime visualization is built.

In contrast with the already published, three-dimensional reconstructions of the *Philips Pavilion* that are limited to a representation of the building as a mapping of the twisted and bent plane surfaces (hyperbolic paraboloids) of the outward appearance, the goal of the present emulation was to reconstruct the inner chamber of the pavilion, in accordance with the original performance of the *Poème électronique* being held in the pavilion's interior.

There were additional constructions, struts and supports, that were not visible in any available documentation, and this created a problem; these important construction elements can only be ascertained and incorporated into the structure of the building by means of photographs. Arising from this problem is the placing of additional walls and entrance and exit doors. Figure 2 shows all supports, struts, and pillars with the additional walls in the entry area as well as the exit.

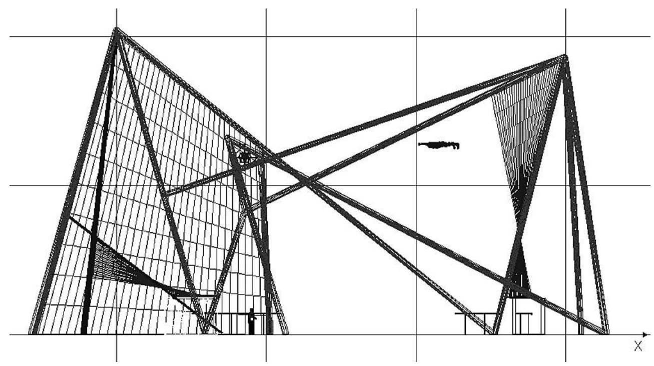

Figure 2: Construction elements of the Philips Pavilion

Another difficulty arose from the fact that no illustration of the exit area of the inner chamber exists. The photo in figure 3 shows a door inside the pavilion. One can see here how the sculpture is placed on the roof over the door near the entrance in the auditorium. However, this photo was mistakenly published in reverse, and this perspective could lead to the assumption that the picture deals with the exit, which is not the case.

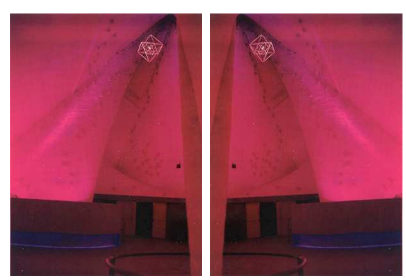

Figure 3: Reversed and corrected photo of the entrance

Plans of the technical facilities of the pavilion allowed for further conclusions to be made regarding the physical form of the inside. Combined with existing photographs of the inner chamber, this made the present reconstruction possible. The construction of the wall surfaces above the exit remains speculative, as well as the exact placement of all loudspeakers (most of which are recognizable, but, since there is no picture available, not those near the exit. Their exact placement was not documented.).

To show that the reconstruction of the pavilion with the existing documents is not an easy task, this example illustrates the difficulty, by interpreting details of a realized draft of the pavilion.

A design engineer in one of the published emulations included a stairway in front of the entrance, which, because of the height relationships in and around the pavilion, doesn't make any sense.

One method of checking the correctness of the reconstruction of the *Philips Pavilion* is by direct comparison of virtual photographs from the simulation with historical photos. One can then attempt to recreate a scenario by adjusting the location of the photographer and the optical setting of the camera to match that of the historical photo. It remains to be said that any attempt to emulate the Pavilion remains an approximation for reasons that the sources only partially represent this no-longer-existing building.

Figure 4: Virtual "inner chamber (Entrance)" in comparison with an archival photo (Fig. 3)

An additional challenge is the timing and placement of the projected images and light effects by Philippe Agostini that constitute the third aspect of the Gesamtkunstwerk *Poème électronique*.

3. SPATIALIZATION

The main goal of *VARESE* is to reproduce a threedimensional soundfield. Five soundsources (according to the original tracks) should be moved along sound-paths in a virtual *Philips Pavilion*. The listener should be able to move freely inside the pavilion and explore the spatial composition.

Because of the high interactivity of the application, it is meant as a single-user environment. A binaural reproduction system based on ambisonic principles recreates the soundfield via headphones.

3.1. Ambisonics

The main part of *VARESE* is based on the Ambisonics theory, which was developped by Gerzon [7] in the early 1970s.

The main idea is to compare a soundfield created by several soundsources with the soundfield created by the reproduction system and equate them.[8][9]

The Ambisonic theory takes only plane waves into account. A three-dimensional plane wave can be developped into an infinite Fourier-Bessel series. Since the high-order elements of the series tend towards zero, an approximation of the series can be obtained by encountering only the elements of lower orders. Thus a set of ambisonic channels can be defined as:

$$B^{\sigma}_{m,n} = Y^{\sigma}_{m,n}(\varphi, \vartheta)s(t) \tag{1}$$

While $s(t)$ is the sound pressure function, $Y^{\sigma}_{m,n}$ defines a weighting factor with a directional characteristic. These weighting factors are known as "spherical harmonics" and consist of linear combinations of sines and cosines of the angles of incidence. A set of weighting factor up to an order M (with $0 \leq n \leq m \leq M$) is called an *encoding rule*. For threedimensional encodings, the number of ambisonic channels N increases with $N = (M+1)^2$.

Due to the linear nature of the "spherical harmonics", a number of input signals can be encoded into an "ambisonic soundfield" by superposition:

$$\vec{B} = \mathbf{Y} \cdot \vec{s}(t) \tag{2}$$

Applying this encoding to both the "original" soundfield (consisting of a number of input signals $\vec{s}(t)$) and to the reproduction soundfield (created by a number of speaker feeds $\vec{p}(t)$), we get for the case of an ideal reproduction:

$$\mathbf{Y} \cdot \vec{s}(t) = \vec{B} = \mathbf{C} \cdot \vec{p}(t) \tag{3}$$

Since $\mathbf{C}$ is typically a non-square matrix, we can calculate the speaker-feeds with the aid of the pseudo-inverse:

$$\vec{p} = \mathbf{C}^{\mathrm{T}} \cdot (\mathbf{C} \cdot \mathbf{C}^{\mathrm{T}})^{-1} \cdot \vec{B} = \mathbf{D} \cdot \vec{B} \tag{4}$$

One big advantage of the technique is, that the intermediate ambisonic soundfield $\vec{B}$ is independent both of the number and positions of the input soundsources and of the speaker-layout used for reproduction. It is possible to manipulate the entire soundfield by manipulating the ambisonic encoded soundfield, again being independent of both the input and the output setup.

The computational load of encoding (and decoding) an ambisonic soundfield is rather low, as it only involves multiplications and accumulations. Additionally it is possible, to scale the system at the cost of reproduction quality by chosing an appropiate ambisonic order M.

Due to the limited number of reproduction channels and the finiteness of the order M, the soundfield can only reproduced perfectly within a small sweet spot.

Furthermore it is also important, that the loudspeaker setup is aligned as uniformly as possible along a spheres surface, in order to prevent ill-conditioning or even singularities of the decoder matrix $\mathbf{D}$.

3.2. Virtual Ambisonics

Ambisonics is generally used to recreate soundfields with multiple loud-speakers. As shown in [10], it is also an efficient means for binaural reproduction of dynamic artifical soundfields.

The main idea is to reproduce an ambisonic encoded soundfield to virtual loudspeakers. These virtual loudspeakers are then mixed down to a binaural signal, by convolving the loudspeaker-signal with HRTFs, according to the speaker-position. By combining the two techniques it is possible to overcome the shortcomings of both approaches.

Using virtual loudspeakers allows us to position any needed number of speakers on ideal positions, which is mostly not possible in real world. The use of headphones also guarantees, that the listener is always localted exaclty in the sweet-spot.

Using HRTFs directly is normally not feasible, as filtering signals with HRTFs is a very computational task. Furthermore, using the HRTF approach with moving soundsources would yield the problem of highquality time-varying interpolation between different HRTFs.

Using the virtual Ambisonics approach, the HRTFs are only applied to render the virtual speakers, which are immovable in relation to the listener's head. Therefore time-invariant HRTFs can be utilized.

The movements of single soundsources and the entire soundfield are calculated in ambisonic domain.

Using a headtracker it is thus possible to stabilize the soundfield in relation to the head-rotation. This improves the possibility to localize soundsources significantly.

4. IMPLEMENTATION

The implementation of *VARESE* has been done entirely in Miller S. Puckette's realtime computer music environment pure-data [11], because of the tight integration of realtime audio and graphics processing.

4.1. Audio engine

The structure of the audio-engine is shown in figure 5. Each soundsource is encoded into ambisonic domain according to its position relative to the listener. Furthermore, early reflections are calculated using virtual soundsources. The ambisonic encoded soundfields of each (virtual) soundsource are superimposed to form a representation of the entire soundfield. A directional reverb is generated for the entire soundfield.

The ambisonic representation of the soundfield can then be rotated according to the real head position of the listener (as detected by the head-tracker).

The ambisonic soundfield is then decoded to the virtual loudspeaker setup. Finally, the virtual loudspeakers are filtered with the apropriate Kemar-HRTFs and played back via the headphones.

4.1.1. Moving sounds

Ambisonics is used only to encode directional information. Since the positions of both soundsources and listener are given in cartesian coordinates, they are transformed into spherical coordinates, with the listener being in the origin of the coordinate system.

Azimuth and elevation of the soundsource are used to calculate the ambisonic channels. To achieve distance sensation, the input signal is lowpass-filtered and damped in relation to the distance between soundsource and listener. Furthermore, the signal is delayed to simulate the finite speed of sound in the air.

If a soundsource is moving, there has to be some interpolation between two successive positions. Since the movement is generally smooth and position changes are rather small, it is sufficient to

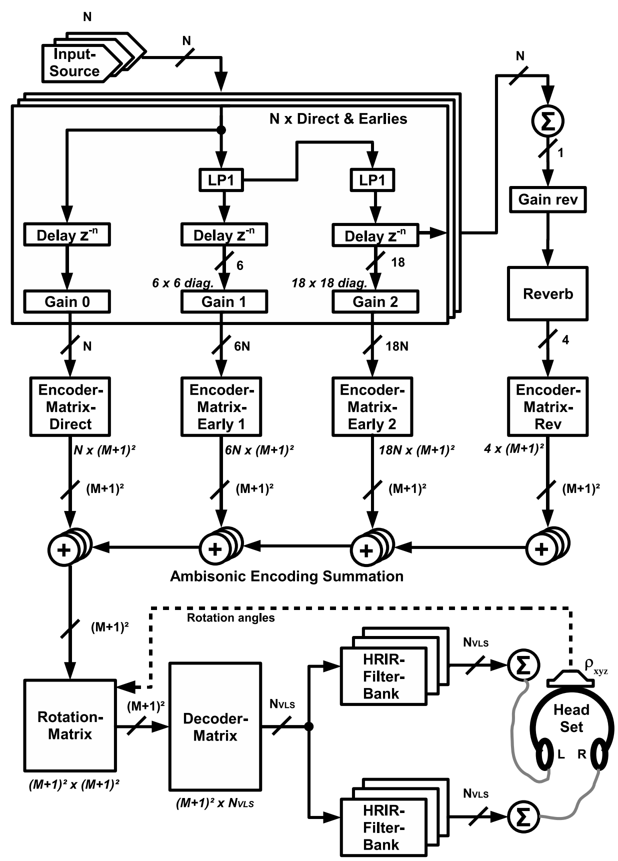

Figure 5: Structure of the audio-engine[10]

interpolate between the ambisonic weights rather than the angles of incidence.

Changes in distance can be realized with a variable delay line, which leads to a doppler effect. While this doppler effect is an exact simulation of a real world phenomenon it generally leads to an "unnatural" sensation, since mass-less virtual objects tend to be moved faster than in reality. While the resulting pitch shift might be acceptable for signals like speech, it is generally too much for musical signals.

Therefore (and because the original implementation of the moving sounds in the *Poème électronique* lacked the doppler shift too) a doppler-less implementation of a variable delay line is used. Such doppler-less vari-delay can be made, by cross-fading to constant delay lines.

4.1.2. Room simulation

The original *Philips Pavilion* had a very dry acoustics. Nevertheless, to improve localization it is necessary to add virtual acoustics by means of early reflections and reverb.

To simplify the task of reverberation, the complexity of the *Philips Pavilion* is reduced to a simple box. Based on this assumption, early reflections are calculated using virtual mirror sources up to the order of 2.

To keep the computational load low, diffuse reverberation is applied to the entire ambisonic encoded soundfield. A more natural reverb can be made by applying directional information. This is done by first decoding the ambisonic soundfield to five virtual speakers that are setup as a pyramid. These speaker-feeds are reverberated using a feedback delay network and then encoded back into ambisonic domain. Since the reverberation should not be highly localizable, it is sufficient to encode the reverberation soundfield with lower order Ambisonics. An Ambisonics system of 2nd order has proven to be satisfying.

4.1.3. The need for speed

Rendering a three-dimensional soundfield is rather costy in terms of CPU. Since the number of ambisonic encoded channels is relative to the square of the ambisonic order, a lot of CPU-power can be saved by reducing the order of the ambisonic system. According to the available ressources, one can choose between 3rd, 4th and 5th order Ambisonics.

Furthermore, early reflections and reverberation can be turned off completely, again freeing ressources.

4.2. Graphics engine

The implementation of the visualization is straight-forward: The external library Gem[12] allows to create and control 3D-scenes within the environment of pure-data, based on SGI's openGL. With modern hardware, openGL-instructions are calculated entirely on the graphicscard instead of the CPU, thus freeing ressources for doing the critical audio processing and controlling.

The "Philips Pavilion" is integrated into the scenes, by displaying a 3D-model, which is stored in Alias/Wavefront's OBJ-format.

The virtual soundsources are represented by simple shapes, which are moved according to the "sound-paths".

A representation of the listener is done by using a humanoid model, which can also be moved around. The head of the model can be turned according to the information gathered from the head-tracker.

Normally, the scene will be looked at from the eyes of the listener (also known as "1st person view"), to make the visual sensation match the acoustical one.

However, there are several other viewpoints to give an overview over the whole scene, e.g. a "bird's view". Naturally, these visualizations do not directly match the acoustical sensation.

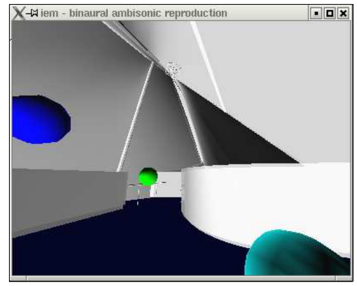

Figure 6: Visualization from the 1st-person-view

5. INTERACTION

The movements of both soundsources and the listener can be controlled either with pre-recorded "movement paths" or interactively by the user.

Navigating through a three-dimensional space is problematic, as normal human input devices, like the mouse, are only useful to navigate on a two-dimensional plane. The game industry has shown, that joysticks with more degrees of freedom might be used for this task. While using joysticks might be good (because we are used to it) to explore a 3D-scene from a 1st-person-viewpoint, it has proven that it is rather tedious to use it to move objects precisely along given paths, e.g. to move the soundsources according to compositional guidelines of the *Poème électronique*.

Since the mouse is the most common device for "drawing" and people are used to use it accurately, it has been chosen as the main input for movement. In one of the orthogonal viewpoints (e.g. "from above" or "from left") dragging the mouse on the visualization window moves one or more soundsources along the two obvious axes. To be able to simulatenously move along the third axis ("into the screen"), an additionally connected MIDI-controller can be used.

This gives full control over accurate movements using both hands of the user.

Additionally each object can be controlled by setting the coordinates via "numberboxes" (widgets that allow to enter values directly).

The sound-paths are stored as textfiles, containing equistantly sampled positions of the objects. This makes it possible to edit the sound-paths noninteractively in external editors.

6. CONCLUSIONS

In this paper the reconstruction of Varèse's *Poème électronique* inside a virtual *Philips Pavilion* has been described.

Based on commonly available hardware it allows to explore the original *Gesamtkunstwerk*, as it might have been meant by it's creators. The system is open to realize the user's own interpretation of the piece, in order to take future research and individual interpretations into account.

7. ACKNOWLEDGEMENTS

This project has been supported by the Institute of Electronic Music and Acoustics, of the University of Music and Dramatic Arts Graz, the Gesellschaft zur Förderung von Elektronischer Musik and by Palacký University Olomouc. Kees Tazelaar digitized the original production tapes and Prof. Konrad Boehmer kindly gave permission for their use at the IEM. J. v. d. Bigelaar helped the research at the Archive of Philips in Eindhoven.

8. REFERENCES

[1] M. Treib, "Time Calculated in Seconds : The Philips Pavillon, Le Corbusier, Edgard Varèse", Princeton, Princeton University, 1996, 282 p

[2] R. D. Lukes, "The Poème électronique of Edgard Varèse", PhD diss., Harvard U., 1996, 385 p

[3] "Kosten Philips Paviljoen", 18th February 1959, Philips Company Archives etc.

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

[5] S. L. de Bruin, "C. Die elektronische Steuerung, Verwirklichung des "Elektronischen Gedichts"", in *Philips Pavillon auf der Brüsseler Weltausstellung.*, Philips Technische Rundschau 20, 1958-59, p. 81

[6] Philips Company Archives, Eindhoven, Fundation Le Corbusier, Paris, etc.

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

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

[10] M. Noisternig, T. Musil, A. Sontacchi and R. Höldrich, "A 3D Real Time Rendering Engine for Binaural Sound Reproduction", in *Proceeding of the International Conference on Auditory Display*, Boston, MA, USA, July 6-9, 2003

[11] M.S.Puckette, "pure-data", http://crca.ucsd.edu/˜msp/software.html

[12] J.M.Zmölnig et al., "Gem", http://gem.iem.at