

# VIRTUAL AUDIO REPRODUCTION ENGINE FOR SPATIAL ENVIRONMENTS

*Thomas Musil*  
*Johannes M. Zmölnig*

Institute of Electronic  
Music and Acoustics,  
University of Music and  
Dramatic Arts, Graz,  
Austria.  
musil@iem.at  
zmoelnig@iem.at

*Vit Zouhar*

Department of Music,  
Palacký University,  
Olomouc, Czech  
Republic.  
vit.zouhar@upol.cz

*Robert Höldrich*

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

## ABSTRACT

Topic of this paper is an interactive sound reproduction system to create virtual sonic environments and visual spaces, called *Virtual Audio Reproduction Engine for Spatial Environments (VARESE)*. Using *VARESE*, Edgard Varèse's *Poème électronique* was recreated within a virtual *Philips Pavilion*. The system draws on binaural sound reproduction principles including spatialization techniques based on the Ambisonics theory. Using headphones and a head tracker, listeners can enjoy a preset reproduction of the *Poème électronique* as they freely move through the virtual architectural space. While *VARESE* was developed specifically for its use in reconstructing *Poème électronique*, it is flexible enough to function as a standard interpreter of sounds and visual objects, enabling users to design their own spatializations.

<http://iem.at/projekte/newmedia/VARESE>

## 1. MOTIVATION

From 1956 until 1958, 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 electro acoustic work by Varèse and multiple projections of a film by Agostini.

In January 1959 the *Philips Pavilion* was dismantled; 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 [10][5].

### 1.1. Physical reconstructions of the *Poème électronique*

Although there were attempts to rebuild a copy of the original pavilion in Eindhoven, not one project was realized in the end [7]. 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 color-projections of the 1958 performances. The first such project was presented by the Asko Ensemble and Bart Lootsma on 11<sup>th</sup> 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 Scienza festival, on the 1<sup>st</sup> and 2<sup>nd</sup> of June in 1999, at Giardini della Filarmonica in Rome, a production that also featured a film projection.

For presentation in Omniversum in The Hague 2003, Kees Tazelaar (from the Studio of Sonology in The Hague), Willem Hering, Piet Lelieur and Robin Sip prepared full reconstruction of the *Poème électronique* using original film, color projections and sound.

Kees Tazelaar also created a reconstruction at the Musica Viva 2004 festival in Lisbon. His setup included 32 channel spatialization. Tazelaar used digitizations of all original production tapes, which had been made by de Bruin and Varèse in Eindhoven.

### 1.2. Virtual reproduction

In 2002 our team at the Institute of Electronic Music and Acoustics Graz (IEM) has started experiments with 3D simulations of Varèse's *Poème électronique*. It was a logical next step in the application of spatialized and binaural systems developed at the IEM during the past few

years[9].

After initial experiments involving a large 24 channel 3D loudspeaker setup, the system was miniaturized into the 2-channel binaural system *VARESE*.

In 2004, the graphics model of the *Philips Pavilion* was reconstructed by Rainer Lorenz[12] and incorporated into the newly designed audio engine, leading to the current version as demonstrated in this paper.

## 2. SPATIALIZATION

The main technical goal of *VARESE* is to reproduce a three dimensional sound field. Several sound sources (five according to the original tracks of the *Poème électronique*, but likely more for other compositions) should be moved along sound paths inside a virtual building (e.g. inside the *Philips Pavilion*).

The listener should be able to move freely inside the building and explore the spatial composition. Because of the high interactivity of the application, it is meant as a single-user environment.

An interactive binaural reproduction system based on ambisonic principles recreates the sound field via headphones.

### 2.1. Virtual Ambisonics

The main part of *VARESE* is based on the Ambisonics theory ([1]), which was developed by Gerzon [3] in the early 1970s.

While Ambisonics is generally used to recreate sound fields with multiple loudspeakers, it is also an efficient means for binaural reproduction of dynamic artificial sound fields, as shown in [6].

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

The virtual nature of the loudspeakers allows to use an ideal loudspeaker layout, with the listener always located exactly within the sweet spot. Since the HRTFs are only applied to render the virtual speakers that are immovable in relation to the listener's head, time-invariant HRTFs can be utilized. The movement of sound sources is already handled in (computational efficient) ambisonics domain.

Using a head tracker it is possible to stabilize the sound field with respect to the head-rotation. This significantly improves the ability to localize sound sources.

## 3. IMPLEMENTATION

The implementation of *VARESE* has been done entirely in Miller S. Puckette's real-time computer music environment pure-data [8], because of the tight integration of real-time audio and graphics processing.

### 3.1. Audio engine

The structure of the audio engine is shown in figure 1. Each sound source is encoded into ambisonic domain according to its position relative to the listener. Furthermore, early reflections are calculated using virtual sound sources. A directional reverb is generated for the entire sound field.

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

The ambisonic sound field is then decoded to a virtual loudspeaker setup. As an experimental setup it is possible to apply order-based weights to the ambisonic encoded sound field. This is a known technique to widen the sweet-spot of the ambisonic reproduction to a sweet-area at the cost of blurring the image in the original sweet-spot a bit (e.g. *in-phase*-decoding). While a listener with headphones is always located within the sweet spot, this allows *VARESE* to be used as a binaural mastering tool for 3D-sound fields, which are later reproduced via a "real" multi-channel loudspeaker setup.

Finally, the virtual loudspeakers are filtered with the appropriate HRTFs, mixed down to a stereo signal and played back via headphones.

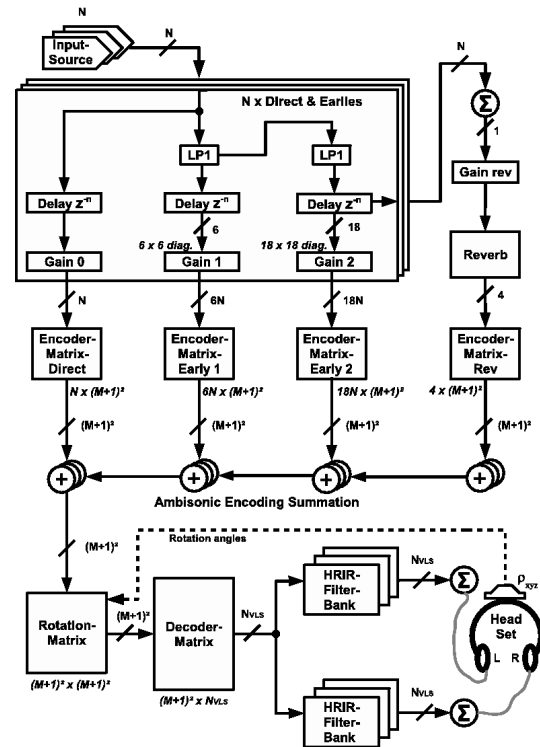


Figure 1. Block diagram of the structure of the audio engine[6]

Reproducing the sound field with a real loudspeaker setup can be done by simply exchanging the decoder unit (in fact, only the HRTF-filtering has to be turned off).

### 3.1.1. Using HRTFs

Per default, VARESE uses the diffuse-field compensated Kemar HRTF-set[2]. However the used HRTF-set can be exchanged, e.g. by measurements of the listener themselves. The filtering is done in frequency domain. For evaluating purposes the length of the used HRTFs can be changed.

### 3.1.2. Moving sounds

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

To achieve distance sensation, the input signal is low pass-filtered and damped in relation to the distance between sound source and listener. Furthermore, the signal is delayed to simulate the finite speed of sound in the air.

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. The resulting pitch shift is generally unacceptable for musical signals. Therefore a Doppler-less implementation of a variable delay line is used.

Additionally, the sound sources are damped with respect to the distance between listener and source according to the  $\frac{1}{r}$ -law. To allow the user to control the extent of this damping, an exponential factor  $\alpha$  has been introduced, so the actual damping is relative to  $(\frac{r_h}{r})^\alpha$ .

### 3.1.3. Room simulation

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 sound field. To make the reverb sound more natural, directional information is applied to it, by first decoding the ambisonic sound field to five virtual speakers. 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 sound field with lower order Ambisonics. An Ambisonics system of 2<sup>nd</sup> order has proven to be satisfying.

### 3.1.4. The need for speed

Rendering a three-dimensional sound field is rather costly 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 resources, one can choose between 3<sup>rd</sup>, 4<sup>th</sup> and 5<sup>th</sup> order Ambisonics.

## 3.2. Graphics engine

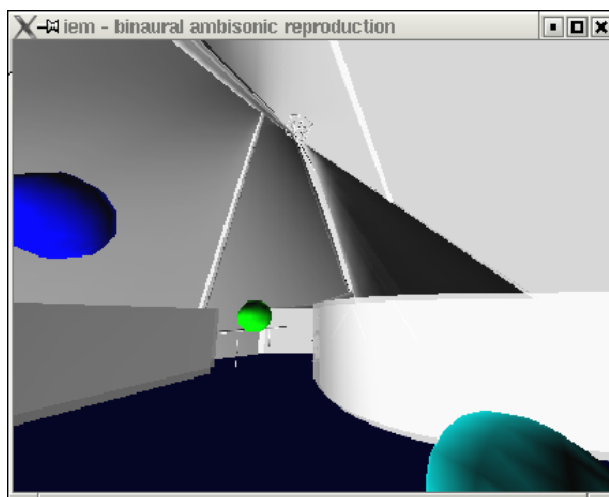
The implementation of the visualization is straight-forward: The pure-data external Gem[11] allows to create and control 3D-scenes within the environment of pure-data, based on SGI's OpenGL. Since modern graphics cards provide hardware acceleration of OpenGL instructions, this again frees resources for 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 sound sources 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.

While normally the scene will be looked at from the eyes of the listener to make the visual sensation match the acoustical one, there are several other viewpoints to give an overview over the whole scene.



**Figure 2.** Visualization of the Philips Pavilion from the 1<sup>st</sup> person view

## 4. INTERACTION

The movements of both sound sources and the listener can be controlled either with pre-recorded “movement paths” or interactively by the user.

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 the orthogonal viewpoints dragging the mouse on the visualization window moves one or more sound sources along the two obvious axes. To be able to simultaneously 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.

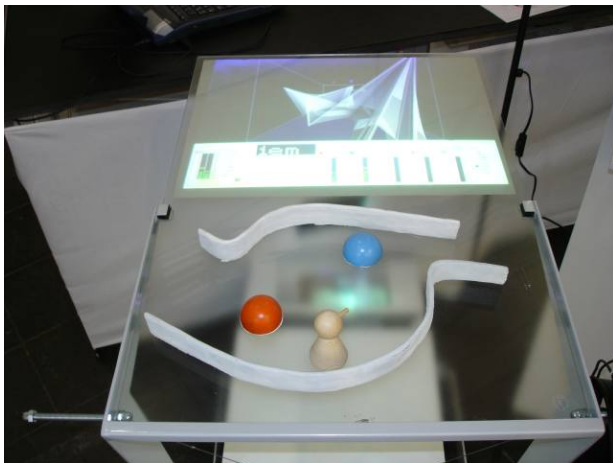
The sound paths are stored as text files, which makes it possible to edit the sound paths non interactively in external editors.

#### 4.1. Haptic interaction

While the mouse is available on virtually any “multimedia-enabled” computer, usage has shown that it often lacks enough feedback to give the user the experience, that they are *really* moving an object in space. The mouse as a *pointing device* does not represent “objects as such” very good.

Therefore we are currently experimenting with more intuitive haptic interfaces, which give the user the possibility of simultaneous interaction with several objects while still providing an overview of the complete scene.

Both sound sources and listener are represented by simple geometric forms which can be moved on a plane surface (see fig.3) The objects are tracked via video tracking. The x/y-position of the objects directly represent the position in the virtual world. By twisting an object one can control the vertical excursion of the corresponding sound source.



**Figure 3.** A haptic interface for controlling the movement of sound sources (presented at the CeBIT 2005 in Hannover) [4]

#### 5. CONCLUSIONS

In this paper a system for audio and video reproduction of spatial environments has been described. Based on commonly available hardware it allows to explore otherwise unavailable spatial compositions, like Edgard Varèse’s *Poème électronique*.

As a specialized research tool for Varèse’s piece, it is open to realize the user’s own interpretation in order to

take future research and individual interpretations into account.

As a general reproduction tool it can be used as an aid in spatial composition, both in auditory and visual domains.

#### 6. REFERENCES

- [1] 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
- [2] B. Gardner, K. Martin, “HRTF Measurements of a KEMAR Dummy-Head Microphone”, <http://sound.media.mit.edu/KEMAR.html> (March 1, 2005)
- [3] M. A. Gerzon, “Ambisonic in multichannel broadcasting and video”, *J. Audio Eng. Soc.*, vol. 33, pp. 859-871, 1985
- [4] “Mixed Reality Interface (MRI)”, <http://www.kommerz.at/mri/> (March 13, 2005)
- [5] R. D. Lukes, “The Poème électronique of Edgard Varèse”, PhD diss., Harvard U., 1996, 385 p
- [6] M. Noisternig, T. Musil, A. Sontacchi and R. Höldrich, “A 3D Real Time Rendering Engine for Binaural Sound Reproduction”, in *Proc. of the Intl. Conference on Auditory Display*, Boston, MA, USA, July 6-9, 2003
- [7] “Kosten Philips Paviljoen”, 18th February 1959, Philips Company Archives etc.
- [8] M.S.Puckette, “pure-data”, <http://crca.ucsd.edu/~msp/software.html> (March 1, 2005)
- [9] 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
- [10] M. Treib, “Time Calculated in Seconds : The Philips Pavillon, Le Corbusier, Edgard Varèse”, Princeton, Princeton University, 1996, 282 p
- [11] J.M. Zmólnig et al., “Gem”, <http://gem.iem.at> (March 1, 2005)
- [12] V. Zouhar, R. Lorenz, T. Musil, J.M. Zmólnig and R. Höldrich, “Hearing Varèse’s Poème électronique inside a virtual Philips Pavilion”, in *Proc. of the Intl. Conf. on Auditory Display*, Limerick, Ireland, July 6-9, 2005