ViMiC Developments
Technical Report

Nils Peters

June 29, 2007

Chapter 1

1.1

1.1.1 ViMiC Visualization tool

A visulaisation tool was programmed in PD using the graphics Environment for Multimedia (GEM) to have a visual control and feedback over the ViMiC environment. This tool helps to monitor the spatial scenery by displaying all source positions (red boxes) and microphone positions (blue boxes) in relative position to the floor of the virtual room (figure 1.1). All positions are updated in real-time and the user can switch between different viewing angles.

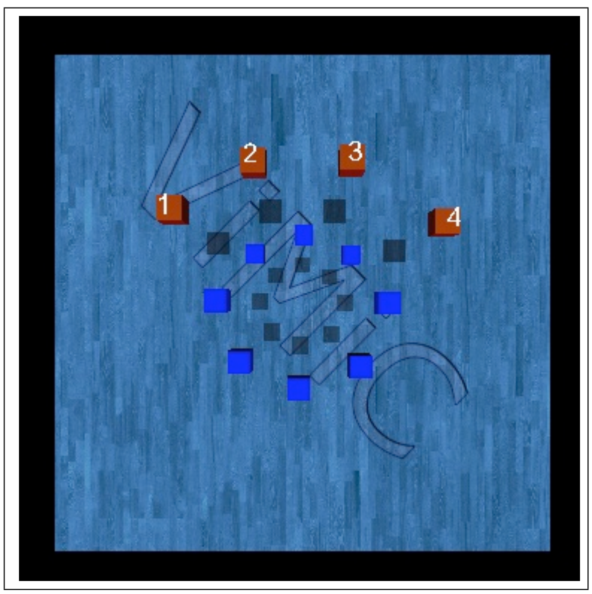

Figure 1.1: ViMiC 3D visualization tool; virtual microphones (blue), sound sources (red)

1.1.2 Control plug-in

An engineer in a mixing studio might prefer to control ViMiC spatialization from within his/her preferred sequencing program. There is also the need to store automation data together with pre-recorded audio material to create complex and dynamic soundscapes. The ViMiC unit can be controlled from most commercially available Digital Audio Workstations. A plug-in was designed to run on VST, RTAS or Audio Units host applications. The control plug-in based on the Pluggo Runtime Environment for Max/MSP. A separate plug-in unit can be loaded for each audio track to be spatialized with ViMiC. The DAW software automation can be used to control the values of all ViMiC parameter data. The ViMiC control plug-in communicates with the dedicated auditory rendering system through OSC control messages and a UDP network. The pre-recorded (and also unprocessed) signal is streamed from the DAW to the ViMiC rendering unit through a digital multichannel audio connection.

Figure 1.2: ViMiC control as a VST plug-in

Controlable parameter

There are a number of adjustable parameters, all normalized between zero and one, which affects this auditory virtual environment, including:

- Position of the sound sources [X, Y, Z]
- Ratio of direct early and late reflections
- Absorption properties of the reflecting surfaces
- Diffuse reverberation (decay time, density, EQ)
- Room size [X, Y, Z]
- Microphone angles and position [α, φ], [X, Y, Z]
- Microphone directivity patterns

1.2 ViMiC reverberator

In the process of improvement the old *Reverberator* module was replaced by a more flexible and better sounding algorithm.

1.2.1 Introduction

To create or simulate musical performance in (virtual) spaces, there is a need to have efficient algorithm to render reverberation. When a room is seen as a time invariant filter it can be described by a impulse responds h(t). Figure 1.3 shows a simplified and partitioned room impulse responds (RIR). [Theile, 2001] shows in figure the perceptional aspects for each of this parts to determine the simulated room. In this project the reverb tail is going to be simulated.

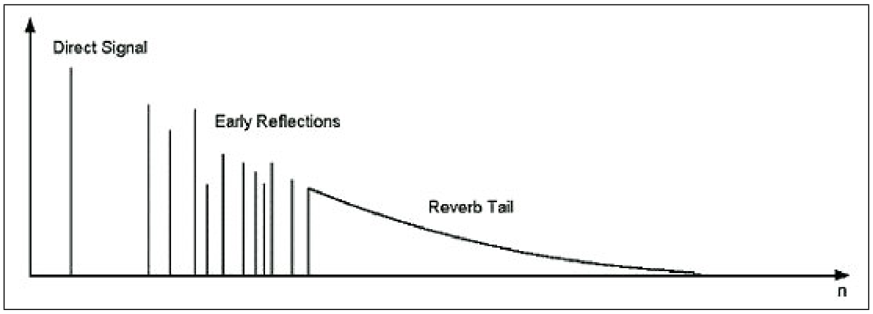

Figure 1.3: simplified room impulse responds (RIR)

1.2.2 Method

The use of feedback delay networks (FDN) for artificial reverberation can be justified fundamentally by a stochastic model of late reverberation decays assuming that sufficient density of acoustic model in the frequency domain and of reflections in the time domain are achieved. Under these assumptions, later reflections can be modeled as a Gaussian exponentially decaying random process, characterized by a spectral envelope. Although the FDN can be seen as a stochastic model it also has an physical interpretation in the sense that the delay-length represents different propagation paths in the room and the mixing matrix controls "diffuse" vs. "specular" reflections. This approach cannot guarantee the same degree of accuracy than a convolution with a measured impulse response, but provide a more efficient parametrization for dynamic control of the synthetic reverberation effect like reverberation time, modal density or reverb coloration.
A FDN proposed by [Jot, 1992] is shown in figure 1.4. Any FDN can be represented as a set of digital delay lines whose inputs and outputs are connected by a feedback/mixing matrix which defines the FDN structure.

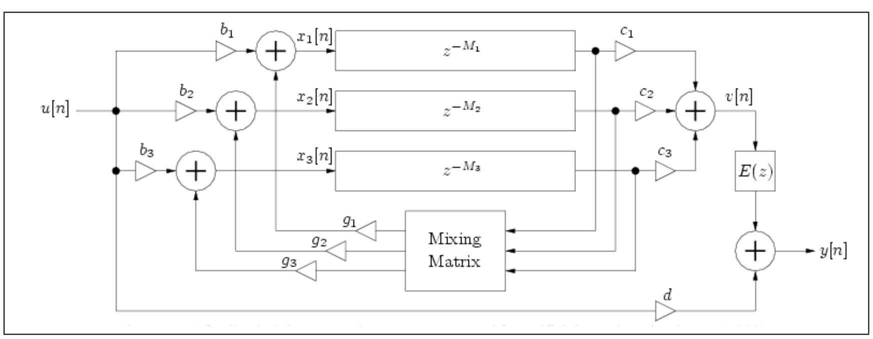

Figure 1.4: feedback delay network proposed in [Jot, 1992]

This matrix provides diffusion by mixing the energy of every input to every output. The matrix is supposed to be unitary (or orthogonal) to be maximal efficient in creating uncorrelated signals. A System is said to be unitary if its matrix transfer function is unitary for any complex variable z on the unit circle. Another condition is that the matrix is lossless. Matrices which fulfill these requirements are classes titled Householder matrices, Hadamard matrices or Circulate matrices [Smith, 2006].

The decay control is provided by the weights in every feedback path and depends on the chosen reverberation time t_60 and the length M_i of the delayline in the path:

$$g_i = 10^{-3M_iT/t_{60}(0)} \tag{1.1}$$

1.2.3 Implementation

The reverb is implemented as a FDN with 16 delay lines, seen in the figure 1.5. A 16x16 Hadamard matrix is used (see fig. 5). The advantage of this matrix is that it can be implemented with a butterfly algorithm using N log2 N = 64 additions. Afterwards the outputs of the matrix has to be scaled by $1/\sqrt{16}$ which can be combined within the multiplication g_i for the decay time. I made some modifications before and after the FDN. At the beginning I added a static predelay to shift the reverb tail in time in relation to the direct sound input (see figure 1.3). The delay depends on the volume of the simulated room and is calculated by the following equation [Jot et al., 1997].

$$t_{onset} = \sqrt{V} \text{ (ms)} \tag{1.2}$$

Due to the fact that there is only one input but 16 output, the coloration filter which Jot was placed behind the FDN is now at the beginning. This is much more efficient, considering the user does not want to filter every output independently. To be able to use the reverb in a multichannel loudspeaker setup every delayline has an output which is feeding an additional object to merge the 16 outputs together to the number of loudspeaker (2, 5, 8 or 16 speaker).

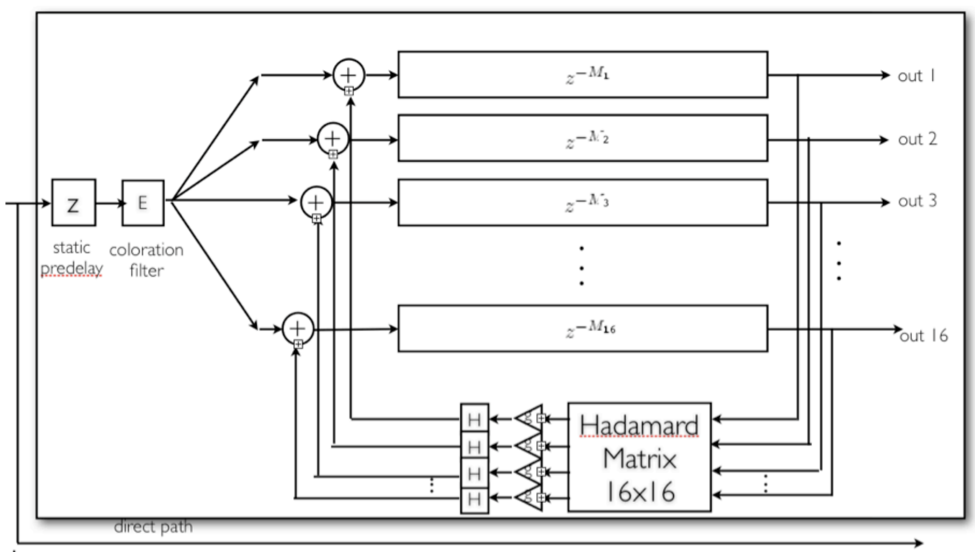

Figure 1.5: Implemented FDN

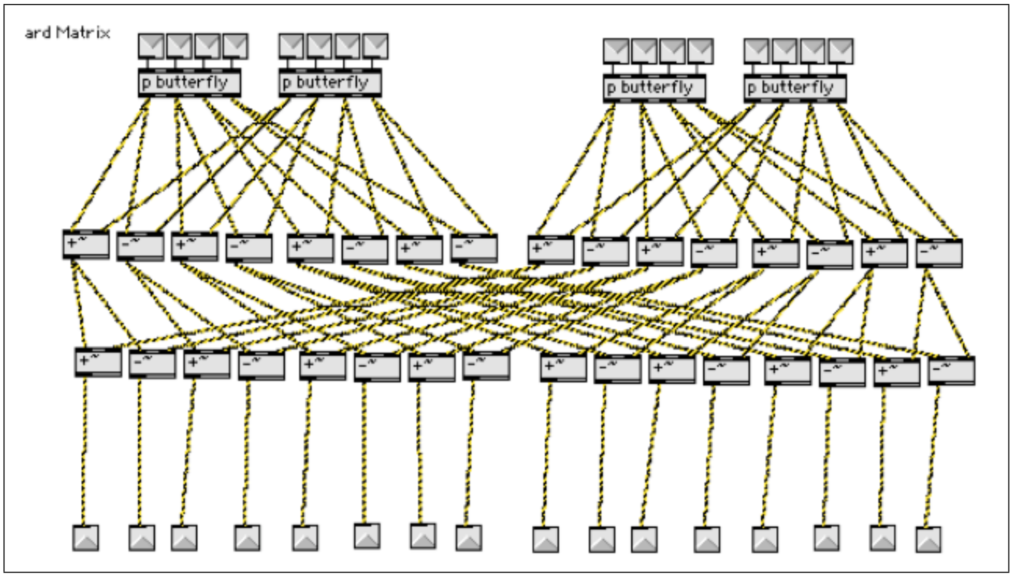

Figure 1.6: Butterfly algorithm to calculate the Hadamard matrix

1.3 Binaural rendering for headphone listening

ViMiC was extendend with an option to render the simulated soundfield especially for reproduction based on headphones. The purpose of this extension is to give composers the possibility to create spatial music without being surrounded by a number of loudspeakers which would usually be needed to auralize spatial elements of their pieces. It is also planned to use this headphone based version for the musicians during the concert to give them the possibility to hear themselves playing in the spatialized soundscape - comparable to that sound the audience will perceive.

1.3.1 Method

Humans can localize sound sources in a 3D space with good accuracy using several cues. Assuming that the listener receives the sound material via stereo headphones, it is possible to reproduce most of the cues that are due to the filtering effect of the pinna-head-torso system, and inject the signal artificially affected by this filtering process directly to the ears [Rocchesso, 2002, p. 149]
Classic psychoacoustic experiments have shown that, when excited with simple sine waves, the hearing system uses two strong cues to estimate the apparent direction of a sound source. Namely, interaural level and time differences (ILD and ITD) are jointly used to that purpose. ILDs are mainly useful above 1500 Hz, where the acoustic shadow produced by the head becomes effective, thus reducing the intensity of the waves reaching the contralateral ear. For this high-frequency range and for stationary waves, the ITD is also far less reliable, since it produces phase differences in sine waves which often exceed 360”. Below 1500 Hz the ILD becomes smaller due to head diffraction which overcomes the shadowing effect and it is possible to rely on phase differences produced by the ITDs. ITD’s and ILD’s are part of the so called *head-related transfer functions* (HRTF), which describes the filtering process caused by the ears before a sound can reach the ear drums. These HRTFs evaluated at discrete azimuth and elevation angles are sufficient for the synthesis of realistic three-dimensional sound events for headphone [Huopaniemi and Karjalainen, 1997].

These HRTFs have to be measured in a anechoic chamber. In [Rocchesso, 2002, p. 154] such procedure is described. Unfortunately HRTFs differ in between people and yield to artifacts in the spatial perception. Furthermore the fitness of nonindividual HRTFs is almost unpredictable. Fortunately there are several public HRTF databases available which allows to choose from a variety of filter-sets:

Database	URL
Acoustic Information System Laboratory	`www.ais.riec.tohoku.ac.jp/lab/db-hrtf/`
AUDIS HRTF databases	`www.eaa-fenestra.org/Products/Documenta/Publications/09-de2`
CIPIC databases	`http://interface.cipic.ucdavis.edu...` `.../CIL_html/CIL_HRTF_database.htm`
Itakura Laboratory databases	`www.itakura.nuee.nagoya-u.ac.jp/HRTF/`
Kemar databases	`http://sound.media.mit.edu/KEMAR.html`
Listen databases	`http://recherche.ircam.fr/equipes/salles/listen/`

Ideally, filtering a dry monaural signal with the HRTF for sound sources coming from $30°$ and reproducing the result via headphones will let the sound coming from $30°$.

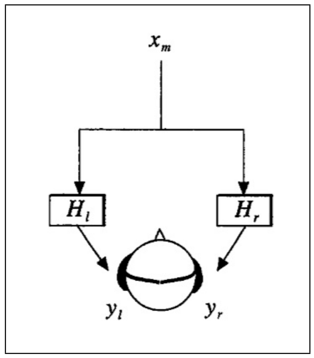

Figure 1.7: Transferfunction for binaural processing, [Huopaniemi and Karjalainen, 1997]

1.3.2 Implementation

When a sound source movement is simulated by the use of HRTFs, the applied filter have to be constantly changed and/or interpolated according to current position of the sound source. To overcome the problem of expensive high-quality, time-varying interpolation between different HRTFs, ViMiC is taking benefit of its virtual micophone approach. Instead of rendering every sound source according to the current position, the soundfield which is captured by the virtual microphones is filtered by a set of HRTF with the corresponding directions: If a microphone is placed at $45°$ azimuth in the ViMiC simulation, it is filtered with an HRTF for $45°$ (See figure 1.3.2) . As long as the microphone is not displaced the same filter can be applied, independently from the number of sound sources which are spatialized.
This implementation is going to be presented at the *ICAD - International Conference on Auditory Display 2007*.

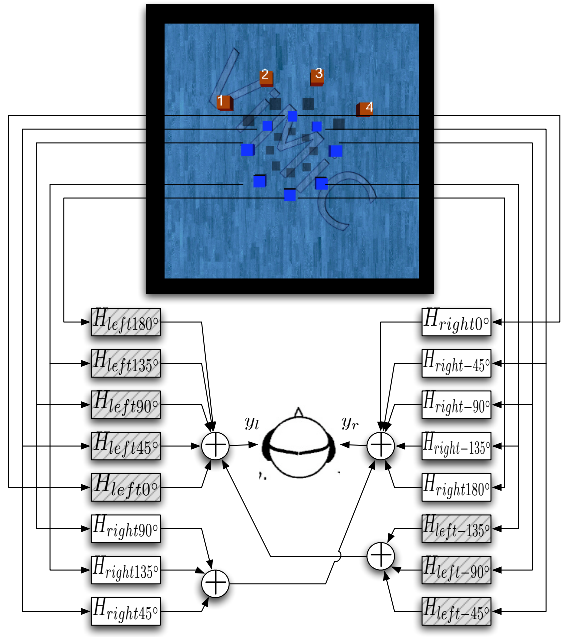

Figure 1.8: HRTF filtering for 8 virtual microphones

1.4 Integration of Open Sound Control

When different kinds of interfaces and sensor devices were connected to control the spatialization system in real-time it became clear that *Open Sound Control* (OSC) would ease the communication in between them a lot. Open Sound Control [Wright et al., 2003] is a protocol for communication among computers, sound synthesizers, and other multimedia devices that are optimized for modern networking technology. Especially the good temporal accuracy when controlling many parameters simultaneously and the possibility to communicate with other OSC-compatible devices facilitates the dataflow. Therefore a OSC namespace was developed and included into the ViMiC rendering unit. Figure 1.9 shows a general strategy to control ViMiC via sensor devices. In this context the implementation of the *Gesture Description Interchange Format* GDIF should be mentioned which is a format for standardizing gesture related information and under development for this project [Marshall et al., 2006].

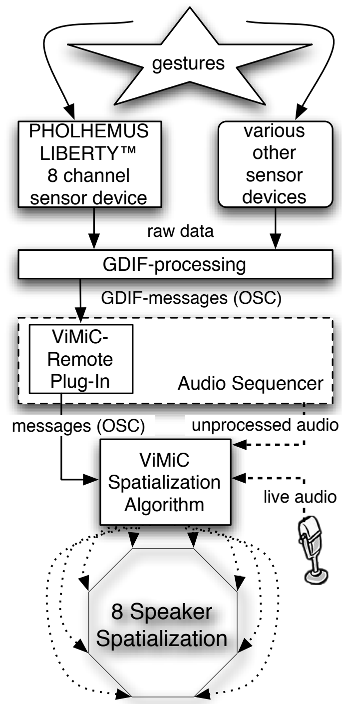

Figure 1.9: OSC dataflow

Bibliography

J. Huopaniemi and Matti Karjalainen. Review of Digital Filter Design and Implementation Methods for 3-D Sound. In *102nd AES Convention, preprint 4461*, Munich, Germany, 1997.

Jean-Marc Jot. *Etude et Réalization d'un Spatialisateur de Sons par Modèles Physiques et Perceptifs*. PhD thesis, France Telecom, Paris 92 E 019, 1992.

Jean-Marc Jot, L. Cerveau, and Olivier Warusfel. Analysis and synthesis of room reverberation based on a statistical time-frequency model. In *Preprint 4629 of the 103th Audio Engineering Society Convention*, 1997.

MT Marshall, N. Peters, AR Jensenius, J. Boissinot, MM Wanderley, and J. Braasch. On the development of a system for gesture control of spatialization. *In Proceedings of the 2006 International Computer Music Conference*, pages 6–11, 2006.

Davide Rocchesso. Spatial effects. In Udo Zölzer, editor, *DAFX : Digital Audio Effects*, chapter 6. John Wiley & Sons, Ltd, 2002.

Julius O. Smith. *Physical Audio Signal Processing: Digital Waveguide Modeling of Musical Instruments and Audio Effects*. Center for Computer Research in Music and Acoustics (CCRMA) Department of Music-Stanford University-California, 2006.

G. Theile. Multichannel Natural Music Recording Based on Psychoacoustic Principles. In *extended version of the paper presented at AES 19 thInternational Conference*, 2001.

M. Wright, A. Freed, and A. Momeni. OpenSound Control: state of the art 2003. In *Proceedings of the 2003 conference on New interfaces for musical expression*, pages 153–160, Singapore, 2003.