I/VOID/O: real-time sound synthesis and video processing in an interactive installation*

Daniel Luís Barreiro[1], Sandro Canavezzi de Abreu[2], André Carlos Ponce de Leon Ferreira de Carvalho[3]

[1]Faculdade de Artes, Filosofia e Ciências Sociais
Universidade Federal de Uberlândia (UFU)
Av. João Naves de Ávila, 2121 – Bloco 1V – 38408-100 – Uberlândia – MG – Brazil

[2]Faculdade de Arquitetura, Urbanismo e Design
Universidade Federal de Uberlândia (UFU)
Av. João Naves de Ávila, 2121 – Bloco I – 38408-100 – Uberlândia – MG – Brazil

[3]Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo (USP)
Av. Trabalhador São-carlense, 400 – PO Box: 668 – 13560-970 – São Carlos – SP – Brazil

dlbarreiro@gmail.com, sandroid@gmail.com, andre@icmc.usp.br

Abstract. *This paper presents some technical and aesthetic aspects involved in the conception of the interactive installation I/VOID/O concerning both the visual and sonic processes generated in real-time with Max/MSP/Jitter. It mentions the main characteristics of the installation and how the patches were implemented in order to provide coherent relationships between sound and image with the aim of offering an immersive experience for the people who visit and interact with the installation.*

1. Introduction

This paper presents real-time interactive processes with sounds and images implemented in the installation *I/VOID/O*, by Sandro Canavezzi de Abreu, with soundscapes by Daniel Barreiro. The installation was exhibited in the event *Emergência - Emoção Art.ficial 4.0*, at Itaú Cultural, Sao Paulo, from 1st of July to 15th of September 2008.

In this installation, images are captured inside a metallic sphere with a mirrored internal surface. Four cameras are used inside the sphere, one of which is placed at the tip of a stick that can be manipulated by the people who visit the installation. Two other cameras are linked to capture stereoscopic images, which are presented only in the last

* The sound synthesis implementations that are discussed here were part of a Pos-Doc research carried out with the support of Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, Brazil.

stage of the interactive process (explanation regarding the stages, understood as levels of immersion, are presented later on in this paper). Another camera is also used to capture images inside the sphere, but not stereoscopic ones. A fifth camera is positioned on one of the walls of the installation to capture the image of the person who interacts with the sphere (see Figure 1).

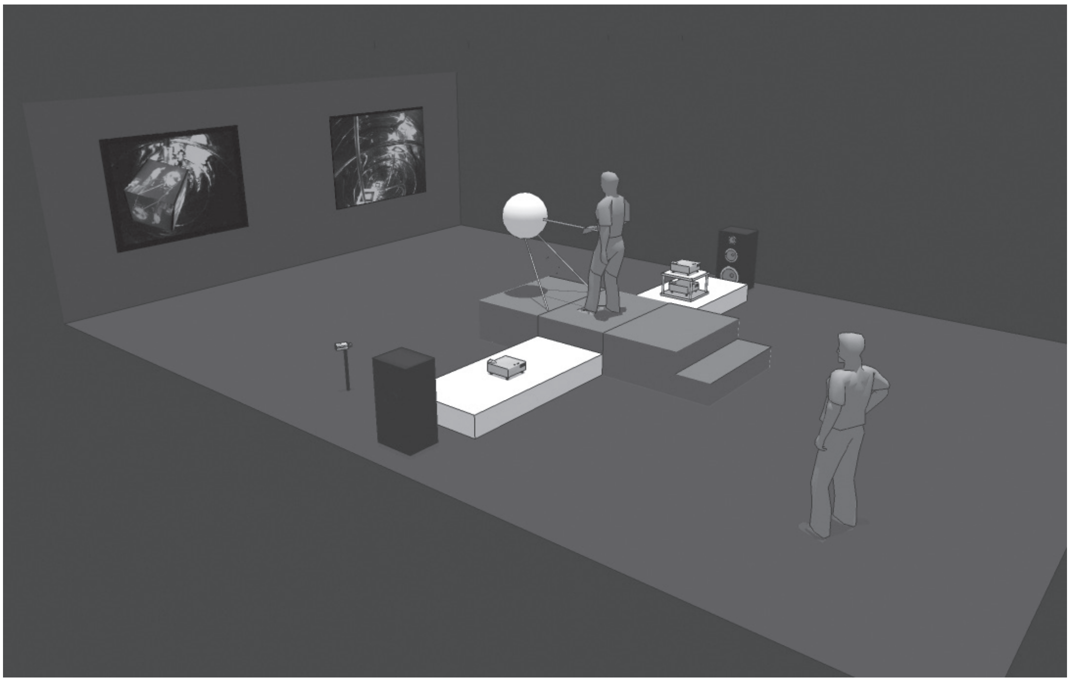

Figure 1: Diagram of the installation

The visual content of the installation is altered according to the level of immersion of the interactive process (in a total of four main levels and two transitory ones). Two video projection screens are used, placed side by side. Images, in black and white, are projected on the left screen. The right screen is only operational in the last stage (last level) of the interactive process with images in red and blue that enable 3D vision with the aid of red/blue glasses. The sounds – generated in reaction to the transformation of the images and the passage through different levels – are distributed over four loudspeakers (each one placed on one of the four walls of the installation), which contributes to generate an immersive environment.

Five computers are used in order to distribute the computational tasks and due to specific demands related to the projection of the images. They carry out the following roles:

- one computer for the real-time sound synthesis processes;

- one computer for the augmented reality system (software *artoolkit* running in Linux);

- one computer for the global management of the operations and movement tracking;

- two computers for the projection of the stereoscopic images (we decided to juxtapose the red/blue images from two different projection sources in order to obtain higher colour fidelity).

The images captured inside the metallic sphere are processed in different ways in each one of the four main and two transitory levels. The images are also analysed using various computational tools and the analysis data are sent to the computer that carries out the audio processes using the OSC protocol (Open Sound Control – see Wright et al. 2003)[1] via Ethernet connection.

The installation is seen as a system that can present different behaviours as the levels unfold. It articulates different relationships between images and sounds on each level (an example can be seen on http://br.youtube.com/watch?v=fnffoU7pX2o).

The following topics present an outline of the aesthetic proposal that informed the conception of the installation, followed by a description of the images and the immersive levels that occur in *I/VOID/O*. A following topic mentions the processes used for generating the sounds in real-time and their relationship with the images. After that, a brief evaluation of the sonic results is presented, followed by final considerations which point out to possible future developments of our work with interactive installations.

2. Aesthetic Proposal: *I/VOID/O*, a metainterface

The installation *I/VOID/O* (input/VOID/output) approaches the observation of an object as a physically interactive phenomenon. The *interactor* (person who interacts with the installation), while watching the content of a 'black-box' (this sentence is itself an oxymoron, in cybernetic terms [Wiener 1962]), interferes in such a way with the observed object that he/she ends up (re)creating it. However, that which is created, the internal reality that is observed, is nothing more than the *interface* itself. As a consequence, the interface is related to itself, a *meta-interface*, which redesigns itself continually from the initial *input* provided by the user: his/her observation.

This feedback process, which determines the non-linearity of the system, is the logical background that permeates the whole interaction. This feedback, however, is open, i.e., the feedback parameters are dynamic and are used for reorganizing the system. This dynamicity occurs as a result of the permanent confrontation between the analog and the digital domains which are present in the interface (here understood as a "field of tension" or *Schnittstelle* [Zielinski 1997]). The tension generated by this confrontation, the variation of levels of entropy in the system (in which one domain destabilizes or controls the other in an attempt to assimilate it mutually), is the material of the interface.

Light (image as reflected on the mirror and image captured by cameras) is translated in discrete parameters that change the behaviour of both sound and image in a continuous and vertiginous 'loop'. In this dynamic process, the interactor is taken to move between dimensions, or levels of immersion, touching different realities and internal universes of the 'black-box'.

This motion erodes the isolated vision of an external super-observer (as in the Classical Objectivity) and generates a cognitive friction in the observer/interactor. This

[1] http://archive.cnmat.berkeley.edu/OpenSoundControl/

friction is the result of spatial paradoxes created by the interface. These paradoxes are created by the reflection of the images on the concave mirror: they seem to release from the mirror surface and ephemerally float in the space, like a ghostly hologram. Also, counter-intuitive distortions, inversions and reversions of the images, fused in continuous visual *feedbacks* that tend to infinity, pose questions upon our understanding of the internal space of the sphere.

This cognitive friction points out to what cannot be directly observed; it points out to the shadow, the interval between dimensions – an interval that is not only void; an interval that is structural and, therefore, that organizes and supports the different dimensions[2].

I/VOID/O, therefore, approaches the observation process itself as its constitutive material. In this installation the object under observation is *observation* itself – which explains the use of several cameras and different image processing techniques that enable the appreciation of several forms of observation. The observation process incorporates *feedback* as a self-destructive process – the interface builds itself only when it destroys itself. 'Looking' is forged in order to enable observation. However, when one observes, he/she sees him/herself, and therefore stops observing, and so on and so forth. *I/VOID/O*, therefore, is about the impossibility of observation without interference. And more: it is about observation as creation and death, cyclically.

3. Images and immersive levels in *I/VOID/O*

The immersive levels in *I/VOID/O* are sequential levels that are reached and surpassed during the interactive process. Each level presents a different way of observing the interior of the sphere. The succession of levels corresponds to an increase in 'observation ability' in the manipulation of the interface.

3.1. Level I

In the first level, the images seen by the interactor are disconnected from his/her movements. The rupture of temporal linearity in the images results from the programming done in *Jitter*, which uses a video *buffer* that is updated every three seconds (in order not to overload the use of RAM). These three seconds are read randomly, i.e. the bits are not read linearly. The visual result is the temporal fragmentation of the image, which does not present the continuity that can be found in the images of movement that we see in our daily lives.

The rupture between the images and the interactor's reaction can lead him/her to seek some kind of coherence (or a more evident reactivity) by producing stronger and sudden movements[3]. When this happens, the level of entropy in the system increases and the interactor ends up trapped in this level (Level I). Entropy is understood here as

[2] The ideas of 'super-observer' (mentioned earlier), 'cognitive friction' and 'the void between dimensions' are poetic adaptations of concepts about observation and interactivity articulated by Roessler 1998.

[3] This kind of behaviour from the interactor had been observed in previous versions of the I/VOID/O, such as the one exhibited in 2005 at the Festival Internacional de Linguagem Eletrônica – FILE 2005.

‘disorganised energy’. In this case, the disorganisation is the result of unbalance in the interactor’s movements, which is calculated as follows: the amount of movement to the left (measured by the difference in the amount of pixels that change between two video frames, from right to left) is subtracted from the amount of movement to the right. When this difference reaches a pre-determined threshold within a certain time span (three minutes), the interactor restarts at the same level (in case he/she is in Level I) or he/she returns to the previous level.

In order to advance to another level, the interactor has to produce more controlled movements in an attempt to explore the details of the images. As a consequence, his/her movements do not disturb the system in excess and he/she advances to the next level. The deceleration of the movements can happen when the interactor starts to search for details in the image, or when he/she tries to understand the internal events inside the sphere. At the exhibition there was also an assistant who would inform the interactor about the possibility to decelerate his/her movements and the resulting reaction of the system.

It is important to mention that computer vision algorithms were used ('cv.jit' library for Max/MSP/Jitter[4]), which track the direction of movements in the images. It was necessary to add some other logical and arithmetical operations in order to quantize the variation of movement within a certain time span.

3.2. Level II

While in Level II, the interactor can notice a greater degree of coherence between his/her movements and the images that are projected on the screen. In Level I, the direction of the movements practiced by the interactor does not present any relationship with the images, due to the fragmentation of the images mentioned earlier. In Level II, on the other hand, the direction of the movements is recognised by the interactor in the images that he/she sees because the camera moves inside the sphere according to the movements that he/she makes and the images present what is captured by the camera.

However, the degree of coherence is not at its full: the concave mirror of the internal surface of the sphere generates visual paradoxes that present themselves as challenges for the understanding of the space that is being explored.

3.3. Levels III and IV

In Level III, the interactor is able to observe the interior of the sphere more accurately: images do not come from the camera located at the tip of the stick anymore. They come from another camera positioned on the internal surface of the sphere that provides a static point of observation pointing towards the centre of the sphere. It would be logical to infer that the image captured in such a way should show the stick and the camera (placed on its tip) moving inside the sphere. This, however, is not what happens. What one sees is a floating cube on the tip of the stick. On this cube, the interactor can even see him/herself, since his/her image is projected onto the surfaces of the cube (this

[4] For information on Max/MSP/Jitter, see http://www.cycling74.com. For information on cv.jit, see http://www.iamas.ac.jp/~jovan02/cv/

image of the interactor is captured by the camera placed on one of the walls in the space of the installation). This cube presents a phantom aspect: although it seems real, one cannot see its reflection on the internal surface of the sphere. This is due to the fact that the cube is not *really* there: it is rendered and synchronised to the stick, which gives the impression that the cube is attached to the stick. The image of the cube characterises the level of immersion related to the 'Cartesian illusion' that produces a certain degree of coherence in the internal space of the sphere. The idea of ‘Cartesian illusion’ is understood here as a construct that creates and organises a homogeneous and coherent space, which is not able to deal with congruent and parallel dimensions as understood by the topology and space of phases.

Besides the image of the cube, the interactor can occasionally visualise another perspective of the ‘Cartesian illusion’ depicted in Level IV (a transitory level): a 3D image (rendered in the form of lines) reveals distortions on the spherical space caused by the movements of the camera – the 3D grid is rendered in real-time and its vertices are continuously repositioned in relation to the intensity and the direction of the movements of the camera.

3.4. Levels V and VI

Level V is a transitory one. In this level, the interactor observes the image of the cube (rendered and synchronised to the movement of the camera) being continuously enlarged until it takes up the whole area of the projected image. This enlargement happens within six seconds and at the end of this interval, the projection is interrupted – which instantaneously activates the projection of Level VI on the right screen. Wearing red/blue glasses, the interactor can see the internal images of the sphere projected on the right screen with stereoscopic view (which provides a sense of depth to the images). When this level is surpassed, there is a return to Level I again, with images in black and white projected on the left screen.

4. The sounds in *I/VOID/O*: synthesis and sound processing in real-time

On the sonic domain, the interactive processes implemented in *I/VOID/O* are based on synthesis and sound processing techniques carried out in real-time using data from the analysis of the images and parameters that are changed randomly. Techniques of granular synthesis (see [Truax 1988], [Lippe 1994], [Keller and Rolfe 1998] and [Keller and Truax 1998]) and additive synthesis (see Dodge and Jerse 1997) are implemented in Max/MSP in a patch especially designed for the purpose of the installation.

The processes are carried out by several Max/MSP modules (*subpatches*) embedded in the main *patch* (see Figure 2). Apart from the subpatches that receive data related to the analysis of the images via OSC protocol, the main patch presents a subpatch that controls the amplitude of the sounds and their distribution over the four loudspeakers (“p volume_control”) and also another subpatch in which the synthesis and the sound processing modules can be found (“p sound_source”). Inside “p sound_source”, the modules are grouped in three different subpatches – one that generates the soundscapes for Levels I and II; another for Levels III, IV and V; and a third one that generates the soundscape for Level VI.

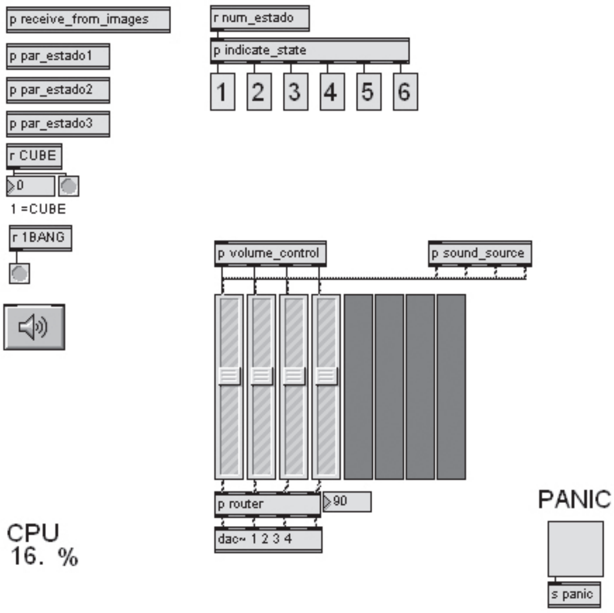

Figure 2: Audio – main *patch*

Whenever the interactor advances levels in the installation, the computer that carries out the sonic processes receives this information via OSC protocol. Changes in the visual domain are, therefore, synchronized to changes in the sonic domain.

4.1. The sounds of Levels I and II

Levels I and II are based on sounds generated by granular synthesis. Pre-selected audio files are 'sliced' in segments of very short durations (grains) and juxtaposed within time spans of variable lengths. The spectrum and the texture of the new sounds vary according to the parameters (and the variation of these parameters through time) used for the granulation process.

The most important parameters are grain size and grain rate. In the implementation of Max/MSP patch, grain size is the result of an initial grain size value added to a grain size random variation. Grain rate is also implemented as the result of an arithmetic operation involving two values defined separately – a value related to the time span between the onset of successive grains added to the result of a random variation.

In this subpatch two granulators are used in parallel, each of which generates up to 20 streams of grains from three different sound files selected beforehand. The choice of the sound files and the option for using three of them was determined empirically after trying out different possibilities and deciding for the alternative that seemed to offer the most interesting sonic results (according to the opinion of the authors). Figure 3 displays an image of one of such granulators[5].

[5] These granulators were especially designed for the purpose of the *I/VOID/O* installation using features of granulation patches previously designed in Max/MSP by Erik Oña and Peter Batchelor who gave

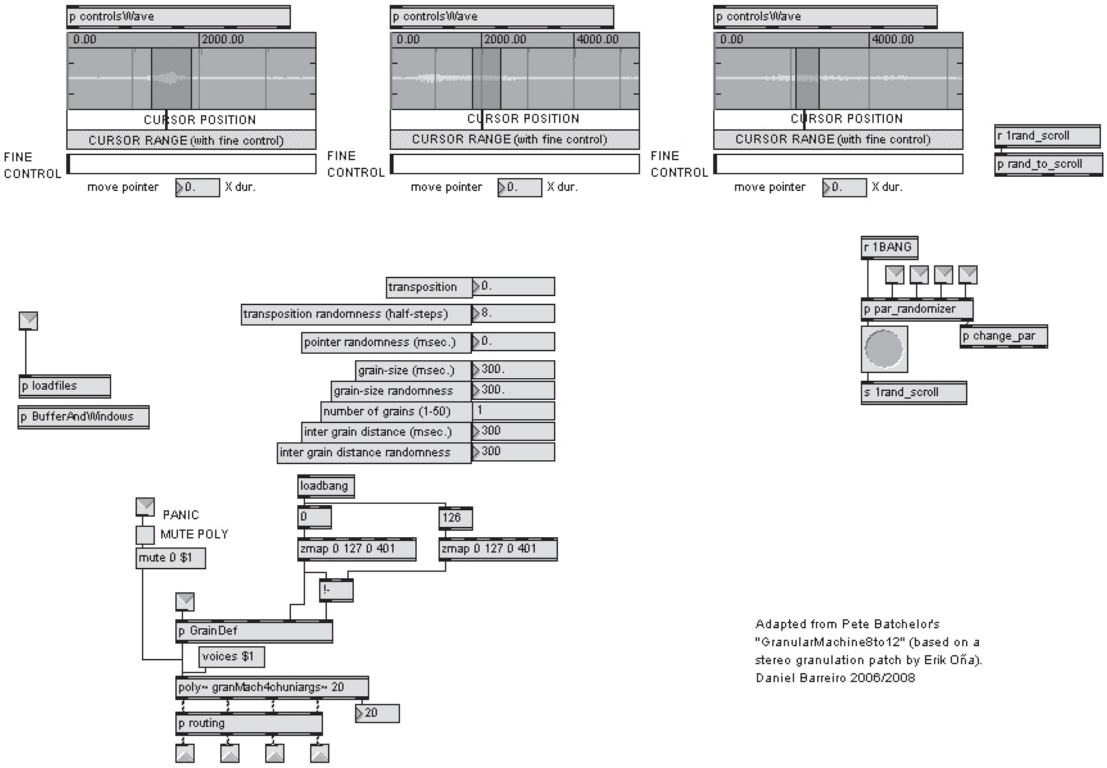

Figure 3: Granulator sub-patch

The granulation parameters change whenever there is some kind of movement – traced by the analysis of the images. The selection of the sound file to be granulated in each one of the granulators, on the other hand, is made randomly. Changes in the granulation parameters and the distribution of the sounds in the four loudspeakers are also defined by random processes within certain ranges determined beforehand. These changes occur independently in each one of the granulators. Although the sound file selection process, the changes in the granulation parameters and the distribution of the sounds in the space are determined by random operations, the resulting sonic stream is also dependant on the movements in the image. Therefore, there are some links between the interactor's actions and the sonic behaviour of the system. The random processes, however, prevent these links from being too strictly, which avoids the so called 'mickey mouse effect'.

The option for generating sounds by means of granular synthesis was mainly motivated by the morphology encountered in the sounds generated using such a technique. In Level I, the grain size and grain rate values are determined within certain ranges that prevent the system from outputting continuous sounds as the result of the granulation process. The sounds tend to present a granular character and just a few seconds of duration. Since the images in this level derive from a 3-second video *buffer* that is read randomly – in a process that shows similarities with sound granulation – both images and sounds present a non-continuous character.

one of the authors of this paper their patches and personal permission to use them whenever appropriate.

In Level II, the sounds generated by granular synthesis (using the process described above) pass first through a 512-band EQ implemented with Fast Fourier Transform (see Settel and Lippe 1994, 1995, 1998, 1999), and then by a reverb.

As a consequence, the passage from Level I to Level II is marked by changes in the behaviour of both the images and the sounds, as they become more continuous and closely related to the movements of the stick in comparison to the previous level. Regarding the sounds, although they are still generated by granulation, the continuous character derives from greater grain size values and smaller grain rate values. Also, the reverb applied to the sounds imposes a more continuous and resonant character to them. The sounds also present a different spectrum in comparison to the previous level, as they are changed by the EQ.

In this level, the internal space of the sphere can be more thoroughly explored – not only visually (regardless the strange forms resulted from the reflection on the curved internal surface) but also sonically by the reverberations (although synthetically produced) that happen in Level II. Also, the sounds that are generated resemble those of a metallic object – which potentially produces a connection between the sounds and the visual aspect of the sphere.

Figure 4 shows the configuration of the EQ at a certain moment in Level II (the horizontal axis represents frequency and the vertical axis represents amplitude for each of 512 bands). It can be noticed from Figure 4 that some frequency bands are completely attenuated, whereas others are reinforced in different degrees. The amplitude of each frequency band is defined randomly and set to a new value after time spans greater than 1000 miliseconds. The actual amplitude applied to each frequency band, however, does not change abruptly from one setting to the next, as the values are slowly interpolated, which is carried out by the vectral~ object in the pfft~ subpatch. Therefore, although the resonant frequencies and their amplitudes are determined by a random process, the actual changes are smooth and, therefore, a resonant sonic structure can still be obtained.

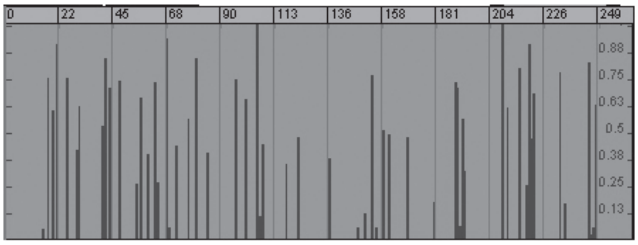

Figure 4: EQ frequency bands (FFT)

4.2. The sounds of Levels III, IV and V

In Levels III, IV and V, sounds are generated by additive synthesis (superposition of sine waves), using eight synthesis modules that superimpose six sine waves each (see in Figure 5 an image of one of these modules).

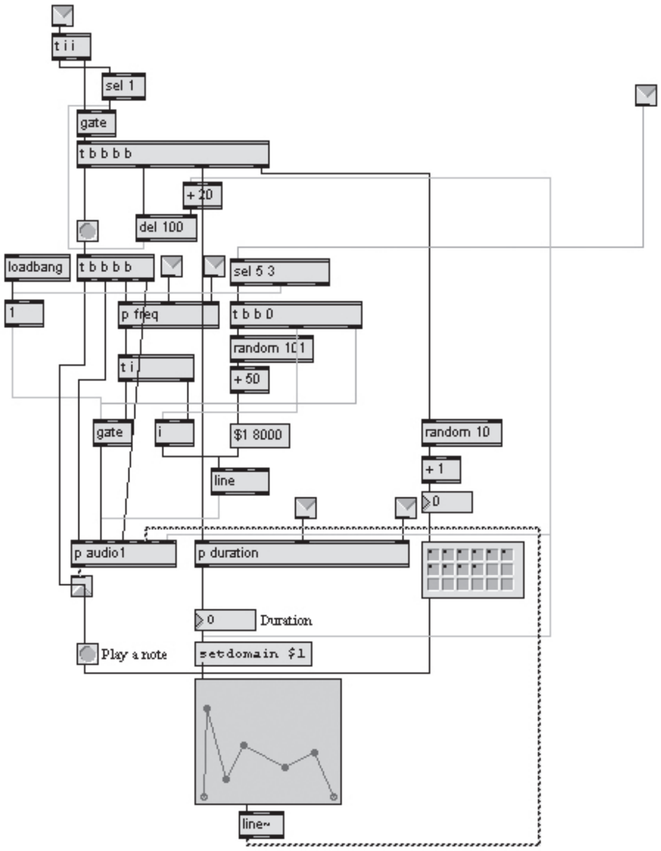

Figure 5: Additive synthesis module

The relationship between the frequencies of the six sine waves in each composed sound and their dynamic envelope are determined by the random selection of eight different *presets*.

The option for a synthesis technique other than the one used in previous levels has a connection with the change in the visual domain that happens in Level III. In this level, images are captured from a static camera pointing towards the centre of the sphere (and not the camera from the tip of the stick anymore). The form of the cube as a Cartesian object is associated with sounds obtained by means of additive synthesis.

Whenever there is a transition to Level IV (which can be visualised as lines rendered in 3D), the sounds generated by additive synthesis are subjected to slight variations using amplitude modulation.

When the interactor advances to Level V (a transitory stage in which the cube expands until it takes up the whole screen) the synthesised sounds perform a glissando towards the low register, which links the sounds of this level with the low registered sound from Level VI.

4.3. The sounds of Level VI

Level VI is based on a single long sound in the low register produced by granular synthesis. The process that was used is similar to the one described for Levels I and II, with the difference that the parameters are configured in such a way as to produce a continuous sound without a strongly noticeable granular character. This long sound rotates in the room, moving faster each time it turns around the space. After 2 minutes, both sound and image return to their configuration in Level I.

5. Considerations on the sonic results obtained

The aim of this implementation was to generate sounds in real-time that could work in a coherent and well-integrated way with the images.

Both the sounds and the images generated at a certain level present global features that happen in all occurrences of that specific level. As a consequence, the whole cycle of six levels of immersion maintains some kind of consistency in several of its occurrences.

It was possible to verify that the random processes used for varying the synthesis and processing parameters did not compromise this global sonic coherence, since they operated only in the definition of the micro-elements (details) of the sonic structures (see [Keller 2000] and [Keller and Capasso 2006]).

The synchronicity between sound and image in the passage from one level to the next and the use of analogies between the behaviour of the images and the morphology of the sounds contributed to provide coherent relationships between both, helping to create the immersive environment of the installation.

6. Final Considerations

This paper presented the main characteristics of the interactive installation *I/VOID/O* and the way synthesis and sound processing techniques were implemented in real-time in order to work with the images and the general aesthetic motivation of the installation.

Granular and additive synthesis techniques were used. The parameters of synthesis were altered using data from the analysis of the images, messages indicating the beginning of each level of immersion and also random processes.

For our future interactive projects, it would be interesting to explore the artistic potential of other computational processes and models that we have been studying, such as swarm intelligence, and the use of other kinds of interactive interfaces, such as sensors and the *wiimote*.

7. References

Dodge, C. and Jerse, T. A. Computer Music: Synthesis, Composition, and Performance. Wadsworth Publishing Company, 1997.

Keller, D. (2000). “Compositional processes from an ecological perspective”. Leonardo Music Journal, 10, pp.55-60.

Keller, D. and Capasso, A. (2006). New concepts and techniques in eco-composition. Organised Sound 11(1), pp.55-62.

Keller, D. and Truax, B. (1998). “Ecologically-based granular synthesis”. Proceedings of the International Computer Music Conference. Ann Arbor, MI: ICMA. http://www-ccrma.stanford.edu/~dkeller/pdf/KellerTruax98.pdf

Keller, D. and Rolfe, C. (1998). “The corner effect”. Proceedings of the XII Colloquium on Musical Informatics. Gorizia: AIMI. http://www-ccrma.stanford.edu/~dkeller/pdf/KellerRolfe98.pdf

Lippe, C. (1994). “Real-Time Granular Sampling Using the IRCAM Signal Processing Workstation”. Contemporary Music Review 10(2). United Kingdom: Harwood Academic Publishers, pp.149-156.

Roessler, O. E. Endophisics: The World as an Interface. World Scientific Publishing Company, 1998.

Settel, Z. and Lippe, C. (1994). “Real-Time Timbral Transformation: FFT-based Resynthesis”. Contemporary Music Review, Vol.10(2). United Kingdom: Harwood Academic Publishers, pp.171-180.

Settel, Z. and Lippe, C. (1995). “Real-time Musical Applications using Frequency-domain Signal Processing”. In: IEEE ASSP Workshop Proceedings. New York: Mohonk. http://ieeexplore.ieee.org/iel2/3495/10326/00482997.pdf

Settel, Z. and Lippe, C. (1998). “Real-time Frequency Domain Signal Processing on the Desktop”. In: Proceedings of the ICMC1998. San Francisco: ICMA, pp.142-149.

Settel, Z. and Lippe, C. (1999). “Audio-rate control of FFT-based processing using few parameters”. In: Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99), Trondheim.

Truax, B. (1988) “Real-Time Granular Synthesis with a Digital Signal Processor”. In: Computer Music Journal, Vol. 12(2), pp. 14-26.

Wiener, N. Cybernetics, or Control and Communication in the animal and the Machine. Cambridge: MIT Press, 1962.

Wright, M., Freed, A. and Momeni, A. (2003). “OpenSound Control: State of the Art 2003”. In: Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03). Montreal, Canada, pp.153-159.

Zielinski, Siegfried. Interfacing Realities. Rotterdam: Uitgeverij De Baile and Idea Books, 1997.