3D Audio Effects for Multi-Channel Reproduction

Antonio Escamilla P.

MASTER THESIS UPF / 2010

Master in Sound and Music Computing

Master thesis supervisor:
Pau Arumí, Toni Mateos
Department of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona

UNIVERSITAT POMPEU FABRA

Abstract

Immersive audio is today's trend in the Media Industry, devoted to offer a new perceptual experience for the users and a lot more possibilities for sound designers. Going beyond the actual audio exhibition systems commonly used at home, in movie theaters and even in sound installations, the 3D audio technologies are designed to take full advantage of the use of space, but being independent of any particular speaker configuration or layout, to create a natural listening experience in which the user perceive sound coming from locations anywhere around him. This project is oriented to explore and experiment with the technology platform developed in the Audio Group of Barcelona Media Innovation Center to create audio effects that will serve as tools for music creation and performance from a 3D perspective. Four different case studies are presented as methodologies and research source; using signal processing techniques, efficient real-time audio-oriented processing software and spatialization algorithms to render 3D audio. A general framework is presented as strategy to develop audio effects that are designed to be reproduced in three-dimensional audio exhibition systems.

Acknowledgements

I would like to thank Pau Arumí and Toni Mateos for the opportunity they provided me to be part of the Audio Group of Barcelona Media and for their interest and support during the realization of this project.

I am also grateful, to all members of the Audio Group since they all have helped to develop the 3D audio tools that made this project possible, but especially to Natanael Olaiz for his continuous help and guidance, and to Angelo Scorza and David Garcia for their contributions.

Contents

List of Figures

List of Tables

1. Introduction

A 3D audio system is a system capable of rendering a natural listening situation based on sound sources located anywhere around a listener. To design such a system there are two basic approaches. The first is to completely surround the listener with a large set of speakers, which allow the exact (or approximated) sound field reconstruction of an auditory event. The second is to reproduce only at the ears of the listener the acoustic signals that would occur in the hearing experience to be simulated.

In a multichannel setup, the sounds are actually created by the loudspeakers but the listener's perception is that the sounds come from arbitrary points in space. Thanks to this phenomenon the system enhance the sense of immersion for a group of listeners by reproducing the sounds that would originate from several directions around the listeners, thus simulating the way we perceive sound.

Spatial auditory systems are nowadays the next step, in the media industry, to grab the attention of a public by means of a greater involvement and emotional impact. To achieve these sensations, special effects and unexpected features for a creative-virtual sound field design should then be developed as well as tools to: recreate a real-life situation and offer better realism in the reproduction.

1.1. Scope and Orientation

The main goal of the present work is to explore and exploit the capabilities of a 3D audio reproduction system, related mainly with the use of space, by extending some properties of already existing stereo audio effects to design 3D audio effects that go beyond the limits imposed by the actual stereo or 5.1 exhibition systems. The research work takes part inside the "Barcelona Media Innovation Center" 3D Audio Group, which among other fields, studies the

use of multiple speaker setups in music production, reproduction and performance and also the technology that supports it.

The approach consist in using the technology platform that the 3D Audio Group has, namely, the software development tools for source spatialization and audio processing, and the 22.2 speaker exhibition system, to develop a set of audio effects and processing block-sets that may offer a new perceptual experience to the users, experimenting with possible meaningful applications of 3D audio as a new upcoming technology. A challenging issue that one encounters when dealing with many different 3D exhibition systems, is that each system may have a different setup, i.e. (number of speakers and its locations), causing that a particular audio "processing" designed for a particular exhibition system not to be useful or usable in some others, therefore, the philosophy within this research of rendering 3D sound independently of the exhibition system.

2. State of the Art

This chapter is structured as follows. The first section of this review, present a summary of existing and well known audio effects implementations widely used in music production and performance. In particular time based modulations, delays and reverberations algorithms are described since these types of effects are more suitable for 3D audio exhibitions systems. The following section is devoted to the spatialization algorithms of sound sources and technologies that can be considered as independent of the final speaker setup. The final section gathers information about relevant research works in the area of 3D music performance, immersive spatial audio installations and sound spatialization by means of various loudspeakers.

2.1. Time Domain Audio Effects

2.1.1. Delay Lines

Since the very beginning of the digital signal processing and computer music era, the delay line has been extensively used as a basic block in software synthesis languages due to its easy implementation of a constant delay in comparison to the analog methods. Delays lines are used to implement audio effects, such as pitch shifting, reverberations or to model wave propagation in musical instruments, where not only constant delays are needed but also dynamic variations of the delay length. A circular buffer, which is accessed by a writing pointer and a different reading pointer, allows to have a variable-length delay line by changing the relative distance between these two pointers sample by sample. Like any linear time invariant operation, a delay can be considered in a transform domain. The transfer function of this system in the z-domain can be devised in the Figure 1 and be written as:

$$Hd(z) = \frac{Y(z)}{X(z)} = \frac{z^{-D}X(z)}{X(z)} = z^{-D} \tag{1}$$

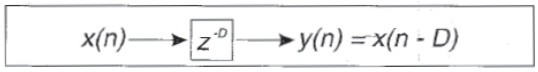

Figure 1. An ideal discrete- time delay system.

Depending if the delay network is based on a feedforward approach or in a feedback loop, a FIR or IIR comb filter realization will be respectively obtained; the main differences between both implementations are that the frequency peaks get narrower as the magnitude of the loop gain come closer to 1, and that the gain grows very fast for the infinite response implementation. A general form network using both topologies is called the Universal Comb Filter and it's simply the use of only one delay line unit connected through feedforward and feedback loops each one with its proper gain. The network, as shown in Figure 2, is simply an allpass filter whit the M sample delay operator Z^{-M} and an additional multiplier FF. The special cases for differences in feedback parameter FB, feedforward parameter FF and a blend parameter BL vary the functionality of the system [28]. An extension of the above universal comb filter to a parallel connection of N comb filters and how to use it for echo, slapback and reverb effects are deeply explained in [28][24].

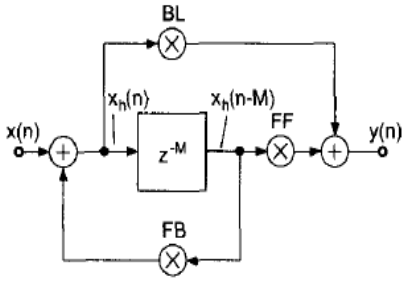

Figure 2. Universal Comb filter.

2.1.2. Fractional Delay Lines

In a digital world the sample rate must satisfy the Nyquist criterion and as such it impose the minimum delay that can be achieve by a delay line, since it operates by delaying the signal an integer number of samples. The fractional delay unit [27][32] is then the digital domain approximation of a continuous time analog delay line; it means, assuming uniform sampling, a delay that is a non-integer multiple of the sampling interval. In order to allow for fractional lengths, some form of interpolation has to be applied at the reading point inside the delay line buffer and for this; at least three properties should be ensured: flat magnitude frequency response, linear phase response and transient-free response to variations of the delay [7]. A block diagram for a delay line interpolation can be devised in Figure 3, which has as starting point a delay line of length M samples. As in any delay line, the new input sample is accepted in the first position of the delay line (sample 0), making it necessary to discard the oldest value. The output of the delay is not the last sample but rather an arbitrary tap in between, and because its smoothing movement around a tap center, an interpolation method between samples is necessary; sees Figure 4. For audio applications in particular, proposed algorithms for the samples interpolation can be found in [16] and [7]. All of these methods perform interpolation of a fractional delayed output signal with different complexity and performance, which are summarized in the work by [7].

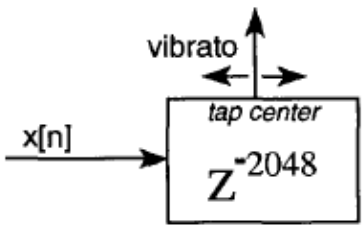

Figure 3. Basic block of a variable length delay, as used for vibrato.

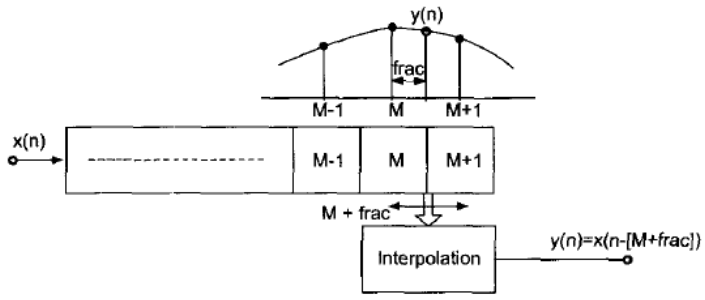

Figure 4. Fractional delay line with interpolation.

2.1.3. Audio effects based on Delay Lines Modulation

What became a standard topology used in the industry was proposed by [16] for the implementation of a few of the most common audio effects based on fractional delay units [26]. Namely the Chorus and the Flanger effects can be obtained with the circuit shown in Figure 5. It is based on an all-pass interpolation towards a general all-pass comb filter. The main modification is the introduction of a negative feedback path into the delay line, whose tap point is separate from that of the feed-forward path but fixed at the center of the modulating delay in the feed-forward path.

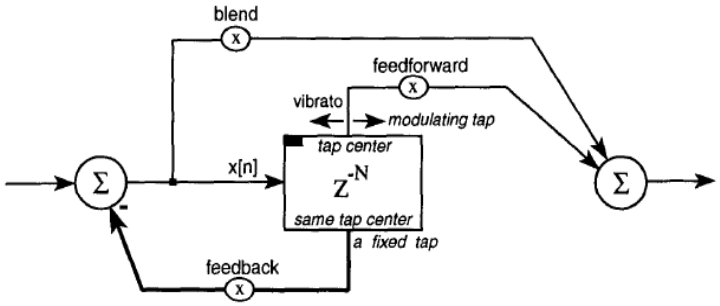

Figure 5. Standard effects structure based on an all-pass interpolation.

For the Flanger effect [18], the consequence of adding two signals at the input is to introduce a comb of moving troughs into the signal spectrum, and the deeper and selective these troughs are, the better. The opposite happens with the Chorus, since these troughs are not wanted, situation that force the inclusion of unequal amounts of original signal and delayed replica to limit their depth. As is explained in [16] the choice of not feedback the modulation

is because it induce pitch changes, situation that produces each time more pitch change and too fast, even for the chorus or the flanger.

The Table 1 indicates the settings of the parameters for obtaining different effects. All listed effects can be implemented using the same network by using the gains parameters and also the appropriate tap center for the modulation. See Table 2.

Table 1. Parameters settings. [16]

Effect	Blend	Feedforward	Feedback
Vibrato	0.0	1.0	0.0
Flanger	0.7071	0.7071	−0 7071
Industry standard chorus	1.0	0.7071	0.0
White chorus	0.7071	1.0	0.7071
Doubling	0.7071	0.7071	0.0
Echo[82]	1.0	≤1.0	<1.0

Table 2. Approximate effect delay range (ms). [16]

Effect	Onset	Nominal	Range End
Vibrato[83]	0	Minimal	5
Flange	0	1	10
Chorus	1	5	30
Doubling	10	20	100
Echo	50	80	∞

2.1.4. Granulation

Sound granulation is based on building complex sounds from small units called grains and philosophically inspired on the theory developed by Dennis Gabor in 1947, who created the idea of the quantum of sound: In which a granular representation could describe any sound. The pioneers in methods for synthesizing complex sounds using this small audio fragments in a computer music framework were Iannis Xenakis (1971) and Curtis Roads (1978), both of them relying on the production of a high density of small acoustic events

called 'grains'. A sound grain lasts a short time, which get close to the minimum perceivable event for duration, frequency and amplitude discrimination, typically between 10 and 30 ms [2]. Key to all granular techniques is the grain envelope. For sampled sound, a short linear attack and decay prevent clicks being added to the sound. Changing the slope of the grain envelope, in classic micro-sound practice, changes the resulting spectrum, sharper attacks producing broader bandwidths, just as with very short grain durations [6].

Despite the mathematical complexity of some of the granular models used today, granular synthesis is somehow attractive to composers because of its conceptual simplicity: small fragments of sounds are superimposed to construct more complex sound material. Since 1987, Barry Truax [3] have used this technique extensively to process sampled sound as compositional material, at the beginning being limited only to short "phonemic" fragments, but later longer sequences of environmental sound were used by him in different many pieces. In each of his works, the granulated material is time-stretched by various amounts, producing a number of perceptual changes that seem to originate from within the sound [2].

If we consider two signals: x(n) and y(n) as the input and output respectively, then a grain $g_k(i)$ can be extracted from x(n) by using a windows function $w_k(i)$ of length L_k as it follows:

$$g_k(i) = x(i + i_k)w_k(i)\, i = 0,\ldots\ldots,L_{k-1} \tag{2}$$

where, i_k indicates the time instant where the segment is extracted and L_k determines the amount of signal to be extracted. The type of window used will affect the frequency content and will determine the grain limits by proper fade-in and fade-out computation.

For synthesis the formula is given by

$$y(n) = \sum_k a_k g_k(n - n_k) \tag{3}$$

where a_k is an amplitude coefficient and n_k is the time instant where the grain is located at the output .

With the previous definition as starting point, many different implementations with different sound characters can be found in the market, for example, depending on the strategy to pick the synthesis instants; where, deterministic functions are used for synchronous methods, meanwhile asynchronous methods are based on stochastic functions. It is also possible, by using a time transformation such as modulation or time stretching, to modify the grain waveform and thus the sound textures. A complete set of parameters, strategies and perceptual results for different algorithms can be found in [28].

2.1.5. Doppler Effect

The Doppler effect is perceived as a pitch change in the sound produced by a source due to the motion of the source and/or listener relative to each other. We usually appreciate the pitch shift due to the Doppler effect in non-musical situations, like a train or ambulance passing by, and it can be so relevant that it can impose a perception of motion between the source and listener even when other cues may indicate a static relation between emitter and receiver. The physical concept can be explained easily with the following formula, normally found in basic physics books,

$$\omega_l = \omega_s \frac{1 + v_{ls}/c}{1 - v_{sl}/c} \tag{4}$$

where ω_s is the radian frequency emitted by the source at rest, ω_l is the frequency received by the listener, v_{ls} denotes the speed of the listener relative to the propagation medium in the direction of the source, v_{sl} denotes the speed of the source relative to the propagation medium in the direction of the listener, and c denotes sound speed. A very clear analogy is described in [19] to understand the physical concept: “The air can be considered as analogous to a magnetic tape which moves from source to listener at speed c. The source is analogous to the write-head of a tape recorder, and the listener corresponds to the read-head. When the source and listener are fixed, the listener receives

what the source records. When either moves, the listener observes a Doppler shift, according to the previous equation".

In previous sections the time-varying delay line was presented along with its consequence: a pitch shift; that is why most of the implementations use this basic unit to simulate a Doppler effect. The effect can be achieved by calculating a modulation of the phase of a previous recorded audio according to the delay definition:

$$\begin{aligned} y(n) &= x(n - D(n)) \\ D(n) &= M + d * X_{mod}(n) \end{aligned} \tag{5}$$

where D(n) is the modulating factor, that is now dependant on a modulating signal $x_{mod}(n)$. Here the pitch of the input signal x(n) is following the envelope of the modulating signal, as is depicted in Figure 6[28]. The performance of the Doppler shift will depend on the interpolation method used in the fractional length delay, as some artifacts may appear if the modulation products affect the transposed signal.

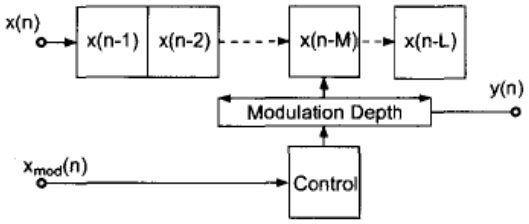

Figure 6. Pitch controlled by envelope of signal $x_{mod}(n)$.

2.2. Spatialization Techniques

2.2.1. Ambisonics

Ambisonics is a 3D audio multichannel technique based on the description of the acoustic field on a point in space through the decomposition of the pressure in a set of functions called spheric harmonics [21]. At the beginning, it was limited to a first order expansion implying the encoding of the spatial

properties of the sound field using the pressure and three components of the pressure gradient. It can be demonstrated that all the information of an acoustic field in a given point can be totally codified with order 0 and 1 microphones capturing in a coincident configuration. These four signals are called B-format, and are refereed as {W, X, Y, Z}. As shown in Figure 7, the W signal corresponds to an omni-directional pressure signal, while the X, Y and Z signals, capture information on a bidirectional form, each one of them oriented on one of the three directions of a three dimensional Cartesian space. Using a spherical coordinate system:

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} \\ \theta &= \arctan\left(\frac{y}{x}\right) \\ \varphi &= \arctan\left(\frac{z}{\sqrt{x^2 + y^2}}\right) \end{aligned} \tag{6}$$

where the angular dependency of the sensitivity of each transducer can be expressed as follows:

$$\begin{aligned} X &= \cos(\theta)\cos(\varphi) \\ Y &= \sin(\theta)\cos(\varphi) \\ Z &= \sin(\varphi) \end{aligned} \tag{7}$$

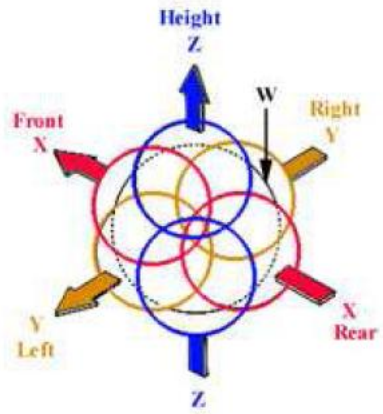

Figure 7. Polar patterns of the B-format encoding.

For obtaining signals in B-format, the SoundField microphone [33] is being used widely in the media industry; providing analog and digital versions, combined with unit controls for decoding and manipulation of signals on a post-production level. The microphone in constructed with four sub-cardioid capsules, arranged on the vertices of a tetrahedron and whose sensitivity is described by the following function:

$$P_{output} = A\left[\frac{3}{4}p + \frac{1}{4}\hat{n}\int dt\vec{\nabla}p\right] \tag{8}$$

where n is an unitary vector in the direction of maximum sensitivity, p is the sound pressure and p_{output} is the output signal of the transducer.

The direct form of the captured signal, are called the A-format, from which the B-format is obtained using a linear transformation. In the case of the SoundField device, the transformation is computed inside the control unit, providing the B-format at its output.

$$\begin{pmatrix} W \\ X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} L_F \\ R_B \\ L_B \\ R_F \end{pmatrix} \tag{9}$$

In the reproduction stage, the B-format has to be decoded according to the specifications of the terminal, (number of speaker and its positions) and be configured in such a way, that the sound reproduction in the given system can produce signals in the "sweet spot", as similar as possible to the {W, X, Y, Z} recorded signals. By doing this, a maximum fidelity reproduction is assured for the encoded acoustical scene. The Ambisonics approach has the advantage that the B-format content remains independent of a particular reproduction layout at the exhibition stage; resulting in a flexible and versatile approach compared over channel-per-speaker techniques.

The main problems with Ambisonics are: that the encoded audio field is correctly heard on the condition that the listener is located on the center of

the speaker array ("sweet spot"), being it too small for first order Ambisonics[12], and that the angular resolution is not very high, due to the necessity of reproducing a sound object in all the speakers to locate the source in a particular point in space. Even so, first order Ambisonics is ideal for coding reverberant and diffuse fields, as well as ambient sounds.

2.2.2. Vector Based Amplitude Panning

The technique is an extension of the panning method used in stereo reproduction, where the source is located between two speakers by adjusting the amount of signal that its sent to each loudspeaker (usually with a pan pot, in a mixing desk); and that relies on the perception of a phantom image between the speakers that the brain creates.

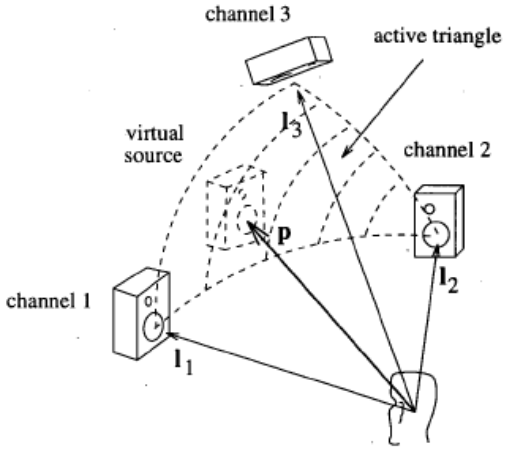

Figure 8. Triangle scheme for the three-dimensional source location.

The 3D generalization was developed by [29], considering three loudspeakers forming a surface of a three-dimensional unit sphere equidistant to the listener, to locate a source in space. Each loudspeaker position is defined by a unit vector with origin in the center of the sphere l_i, i=1,2,3, defining the directions of the loudspeakers. The direction of the virtual sound source is defined by the unit vector $p=[p_1,\ p_2,\ p_3]^T$, as can be devised on Figure 8. Analogically to the stereo panning, the virtual source vector p can be expressed as a linear combination of the vectors l_1, l_2, l_3 using g_1, g_2 and g_3 as gain factors.

$$p = g_1 l_1 + g_2 l_2 + g_3 l_3$$
$$p^T = g L_{123} \tag{10}$$

and solving for g,

$$g = p^T L_{123}^{-1} = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix} \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix}^{-1} \tag{11}$$

The previous eq. makes a projection of vector p to a vector based defined by L_{123} in a similar way as in the two-dimensional case [30]. For the proper usage of the computed gains, a scaling factor must be computed

$$g_{scaled} = \frac{\sqrt{c}\, g}{\sqrt{g_1^2 + g_2^2 + g_3^2}} \tag{12}$$

One important generalization of the system is explained by[29], arguing and proving, that when the three loudspeakers are placed in an orthogonal grid, the gain factors calculated with the three dimensional VBAP are equivalent to the absolute values of gain factors calculated in the three dimensional Ambisonics system.

The extension of the technique for more than three loudspeakers should be done triplet-wise and for each of those triplets the following conditions should be met: First, the loudspeakers of a triplet should not be all in the same plane with the listener and, second, the triangles should not overlap and should have as short sides as possible. See Figure 9.

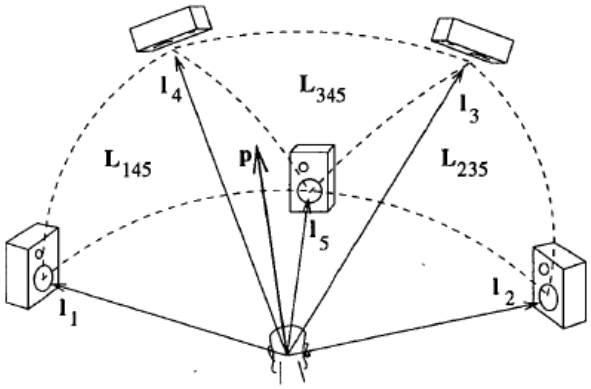

Figure 9. Three-dimensional VBAP system with five loudspeakers.

The advantage of this method is that it provides a bigger sweet spot, since all the loudspeakers that are not close enough to the source don't produce any signal, which is ideal for locating point-like sources. But in the other side, The VBAP technique finds it difficult to reproduce reverberations, diffuse fields or ambient sounds. A complete full discussion of the capabilities of VBAP and a methodology of spatial sound evaluation for VBAP can be found in [30].

2.2.3. Wave-field Synthesis

The Wave-field Synthesis technique is based on the Huygens principle, according to which, the acoustic field generated by a source can be reconstructed inside a volume free of sources if the pressure and gradient pressure can be known in the boundary of the two volumes. Considering a space Ω1, in which all sources are inside, and the complementary space Ω2, the WFS theory allows expressing the pressure field in all points inside Ω2 as if it were generated by sources located in the boundary between both spaces ∂s. The general formula is:

$$p(\vec{r},t) = \int\int_{\partial\Omega_0} \partial S_0 \hat{n} \left\{ \int_{t1}^{t2} dt_0 \left[g(\vec{r}-\vec{r}_0,t-t_0)\vec{\nabla}p_0(\vec{r}_0,t_0) - p_0(\vec{r}_0,t_0)\vec{\nabla}g(\vec{r}-\vec{r}_0,t-t_0) \right] \right\} \quad (13)$$

where $\hat{n}$ is the unit vector orthogonal to the surface ∂S of the frontier, $\overrightarrow{r_o}$ is the variable of integration and t_1 and t_2 are the extremes of the time interval of the duration of sound. From a mathematical point of view, the problem is solved using Green functions and the pressure field at any time t at position $\vec{r}$

inside Ω_2[5]. Therefore, to obtain the necessary information for the reconstruction of the sound field, microphones arrays should be necessary to capture the pressure and gradient pressure on the boundary surfaces ∂s. For the reconstruction of the field, these signals need to be played back by means of baffled loudspeakers (closed back, monopole radiation) and non-baffled loudspeakers (sound radiation on both sides, dipole radiation) respectively, located at the same points where the microphones were put during recording, see Figure 10.

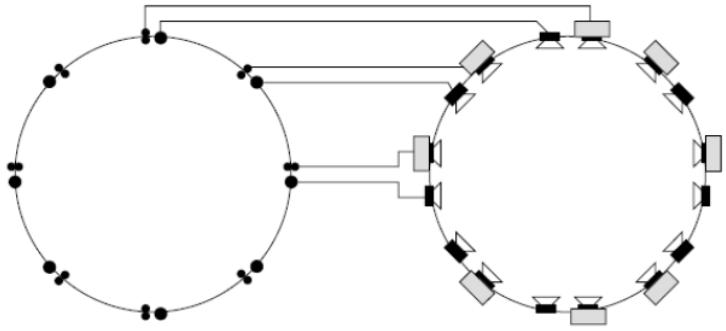

Figure 10.Acoustic field recording and reproduction in a WFS system.

Theoretically WFS it is a better technique than the two previously mentioned, but it is in most of the cases impractical and unrealizable. The most restricting boundary condition is that the system produces the sound field accurately only if the loudspeakers are at distance of maximally a half wavelength from each other, condition that impose that the centroids of the loudspeakers should be a few centimeters from each other to be able to produce high frequencies appropriately[30].

2.2.4. Binaural Coding

Based on psychoacoustics, binaural recordings preserve the necessary spatial cues used by the brain to perform localization and size estimation of sound sources as well as source distance and environment perception. The technology is based on the recording of sound events with two omni-directional pressure transducers located in the ear canals of a subject or inside a dummy head, so that each signal is then unambiguously played back using headphones.

Binaural coding of mono sources is achieved by convolving the signal file with a Head-Related Transfer Function (HRTF) filter database [17]. A HRTF describes how a given sound wave input (parameterized as frequency and source location) is filtered by the diffraction and reflection properties of the head, pinna, and torso, before the sound reaches the transduction machinery of the eardrum and inner ear.

The hearing experience of binaural recordings improves when its combined with head-tracking systems, so that rotations of the user's head are compensated in real time to simulate a real auditory scene; so that, as in real life when a subject rotate his head, also do the acoustic scene[31]. One drawback of binaural recordings is that the user may suffer from non-individualization of the HRTF's, resulting in a mismatch between the HRFT used during the recording and the particular filter characteristics of the pinna and head of the user; situation that causes difficulties to localize sounds, confusions between front and back and non-externalization of some auditory events[12].

2.2.5. Source Extent Rendering

A review of the psychoacoustic phenomena involved in the perception of a sound source extent is presented in [12]. It is shown that the perception of the size of a single source is controlled by the pitch, duration, loudness and type of signal emitted by the sound source. For multiple sources several experiments are studied, to demonstrate how the position and level of coherence between sources affect the perception of the extent and to link reverberation with a decrease of the IACC, resulting in wider apparent source width and an increased impression of spaciousness [13].

Several methods and techniques have been used to artificially control the sound source extent [14]. For stereo systems, thanks to the work by Blumlein in 1933, many audio engineers have use decorrelation between the two speakers channels to produce nice, delightful and spacious musical productions. The Mid-Side (MS) recording technique have been used for this purpose; it uses one cardioid capsule pointing the scene and a bi-directional capsule perpendicular to it, capturing the sides of the sound event. Varying

the gains of each signal, before a matrixing process, affects the inter-channel correlation of the stereo image making it convenient for controlling the width after the recording was made.

For a 3D audio scene, sampling a broad natural sound with an array of microphones can extend the previous concept; case in which a set of decorrelated signals is obtained. Later, these signals are spatialized at their proper locations to render a broad sound source in the reconstructed 3D sound scene. As not always such an array of microphones is available to be used and also not always is possible to record an audio event, an alternative method is to apply decorrelation on an existent mono source to obtain a set of decorrelated copies and use them at discrete source locations [11]. A state of the art review on decorrelation techniques used for this purpose is presented in [12], where according to the author, the Dynamic Decorrelation technique imparts a quality of liveliness to a sound field that is missing in the commonly used FIR implementations and other basic decorrelation techniques.

2.3. Music, Effects and Space on 3D exhibition systems

The use of space has been an important aspect of electroacoustic music in its different forms especially during the last decade due to the availability of tools for spatial composing, software with automation tools and the availability of multi-speaker setups. With the systematic inclusion of space into musical composition, music finally leaves behind the old conceptual constraints of music as a time-art. The computer emerges as the ideal tool for development and puts new concepts at the hands of creative users. Some of the relevant work involving music, sound effects and spatialization techniques to diffuse sound in three dimensions are now presented.

2.3.1. Spatio-musical composition strategies

> *"Creating spaces is strongly connected to the experience of our surroundings, and in this respect, spatial mimicry made possible with Ambisonics can be a useful approach. When you consider very tiny sounds building up an impression of space, then you can begin to imply a space within which these sounds should live. If you use abstract sound material,*

> *it can be difficult for the listener to find the spatial context. If you don't want to use reverberation, you don't have a clear spatial context to start off with, either. But gradually, as the sound material unfolds, its behavior, its motion behavior—the relation between many things happening at once—imply space, even though you are not using reverberation or clear sound identities. This is something I find very interesting because I don't like using reverb. When I do, I try to calculate a realistic room model using Ambisonics reflections".* *Natasha Barrett* [9].

Natasha Barrett is a prominent and prolific composer with relevant work-pieces in areas such as sound installations, instruments and live electronics, dance, theatre, and animation projects, but all her energy seems to stem from her acousmatic composition. In [22] the author discusses different approaches to space and how the composer can work with these ideas. She argue that the composer must understand what happens to the sound-object in the listeners space so that the perceived space appears real (spatial illusion), knowing that the audience is listening to an illusion in a stereo or multichannel space produced through the phantom images from two or more loudspeakers. In most of his pieces the perceived space appears real through maintaining 4 real "spatial laws": - The effects of sound transmission: using the absorption coefficient as a usable composition parameter. – The properties of a reverberant field: using surface reflections forming the reverberant field to give an idea of the characteristics of the space. –The object image size and multiple object relationships: the sizes of images give a good clue to the proximity of sound-objects in relation to each other, as well as in relation to the listener. - Doppler shifts and gestural-spatial definition: the motion of a sound-object is the main clue to the size of an enclosure [22].

2.3.2. Other Relevant pieces

Many other musical pieces make use of space by means of multi-speaker arrays. A compilation of recent computer musical exhibitions is made by [4][10], where the discussion in focused on 4 aspects, namely: Time delay, The Role of Dynamics in the Simulation of Distance, Inner Sound Space Organization and Movement in Space. The publication uses examples from acoustics, psychoacoustics and music to conclude how these 4 previous

mentioned features are being used as compositional and performing tools. An explanatory block diagram is shown in Figure 11. In [23] two pieces composed by the paper author are explained and analyzed: “Words on the streets are these” is a interactive composition where granulation and modulation can be used to create more abstract textural material, and where through filtering the listener can create rhythmic loops that ring in various inversions of inharmonic chords. “Still Life” is a concert work for string quartet and live electroacoustics, whose aim was to explore some of the potential interactions between instrumental gestures and sounds, acoustic space, and the ‘technology-performer’, using transformation and spatialization technology as a mediator. Both pieces used PureData and a set of patches to spatialize and transform sound. Finally a well documented report of the exhibition of a piece called: “Space within Space: An Acousmatic Evening” can be found in [15]. Within the topics that are explained, fine details are commented on how the speaker setup was chosen, the programming structure, the acoustic hall and its role inside the exhibition and a more theoretical discussion on the Art of Diffusion.

2.3.3. Spatial Grains & Textural Composition

An interesting research work and application combing granular synthesis and spatialization is presented in [8]. Through the investigation the term Spatial-grain is used to refer to the spatial information that is already encoded in an Ambisonics recording and that can be retained at granular level. Considering an Ambisonics recording of a source moving from a point to another (a left to right movement, for example), that is later granulated, or broken up into tens of thousands of spatial-grains each lasting 10-50 msecs.

Since each spatial-grain retains the original recording’s spatial-domain information, it is possible to mix spatial-grains of the source in a particular location with spatial-grains of the source from other location in the trajectory. The result would be granulated source sound coming from both the left and the right. As an extension of the first hypothesis, the authors consider the synthesis of non-points of sound where for example from a line trajectory a surface of spatially located grains is created using some of the different approaches for synthesizing apparent source extent [12], as for example VBAP

spread or decorrelation techniques. After all this process, some simple manipulations of spatial-grain clouds could offer exciting possibilities for spatial sound design.

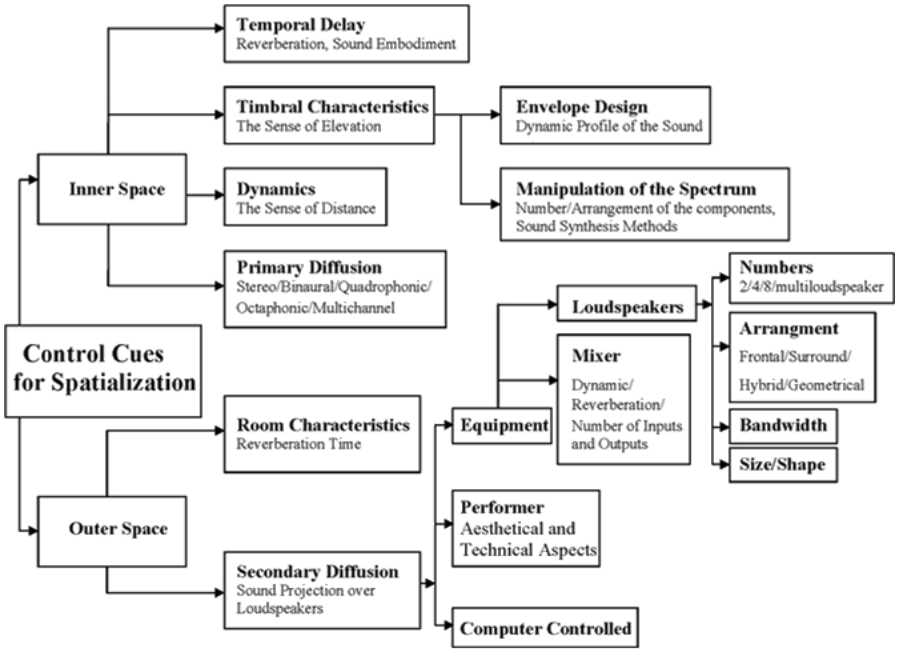

Figure 11. Control cues for using spatialization as compositional and performing tools [4].

A more conceptual term used in electroacoustic is Textural Composition, "used for the practice of working with musical texture and subjugating other musical qualities for the express purpose of creating a sonic space on the boundary between sound-objects and soundscape" [20]. For the spatialization of textures: loudspeaker amplitudes, inter-aural time differences, and artificial reverberation with uniform random variables create the mobile environment without spatializing each sample. As it is explained in [20], for compositional implementations, three movements are mixed hierarchically to define texture trajectories as seen on Figure 12. – One simpler approach is to give a constantly changing virtual angle that moves around the circle at varying speeds. –Then a second type of movement is built when a copy is slowly oscillated in and out of phase with the original stream by means of a variable delay. – Finally, the use of random values are used to control the send level to a reverberation that is statistically assigned to 4 locations on the circle, providing the sensation that sounds are moving outside the circle.

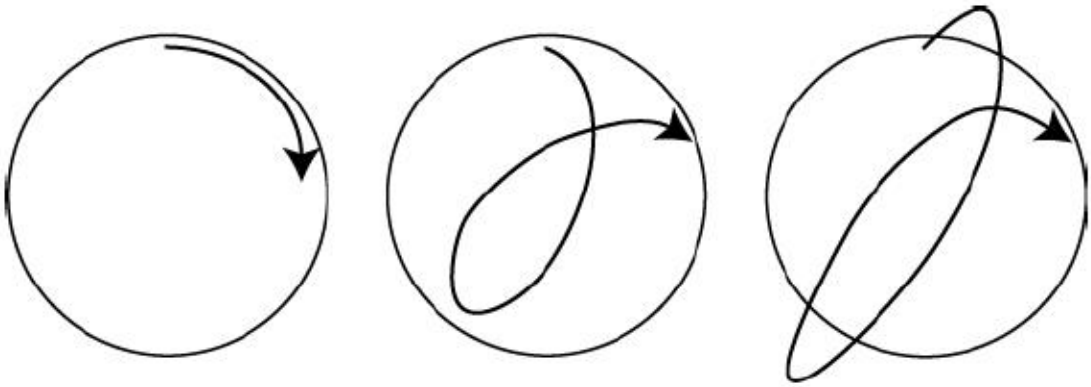

Figure 12. Effects of spatialization.[20]

3. 3D Implementation, Exploration, and Contributions

A set of different applications and case studies were implemented using the previously mentioned tools and theoretical background, with the goal of exploiting the sound spatialization capabilities that multi-speaker array systems offer. This chapter describes the different strategies, tools and methodologies that were followed.

3.1. A Basic 3D Reverberation Model

Natural reverberation can be considered as a phenomenon occurring when sound waves propagate inside a room; therefore it contains information about the dimensions, shape and size of an enclosed space as well as information of materials and textures of walls and objects inside. For the purpose of this survey we will consider a simple rectangular geometry containing only an omni-directional point source, since more complicated and realistic case of studies can be considered a generalization of our simplistic realization but also because a complete 3D reconstruction of the acoustic field inside a real room will be out of the scope of this thesis.

The implementation is based on the reproduction of the direct sound, given the location of a sound source and the location of a listener within a room, along with a certain quantity of the first early reflections that the source-room-listener setup defines. What makes this approximation different from any stereo reverb is the possibility of particularly spatialize and treat the direct sound and each of the early reflections as completely different sound sources coming from different places around the listener, as really happens in a real life situation. This independency between reflections and the possibility of rendering the reverberation effect with a specific number of reflections are

useful to experiment with the capability of the VBAP algorithm and the audio engine used in Barcelona Media, to create the sensation of space imposed by the room geometry; being this the primary purpose of this approach and the reason of its implementation.

3.1.1. The Image Source Model

The image-source model is a well-known method that can be used in order to generate a room impulse response, i.e., a transfer function between a sound source and a listener, in a given environment. It is based on the principle, that a specular reflection can be constructed geometrically by mirroring the source in the plane of the reflecting surface. As is shown in Figure 13, a sound source S is reflected symmetrically at the boundaries of a rectangular room, then, the obtained images are reflected about the images of the boundaries. As a consequence, all lines connecting an image of the source with a listening point R corresponds to a particular path between the source and the listener after a number of reflections.

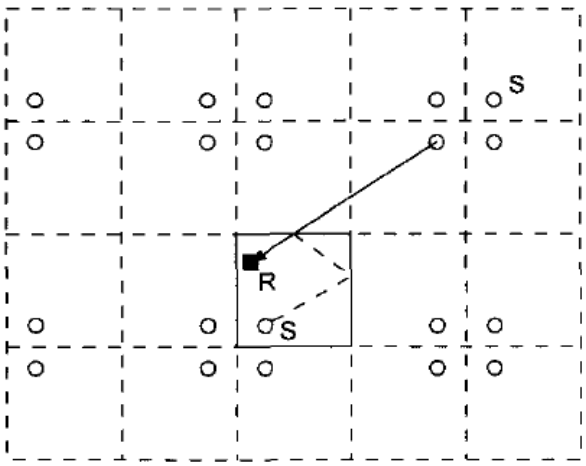

Figure 13. The Image-source method.

For particular practicality, we are not interested in the impulse response computation as the method suggest, since we are not interested in obtaining a reverberation by convolution. From the model we only require the delay time and magnitude of each echo as it is heard from a particular position in the room, and the direction of each image source relative to the listener position.

3.1.2. Implementation Details

As depicted in Figure 14, three main blocks characterize the functionality of the proposed sound effect. The audio processing is done by a CLAM network, which implements a gain control, a simple delay line and a VBAP spatialization block-set for each of the echoes computed by the Image Source method. These chain of blocks fully describe the behavior at any time of each of the early reflections in a constantly changing setup, (where the sound source or the listening point may be changing), as follow. The gain control sets the attenuation that a reflected sound has suffered, the delay line define the time that the reflection traveled before reaching the listener and the VBAP localize the image source in space using as parameters azimuthal and elevation angles, as in spherical coordinates.

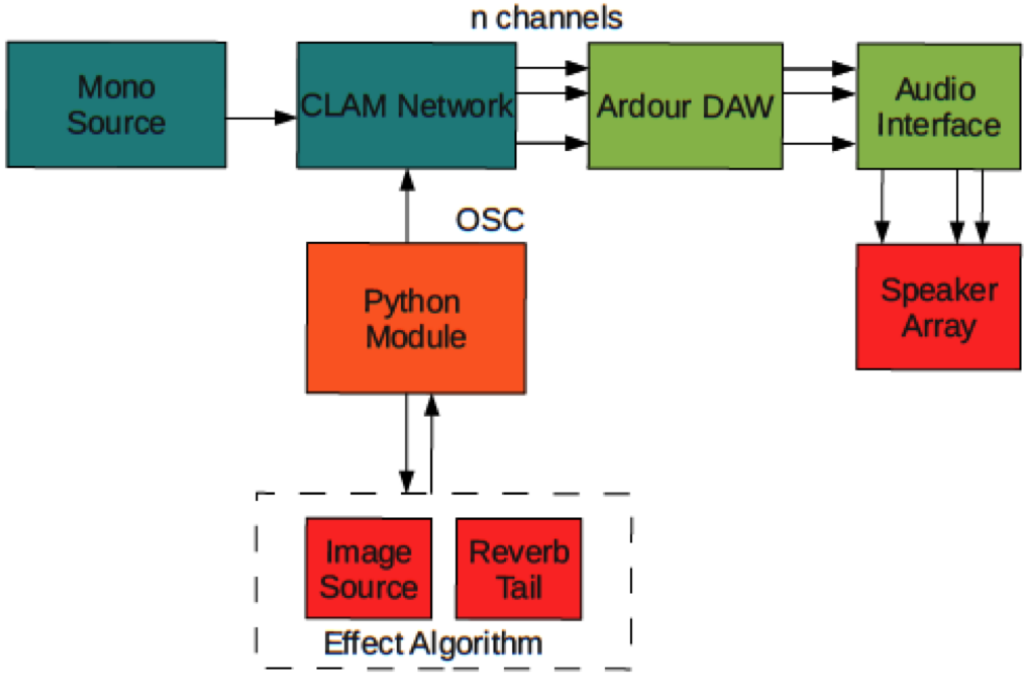

Figure 14. Block diagram of a 3D reverberation.

The second important block in Figure 14 is the python module. Implemented as a control of the audio scene, it dynamically changes the input parameters of each of the blocks in the CLAM network for each of the echoes, according to the scene changes, by sending OSC messages to CLAM. Therefore, is within the python module that the scene is configured and somehow animated by

defining room dimensions, number of early reflections computed by the model, reflection coefficient of walls, source position, listener position and all possible changes of the parameters in time that characterize the particular audio scene that is going to be simulated.

To make the effect independent of the final exhibition system a third module is needed to properly configure the CLAM network. Considering, for example, a reverberation that uses the first 50 early reflections obtained from the model will imply having a CLAM network with that number of block chains and a pre-established VBAP block configured for a particular speaker layout, which would be unpractical. For this purpose a configuration python script was designed to be run before the main module to create the CLAM network that will be used, with all the connections according to the number of the early reflections that are going to be spatialized and most important of all, with the setting of the VBAP blocks for the particular speaker layout that will be used in one particular demonstration. With this, hearing the reverberation in a different speaker setup would be as easy as running the configuration script with a new speaker layout with the possibility of changing the number of sources that the CLAM network will be handling.

3.2. 3D Extent of Audio Effects

As was previously presented in section 2.1.2, the Fractional Delay Line is the basis on which to build more complex stereo effects such as the Chorus and the Flanger. Therefore, it becomes necessary to implement a delay-line interpolator inside CLAM that can be used later on, to design 3D realizations of these mentioned effects.

Based on the survey presented by Dattoro [16], an all-pass interpolation was chosen for final implementation over the more simple approach of a linear interpolation. Dattoro argues that despite the implementation cost versus performance of linear interpolation is difficult to beat it does have significant drawbacks such as amplitude and phase distortion, and aliasing. Figure 15 shows the actual circuit used to implement the all-pass interpolation, where is easily observable the dependency on not only the last input sample, but also the last output sample value.

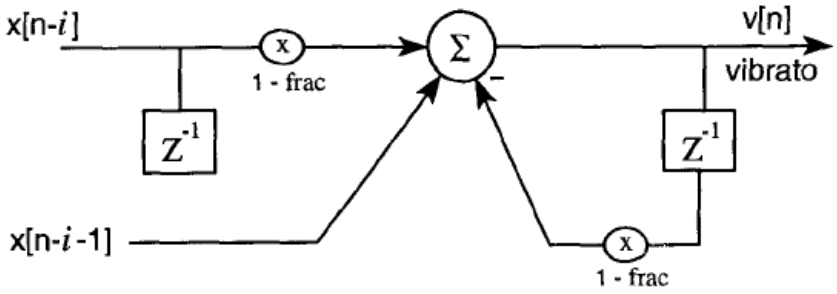

Figure 15. All-pass interpolation circuit.

The advantage of having a Fractional delay line resides on the possibility of obtaining a soft delay modulation thanks to the continuous variation of the delay that can be achieved. This means that a low frequency oscillator (LFO) can easily drive the delay that it's going to be applied at each sample time. The phenomenon can be devised in Figure 16 where a sinusoid signal is delayed a fractional number of samples.

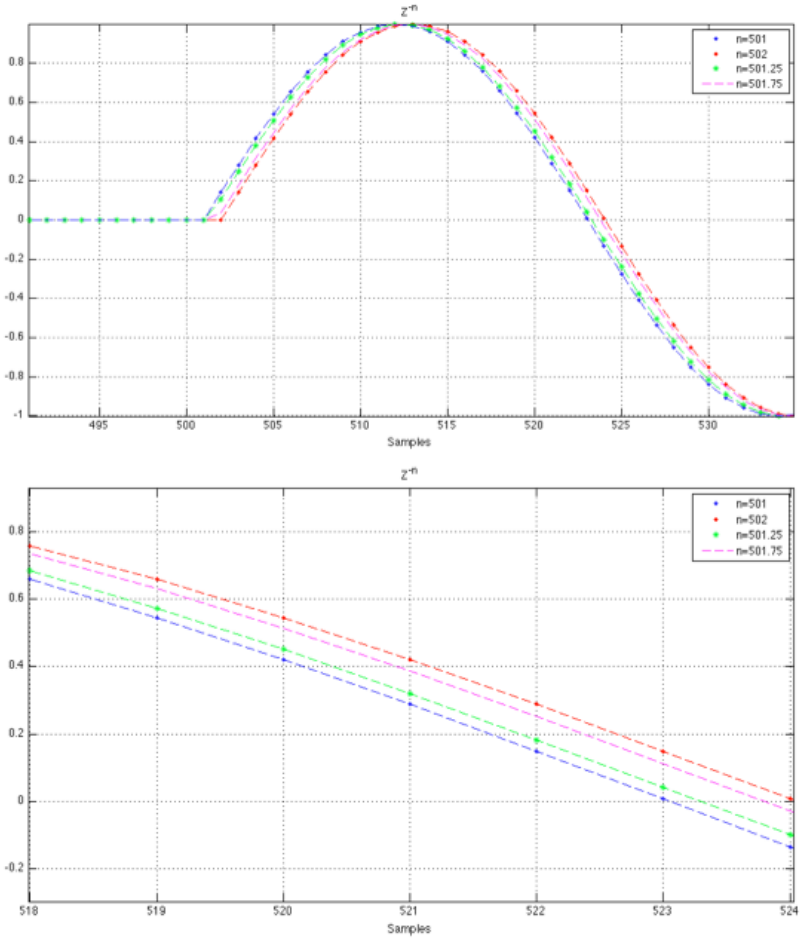

Figure 16. Signal delayed a fractional sample number.

Once the fractional delay is implemented, obtaining different effects (based on the modulation of the delay line) is done with the circuit topology depicted in Figure 5. Being a single topology for various effects, what defines the behavior of the circuit as one effect or another is the gain of each of the three loops on it (feedback, feed-forward and blend), the delay range (modulation width) and frequency of the LFO. Table 1 and Table 2 gather the values of the parameters for each of the effects that are obtained with this topology. Finally a few processing block-sets were developed inside CLAM to obtain each effect separately and in stereo configuration, as can be seen in Figure 17 where from a mono source a stereo output is generated with the sum of a Chorus and a Flanger.

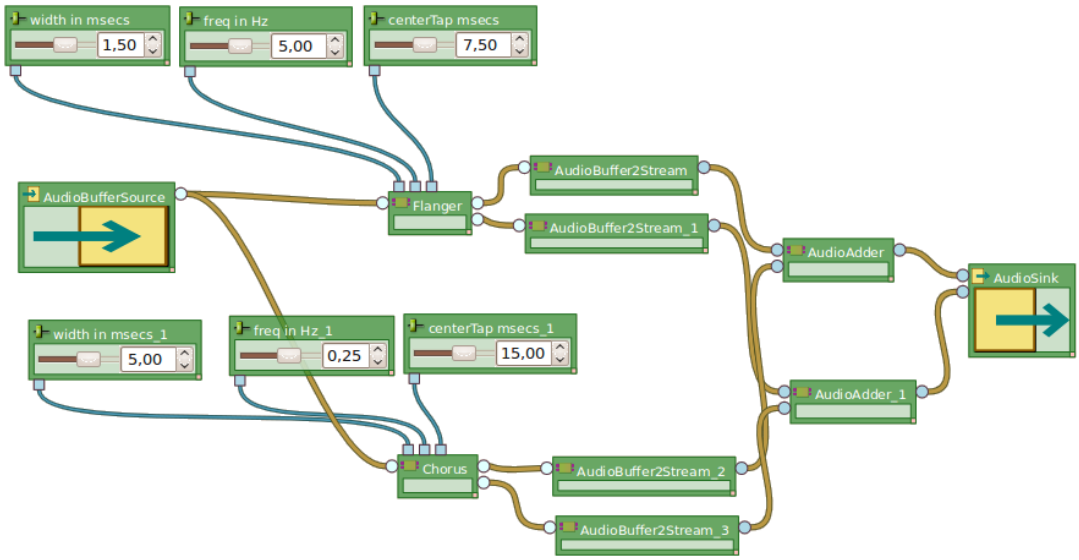

Figure 17. CLAM network using the Chorus and Flanger processing blocks.

3.2.1. 3D Flanger

During the flanger implementation in stereo many ideas were considered about how a 3d version should sound and especially which implementation will take more advantage of a multi-speaker system. One clear sensation that is perceived when hearing a stereo flanger is that something is moving, concept that is supported by the original conception idea: “The flange effect originated

when an engineer would literally put a finger on the flange, or rim of one of the tape reels so that the machine was slowed down, slipping out of sync by tiny degrees. A listener would hear a "drainpipe" sweeping effect as shifting sum-and-difference harmonics were created. When the operator removed his finger the tape speeds up again, making the effect sweep back in the other direction". Based on that perceptual idea, a 3D Rotating Flanger was design with the characteristic of being continuously spinning around the listener at the same frequency rate that the delay modulation.

For the implementation in CLAM, starting from the previously implemented stereo flanger, the LFO signal calculated for the delay modulation is used to generate an output control by taking the phase of the LFO signal and converting it in an azimuthal angle in degrees. This control output is later connected to a spatialization block-set inside CLAM named PBAP (Proximity Based Amplitude Panning) that uses the location of a source in spherical coordinates (azimuthal and elevation), and two angular parameters (angular decay and angular thickness) to control the individual gains of the channels in the actual exhibition system. The CLAM network template used for the generation of the LADSPA plug-in is shown in Figure 18, while a screenshot that shows the plug-in interface and the input control parameters is depicted in Figure 19.

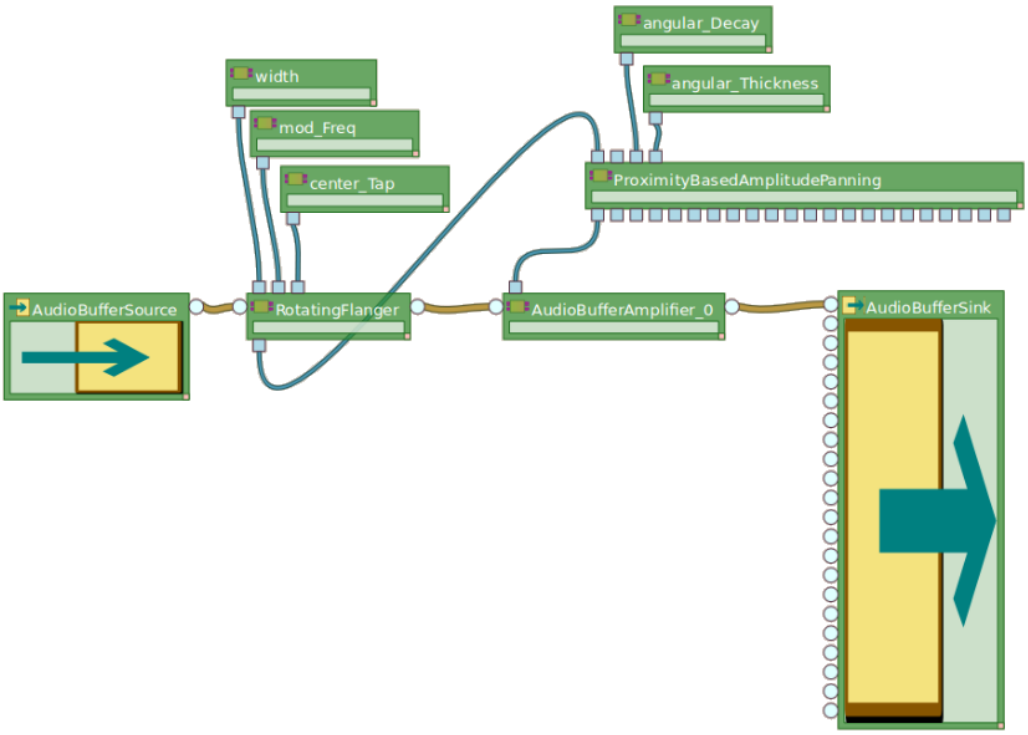

Figure 18. Network template for the 3D rotating flanger plug-in.

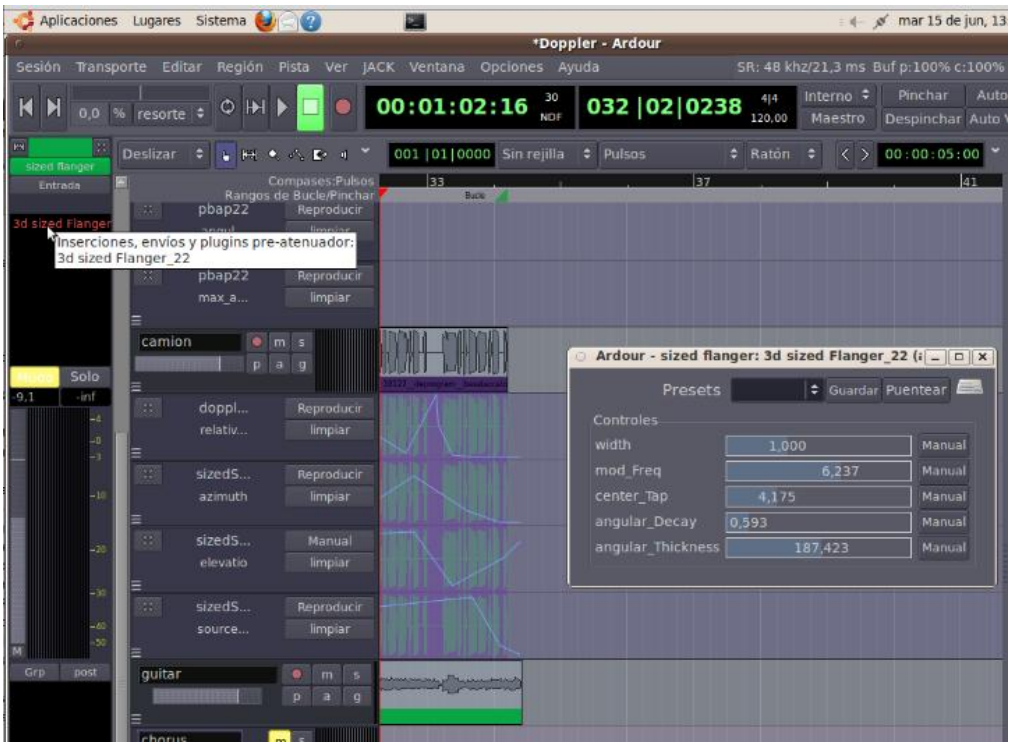

Figure 19. Screenshot of the 3D flanger LADSPA plug-in inside Ardour.

3.2.2. 3D Chorus

In the case of the Chorus, the hearing perception is quite different from the Flanger. In a Chorus effect there is a sense of steadiness since the frequency of the delay modulation is quite small, (around 0.5 Hz) with large excursions (around 30 mseg) therefore, what the user finds is a bigger perception of space. Sticking to that idea, a 3D version of a Chorus was implemented by just spatializing different copies of the mono Chorus effect around the listener (as many as channels the multi-speaker system have), reinforcing the perception of space. Since adding different sources with exactly the same audio content around the listener wouldn't be of any help, two control parameters are added to the processing block-set. One is a "phase offset" for the delay modulation (LFO) signal, which is assigned by default accordingly to the azimuthal location of each of the speakers in the exhibition system. The other is a "Parameter Variability" control, used to deviate the basic setting of the Chorus for that particular channel, namely the width and frequency of the LFO. The template CLAM network used for the LADSPA plug-in compilation is depicted in

Figure 20. Figure 21 shows a screenshot of Ardour when using the 3d Chorus LADSPA plug-in.

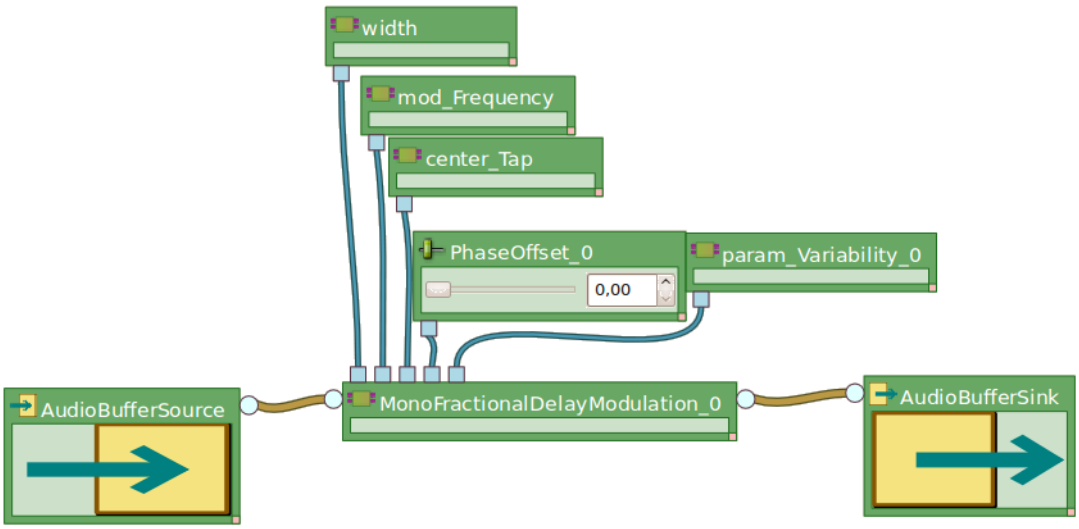

Figure 20. Network template for the 3D Chorus LADSPA plug-in.

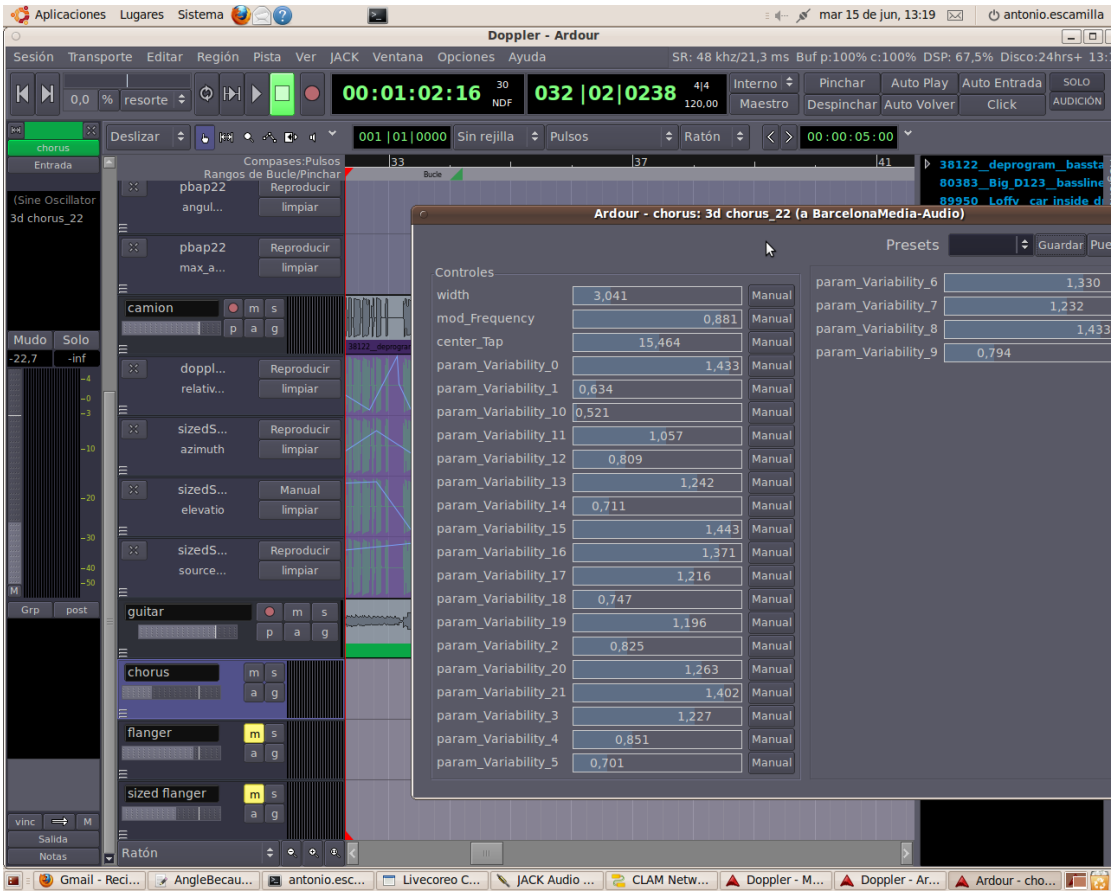

Figure 21. Screenshot of the 3D chorus LADSPA plug-in inside Ardour.

3.3. 3D Moving Source Model

The simulation of moving sources is of big importance in many audio applications, including musical applications, where moving sources can be used to generate special effects creating new hearing experiences. Motion of instruments while they are being played can also cause small variations in the sound, and affect the expressiveness of the performance. Although the effect of this motion on sound has not yet been clearly established, it probably contributes to the rendering and should be taken into account in attempts to synthesize musical sounds. Video games and virtual reality are other fields, where moving sources play an important role. To simulate motion, the speed and trajectories are crucial to creating realistic acoustical environments, and developing signal processing methods for reconstructing these contexts is a great challenge.

Four important perceptual cues can be used to design a simple model for a moving sound source. Most of these cues do not depend on the spatialization process involved, but they are nevertheless greatly influencing the perception of sounds, including those emitted by fixed sources.

3.3.1. Doppler Effect

The Doppler frequency shift can be controlled by a variable fractional delay line. In the case of a sound source emitting a monochromatic signal and moving with respect to a fixed listener, Smith et al. [19] obtained the following expression:

$$\frac{dT(t)}{dt} = -\frac{Vsl}{c} \tag{14}$$

To implement a continuously varying delay, a delay growth parameter g was added to the fractional delayline. When g is 0, we have a fixed delay line, corresponding to dT(t)/dt = 0. When g > 0, the delay grows g samples per sample, which can also be interpreted as seconds per second. An illustrative example can be devised in Figure 22, where the delay and amplitude variations are shown for a sound source moving towards a fixed listener.

3.3.2. Distance Attenuation and Air Absorption

The simplest acoustic effect is due to the dispersion of sound waves into the environment. From the physical point of view, the sound pressure relates to the sound intensity, and in a more complex way, the loudness. Assuming no other attenuation, the intensity decays proportionally to the square of the distance between the source and the listener. It is important to understand that only the relative changes in the sound pressure should be taken into account, since the absolute pressure has little effect on the resulting perceptual effect.

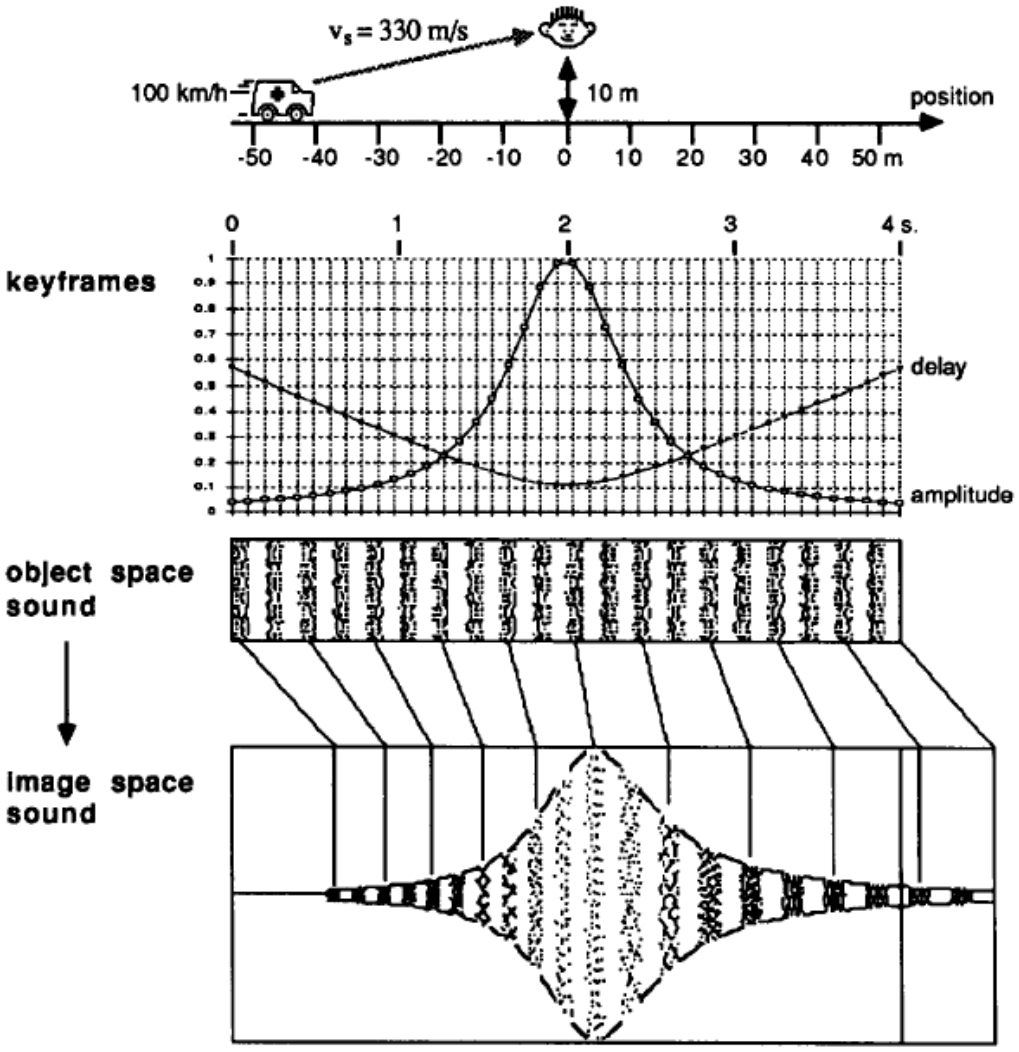

Figure 22. Key-framed values of amplitude and delay for a moving sound source, and the corresponding mapping of sound signal from object to image space.

In a moving source situation the sound timbre changes with distance, phenomenon that can be related physically to the absorption or the air. The main perceptual effect of air absorption on sounds is due to a low-pass filtering process, the result of which depends on the distance between source and listener. A classical high-shelving second-order IIR filter was used to model the timbre variations due to the air absorption, as can be seen on Figure 23. To simulate air absorption, the control parameters (cutoff frequency and gain) have to be linked to the listener-to-source distance, especially for moving sources that cover large distances where the effect due to air absorption should be more noticeable.

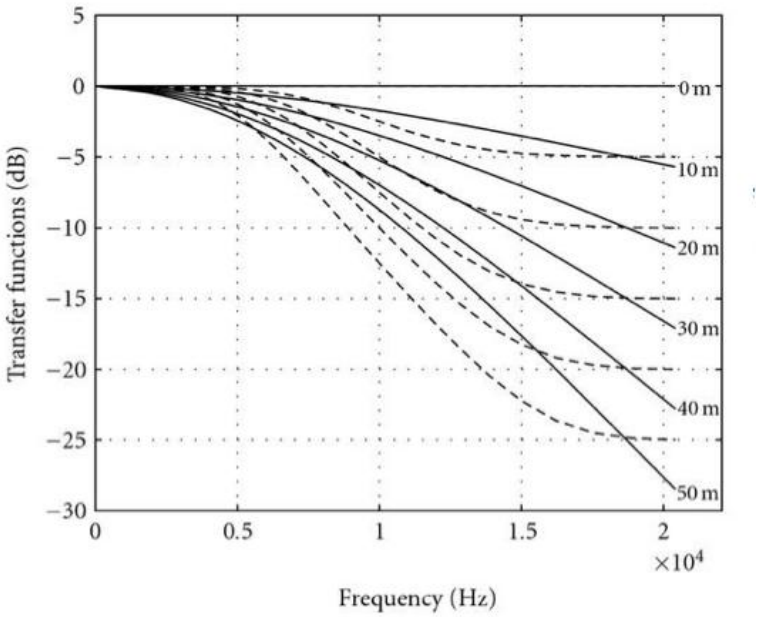

Figure 23. In solid lines the air transfer functions, in dotted lines the simulated filter transfer functions.

3.3.3. Source Extent

The last cue to be simulated is the perception of the extent of a moving source. Here the scenario is that given a sound source with a fixed size and depending on the distance to the listener, its extension should be perceived differently. That means that when the source gets closer to the listener, the sense of extension of the source, perceived by the listener is increased, and just the opposite when the source go away. The PBAP (Proximity Based Amplitude Panning) processing already implemented in CLAM is used for this purpose. As input control parameters, the PBAP unit receives the orientation

of the source in azimuthal and elevation angles, the maximum angular width, which defines the range in degrees of the speakers around the source location that are active, and the angular thickness k, which is used to compute a directivity pattern to set the gain of each of the channels in the exhibition system. To obtain a value of k from the source size and source distance, a block-set was implemented to compute a linear interpolation with the aperture angle imposed by the source when situated at a certain location to obtain a values of k, between 0 (onmidirectional) and 2.5 (hipercardioid). The CLAM network template used to generate the multichannel version according to a specific speaker layout is shown in Figure 24.

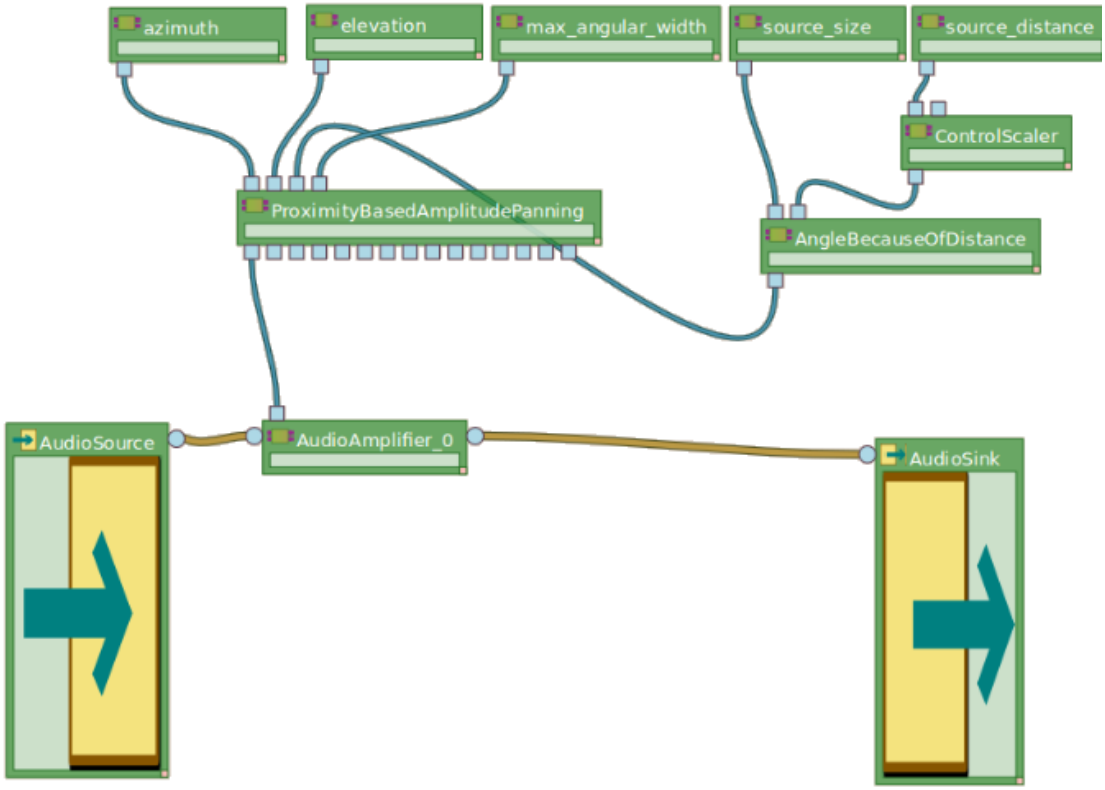

Figure 24. Network template used for the generation of the multichannel version of the process.

3.3.4. Task Summary

Each previously discussed phenomenon was implemented as an individual processing block-set for CLAM procuring to maintain the modularity that characterize this audio processing library, nevertheless all cues add an specific task to the final model that simulate the motion of an acoustic source by processing a sound file corresponding to the acoustic radiation emitted by a fixed source.

For an efficient processing of audio, the concatenation of perceptual cues and the sized-source plus spatialization modules were first implemented on CLAM networks (Figure 25) that were later used to generate LADSPA plug-ins to be used in Ardour as a sound-designing tool with automatization capabilities. As can be seen in Figure 26, the model is fully achieved by using consecutively a "Doppler&Distance" plug-in that produce a mono signal with the Doppler, distance attenuation and air absorption simulation(Figure 27), then a second plug-in "SizedSourceN" which from a mono source and by using input parameters of source distance, size and location angles produce N audio channels according to the actual speaker layout of the exhibition system (Figure 28) and last a "SpreadSourceDecoder" that simply decorrelates the N input channels before being sent to the physical outputs.

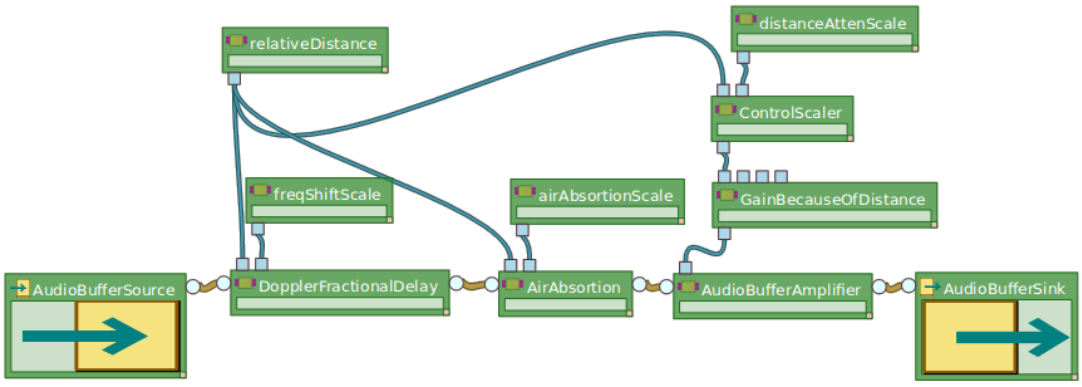

Figure 25. All audio processings in the model in a CLAM network.

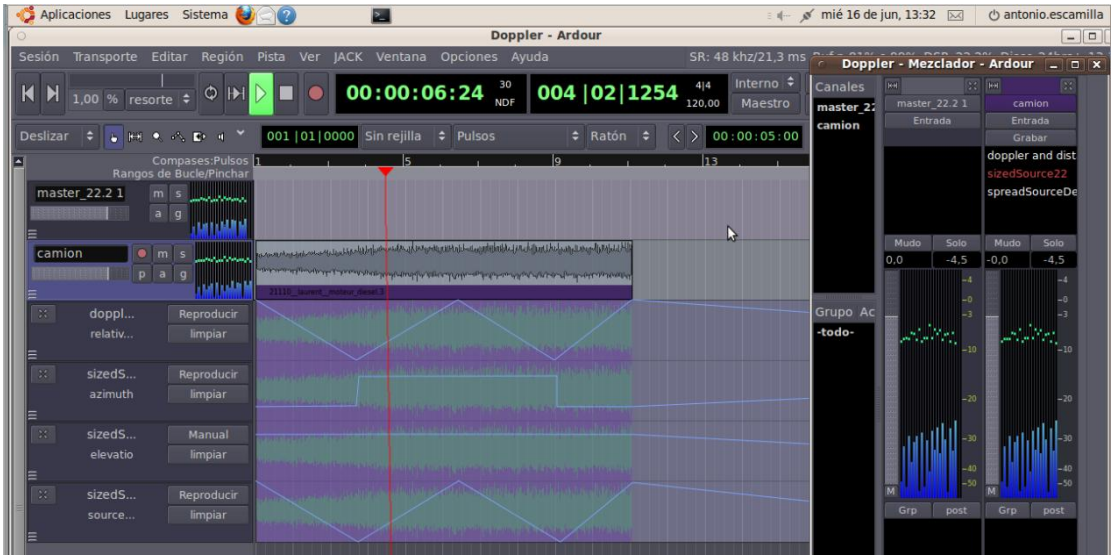

Figure 26. Plug-ins used in the moving source simulation inside Ardour.

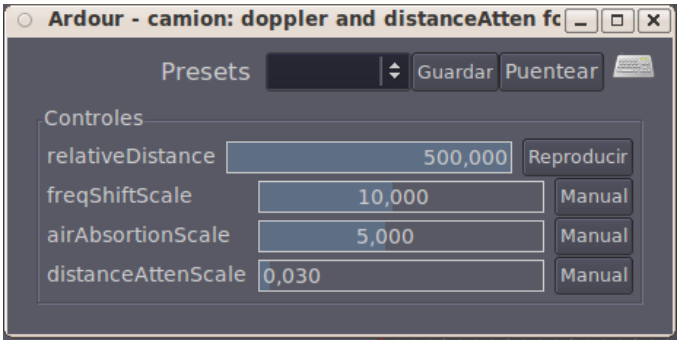

Figure 27. Interface of the Doppler, Air Absorption and Distance Attenuation LADSPA plug-in.

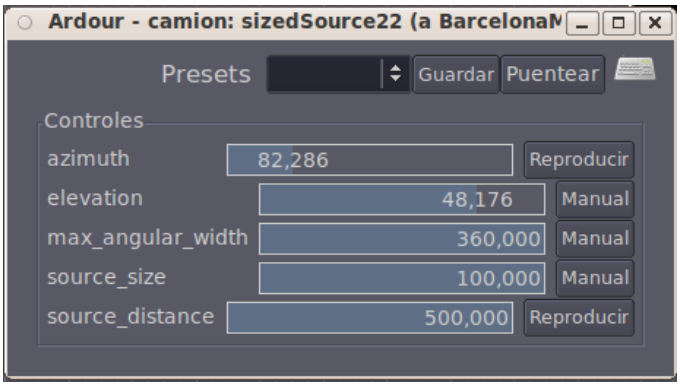

Figure 28. Interface of the SizedSource22LADSPA plug-in.

3.4. Generative Audio Scenes

One last task in this survey is presented below. Its scope is the automatic generation of choreographies in which the motion of sound sources in a virtual 3D audio scene is determined by a set of laws that are previously programmed and that may be assigned to each of the sources present inside the scene. To add variability and sense to the animation, some laws use energy-based features extracted in real-time from the audio to modify in different ways the movement pattern that has been assigned to the sound source. The visual feedback of the sound source's motion and its 3D spatialization is done in "LiveCoreo", a computer tool developed by the 3D audio group of Barcelona Media, capable of rendering 3D audio while recording or playing a motion choreography of the sound sources on it.

The implementation of this task required the interaction between different audio tools and the capabilities of python to control and synchronize all of them. Ardour is used as sound reproduction tool and MTC (MIDI Time Code) generator; here we suppose that each sound source in the 3D scene should have assigned a mono track in Ardour, meaning that the sound content associated in time with a given source should be included inside Ardour in the track assigned to that source. For the real-time extraction of energy-based descriptors from the audio being reproduced in Ardour, a CLAM network was implemented. This "processing" inside CLAM obtain audio information from Ardour via JACK and in a frame-basis approach extract the energy, the energy changes and a measure of the inter onset intervals when these are detected. These three descriptors are computed each frame and sent via OSC to the Python controller script. Refer to Figure 29to see the CLAM network used.

As was said before, LiveCoreo is used for visual feedback of the sources animation but most important of all, for the 3D sound rendering. Inside LiveCoreo an audio-visual scene is composed of a listener with an associated scene view and different types of sound sources each of them characterized by the algorithm that renders its 3D sound. In LiveCoreo once a scene is configured (off-line), all objects that are present, can be controlled via OSC messages in execution time.

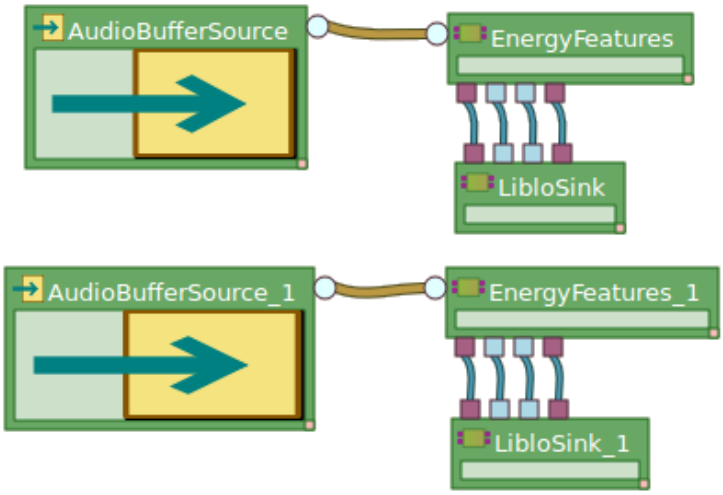

Figure 29. CLAM network used for the extraction of audio descriptors and it's sending via OSC.

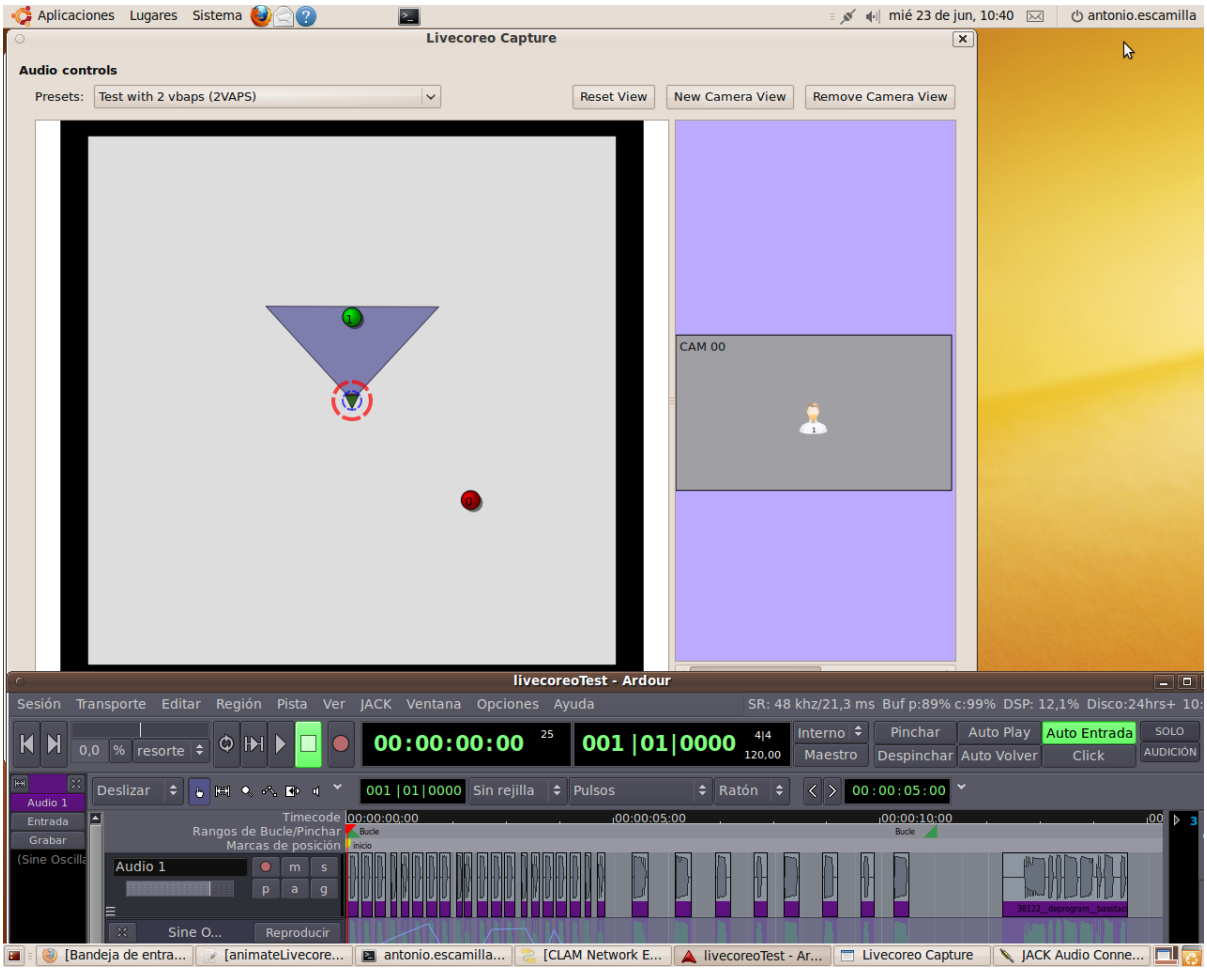

Figure 30. Screenshot of Ardour and LiveCoreo generating a 3D audio choreography.

A Python script controls the whole functionality of the application, which has the important task of gather, process and send the information required by the others tools. To summarize the data flow, it is important to say that the python script receive MTC messages from Ardour as source of synchronization, receives OSC messages from CLAM with the audio descriptors and sends OSC messages to LiveCoreo to define each frame the location of the sound sources in the scene. Besides data handling, the script is used to define the scene, the source objects and the motion laws associated to each source in an object oriented approach. Figure 30 presents a screenshot of LiveCoreo and Ardour, where two sound sources and a listener define the scene choreography; in this particular case the separation of the onsets on track 01 determines the frequency of the circular movement of source 01, and the energy changes of the same audio track (when above certain threshold) produce a instantaneous jump on the source position.

3.4.1. Audio scene composition

An object-oriented structure was defined to easily construct the 3D audio scene in which the sound choreography will take place. To do that, the user should define the set of objects that compose the scene and assign the relations between objects that produce the generative audio scene. For example, any "scene" object must have a listener and a set of sound sources, then, for each sound source object, a number of motion laws should be assigned to it; here each of these laws adds a particular change in position producing as final result a superposition of simple movements that turn into a complex and audio-feature dependant motion.

Different motion laws were considered for the choreography generation, some are simple in their definition and in the movement they produce, but with the characteristic of being dependent of the audio descriptors extracted in real time from the respective audio track. Some others laws show a chaotic behavior and are extremely dependent of changes in their initial position, even though, these laws are completely deterministic. Circular and straight line three-dimensional movements are examples of the first group of laws, meanwhile, a double inverted pendulum and different strange attractors, such us the Lorenz and Rösler dynamics, are those laws that exhibit chaotic

behaviors. Each law is an object that may be instantiated and kept in an array of laws, that is assigned to a sound source to rule its motion on space. Figure 31 and Figure 32 show a generative audio scene, where a six channel multi-track audio session is being spatialized according to six different sets of motion laws that are assigned by the 3D sound scene designer using subjective and artistic criteria.

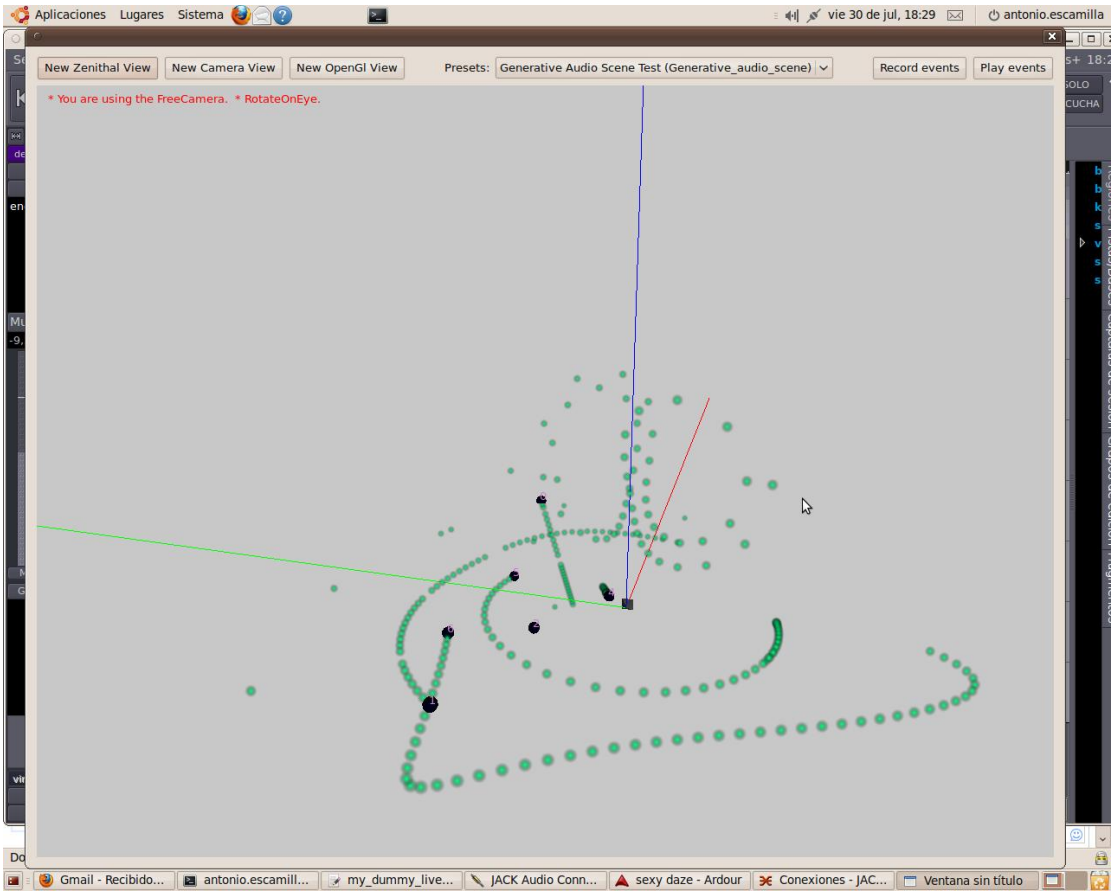

Figure 31. LiveCoreo screenshot of a 3D sound choreography.

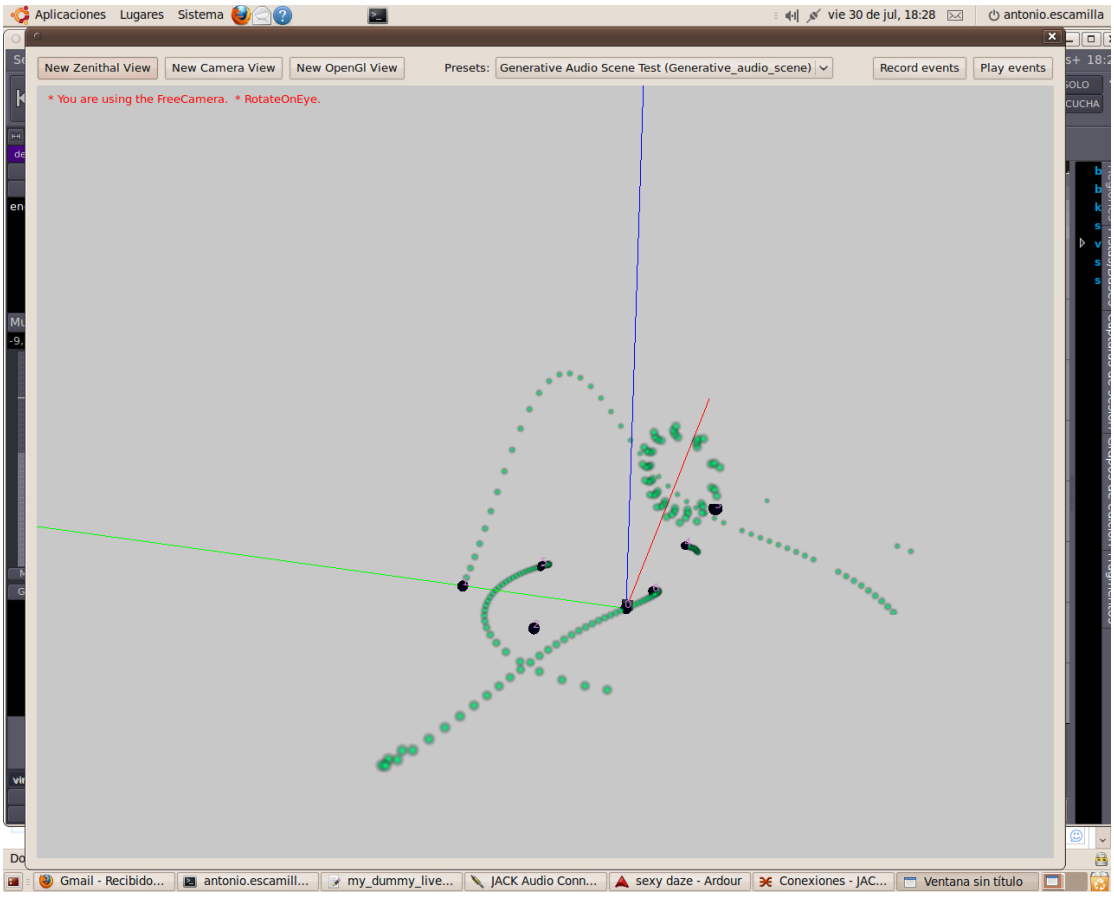

Figure 32. Screenshot of LiveCoreo with six sound sources on a 3D generative audio scene.

3.5. Contributions and existing implementations

CLAM is a robust framework for audio processing, which proved to be efficient and reliable for rendering 3D audio. Using this tool, the mentioned VBAP and PBAP methods, were implemented by the audio group of Barcelona Media as part of the research on “spatialization techniques”. In all the 3D audio effects presented in this survey, the VBAP or PBAP spatialization is used as a block-set that is not only already integrated in the CLAM framework, but also extremely flexible, since both can be configured to work with a specific speaker layout. All the other block-sets used inside CLAM and that were not implemented by the author are part of the development environment that CLAM offers for audio processing and analysis. For example, an OSC server, a simple sample delay line or arithmetic operations over control signals, to mention just a few.

In the first case of study, the image source model was not a contribution of the author; again it is part of the research that Barcelona Media have done in the 3D audio field and it was simply used as a function that is manipulated by the scene controller according to the demands of the user that program it.

All other described features were developed in the framework of this thesis and are contributions, experimentations and ideas product of discussions between the author and the thesis supervisors. For the 3D reverberation the scene controller and the scene configuration scripts were designed to successfully define how the audio would be rendered in 3D, according to the initial chosen parameters and the scene changes in real time. For the second case, all the audio effects are original implementations based on the fractional delay line block-set that was programmed by the author, using CLAM. All of these effects are now fully usable block-sets inside the CLAM's "NetworkEditor" and can be used for musical creation and experimentation as part of an audio "Network" that could be compiled as a LADSPA plug-in.

Almost the same scheme was followed in the Moving Source Model and the Generative Audio Scenes. In the case of the moving source, the model implementation, (which is a sequence of physical phenomena simulated by CLAM block-sets),was programmed by the author to be used as part of a CLAM network which define the final behavior of the sound effect in a 3D audio scene. All the python scripts needed for the network's generation and also the network templates used for the plug-in compilation are part of the original work presented in the project. For the Generative Audio Scene, all software and scripts implementations, signal processing and features extraction done outside the "LiveCoreo" application, were contributions of the author. Even the final 3D audio demonstrations, scene design and movement assignment for each sound source, (using a couple of multi-track audio sessions recorded and produced by an renowned electronic musician), were part of the thesis project.

4. Results and Experiments

In this chapter we will discuss some experiences and results extracted from the different case studies that were implemented. As it was stated in the literature review, in the section on Music, effects and space on 3D exhibition systems: with the systematic inclusion of space into musical composition, music finally leaves behind the old conceptual constraints of music as a time-art, which makes almost unnecessary to make detailed evaluations considering rigorous models if these do not take into account the perceptual experience that may be produced once the audio is spatialized. Therefore we should consider the speaker's layout and the system calibration as important as the audio processing itself when evaluating the procedures. One thing to have in mind is that when the numbers of speaker are increased beyond a stereo or 5.1 surround system the degrees of freedom of the system linearly growth becoming even more relevant than the physical model that define the sound processing.

The reverberation case study was rendered with approximately 50 early reflections computed by the Image Source model and with different room sizes to perceive the differences of simulated space when the number or early reflection were maintained fixed. Some other listening test considered dynamics changes of the source location to experience how the different locations may or not give clues about the geometry of the room that is being simulated for different numbers of early reflections. The second case of study may offer more perceptual discussion since even before the 3D experimentation diverse differences were obtained with just slightly variations on the center tap parameter in the Chorus and Flanger effects. This phenomenon is consequence of the use of the Fractional Delay Line with two outputs, one being the fractional modulated delayed output and the other the fixed central deviation of the modulation, as shown in Figure 5. Therefore

when the center tap is not big enough to enable the complete delay excursion defined by the width parameter, the effect is altered in a strange way, where the sense of the modulation is lost. Both 3D realizations of the delay modulation effects presented seemed to work fine once spatialized: In the case of the 3D flanger care must be taken when the frequency of the modulation is set to high, (around 15 Hz), since it may cause the effect to be jumping to fast between not consecutive speakers losing the desired sensation of "spinning" around the listener. In the case of the 3D Chorus a compromise between an exaggerated effect and a simple sensation of wider space should be carefully defined, by using well the "parameter variability" slider available for each chorus source in the 3D effect. Otherwise it will sound like a weird modulated reverberation or a not meaningful realization, respectively.

The moving source model was more dependent to the speaker's layout and calibration of the system than the previous two case studies. When considering a sound emitting object that moves at a certain velocity and with a certain trajectory, the perception on the location of the source as long with the frequency variations due to the Doppler effect may not be that clear in systems where the speaker positions are not equidistant of the sweet spot or may only be perceived well in a small central spot inside the speaker's geometry. One perceived failure of the model implemented was the difficulty to produce perceptible frequency shifts for sounds with low frequency content.

The automatic generation of choreographies was perceptually tested with a couple of multi-track songs and the results were quite good. Once the whole application was programmed and tested from a technical point of view, it was simple to use but flexible enough to be creative. While preparing and programming a 3D scene with the "application" some characteristics showed to affect more the perceptible experience and required a detailed management, as for example in the case of the amplitude attenuation by distance, to avoid strong and fast transitions of volume.

5. Conclusions and Future Work

A general framework was presented as strategy to develop audio effects that were designed to be reproduced in three-dimensional audio exhibition systems. For this purpose a set of computational tools were used to take full advantage of the spatialization capabilities of the 3D audio framework, to improve the sense of immersion and involvement over an audience in comparison with traditional stereo systems.

One clear target of this survey was to experiment with the already existing technology platform of the Barcelona Media 3D audio group to encourage the use of 3D audio systems in music creation and performance by presenting a few applications along with the programming procedures and methodologies that led to its obtainment.

All the developments presented were successfully maintained independent from the final exhibition system, namely: number and location of speakers. This feature adds robustness and functionality to the obtained audio effects since there is no need to comply with a given standard, which punctually may define a specific speakers layout as in the case of 5.1 systems.

CLAM, the C++ Library for Audio and Music, has not only proved to be an excellent Framework for research and application development in the 3D Audio domain, but an essential tool for the LADSPA plug-in generation. The OSC support inside CLAM fully extends its capabilities since most of the configuration parameters inside any CLAM network can be dynamically changed in real-time from other software tools via OSC, as for example from Python, where a script was often used as scene controller.

5.1. Future Work

The rendering of reverberation effects was not fully studied under this thesis since it is already an own investigation field by itself, subject to more detailed research dealing with a theoretical physical background, which is already undertaken by other team members inside the Barcelona Media 3D audio group. As was mentioned before, other techniques, as Ray Tracing and Impulse Response Convolutions are more effective, to really reconstruct the "sound" of rooms with non-regular geometries, than the Image Source Method used within this survey.

In the framework of this project, only delay modulations effects were extended from stereo into a 3D spatialized version, which leaves an open door to many more audio effects that definitely may sound better or at least impressive in exhibition systems with many loudspeakers. Sound granulation could be used to obtain sound textures that once spatialized may cause perceptual experiences never heard before in stereo or 5.1 reproduction systems.

The moving sound source model presented did not consider the effect of an acoustic room in which the sound source may be moving; therefore a future improvement in the model should consider the rendering in real-time of a reverberation in which the source is constantly moving with not predefined trajectories. This junction of models can be used in sound installations and sound synthesis engines where generated sounds resembling music instruments can gain realism and impact by simulating movements where the creativity is the limit.

6. References

[1]A. Huovilainen. Enhanced digital models for digital modulation effects. *Proceedings of the 2005 Digital Audio EffectsConference*. Madrid, Spain, 2005.

[2] B. Truax. Composition and Diffusion: space in sound in space. *Organised Sound*. 3(2): p. 141-146. 1999.

[3] B. Truax. The Aesthetics of Computer Music: a questionable concept reconsidered. *Organised Sound*, 5(3), 119-126. 2000.

[4] B.Zelli. Space and Computer Music: A Survey of Methods , Systems and Musical Implications.*Toronto Electroacoustic Symposium*, November, 2009.

[5] G. Cengarle. Application of sound intensity to sound recording and reproduction: the Acoustic Quadraphony. *Master degree thesis*, University of Trieste (Italy) 2007.

[6] C. Roads. *Microsound*. MIT Press, Cambridge, Massachusetts, 2001.

[7]D. Rochesso. Fractionally addressed delay lines. *IEEE Trans. on Speech and Audio Processing,* 8(6):717-727. November 2000.

[8] E. Deleflie and G. Schiemer. Spatial Grains: Imbuing Granular Particles with Spatial-Domain Information. *Proceedings of ACMC09, The Australasian Computer Music Conference*, Queensland University of Technology. 2-4 July 2009.

[9] F. Otondo. Creating spaces: an interview with Natasha Barrett. *Computer Music Journal,* 31(2): 10–19. 2007.

[10] F. Otondo. Contemporary trends in the use of space in electroacoustic music, *Organised Sound*, 13(1), 77-81. 2008.

[11] G. S. Kendall. The decorrelation of audio signals and its impact on spatial imagery. *Computer Music Journal, 19(4):71/87, 1995.*

[12] G. Potard. 3D-audio object oriented coding. *PhD thesis*, School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, 2006.

[13] G. Potard and I. Burnett. Decorrelation techniques for the rendering of apparent source width in 3d audio displays. *DAFX 2004 Digital Audio EffectsWorkshop.* Naples, Italy, 2004.

[14] H. Vaggione. Composing musical spaces by means of decorrelation of audio signals. *DAFX 2001 Digital Audio EffectsWorkshop.* Limerick, Ireland, 2001.

[15] J. Andean. Space within Space: Report on a Concert. *Presented as part of the Large Project towards the degree of Master of Music.* Centre for Music & Technology, Sibelius Academy. Finland, November 2008.

[16]J. Dattoro. Effect design, part 2: Delay-line modulation and chorus. *J. Audio Eng. Soc.,* 45(10):764-788. October 1997.

[17] J.M. Jot, V. Larcher, and O. Warusfel. Digital signal processing issues in the context of binaural and transaural stereophony. In *98th AESConvention*, Paris, 1995.

[18] J. O. Smith. An allpass approach to digital phasing and flanging. *Proc. Int. Computer Music Conf.* pp. 103-109, IRCAM. France, 1984.

[19]J. O. Smith and S. Serafin. Doppler Simulation and the Leslie . *Proc. of the 5th Int. Conference on Digital Audio Effects (DAFx-02)*. Hamburg, Germany. September 26-28, 2002.

[20] K. Hagan. Textural Composition: Implementation of an Intermediary Aesthetic. *International Computer Music Conference.* Belfast, Ireland. 2008.

[21] M. A. Gerzon. The design of precisely coincident microphone arrays for stereo and surround sound. *Preprint of the 50th Audio Engineering Society Convention.* London 1975.

[22] N. Barrett. Spatio-Musical Compositional Strategies. *Organised Sound,* 7(3):313–323. 2002.

[23] N. Fells. On space, listening and interaction: *Words on the streets are these* and *Still life. Organised Sound.*7, no.3: 287-294. 2002.

[24] P. Dutilleux. Filters, Delays, Modulations and Demodulations – A Tutorial. *Proc. DAFX-98*, pp. 4-11. Barcelona, Spain, Nov. 1998.

[25] P. Fernandez-Cid and F.J. Casajlis-Quiros. Enhanced quality and variety of chorus/flange units. *Proc. DAFX-98 Digital Audio EffectsWorkshop*.pp. 35-39. Barcelona, November 1998.

[26] S. Disch, U. Zolzer. Modulation and delay line based digital audio effects. *Proceedings of the 2nd Digital Audio EffectsConference*. Trondheim, Norway, 1999.

[27] T.I. Laakso, V. Vällimälki, M. Karjalainen, and U.K. Laine. Splitting the unit delay. *IEEE Signal Processing Magazine,* 13:30-60, 1996.

[28] U. Zölzer (Edt). *Digital Audio Effects*, John Wiley & Sons Ltd, 2002.

[29] V. Pulkki. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *J. Audio Eng. Soc.,* 45(6), pp. 456-466. 1999.

[30] V. Pulkki. Spatial Sound Generation and Perception by Amplitude Panning Techniques. *Ph.D. dissertation*, Dept. Elect. Comput. Eng., Helsinki Univ. Tech. 2001.

[31] V. R. Algazi, R.O. Duda, and M. Thompson, D. Motion-tracked binaural sound. In *116th Audio Engineering Convention*, Berlin, Germany, 2004.

[32]V. Välimäki and T. I. Laakso. Principles of fractional delay filters. *IEEE ICASSP'00*, pp. 3870–3873, 2000.

[33] www.soundfield.com