A Dual System for Audio and Image Streams

Fernando Falci de Souza\(^{(1,2)}\), Claudiney Carrasco\(^{(1,2)}\), Adolfo Maia Jr.\(^{(1,2,3)}\)

\(^{(1)}\)Interdisciplinary Nucleus for Sound Communication (NICS)
\(^{(2)}\)Department of Music, Institute of Arts (IA)
\(^{(3)}\)Department of Applied Mathematics (IMECC)
University of Campinas (UNICAMP)

Abstract

We describe here the theoretical concepts as well the proposal of a flexible system (algorithm) for granular sound synthesis from an extension of the Time-Frequency Space, which we have named Extended Space Gabor (ESG), which can be useful to composers and performers of electroacoustic music, electronic music, soundscape designers and artistic installations. This system is able to synthesize soundscapes with granular streams directly integrated to the information content of images or videos. Conversely the system also is able to generate complex images from a library of basic graphics through a construction process controlled by the parameter values of a granular sound stream in real time, that is, based on the information content of the audio. So this is a dual system integrating images and sounds using their information content. We included in the Appendix a prototype program in MATLAB language in order to explore experimentally the ideas shown here.

1. Introduction

We are in the golden era of musical experimentation. The increasingly fast, reliable and immediate communication and the digital technology, both in computing and in sophisticated high-tech instruments, have led to new paradigms in sound and image. Of course, these modern technologies require new models of organization of the both material and informational content, with new challenges to human creativity.

In order to present a model, which we have called a dual system, we restrict our sound universe to the so-called Granular Synthesis.

Granular Synthesis is a digital technique of music composition whose mathematical foundation is inspired on the theory of the Acoustic Quantum proposed by the Nobel Prize physicist Dennis Gabor [Gabor, 1946, 1947]. This theory shows that sounds of any complexity and duration can be decomposed into a set of sound units named "grains of sounds" or "granular sounds," so that, theoretically, the original sound can be obtained through some re-synthesis process. Gabor showed that, formally, any sound can be represented mathematically as a set of “acoustic cells”
with simple spectral content and length of about 10 to 200ms [Gabor, 1946, 1947]. The decomposition of original sound in terms of sound grains has the same paradigm of the Fourier decomposition in to sine waves with different frequencies, that is, basis of a Vector Space, in our case the Space of Sounds accessible to the human ear. Gabor's theories have inspired composers to perform the reverse process, that is, composing sounds, timbres and textures from the manipulation and sequencing of a large amount of sound particles of very short duration [Thomson 2004, Roads 1988].

The Greek composer Xenakis was the first one to use the term microsom to refer to musical works in which the composer controls the sound material in microtime as, for example, in his work "Concrete PH". In this composition, Xenakis reorganizes small pieces cut from a magnetic tape containing a recording [Xenakis, 2001]. Stockhausen also composed some works of granular nature as, for example, "Kontact", in which he uses analog pulse generators to create sound grains [Tissot, 2008].

Although it is possible to compose in microtime using analog technology or even acoustic instruments, granular synthesis is more idiomatic in the digital domain. Here we highlight the contributions of Curtis Roads. From the 70's, Roads conducted many musical experiments with granular synthesis creating a comprehensive taxonomy for classifying both types of grains as well as how to organize them. In his famous book Microsound, Roads presents the history and technique of granular synthesis resulting from his work and of other researchers by the year 2000 [Roads, 2001]. Barry Truax is another major contributor in the area, and has been a pioneer in experiments with granular synthesis in real time. Truax presents also an extensive use of granular synthesis in the composition of soundscapes [Truax, 1988, 1990].

Granular Synthesis, used either to obtain new sounds, musical compositions or
soundscapes, presents often the same challenge. Since the basic material of composition are sound particles of very short duration, thousands of these particles are necessary to compose a musical segment. Specifying values individually for each of the particles is a tedious task and often human impracticable. To overcome this difficulty, it is common in our digital age, to employ a mathematical or algorithmic model that realize an abstraction of sound synthesis parameters, that is, a high-level model allowing the user control the synthesis of a large number and variety of grains of sound through a small set of macro parameters. DiScipio’s models control the synthesis parameters via Fractals Models, Dynamical Systems and Chaos Theory [DiScipio, 1990], Maia and Miranda used Fuzzy and Markov Chains [Maia and Miranda, 2005]. Graph Theory has been proposed as a model of control by Lombardo and Valle [Valle, 2003], and Genetic Algorithms were used in one of our systems named EVOGrain [Souza et al, 2009].

The use of sound flows generated by Granular Synthesis synchronized with video has been experienced in audio-visual installations as, for example, the system "Modell 5" [Langheinrich, 1994]. Despite the strong impact this has on the public, it is a very simple system that provides few compositional resources: the system reads a video file associated with an audio file of equal size, the interpreter can choose the speed reproduction, determine loops in specific sections of the file and control the effects of "time-change" and "pitch-change." The system allows four different sections of the file to be played simultaneously.

Recently we developed an application for Sound Synthesis named GranularStreamer which synthesizes in real-time an arbitrary number of independent streams of granular sounds whose synthesis parameters are coordinates of an Extended Gabor Space. In other words, the EGS is a 6-dimensional parameter space
whose variables are frequency of the waveform, grain time duration, the rate of grains emitted per second, the offset of a wavetable, amplitude and stereophonic position of the grain. This system has the additional characteristic of real time control by external devices for gesture control [Souza et al 2010].

Together with GranularStreamer we also developed Java classes and many applets for the visualization of the granular stream named GranularDrawer. This program produces graphical objects, animations and video content automatically from a flow of granular sound that comes from the GranularStreamer. Figure 1 below shows an example image generated in real time by GranularDrawer for a cloud of grains of sound which duration is decreasing.

![Image generated in real time by GranularDrawer](image)

Figure 1: Image generated in real time by GranularDrawer

On the other hand the inverse is also an important task, that is, creating sound streams directly from images content or from a sequence of images (movie). This is commonly named sonification. Automatic sonification can be very useful for videos that would require a long time for an audiovisual artist if he/she were composing manually and the audiovisual result can present a strong correspondence between sound and image. Since Granular Synthesis provides a diverse range of timbres, our proposal can even be feasible to construct audiovisual landscapes, sound environments, art installations or even to a laboratory for multimodal composition.
2. The dual system: Algorithm and MATLAB Scripts

2.1 Sonification

We have written a Matlab script to construct a first prototype of our dual system. MATLAB is a mathematically (numeric) oriented language which has great computational power as well flexibility enough in order to made some experiments testing the limits of our dual model, visualization of sound and image sonification, which are mappings, and their inverses, from an Extended Gabor Space (EGS) to an Extended Image Space (EIS). Since we are working with granular sounds, that is the most basic unit of sound perceptible by the human ear, we can work with the same concept for images in the digital domain, that is, the pixel. This marriage, although interesting and desirable have shown itself as a failed attempt in our experiment. This is due to the huge mass of data to be processed for one image, even in Matlab with its vector fast operations and in powerful laptops equipped with the last generation of processors. For example, an image with 400 x 300 pixels in RGB color needs a matrix with dimension 400 x 300 x 3, or 360,000 data. If we assign a granular sound with the duration of 50 ms for each pixel, with a sample rate of 44.1 K/sec, we need 2205 samples for sound data. Therefore, in principle, to sonify each image with grains we need to calculate 360,000 x 2205 = 793,800,000 sound data. In our experiments, even with a i7 chipset MacBook Pro laptop, the processing really gets terribly slow. The solution is simply do not demand such a processing task by reducing the quantity of information extracted from the image RGB matrix which is in turn mapped to the granular sound stream. In certain sense we are just giving up from the “atomization” of images but not the sounds. So the general idea is to slice the image matrix in a number of small sub-matrices and get “simple information” from each one of them.
For example, we can use just one sound vector for each sub-matrix. This can be accomplished through several mathematical/algorithmic devices. For example we can take just the central pixel of a square sub-matrices (with odd dimension). In fact these sub-matrices does not need to be rectangular. They can be thought as a set of vector data from a small region of the image. In this way it is possible to shred an image in a number of small different regions and associate a sound data vector to each of them.

In our MATLAB prototype experiment we run a sequence of some tens of images (for example: the moon) and associated few sines waves per image, taking some points of the images randomly, which are in turn used to construct granular sounds with additive synthesis. Clearly, it necessary to optimize the code in order to achieve better results. This is important since we intend, for example, save a movie as a sequence of digital frames and proceed the same way to generate granular sound streams associated to images information content.

2.2 Visualization

In order to increase the correlation between audio objects and visual objects it is important to find a dual relation between the mediums. In most audiovisual systems and performances that can be seen today is possible to note a uni-directional relation, either from images to sounds (sonification) or from sounds to images (visualization), and this systemic limitation also results in a preponderance of labour in one of the modalities, that is, the composers focus his attention on the visual domain leaving giving music secondary importance, or the opposite, the composer focus on the music keeping the graphical content as accompaniment.

As well as it happens in the process of sonification of images, the task of analyze a stream of audio, extract parameters from the analysis and use the parameters
to synthesize graphics in real-time requires an enormous computing power. In our case, where we are using the GranularStreamer to generate streams of grains, the sound generation is happening under control of an algorithm. So instead of having to analyze the stream of sound we can set the GranularStreamer to send information of its parameters and about every grain of sound that is synthesized.

This makes possible to control the sound streams in real-time and have a visual feedback with very small latency following a relation of one basic graphic unit for one grain of sound. Although many studies tend find a physical relation between sound and image, other have proved that the audiovisual correlation is purely an aesthetic choice [Abbado, 1998]. So far, in our experiments of generating visualizations of granular synthesis we are drawing circles to represent grains of sound composed by sine wave and gaussian envelope, in a way that the circle location, size, hue and brightness are related to the grain stereophonic position, time duration, waveform frequency and amplitude.

3. Conclusion and Future Work

To our actual knowledge we don’t know, in the specialized literature or in the Internet, about any other dual system design to play this game between image and sound using their mutual information. The dual system presented above is intended to be flexible and simple one. Clearly there exist many different ways to associate image to sound. Our system takes this task to the “atomist level”, that is, pixels for images and granular sounds in the Gabor’s sense. So it has a great potential to construct audio/video performances starting from the scratch.
a) Gesture Control

An interesting extension of our work is to explore gestural controllers mapping strategies appropriate for this intermodal system (audiovisual) allowing the user, composer or performer to interact intuitively and manipulating elements of both sound as well as graphics.

b) Phonemes Recognition

Another interesting application of our system is the recognizing of human phonemes from the speaking image. This could be useful, for example, for direct and real-time lip reading.

c) Syntactic and Semantic Mappings

The mappings we have shown above for sonification are of syntactic type, that is, they maps pixels info (RGB matrix) as well formal operations on images (or frames) to the audio waveform vector. So the numeric values of audio waveform are related to image RGB matrices by a convenient function. These functions transfer coded information of the very structure of images to an associated structured audio file. Nevertheless no information is transmitted on the meaning of the images. A function with this additional characteristic is a sonification of semantic type. For example suppose we have images of beaches, which we semantically associate with some kind of violin glissandi. In presenting a new image of a beach, the system should be able enough not only to decode in sounds its “syntactic structures” but also to map it to convenient violin glissandi. In this sense similar images should be mapped to similar (near) sounds. This implies to define a suitable distance on the image sets as well on the audio sets.
4. Acknowledgments

This work was supported for a grant from CNPq (Brazilian Agency for Research) under the process 480.273/2010-3, Projeto Universal. F. Falci thanks FAPESP (Fundação de Amparo `a Pesquisa do Estado de São Paulo) for a Dr. Sc. Fellowship and Santander Universities for financial support to a short term visit at IDMIL (Montreal Canada). A. Maia and N. Carrasco thanks CNPq for their research fellowships.

5. References


