COMPOSING AUDIO-VISUAL ART: THE ISSUE OF TIME AND THE CONCEPT OF TRANSDUCTION

Benoît Courribet
CICM – University Paris VIII
MSH Paris Nord
4, rue de la Croix Faron
Plaine Saint-Denis
93210 Saint-Denis

ABSTRACT

In this article we study the relations between images and sounds in the field of crossed-media composition. The article aims at giving some elements of reflection considering the issue of time.

The notion of temporal object is exposed, and so are the mechanisms driving the perception of these objects.

The issue of time is then discussed through the study of relations between audio and visual media.and several examples of strategies for mapping are exposed. Finally, the concept of *transduction* is presented.

1. INTRODUCTION

Both video art and music can be seen as "time-based" art. Technological convergence makes it possible to envisage the creation of cross-media compositions and domains such as visualization of sound are currently being explored.

In this paper we will focus on the artistic process of audio-visual art composition. Though we aim at considering the audio-visual art piece as something more than a composite piece, we will remain interested by aspects and problems with which we are familiar in the musical domain.

We will first describe the temporal nature of audio and video media, then address the issue of time in sound-visual relations. Finally, after presenting some strategies of mapping between audio and video, we will envisage the *mapping* process through the prism of *transduction*.

This paper is part of the research of the "Visualization of Sound" workgroup [1] supported by AFIM [2].

2. TEMPORAL OBJECTS

By analyzing the perception of temporal objects, philosopher E.Husserl aims at studying the temporal nature of consciousness itself [3]. Temporal objects are streams flowing in time. Consciousness of these objects is also embedded in time, as both streams synchronize. These objects include sounds, notes, melodies and, of course, all kinds of audio and/or visual media.

Discussing the consciousness of temporal objects, Husserl presents the process of *retention*: the perception of temporal objects retains what has just been, what has just passed, like the tail of a comet attached to the present. He then distinguishes *primary retention* (perception) with *secondary retention* (imagination). The later has to do with bringing back to memory a past temporal object. Although different by nature, these two mechanisms are interdependent, and form the basis of our consciousness of temporal objects. Husserl also presents the notion of *protention*, which can be thought of as anticipation, like *retention* but directed towards the future instead of the past.

3. ARTICULATING IMAGES AND SOUNDS

Before considering relations between sounds and video, we should focus on studying specificities of these media related to time. We will then present some examples of strategies of mapping.

3.1. The issue of time

Sounds don't exist out of time; they are, by definition, due to a variation of air pressure in time. On the contrary, even if our visual perception is embedded in time, a pause in a video stream, a still image, still carries some visual meaning. Moreover, our perception of audio and video streams is different when their temporal structure is modified: we would still see the same thing if a video was played at one tenth of its original speed, while a sound wouldn't be recognizable. The quantitative operation of slowing down the audio stream is responsible of a qualitative change in one's perception of it. We can also point out the discrete nature of video, which makes it hard to slow down a video stream, since the parameters of discretisation (i.e. the number of frames per second) are close to the limits of our perception of visual continuity.

In the domain of visualization of sound and/or music, it is interesting to point out the issue of the visual representation of time. For most graphical representations of sound, time appears as an explicit dimension of it. Depending on the portion of time represented, our perception of sounds or higher-level structures (mesostructures: scale of groups, phrases composed of sound objects, macrostructures: scale of a whole composition [4]) can be affected, as the processes of *retention* and *protention* described earlier are being diturbed. For example, looking at the waveform of a piece of music while listening to it, makes it possible to anticipate future evolutions, and thus modify our perception of it.

Finally, when considering the possibility to synchronize the audio and video streams, we are confronted to the problem of time scales that appear to be different for audio and video. Hoarcio Vaggione [5] and Curtis Roads [4] have provided extensive reflections and research on the subject of micro-time scales for sound synthesis, sound processing and music composition. They have showed that going below the threshold of a millisecond is possible and can be a powerful tool for timbral and texture-based sound design and composition. On the contrary, it won't be possible to obtain such accuracy with the video medium, though other strategies than strict synchronization can be considered (mapping micro-time evolutions with graphical textures, for example).

3.2. Strategies of mapping

We have decided to concentrate on synchronous mapping between audio and video in order not to disturb the processes of *retention* and *protention*. It means that events should happen simultaneously in the video and audio parts. We will discuss further the possibility and interests of asynchronous mapping.

Regarding the many possibilities of setting sound/visual relations (ranging from a music score to a video clip or a waveform), we have chosen to focus on two schemes. These two schemes are interesting because of the formal coherence induced by the sharing of data between both parts. Whether these schemes are used in real-time or non-real-time situation will not be discussed here.

The first one consists in performing an analysis of the audio stream (pitch tracking, onset detection, envelope following…), then mapping the results of the analysis with visual parameters.

For example, the *loudness* parameter from left and right channels of a stereo track is used to control the zoom ratios of the left and right region of a video clip.

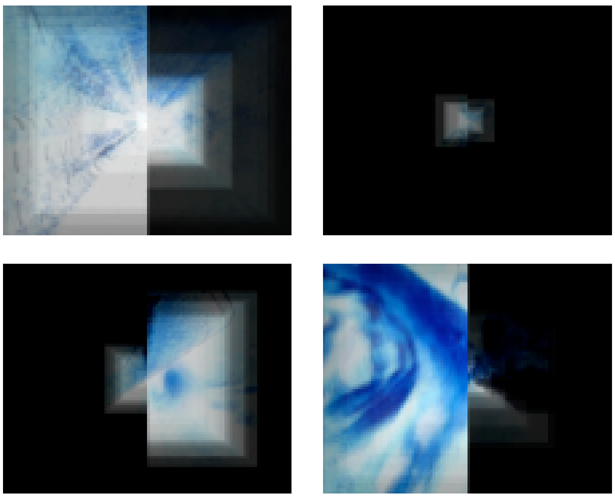

Figure 1: mapping *loudness* to zoom ratios

On this second example, onsets are detected and mapped with a reconfiguration of the geometry of a "flattened 3D" OpenGL scene.

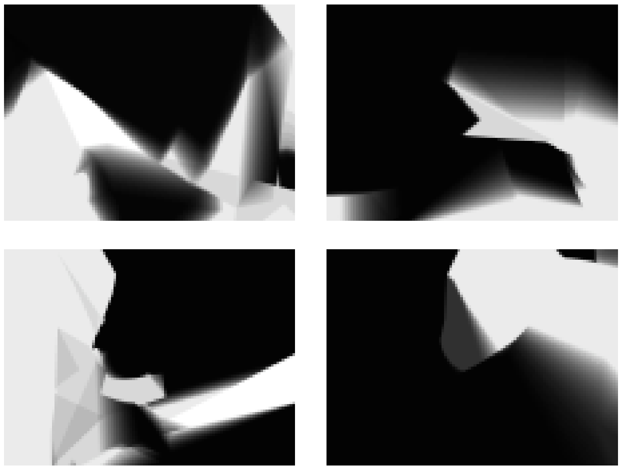

Figure 2: mapping onsets with 3D geometry configuration

The concept of *mapping* is fundamental here: a visualization system will or will not be efficient according to the way that data is shared. Because our perception of sound parameters (frequency, loudness) is non-linear, transfer functions are often used to achieve a perceptually relevant mapping. This scheme will be qualified as "sequential".

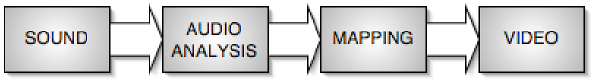

Figure 3: sequential scheme for sound/visual relations

Though this approach can provide efficient visualizations of sound in certain cases (simple structures, minimal sound materials…), the video part is based on the whole audio stream, and therefore is limited to a global relation. For example, it is almost impossible to refer visually to individual sounds.

We will now present another approach, which can achieve higher-level interactions between music and video.

We will refer to *intentional parameters* for parameters that preexist the sound: for example, the graphic shape of a volume envelope to be applied to a sound, or the times at which sounds will be triggered. These parameters could also be used (after mapping) to process and/or generate video. Thus, we can consider that both the audio and the video part are realizations of high-level instructions. This is the main idea of what we will call a "parallel approach".

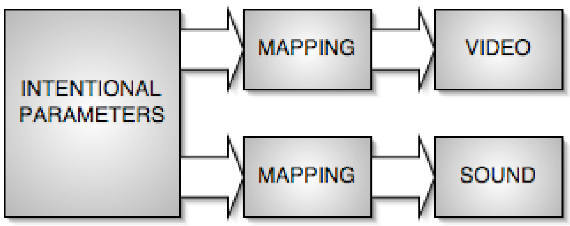

Figure 4: parallel scheme for sound/visual relations

According to the above figure, these intentional parameters will be used both to generate/process audio and video. For example, the author used an audio/video granulation engine for the real-time audiovisual Isol piece with percussionist Julien Tardieu: each time a grain of sound was produced, a 3D structure was deformed on screen.

Here, a real-time analysis would have failed to detect each grain generated by the granulation engine (moreover when grains overlap). We can also imagine using this approach in order to give a visual representation of higher-level structures: mesostructures or even macrostructures

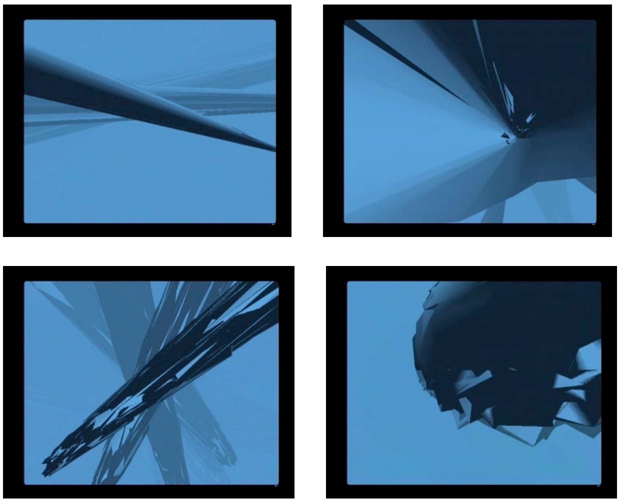

Figure 5: stills from the Isol piece

In the context of interactive composition or multimedia installations, we have also proposed [6] an interface based on terrains of parameters. The *intentional parameters* are organized in 2 dimensional tables (terrains). These tables are then used as textures mapped on a surface (ground) in a virtual reality-like environment (the value of the parameter is represented by the intensity of the color of the terrain).

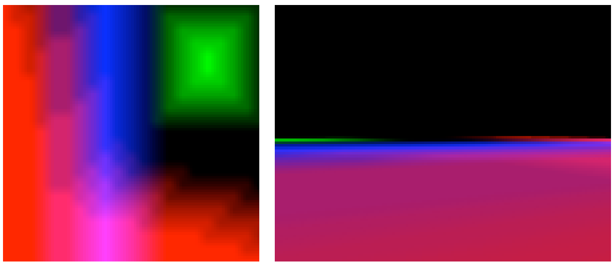

Figure 6: using terrains as control interfaces

The audio part of the piece is realized by the actualization of the parameters to the values corresponding at the virtual location of the user, while the video part consists in the trajectory of the user in the 3D environment.

4. CONCLUSION: TOWARDS THE CONCEPT OF TRANSDUCTION

In this article, we have presented reflections on the topic of sound and visual relations, always considering the issue of time. Some strategies of mapping have been presented. However, these strategies show limitations, due either to the temporal characteristics of the video and audio media, or to the paradigm of the strategy used.

On one hand, it should be very interesting to be able to extend the mapping possibilities to other time scales than the sound object's. On the other hand, using asynchronous mapping techniques would open the door to new possibilities for "playing" with the *retention* and *protention* processes.

Considering that the *mapping* process is foundational for cross-media composition, envisaging it as a *transductive* relation is an interesting idea in order to overcome these limitations.

Transduction usually refers to the process of converting a type of energy into another. However, Gilbert Simondon has proposed, in his studies on *individuation* (the dynamic process of construction of a system) [7], the concept of *transduction* as the mechanism driving the process of *individuation*, it is defined as a relation, both connecting and structuring two terms. The terms do not preexist the relation,. They are *individuated* and structured by it, due to its *transductive* nature.

The audio and video parts of the cross-media composition would be structured dynamically by the *transductive mapping* relation. "Dynamically" do not refer here to the time and duration of the piece, but to the *individuation* time of the composition.

5. REFERENCES

[1] http://gtv.mshparisnord.org/

[2] http://www.afim-asso.org/

[3] Husserl, E. *Leçons pour une phénoménologie de la conscience intime du temps*, PUF, Paris, 1964

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

[5] Vaggione, H. "Articulating micro-time" *Computer Music Journal 20 (1),* MIT Press, Cambridge, 1996

[6] Sedes, A., Courribet, B. & Thiebaut J.-B.: "Visualization of sound as a control interface", *Proceedings of DAFx 2004*, Naples, 2004

[7] Simondon, G., *L'individu et sa genèse physico-biologique*, PUF, Paris, 1964