DYNAMIC INTERMEDIATE MODELS FOR AUDIOGRAPHIC SYNTHESIS

Vincent Goudard
LAM - Institut
Jean Le Rond d'Alembert
vincent@mazirkat.org

Hugues Genevois
LAM - Institut
Jean Le Rond d'Alembert
genevois@lam.jussieu.fr

Émilien Ghomi
LRI – Université
Paris Sud 11
emilien.ghomi@lri.fr

Boris Doval
LAM – Institut
Jean Le Rond d'Alembert
boris.doval@upmc.fr

ABSTRACT

Mapping is one of the most important aspects of software instruments design. We call "mapping" the relation defined between the parameters from hardware interaction devices, and those of the process to be controlled. For software instruments, this relation between the user's gestures and synthesis engine parameters has a decisive role in resulting ergonomics, playability and expressive possibilities of the system. The authors propose an approach based on a modular software design inspired by a multidisciplinary study of musical instruments and their playing.

In this paper, the concept of "Dynamic Intermediate Models" (DIM) is introduced as the centre of the proposed architecture. In such a scheme, DIM modules are inserted between the gestural interfaces and the audio-graphic synthesis and rendering engines. The concept of DIM is presented and explored as an extension of usual mapping functions, leading to an improvement of the interaction between the musician and his/her instrument.

Then, design and programming guidelines are presented, together with some concrete examples of DIMs that have been created and tested. Finally, the authors propose some directions to evaluate such DIMs in the architecture.

1. INTRODUCTION

1.1. Background

The invention of telephone and phonograph, almost contemporary to each other, disturbed our "traditional" relation to sound, voice and music, by allowing transmission and recording of sounds. From then, sounds could "travel" through space and time. Developments of electrical devices and later digital technology increased even more the "distance" between the body and the instrument in music production:

- electricity, by bringing new energy, previously mechanical, to machinery and instrumental devices;
- digital technology by operating a radical decoupling between musicians' actions and effects on the instruments due to symbolic information encoding and processing.

The consequences of these decoupling on musician-instrument interactions have been studied extensively by Cadoz [3] in particular. Without repeating these discussions here, it is crucial to remember that these technologies induced a new definition of social value: from a world where work is highly regarded to a world based on information. Such an evolution has artistic, cultural and social irreversible consequences. Therefore, it is not surprising that, alongside these developments, new artistic sensibilities, new philosophical and scientific paradigms have been invented and experimented.

New questions and new musical devices emerged at the same time, mentioned by several computer music studies [1]. What is an instrument? What is inherent to its nature? How could we create repertoires if instruments keep evolving? How does an interface get its "instrumentality"? How to play yesterday's, today's, and tomorrow's music with these new instruments ?

Usually, research on this topic mostly focuses either on the technical characteristics of the devices, on the software underlying them, or on the sensory-motor interaction between the musician and the "interface". Indeed, technical aspects must be taken into account when studying the "instrumentality" of a new device but, as a counterpart, it can be useful and productive to work on a human-centred approach taking into account the cultural and social aspects of the interaction between some subjects and these new devices.

For acoustic instruments, these relations between input and output are complex and intricate, since they had been refined for centuries, in a co-construction between the morphology of the instrument, the ways of interacting with it, the available sound, and the associated repertoire. For digital sound engines, nothing is actually framed, everyone can code his own implementation of a synthesis algorithm, build his own device to control it, and have to draw links between data coming out from the interfaces and the synthesis parameters.

1.2. Instrumental interaction

In the field of Human-Computer Interaction (HCI), Instrumental Interaction is an operational interaction model introduced by Beaudouin-Lafon [2]. Within that definition, an instrument is a two-way transducer between the user and the object he wants to act on. For Cadoz, who specifically studied musical instruments, its role is to tackle the transformation from gesture to sound in real-time. During the interaction between the musician and the instrument, “specific phenomena are produced, the behaviour and dynamic evolution of which can be mastered by the subject” [3]. Then, he identifies three types of instrumental gestures : excitation gestures, modification gestures and selection gestures.

But, as it has been pointed out by Cance et al. [4], defining precisely what is a musical instrument is a difficult task. Some of the acoustical instruments needed decades to become “instruments”. It depends on cultural and social aspects, since an instrument has not only a functional role, it has also a symbolic one. Then, instrumental quality of a device is not only depending on intrinsic qualities but constructed, through musical play, in a given cultural context.

1.3. Musical gestures, temporal and body scales

At first sight, it seems possible to distinguish the subsequent four phases in which a “musical” gesture is involved [9].

- Composition : production and writing of musical structures (not in real-time).
- Instrument making : building of the instrument and preparation in order to be played (tuning, equalisation, settings and adjustments).
- Playing : production of sounds in real-time. Often, only these actions are considered as musical gestures.
- Listening : perception and interpretation of the acoustic environment.

One can legitimately question the need to give to musical gesture such a broad definition. However, the validity of the categorisation presented above is not proved. The four phases are mostly overlaid, partially erasing the boundaries between categories of actions that seem a priori easy to state. Even when staying in a strictly classical scheme, the distinction established above is not as clear and obvious as it appears. In addition, through new forms of artistic creation (electroacoustic music, interactive music, etc.), the last decades have shown that this classification was fragile.

1.4. Origin of the concept of Dynamic Intermediate Model

For several decades, a lot of researches have focused on human-machine interfaces [2] and the optimisation of layers and mapping methods, although the exact contours and boundaries of the concept of mapping remains unclear. Thus, comparative studies on different types of mapping have been published [11], defining categories (one-to-one, one-to-many, etc.).

In general, we can consider that a classical, static, mapping is reducible to a matrix operation: multiplication of an input vector (the values directly or indirectly provided by the gestural controllers after linear on non-linear scaling) by a matrix describing the mapping function. This operation provides the vector used to generate the control parameters of the synthesis.

Then, the type and characteristics of the mapping is linked to the properties of the matrix. Is it symmetric, diagonal, triangular ?

Other approaches are possible, among which the methods based on gesture recognition techniques (neural networks, hidden Markov models ...) which do not need to explicit the link between gesture and synthesis.

In our case, we followed a different and complementary direction, assuming that a living and coherent synthesis requires too many parameters and evolves too rapidly to be entirely controlled by the musician.

2. DYNAMIC INTERMEDIATE MODELS

2.1. DIM characteristics

In recent years, research projects going in the same direction have been identified, sometimes in conjunction with convincing artistic achievements[1]. This encouraged us to consider the generalisation of these ideas and to propose a completely modular software architecture.

The goals of using DIMs are of different natures and lead to the description of functional and structural aspects. Without being exhaustive, however, we discuss some of the essential characteristics they should show :

- to enrich the sensed gesture with modulations operating at frequencies beyond human gesture frequencies (e.g. fast bounces)
- to control multiple parameters from a single input gesture, since lively audio synthesis often requires more parameters than what one can manage in real-time
- to be easily integrated in a modular software architecture, allowing on-the-fly practice and instrument setting.
- to provide a range of modules corresponding to various needs related to the different gestures involved in music at different time scales (composition, setting, playing ...)
- to offer several levels of monitoring and drive simultaneously various synthesis and rendering engines with the same algorithms, in order to improve the coherency of their output
- to be bi-directional: the DIM should be able to communicate in either direction with the interface, other DIMs or the synthesis engine (cf. Fig. 1) in order to regulate the (non linear) interactions between the different stages.

In a metaphoric way, the role of a DIM in a synthesis process can be compared to the action of a bow, or a piano mechanism, etc. and the modular software architecture enables to strike a violin string with a hammer, or to bow a piano string.

[1] Notably, works by Jean-Michel Couturier [7], Cyril Henry [15] or Mathieu Chamagne.

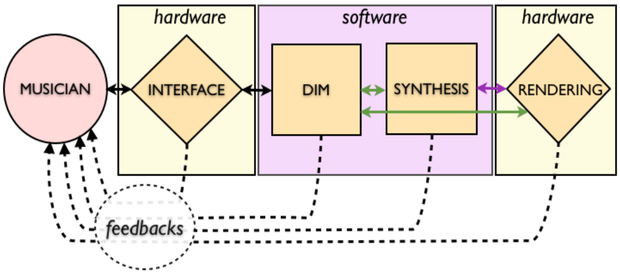

Figure 1. Overview of DIM mapping

2.2. DIM typology

As previously defined, the concept of Dynamic Intermediate Model is, by construction, very large. It must possibly be implemented in various musical tasks envisaged, which result in sound and musical processes at different scales, depending on musical task and gestures. For this purpose, and also in order to provide a wide range of interactions and processings, DIMs can be based on models developed in various scientific fields, enabling very different behaviours. In the OrJo Project[2], we initially focused on three types of DIMs :

- "physical" models
- "topological" models
- "genetic" models

2.3. DIMs as movement generators

The processes at work in a DIM operate on several aspects of the movement transduction expected in a musical instrument.

- **quality of motion :**

To give an acoustic example, a plectrum for a string instrument provides a particular pinch quality and its attack sounds different from that of the fingers. Also, the strings position on the instrument body allows to transform the seemingly linear movement of the hand in a variation of this particular movement, namely a succession of pinches which rhythm is related to the spacing of the strings. The interaction devices are often devoid of surface roughness (for instance a pen-tablet has not the roughness of the horsehair of a bow). A DIM will need to integrate this "surface aspect" of the virtual object.

- **non-linear movements :**

Richness of sound is partly related to the overall nonlinear phenomena in action in a musical instrument and contribute to make the sound rich and subtle. These non-linearities partly due to materials in acoustic instruments (and electronics) often lack in the standardised digital values of software instruments parameters. In the digital realm, non-linearities are actually often found in bugs and codec misuse, all kind of "glitches" that are precisely sought by a vein of electronic music now baring this name, to produce rich sonorities and unexpected soundscapes [5]. The DIM should thus re-introduce saturation, exponential curves, distortion, and other jitter in the transfer functions of digital instruments.

- **Augmented movement :**

By acting on complex models, a one-dimensional variation can be converted into a multi-dimensional polyphonic movement. For example on a string instrument, the many notes of a chords can be played with one touch. This enhancement of movement may act in vertical (poly-phonic), horizontal (poly-rhythmic) dimensions, or on the many dimensions of timbre. We may for example control a "target parameter" that a set of elements would reach through a displacement-logic of their own.

2.4. Mental representation of DIMs

As a computer process, the algorithm may remain abstract for the musician who needs a mental model to predict the outcome of his/her action (ideally "sing" what he/she plays) and avoid the cognitive overload that involve the direct handling of too many parameters. However, this mental model is not so much an intellectual understanding of each of the processes at work, but rather a "body understanding" of the manipulated object, made up of the interaction device and the algorithms that are as one [18]. Rather than "input parameter", we thus prefer the term "handle", or "action item", to describe how we "catch" the virtual object being manipulated.

3. DIM IMPLEMENTATION

We present the implementation of two DIMs in the Meta-Mallette, trying to give a concrete -albeit limited- idea of the use of DIMs as a tool for experimental digital instruments making.

3.1. Meta-Mallette

Meta-Mallette[3] is a software environment for playing computer-assisted music, images (and more) in real time [14]. It allows to load virtual instruments playable with joystick-like interfaces, pen-tablet or more expert interfaces such as the Meta-Instrument[4].

Until now, the instruments developed for the Meta-Mallette directly included all parts of the whole process between the player and audio/graphics rendering. The latest Meta-Mallette version now enables to separate these various elements and reconnect them differently, opening the way for further experimentation, notably on mapping. The work carried out at LAM during the OrJo project precisely aims at developing elements that will ease empirical testing required for any music instrument design.

[2] OrJo is a research project funded by the FEDER and Ile-de-France Regional Council. It gathers the following companies and research labs : Puce Muse, LAM, LIMSI and 3DLized. It aims at developing audiovisual instruments for collective artistic performances.

[3] Meta-Mallette is available for download on Puce Muse website .

[4] The Meta-Instrument developed at Puce Muse is pluggable in the Meta-Mallette, and offers 54 independent sensors sampled every 2ms.

3.2. DIM « Roulette » and « Verlet »

We will try to present an example of the use of DIM in the Meta-Mallette[5]. Let's suppose a setup consisting of a hardware interface, two cascaded DIMs and a Karplus-strong synthesis module.

3.2.1. DIM example 1 : Roulette

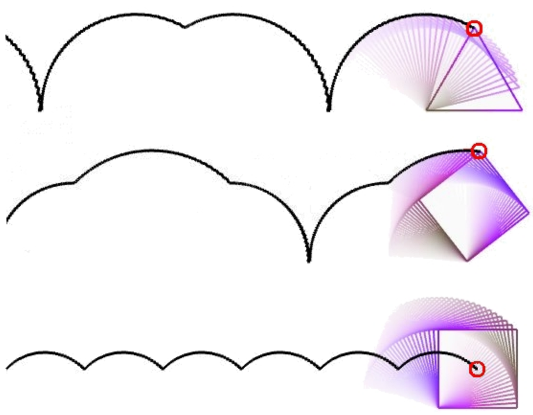

Figure 2. Roulette trajectories engendered by polygon motion.

The first DIM named "Roulette" is a geometric model inspired by Pascal's roulette[6], a regular polygon that can move the following ways :

- tip on one of its corner
- sliding along one of its sides
- pendulum around an axis

These motion primitives have parametrized temporal evolution curves that will change the quality of induced movement. For example, the sliding motion will look like a fall, an burst, or a simple displacement according to the linear or quadratic speed.

Although resembling a physical model, this model differentiates itself by movements which time periods are tightly defined, allowing a rhythmic play to observe a beat. Moreover, the absence of pseudo-physical constraints allows to escape the apparent "rule" anytime.

The handles of the Roulette instrument are:

- the target's position that the polygon will follow
- the selection of movement type to be performed
- the diameter of the polygon and the number of sides
- its position and orientation
- the duration of each movement type

The reaction points (output parameters) of this DIM are:

- the diameter of the polygon and its number of faces
- its position and orientation
- the ongoing progression of movement
- the end of the movement

[5] Interested readers will, however, better look into actual examples available on the LAM webpage. http://www.lam.jussieu.fr/orjo/

[6] Roulette is a generalisation of the cycloid, named after the Treaty of Roulette written by Blaise Pascal in 1659.

3.2.2. DIM example 2 : Verlet

The 2nd DIM named "Verlet"[7] is based on the pseudo physical Verlet algorithm. It can simulate a structure of points connected by elastic links. Here, this skeletal model is contained in a box with which it can collide.

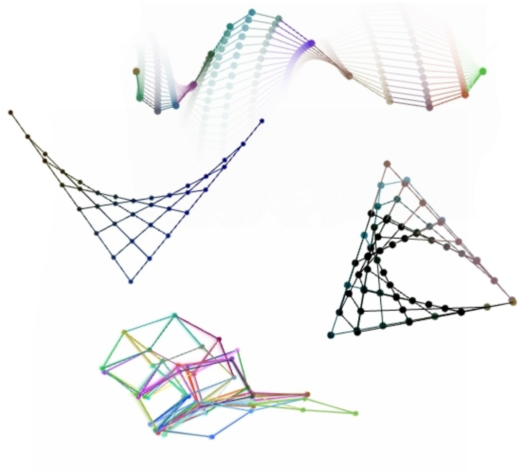

Figure 3. Various Verlet model states

The action points of the Verlet instrument are :

- the position is the orientation of the box
- the length and width of the box
- the length and stiffness of elastic links
- the size and direction of force applied to the model points

The reaction points of the “Verlet” DIM are :

- a matrix containing every point's position
- a matrix containing impact velocities

3.3. DIM interconnection

In the chain of interaction mentioned above, the player acts on the Roulette model which transforms the nature of his/her movement : by moving the Roulette's target (e.g. with a pen tablet), the model reacts with predictable behaviour, yet a behaviour of its own.

Its movement could be directly used to drive the sound synthesis algorithm. Yet, in this example, it is re-used here to drive a 2nd DIM : the position and orientation of the Roulette's polygon controls the position and orientation of the Verlet’s box.

Thus, percussive movements caused by a rocking motion of the Roulette polygon generate a multitude of micro-movements generated by the Verlet algorithm "shaken" by the Roulette motion (fig. 4:1). The instrument-player can thus control with a simple gesture an intuitive process that complexify the (sensed) gesture and generates a complex set of movements.

[7] This numerical integration scheme was developed in 1967 by physicist Loup Verlet. Andrew Benson made an implementation for Max/Jitter, available on http://cycling74.com.

Let's notice that the chaining of these two modules could be done in reverse order, i.e. by controlling a Verlet model which edges will act as a multitude of targets for the Roulette model (fig. 4:2). We could as well connect another Roulette model to the shaken Verlet (fig. 4:3), or make any other connection between DIMs, combining several movement scales.

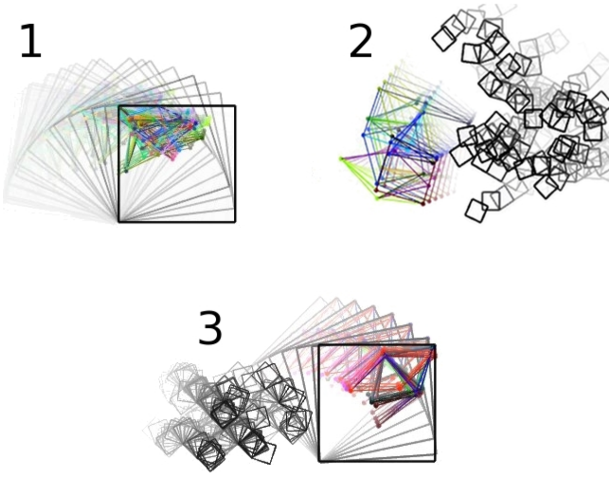

Figure 4. DIMs interconnections :
1 : Verlet edges controling Roulette's targets,
2 : Roulette's polygon controling Verlet's box,
3 : Roulette's polygon controling Verlet's box, which edges control in turn other Roulette's targets.

3.4. Audio synthesis

To sonify the many sources of movement generated by the DIMs, we adapted known synthesis algorithms (Karplus-Strong, FM, granular) to enable the use of matrices to control many synthesis parameters in parallel on multiple voices.

For this purpose, we have jointly developed tools to map values between the input matrix and the synthesis parameters that allow various polyphony strategies (e.g. distribution, interpolation, per-voice detune) in case the number of movement sources and audio synthesis voices do not match. Conversely, various parameterized grids and functions allow ergonomic scaling between movement data and synthesis parameters, easing the task of using traditional music theory rules (tonality, rhythmic division).

The synthesis algorithm used in this example is based on the (plucked string) Karplus-Strong algorithm [12, 13]. We included a "make up gain" in the feedback loop to compensate for the damping function the fundamental frequency, thus allowing to set duration and harmonic filtering independently, and a non-linear pinch parameter which enrich the sound with transients and harmonics. The Meta-Mallette environment allow to plug the data output by the interaction device, the Verlet model, the Roulette model, or any combination of them. This choice of inputs combination to final parameters of sound synthesis is an essential aspect of the adjustment required for the empirical instrument making process.

3.5. Graphical rendering

For now, we have focused our research on graphical representations that help understanding the model and mainly performing a monitoring task. Namely, the representation should help understanding the behaviour, predicting trajectories, by possibly showing the inner assembly. Indeed, DIMs are not necessarily intended to be viewed, especially if they are included in a larger whole. Moreover, there is more than one possible representation for a DIM. Modules providing the graphic synthesis have been separated from the algorithmic model in order to change the visual representation, or simply not use them.

As generators of movement, the output of a DIM may be used for purposes other than audio and graphic synthesis control. In particular, they can drive external elements such as acoustic devices they will excite, motors, light system, or to produce a force feedback. The possibilities are numerous and likely to exceed the framework of the developments undertaken in the current phase of our research.

4. CONCLUSION

The instrumentality of these new devices, as well as of "classical" instruments, does not result from their intrinsic properties only. It is constructed through music playing, interactions between musicians and the design and development of the instruments. The evolution and enrichment of the concept of instrumentality raised by these new practices imposes a pluri-disciplinary approach for the "science of instruments".

However, the OrJo Project, within which this research takes place, involves developing new instruments and also to gauge interest and richness, from an aesthetic and artistic point of view (creation and pedagogy) and as interaction models that could be generalised to other areas of Human-Computer Interaction. By the end of the project, one or more assessment methods should be developed.

4.1. Evaluation of the DIMs

If any scientific research needs to validate the assumptions made, if any engineering work involves assessing the outcome of developments, the task is not easy when one is interested in how a complex system (modular , half-material, semi-software) can become a musical instrument, that is to say considered as such [6].

Unlike many other devices and tools, musical instruments generally require a relatively long learning time. One could consider that this is a defect related to poor ergonomics... On the contrary, one could think that a long learning of the variety and subtlety of tones that can be produced makes the expressiveness of an instrument.

However, being able to choose more or less complex DIMs gives us hope that it is possible to fit the various artistic situations and educational purposes (discovery, learning, amateur orchestra, professional orchestra, etc.). In particular, we have seen considerable differences of assessment between soloist and orchestral practices in the Meta-Orchestra[8]. Interviews were led in a previous project, aiming at a better understanding of the concept of instrumentality with the help of psycho-linguistic methods [4]. This study should be deepened, and compared to works addressing similar problems [17].

Beyond the sound grammar proposed by Schaeffer [16], the Temporal Semiotic Units [8] proposed by the MIM[9] seems to us an interesting tool for assessing the richness and ease to achieve such musical figures with a given instrument.

4.2. Limitations and perspectives

When asking a luthier on the importance of the wood, the glue, the varnish to make a good violin, he/she may answer the crucial factor is the tight and precise assembly of all the elements. The modular approach may encounter limitations when considering the importance of inter- and retro-action between the elements that make a good instrument.

Nevertheless, we believe the developments described in this article will be useful for experimentation. Their availability in an modular environment like the Meta-Mallette[10] will hopefully encourage many digital instrument makers, researchers, composers and musicians to use them and further investigate this still unexplored field.

4.3. Acknowledgement

We would like to thank all partners involved in OrJo project : Puce Muse, LIMSI, and 3Dlized.

5. REFERENCES

[1] Battier, M. « L'approche gestuelle dans l'histoire de la lutherie électronique. Étude de cas : le theremin », in Les nouveaux gestes de la musique, H. Genevois et R. de Vivo (eds). Éditions Parenthèses, 139-156, 1999.

[2] Beaudouin-Lafon, M. « Moins d'interface pour plus d'interaction », *Interfaces Homme-Machine et Création Musicale,* H. Vinet et F. Delalande (eds), Hermès, 123-141, (1999).

[3] Cadoz, C. « Musique, geste, technologie », Les nouveaux gestes de la musique, H. Genevois et R. de Vivo (éds), Éditions Parenthèses, 47-92, (1999).

[4] Cance, C., Genevois, H., Dubois, D. « What is instrumentality in new digital musical devices? A contribution from cognitive linguistics and psychology », Proceedings of CIM09, to be published « La musique et ses instruments » (2011).

[5] Cascone, K, « The Aesthetics of Failure: « Post-Digital » Tendencies in Contemporary Computer Music », Computer Music Journal,24:4, pp. 12–18, The MIT Press, 2000

[6] Castagne, N., Cadoz, C. « 10 criteria for evaluating physical modelling schemes ». in DAFX'03: Proc. of the 2003 conf. on Digital Audio Effects, 2003.

[7] Couturier, J-M., *Utilisation avancée d'interfaces graphiques dans le contrôle gestuel de processus sonores*, thèse, Université de la Méditerranée, Marseille, 2004.

[8] Frémiot, M., et al., *Les Unités Sémiotiques Temporelles – Éléments nouveaux d'analyse musicale*. Editions MIM, Document Musurgia, 1996.

[9] Genevois, H., « Geste et pensée musicale : de l'outil à l'instrument », Les nouveaux gestes de la musique, H. Genevois et R. de Vivo (éds), Marseille : Éditions Parenthèses, 1999, p. 35-45

[10] Ghomi, E. « Utilisation de modèles intermédiaires pour le mapping de paramètres de synthèse », master report ATIAM, (2006)

[11] Hunt, A., Wanderley, M. M., Paradis, M. « The importance of parameter mapping in electronic instrument design ». in NIME '02: Proc. of the 2002 conf. on New interfaces for musical expression, pages 1–6, Singapore, Singapore, 2002. National University of Singapore

[12] Jaffe,D. A., Smith, J. O., « Extensions of the Karplus-Strong plucked string algorithm, » Computer Music J., vol. 7, no. 2, pp. 56–69, 1983.

[13] Karplus, K., Strong, A., « Digital synthesis of plucked string and drum timbres, » Computer Music J., vol. 7, no. 2, pp. 43–45, 1983.

[14] de Laubier, S., Goudard, V. « Puce Muse - La Méta-Mallette », Proceedings of Journées d'Informatique Musicale (JIM 2007), Lyon, 2007

[15] Momeni, A., Henry, C. « Dynamic Independent Mapping Layers for Concurrent Control of Audio and Video Synthesis », Computer Music Journal, Spring 2006, Vol. 30, N°1, 49-66, (2006)

[16] Schaeffer, P., *Traité des objets musicaux, essai interdisciplines, nouvelle édition*, Seuil, 1977

[17] Stowell, D., Plumbley, M.D. & Bryan-Kinns, N. « Discourse analysis evaluation method for expressive musical interfaces », Proceedings of New Interfaces for Musical Expression (NIME'08), Genova, 2008.

[18] Varela, F. J., Thompson, E. T., and Rosch, E. *The Embodied Mind: Cognitive Science and Human Experience*. The MIT Press, 1991.

[8] The Meta-Orchestra is a live audiovisual electronic music orchestra initiated by Serge de Laubier in 2007. It relies on hardware interfaces and a software environment, the Meta-Mallette.

[9] Laboratoire Musique et Informatique de Marseille, http://www.labo-mim.org

[10] Developments led during the OrJo project will be downloadable from the "Meta-Library", a software library currently developed by Puce Muse and available winter 2012. http://pucemuse.com