THE HYPERSAMPLER BASED ON FEATURES EXTRACTION; A BRIDGE BETWEEN THE PLAYER AND INSTRUMENT PARADIGMS

Marco Marinoni
Conservatorio "G. Verdi" di Como
marco.marinoni@virgilio.it

ABSTRACT

This paper concerns the hypersampler implemented for my piece *Il grifo nelle perle nere* for piano and hypersampler, composed in 2008 and premiered during the 54th Venice Biennale's International Festival of Contemporary Music, in the framework of XVII CIM – Colloquium on Music Informatics.

The hypersampler involves a real-time synthesis engine based on processes of feature extraction as an alternative to hyperinstruments' physical control paradigm. Features are derived from the performance of a traditional musician on an acoustic instrument – a piano – and are used as a control for the mapping between the instantaneous power spectrum of the acoustic instrument's sound output (the musical dynamics performed by the pianist) and real-time synthesis engine's parameters.

1. INTRODUCTION

The definition of "hyperinstrument" used here is consistent with that proposed by Machover [6], which I then proposed a possible extension [10]. According to Tod Machover, the basic concept of a hyperinstrument is to "take musical performance data in some form, to process it through a series of computer programs, and to generate a musical result" [6]. The hyperinstrument, in its most simple meaning, as it has been conceived for the first time in 1987 for the work *Valis* [9], is based on musical instruments able to provide a great variety of solutions that musicians play on the computer. The simplest method is through an instrument similar to an existing conventional one, such as a keyboard or a percussion. The hypersampler developed for *Il grifo nelle perle nere* [11] implements a keyboard instrument that be-comes hypertext of another keyboard instrument, the piano.

The parameter that really interested for long time Machover's research is rhythm. In a live performance, this can mean the musicians are required a greater precision than is normally demanded; this may also involve a higher degree of rhythmic complexity, and the creation of delicate relations of synchronicity that would be difficult to play without the aid of computer. However, a theoretical approach to hyperinstruments and their development cannot help but account the dimension of the live performance. Even in its interactions with technology, music is an art that is based on performance and interpretation, so "the 'brain' of a hyperinstrument is the computer system that monitors musical data from the input instrument, redefines the controls on that instrument, and acts in accordance with its programmed musical knowledge" [6]. In this sense, in the piece that underpins this paper, the gestural expressiveness related to pianistic musical dynamics triggers a sonification of the interpretative data. This occurs through a delicate process of feature extraction aimed to the construction of a virtual instrument that is informed in real time by a traditional instrument. The virtual instrument retains its own identity and all the features of a musical instrument in its own right, including the permeability to interpretative data.

Feature extraction is intended by Bullock [2] as "a form of data processing that takes a set of values and returns a more compact representation of those values. The compact representation is called a feature, and the initial set of values could be referred to as the input vector. The process of feature extraction is a form of dimension reduction, because it involves the mapping of an input vector of dimension N onto an output scalar or vector that has dimensionality that is smaller than N". As Bullock stated, a hyperinstrument system based on feature extraction can "minimise the number of prosthetic elements, and provide a seamless sense of interaction for the performer where sound becomes both the source of control and the means of gaining auditory feedback. Using sound as a medium for interaction removes the requirement for sensors, switches and other physical controllers in order to convey gestural information and performer intention".

This approach is consistent with that proposed by Machover in pieces like *Sparkler* (2001) [8], where Machover, Jehan and Fabio developed a hyperinstrument system (an acoustic instrument-plus-laptop combination) aimed at expanding the expressive power of traditional instruments and performers by placing microphones within the orchestra to capture all the instruments, which was then analyzed with a laptop and processed “to shape and manipulate a complex electronic ‘aura’ that was added live to the orchestral sound” [5].

From an interpretive and perceptual point of view, the hypersampler can be regarded as a minimal (two units) Interconnected Musical Network (IMN) intended according to the definition proposed by Weinberg [13], who defines IMNs as “live performance systems that allow players to influence, share and shape each other’s music in real time”. Weinberg also states that “only by constructing electronic (or mechanical) communication channels among players can participants take an active role in determining and influencing, not only their own musical output, but also their peers’. For example, consider a player who while controlling the pitch of his own instrument also continuously manipulates his peer’s instrument timbre. This manipulation will probably lead the second player to modify her play gestures in accordance with the new timbre that she received from her peer”. Both the sensor-based (multimodal) and the feature-extraction-sound based approaches are aimed at developing an interactive network (the hybrid double instrument called the hyperinstrument system) able to combine gestural characteristics of musical interpretation and real-time sound processing into a “constantly evolving collaborative musical product”.

In 1992 Rowe [12] proposed two distinct models of interaction in live electronic music: systems based on *player par adigm*, which provide a musical presence with a personality and a behavior of its own and systems based on *instrument par adigm*, which extend and augment the human performance through direct response to input which is generated by the performer via sound or physical control. One possible way of overcoming the limitation of these two paradigms is represented by the sensor-based approach: it stands on the ground of most of the hyperinstruments. The sensor-based approach investigates the correlation between musical and physical gesture and sonic output through the use of sensors that can be attached to the acoustic instruments and/or performers; their outputs are then scaled and routed into live controlled sound processing algorithms. In the sensor-based hyperinstrument systems, in which “a sensor converts physical energy into electricity in the machine, and may therefore be called the ‘sense organ’ of a system” [4], physical gesture is closely coupled with the audio output but “availability of existing gestural controllers is limited and new controllers can be expensive or time-consuming to develop” [2]. In 2001 Jehan proposed a system developed in Max/MSP[1] which combines audio feature extraction, timbral mapping and synthesis in the context of live electronics performance whereby “continuous changes in articulation and musical phrasing” lead to “highly responsive sound output” [3]. The system developed by Jehan included real time mapping of extracted sound features and sonification of rescaled data in order to get completely new material generated by the performance on traditional instruments.

The research at the basis of the hypersampler started from the purpose of developing a hyperinstrument system intended as an IMN able to overcome the limitations of Rowe’s player-instrument paradigm: a hybrid instrument not including the sensor-based approach, developed following the *instrument-player c ontinuum model* proposed by Bullock [2] in 2008 (see Figure 1) that extends Rowe’s player paradigm and instrument paradigm and takes account of Jehan’s approach to real-time synthesis engines based on the extraction of perceptual features.

The hypersampler includes a piano, a master keyboard (e.g. EDIROL PCR-M1) or MIDI controller (e.g. KORG NANOKONTROL 2) the computer and the technical equipment needed for the implementation of live-electronics. The software environment has been entirely developed in Max/MSP.

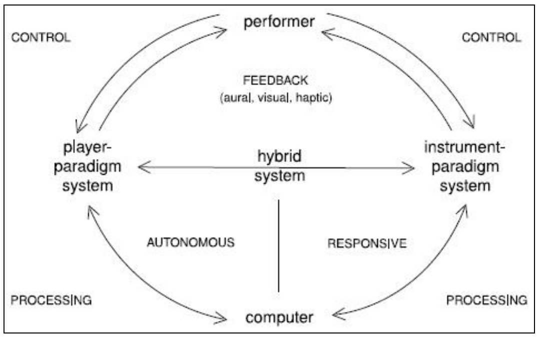

Figure 1: Bullock’s instrument-player continuum model.

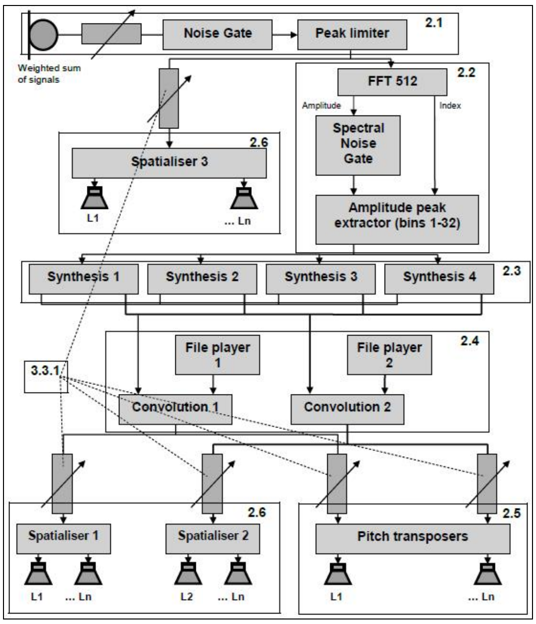

Figure 2: General overview of the hypersampler environment.

[1] http://cycling74.com/products/max/

The instrumental signal is captured using three microphones and then mixed according to different percentages in one monophonic signal. The monophonic signal is routed to the units of processing (see Figure 2), which include: first stage of treatment by noise gate and peak limiter aimed at reducing the dynamic range and making the sound materials more easily treatable; second stage of treatment by processes of feature extraction (analytic level) and transformation of the data so obtained; third stage of treatment by synthesis (synthetic level), informed by the data coming from the previous stage; fourth stage of treatment by real-time convolution (two .wav files are read in real time by buffers; the materials included are recorded sounds of violin, viola, cello and double bass performing held notes produced by various traditional and not traditional performance modalities, such as natural harmonics close to the bridge, "spazzolato", vibrato with bow on the fingerboard, "grattato", etc., covering all frequency ranges and dynamics) of output materials from the synthesis modules, which is a completely arbitrary choice aimed to confer an instrumental identity (the spectral image of a string ensemble in this case) to the hyperinstrument; finally, fifth stage of treatment by pitch transposers and sound projection by spatializing matrices controlled in real time by the hypersampler performer. The instrumental signal processed by first stage of treatment is projected (transparent amplification) too by matrices controlled in real time. For the aims of this paper we will mainly focus on second and third stages of treatment, which form the hypersampler's engine.

It is necessary in this regard to separate what technically defines and implements the theoretical conception of the hypersampler (see paragraphs 2.1 and 2.2), moving from the definitions taken as a starting point and switching to the objectives set out above (see Chapter 1), as opposed to what concerns its implementation in the musical piece, which is not discussed in detail here. We must consider, however, that some parameters and compositional choices cannot be separated from what pertains to the theoretical system, as a piece of music is a unity of design, implementation, and technical realization.

The choices about the envelopes, the spectral content of the files relating to the convolution, the particular type of jitter of frequency and amplitude realized through the modules Clock, the choice to use frequency transposition in the domain of time (here called "detune" and "frequency shifting"), the parametric correlation curves and the spatialization system form part of that required implementation of the idea that underlies the sampler, in order to obtain a musical result that realizes this idea, according to Machover's statement that "the final component of the hyperinstrument is the musical result" [6].

Il grifo nelle perle nere was written in 2008 for the "Concerto per Ipertastiere" included in XVII CIM – Colloquium on Music Informatics. The first performance took place in Venice at the Concert Hall of Palazzo Pisani on Wednesday October 16th 2008 H 5pm, during the 54th Venice Biennale's International Festival of Contemporary Music, with the following performers: Davide Tiso, piano; Marco Marinoni, hypersampler; Alvise Vidolin, sound direction.

2. HYPERSAMPLER'S ENGINE

In this section the typologies of sound processing are described, specifying the data and the variables essential for the realization of the hyperinstrument system. For each processing, the values of the parameters and their significance within the performance are indicated, identifying the ones intended to be controlled real-time by the live-electronics performer.

Jehan's assumption that "the timbre of a musical signal is characterized by the instantaneous power spectrum of its sound output" [3] and Machover's statement according to which "the 'brain' of a hyperinstrument is the computer system that monitors musical data from the input instrument, redefines the controls on that instrument, and acts in accordance with its programmed musical knowledge. These programs range from the highly deterministic (one physical event is tied to a particular musical result), to the computationally flexible (musical result chosen by particular context, and modified according to the specific musical gesture received), to the truly intelligent (analysis by rule of performed music and the machine-choice of an appropriate response)" [6] represented the starting point for the development of the hypersampler's synthesis engine.

The typology of sound tracing developed for *Il grifo nelle p erle n ere* integrates the approach of Jehan with that of Jensenius, which identifies three types of sound tracing: "focusing on sound-production, timbral features or temporal development" [4]. The feature extraction process implemented here uses the third type of sound tracing.

2.1 Spectral noise gate – Amplitude bin extractor

The output signal from peak limiter unit is analyzed using a length N FFT of $\sim x_m$ to obtain the STFT at time m:

$$\tilde{x}'_m(e^{j\omega_k}) = \sum_{n=-N/2}^{N/2-1} \tilde{x}'_m(n)e^{-j\omega_k nT} \qquad (1)$$

where $\omega_k = 2\pi k f_s/N$, and $f_s = 1/T$ is the sampling rate in Hz. The STFT *bin num ber* is k. $N = 512$. Then each FFT bin $\sim x'_m(e^{j\omega k})$ was converted from rectangular to polar form to get bin k's instantaneous amplitude.

$$A_k(m) \quad \stackrel{\Delta}{=} \quad \left|\tilde{x}'_m(e^{j\omega_k})\right| \qquad (2)$$

As shown in Figure 3, only the first 32 bins are used and in particular only the amplitudes of bins that exceed a threshold, in order to cut the residual nondeterministic components of the sound in addition to the deterministic harmonic components. The signal so obtained is then filtered using a second order low-pass filter so as to obtain a low-frequency control signal. That signal is finally 'converted' in Hertz multiplying it by an appropriate conversion factor and sent to the peak extractor unit which identifies the maximum value sent out to the synthesis units by means of the trigger command T-R which is controlled in real time.

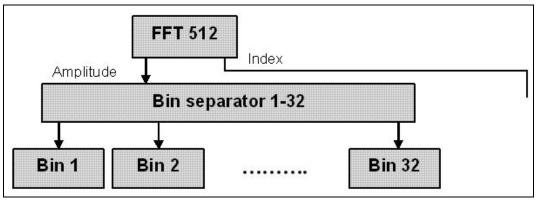

Figure 3: FFT analysis.

In Figure 4 the operations concerning the extraction of the parameter amplitude in one bin and its translation to a frequency scale are described.

The value of the parameter "threshold", that is the minimum amplitude value of single bins sent to the low-pass filter, must be so as to neatly cut the ground noise without compromising or altering the spectromorphological peculiarities of the analyzed signal.

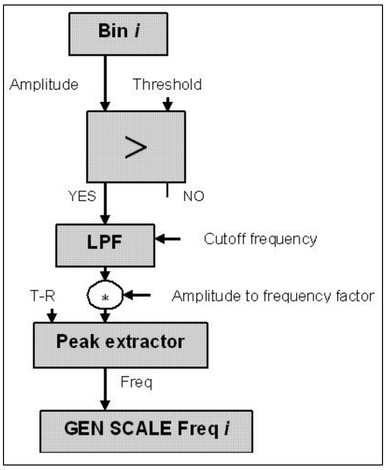

Figure 4: Translation of bin *k*'s instantaneous amplitude to a frequency scale.

The value of the parameter Cutoff frequency of the low-pass filter, is approximately set to 0.4 Hz.

The value of the parameter Amplitude to frequency factor, that is the conversion factor, must be determined in a way that the maximum output values do not exceed the number 4000 and the minimum values never lower the value 20. The Grain Generation Scale (GEN SCALE) is composed of the 32 frequency values so obtained.

2.2 Synthesis

The synthesis engine includes four clock-controlled Grain Generator Units, as shown in Figure 5. In Figure 6 the implementation of a Grain Generator is described.

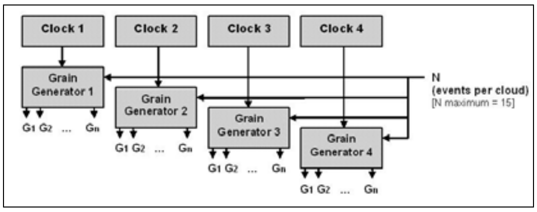

Figure 5: Synthesis module.

The four Synthesis units require four different waveforms (W). In the case of implementation using Max/MSP it is suggested to use the object *gen* (linear b.p.f. wavetable generator) included in *PeRColate – A collection of synthesis, signal processing, and video objects* by Dan Trueman (Princeton University) and R. Luke DuBois (Columbia University)[2] ported from real-time cmix, by Brad Garton and Dave Topper.

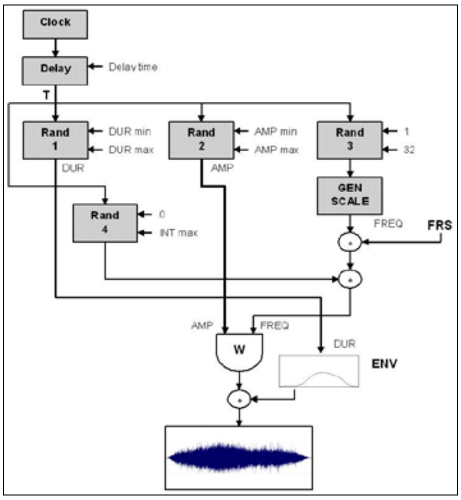

Figure 6: Synthesis engine: Grain Generator unit.

Three different typologies of envelope (ENV) applicable in a mutually exclusive way to the grains generated by the four Synthesis units are required.

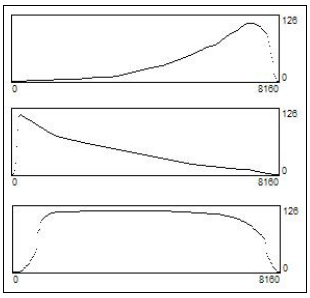

Figure 7: Envelopes which apply to the grains.

[2] http://music.columbia.edu/percolate/

The selection of an envelope is controlled in real time during the performance, with interpolation time equal to 6 seconds. The three types of envelope are shown in Figure 7.

The module RAND 1 controls the parameter Grain Duration, by generation of random floating-point numbers comprised between the minimum value DUR min and the maximum value DUR max. The module RAND 2 controls the parameter Grain Amplitude, by generation of random floating-point numbers comprised between AMP min and AMP max. The module RAND 3 generates random integers between 1 and 32, determining the Grain Generation Frequency among the 32 possible frequencies generated by the module GEN SCALE which form the Grain Generation Scale. The Grain Generation Scale must be changed many times during the performance using the command T-R (see Figure 4). The performer decides, according to his interpretation, how many times the scale is changed during the performance and when, according to the musical score. The module RAND 4 controls the variance of the parameter Transposition Interval expressed in semitones and cent, which causes random variation of the Grain Frequency around the original value, comprised between 0 and the value INT max = 1 semitone, 27 cent. The parameter Frequency Range Shifting (FRS) controls the transposition interval n (in Hertz) applied to the grains so that the grains' frequency is modified as shown by the formula:

$$Freqfin = (Freqinit)*n \qquad (3)$$

The value of parameter T (delay time) of the Delay unit is comprised between 0 and 12700 ms, and is controlled in real time, as well as n. The module Clock (see Figure 8) implements the following parameters.

CT1 = clock time [ms]
Vmin1 = minimum randomly generated number [int]
Vmax1 = maximum randomly generated number [int]
IT1 = interpolation time [ms]
CT2 = clock time [ms]
Vmin2 = minimum randomly generated number [int]
Vmax2 = maximum randomly generated number [int]
IT2 = interpolation time [ms]
CT3 = clock time [ms]

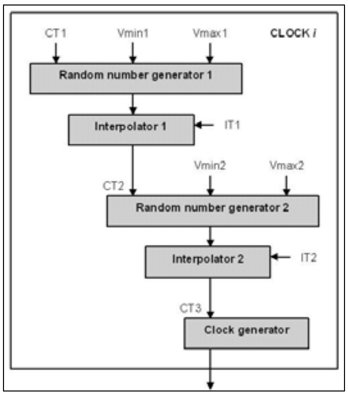

Figure 8: The module Clock.

3. DISCUSSION

In *Il grifo nelle perle nere* a virtual instrument is coupled with a traditional instrument, giving rise to a hybrid between mechanical and computer, using data extracted from the musical interpretation of the pianist to control an independent virtual system, which meets the requirements of a hyperinstrument and realizes the statement of Machover according to which the goal of a hyperinstrument would be "to produce music of unprecedented subtlety, complexity, richness, and expressive power that is intimately, but not obviously, linked to the original intent of the performer/composer" [6].

Machover's approach towards "double" and "triple instruments"[3], in which two or more people are playing a single hyperinstrument, is not unlike the one that underpins *Il grifo nelle perle nere,* where a "double instrument", the hypersampler, is controlled, at different levels, by the pianist and by the keyboard performer: the first, by playing the instrumental dynamics, affects a number of parameters including the choice of the pitch scale from which the synthetic sounds are generated by the computer; these sounds, processed by convolution, are controlled in real time by the second, which in turn can change the number of sounds produced at the unit of time, their density, their positioning within the virtual space, providing the first performer a new musical material on which interact, in a continuous and fertile creative feedback mechanism, since double instrument performers "must relate their musical gestures not only to the resulting sound as in traditional instruments, but also to the gesture and sound of the other performer" [6].

Consistently with Machover's assertions about the importance of the conceptual simplicity of the interface, this system is easily understood by the performer, who has a chance to become aware about the specific relationship of causality (semi-deterministic and bound to the interaction with the live electronics performer) that binds his actions to the production of the sound output by the system and, through a period of practice, refine his performance. In this way, the system is partially controllable by the instrumental performer, which can achieve a level of control over the music that is even greater than it has in general. The computer does not play a part isolated. The performers have the opportunity to check the results and to take on more roles from a musical point of view, depending on the particular di-

3 A *double instrument* is conceived for two musicians that play together on separate physical controllers (one of those can be an acoustic instrument) to breed a hybrid instrument "so that each musician can influence certain aspects of the music, but both players are required to perform in ensemble to create the entire musical result" [6]. For example, in Machover's *Towards the Center* [7] the keyboard player controls the overall sound spectrum –the partials, the harmonic series, the spectromorphologic qualities of sound – while the percussionist controls the behavior of each partial, like a microscope where one observer acts on a smaller portion (controls more extended parts) while another observer acts on a greater portion (controls smaller parts, internal to the parts controlled by the other observer).

rection they decide to give the performance from time to time.

The relationship control / independence between the performers is mediated by the hypersampler, which assumes the role of *double instrument* (a *hybrid instrument* that is partly physical instrument and partly virtual) formed by two performers that work together to control a complex instrument, each of which controlling only part of the final result. This instrument also includes the “partial and expected *unpredictability*” [10] which was mentioned earlier as a distinctive feature of each instrumental practice, traditional or contemporary.

The feature extraction process which is implemented on the traditional instrument does not imply structural change: this increases the level of reproducibility, not bound to the context or the availability of specific technologies, however, placing a question of theoretical order: is the difference between what we call generally live-electronics and what we call a hyperinstrument linked to the use of technologies such as sensors etc., as Machover and the MIT researchers seem to say, or is it a difference of higher order (multiple instruments, different in nature, acoustic and electronic, with specific performers who play performances interconnected, according to Weinberg’s theory, which form a single hybrid instrument, equipped with its own identity, qualitatively different from the sum of the identities of the individual instruments involved: the hyperinstrument system) and the hyperinstruments are but a subset of the broader category of live-electronics?

4. CONCLUSIONS AND FUTURE WORK

In the light of findings from the experience with *Il grifo nelle perle nere*, and assuming with Benzon that music is “a medium though which individual brains are coupled together in shared activity” [1], the hypersampler can be truly considered as a basic Interconnected Musical Network (composed of two units), that is a “live performance system that allow players to influence, share, and shape each other’s music in real-time” [13], being also a feature extraction driven *double instrument* in which an acoustic instrument is “complemented by delicate electronics played and transformed by a keyboard-with-laptop [...] creating shifting textures that ‘fuse’ the various instrumental lines” [5].

From the musician’s perspective, the hypersasmpler behaves intuitively and predictably. The control features of the traditional instrument are used “as input for the model of a different instrument” [3] which is perceptually meaningful. Choices related to the technical implementation appear to be adequate and able to provide the hyperinstrument with a sufficient degree of freedom and individuality, to ensure the possibility of a true musical interpretation.

Future work will include algorithms that extract more control features and the extension of this approach to different instruments such as flute and violin, in order to be able to advocate a generalization of results, with regard to the connection between mapping of input signal and commands to control parameters to audio output.

5. REFERENCES

[1] W. L. Benzon: *Beethoven's Anvil: Music in Mind and Culture*, Basic Books, New York, 2001.

[2] J. Bullock: “Implementing audio feature extraction in live electronic music”, Ph.D. dissertation, Birmingham Conservatoire, Birmingham City University, 2008.

[3] T. Jehan: “Perceptual synthesis engine: An audio-driven timbre generator”, M.S. thesis, School of Architecture and Planning, Massachusetts Institute of Technology, Cambridge, 2001.

[4] A. R. Jensenius: “Action — sound. Developing methods and tools to study music-related body movement”, Ph.D. dissertation, Department of Musicology, University of Oslo, Oslo, 2007.

[5] T. Machover: “Dreaming a New Music”, *Chamber Music*, Vol. 23 No. 5, pp. 46–54, 2006.

[6] T. Machover: *Hyperinstruments: A Progress Report*, Technical report, MIT Media Laboratory, Cambridge, 1992.

[7] T. Machover: *Towards the center*, Score, Editions Ricordi, Paris, 1989.

[8] T. Machover: *Sparkler*, Score, Boosey and Hawkes, New York, 2001.

[9] T. Machover: *Valis: an opera in two parts*, Score, Editions Ricordi, Paris, 1987.

[10] M. Marinoni: “Comporre per gli iperstrumenti. Il sistema-iperstrumento come agente intermediatore tra l'esecutore strumentale e i processi di trasformazione/generazione del suono”, Ph.D. dissertation, Conservatorio di Musica “B. Marcello”, Venezia, 2007.

[11] M. Marinoni: *Il grifo nelle perle nere*, Score, ArsPublica Edizioni Musicali, Camino al Tagliamento, 2009.

[12] R. Rowe: *Interactive music systems: machine listening and composing*, MIT Press, Cambridge, 1992.

[13] G. Weinberg: “Interconnected Musical Networks – Bringing Expression and Thoughtfulness to a Collaborative Group Playing”, Ph.D. dissertation, School of Architecture and Planning, Massachusetts Institute of Technology, Cambridge, 2003.