NEYMA, interactive soundscape composition based on a low budget motion capture system.

Stefano Alessandretti

Independent research
s.alessandretti@gmail.com

Giovanni Sparano

Independent research
giovannisparano@gmail.com

ABSTRACT

Mocap (motion capture) techniques applied to music are now very widespread. More than two decades after the earliest experiments [1], there are many scientists and musicians working in this field, as shown by the large number of papers and the technological equipment used in many research centres around the world. Despite this popularity, however, there is little evidence of musical productions using the mocap technique, with the exception of a few that have been able to rely upon very high budgets and very complex equipment. The following article aims to describe the implementation of "*Neyma*, for 2 performers, motion capture and live electronics (2012)," [2] an interactive multimedia performance that used a low budget mocap system, performed as part of the 56th *Biennale Musica di Venezia*.

1. INTRODUCTION

Neyma is an interactive multimedia performance focused on the sound and the territorial identity of the city of Venice. The work was commissioned by *Biennale Musica di Venezia* and IanniX's development team [2].
The general idea of the project had a dual purpose:

- exploring the sounds of the city,
- exploring its territory.

In Neyma, therefore, 2 performers make up a soundscape [3] and a visualscape [4] simultaneously and in real time through only gestural improvisation with their hands, using non-haptic sensors [5] and direct gestural acquisition [6].
The idea followed 5 basic principles:

- all the original sounds (pre-processing) had to come from Venice,
- all the visual events had to be generated from a map of the city,
- the soundscape and visualscape had to be composed in real time,

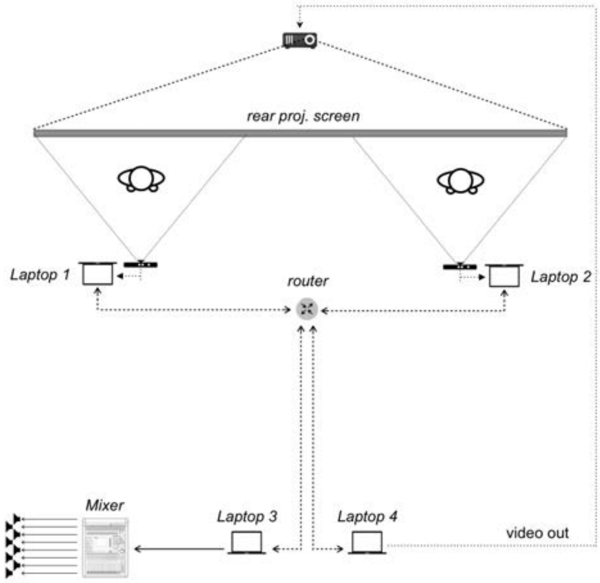

Figure 1. Technical requirements.

- the soundscape and the visualscape had to be made through the hands gestures of the performers,
- the work had to be developed using low-cost or open-source technology and software.

In accordance with these principles the work was performed using the following technologies:

- Max/MSP [7], IanniX [8] [9], Synapse [10] and the Open Sound Control content format (software tools),
- 4 laptops, a large video projector, 2 Microsoft Kinect devices, a mixing desk, a multichannel audio system and a Local Area Network (hardware tools).

The performer's hand movements are mapped using the mocap system formed by Kinect-Synapse-Max/MSP (*performer patch* running on *laptops 1* and *2*) and related data is sent to the main computer via the LAN network (UDP format). *Laptop 3* hosts the data translation/synchronization system (*main patch*) and the audio generation system (*audio patch*). *Laptop 4*, running the IanniX software (*video patch*), receives data from *laptop 3* and generates synchronized visual events (fig. 1).

2. MOTION CAPTURE STSTEM

Each motion capture system is composed of a Kinect device, Synapse application and a Max/MSP patch (*performer patch*). The Synapse app gets the raw input data from Kinect and sends out OSC messages according to a specific syntax:

/<point of the skeleton>_pos_world <float of X position> <float of Y position> <float of Z position>.

Axes are arranged on the basis of the performer's point of point of view. The app can recognize the skeleton of a user, grab some key points from it and send the spatial location out in relative values, with the *pos_world* being the distance expressed in millimeters from the Kinect and the skeleton point determined by the software. Three messages per performer were used: right hand, left hand and torso position (fig. 2).

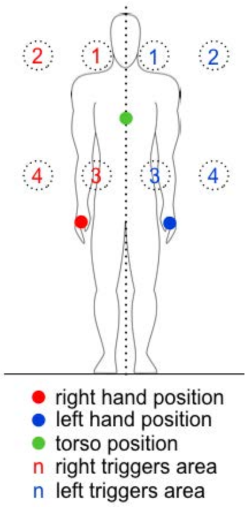

Figure 2. Hand and torso recognition.

In Max/MSP *performer patch*, these messages are translated into:

- the speed motion of the hand,
- the distance of the hands from the torso, useful in obtaining a tracking of hand movements independent from the distance of the performer from the mocap device.

With *performer patch* one can control:

- the spatialization of *drones* through the hand speed motion,
- the activation of *triggers*,
- the recognition of *sequences*. (see §3)

These three controls are automatically activated in specific movements during the performance. *Drones* start automatically and move into an electroacustic space according to the speed of motion of the hands, the triggers being single spheres in 3D space with an adjustable radius, activated by passing hands through the points in which they are placed.

Sequences are chains of triggers: in specific performance sections, the consecutive selection of 2 *triggers* define a *sequence*.
The location of *triggers* and gestures related to *sequences* are initialized before the performance and all lie within an action space that extends in front of the performer (fig. 3).

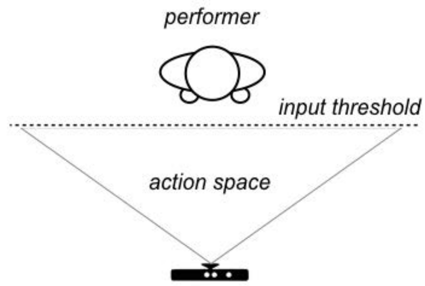

Figure 3. Action space.

The initialization process consists of:

- the adjustment of the input threshold within the hand action space,
- the determination of trigger points which represent the centre of the sphere,
- the determination of sequence points.

All these settings are made by putting the performer's hands in a desired point in space which is then registered into the *performer patch* by an assistant that stores the related presets, the performer placing themselves in the same spot used for the performance.

3. INTERACTIVE AUDIO SYSTEM

The audio processing environment (*laptop 3*) consists in the generation and spatial diffusion of sound events (*audio patch*) and is organized into 4 main modules: a sampler, a bank of automated gain faders (pseudo-random algorithm), a bank of 12 spatializers and a reverberation unit.

In addition to these, there is also a module for the extraction of the amplitude value of the signal consisting of a bank of filters and peak meters that splits the spectrum into 24 bands, detecting each amplitude value (*vocoder*, cf. §4) and sending these to *laptop 4* as the main control variables of visual events (fig. 4-5).

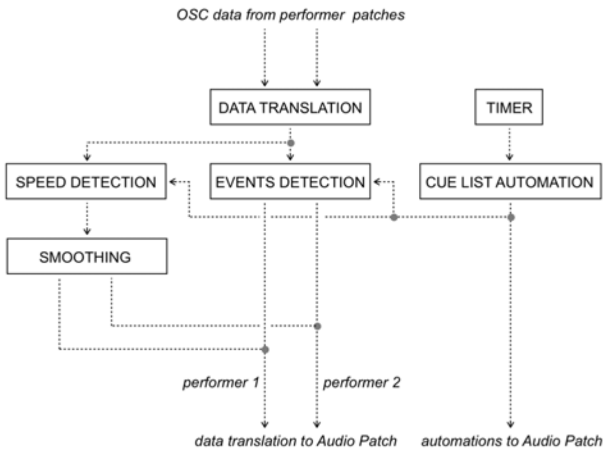

Figure 4. *Main patch* diagram (*laptop 3*).

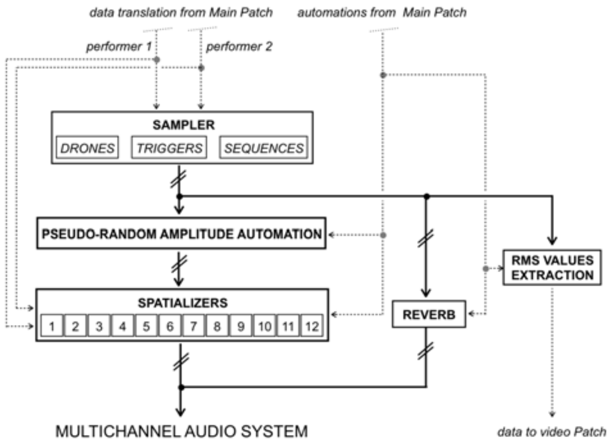

Figure 5. *Audio patch* diagram (*laptop 3*).

Sampler: a bank of 48 file players (24 for each performer) that allows the playback of 3 types of sound events: *drones*, *triggers* and *sequences*. *Drones* are long duration audio files (up to 2 minutes) triggered by the cue list and their function is like a "*basso continuo*". *Triggers* are short duration audio files (up to 12 seconds) activated by virtual buttons around performers while *sequences* are short duration audio files (up to 8 seconds) triggered by the performers' hand gestures (triggers chains, cf. §2).

Pseudo-random automation: a bank of 24 automated gain faders that allows the output level of each sample to be varied randomly, within a preset range. All variables of the module are automated through the synced cue list in the *main patch*. It is a basic system because it allows for the quick setting of all the samples' amplitudes and their automatic control at run time, and at the same time it offers the possibility to simulate a "from near to far" (and viceversa) sound effect.[1]

Spatializers: a bank of 12 spatializers organized according to the type of samples received as input . The motion algorithm is largely based on a matrix (controlling the opening time of the channels) and the speed of movement can be controlled manually (receiving data from the mocap system) or automatically, using the synced cue list in the *main patch*.

Reverberation: a delay line reverb algorithm which allows the adding of a virtual environment and the simulation of the movements mentioned earlier. All variables are automated by the same cue list in the *main patch*.

4. INTERACTIVE VIDEO SYSTEM

IanniX is a graphical open source sequencer that allows graphic representations of a multidimensional score [9]. This score is made up of three different objects: curves, cursors and triggers. For the purpose of this project, only the usage of curves manipulated in real-time through an opportune patch (*video patch*) were considered.
The implemented score was a 2D map of Venice imported in a IanniX project as a set of different curves defined as B-Splines: by moving a point that belongs to a curve, allowing a smooth animation (fig. 6).

Figure 6. Map screenshot.

Selected curves are moved in the third dimension (z-axis) at precise time moments, via *video patch*. This patch controls the location of single or groups of curve-points. The range and sign of movements were arbitrarily defined on the basis of aesthetics.

By zooming, shooting at different angles, hiding and showing groups of curves, it was possible to create a video animation controlled in real-time by a predetermined score and the occurring audio events, the score controlling which curves are visible, the zoom factor and the shot angle. The audio amplitude obtained as well as analysis using vocoding control the size of movements in the *z* dimension and the transparency of the current visible curves. The 24-band vocoder used for the analysis algorithm is a channel vocoder while the centre frequency and bandwidth of each band are listed in the table below and follow the 24 critical bands of hearing on the Bark scale (table 1).

[1] Varying the direct signal and keeping constant the reverberated signal.

Center freq. (Hz)	Bandwidth (Hz)	Center freq. (Hz)	Bandwidth (Hz)
50	80	1850	280
150	100	2150	320
250	100	2500	380
350	100	2900	450
450	110	3400	550
570	120	4000	700
700	140	4800	900
840	150	5800	1100
1000	160	7000	1300
1170	190	8500	1800
1370	210	10500	2500
1600	240	13500	3500

Table 1. Critical bands.

The video score is divided into 4 macro sections in which different curves are pictured and manipulated in real-time. In each section there are 24 selected curve-points which are linked to a specific band of the vocoder.
There is also a relationship between the curves and the sounds used in a single section, the curves being parts of the Venice map in which soundscape audio recordings were made.

The resulting video is a conceptual animation of white lines on a black background in continuous transformation that ends in a bird's eye view of a stylised Venice map (fig. 6-7).

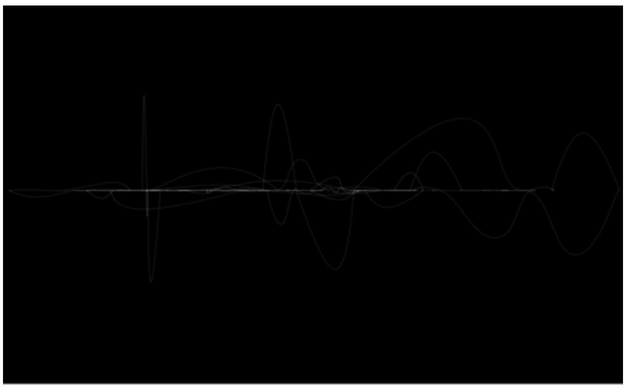

Figure 7. Map screenshot.

5. OSC DATA

The communications between Synapse, *performer patch*, *video patch*, Iannix project, *audio patch* and *main patch* are made possible using the OpenSoundControl content format [12]. The LAn is set up as a mixed peer-to-peer and client-server model network. The Synapse application/*performer patch* and Iannix project/*video patch* pairs are couples of individual nodes in the P2P network in which any communication is purely unilateral: mocap data flows from Synapse to the performer patch and the video score commands from the *video patch* to the Iannix project. The *main patch* acts as a server coordinating messages from the *performer patch* to the *video patch* and the *audio patch* (fig. 8).

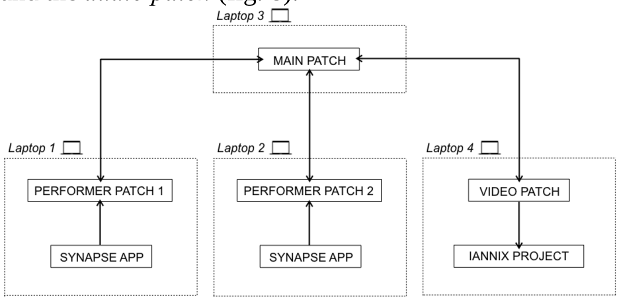

Figure 8. LAN.

6. SOUNDSCAPE COMPOSITION

As indicated above, all the sounds come from the city of Venice, from characteristic spots in sound terms: the Ponte di Rialto, Piazza San Marco, the Campo San Polo, Piazzale Roma, Canal Grande, the Arsenale, San Giorgio Maggiore, SS. Giovanni e Paolo and the Giudecca.
The collected sound samples were then processed using a variety of techniques including granulation, ring modulation, convolution, frequency warping, spectral delaying, filtering and vocoding.[2]
All these sound events were placed into 3 categories: *drones*, *triggers* and *sequences* (see §3); in such a way that each performer has his personal samples library.

Performer 1: 8 *drones* (4 + 4) , 16 *sequences* (8 for each hand), 16 *triggers* (8 + 8, 8 for each hand).

Performer 2: 4 *drones* (2 + 2), 16 *sequences* (8 for each hand) , 24 *triggers* (12 + 12, 12 for each hand).

The gestural improvisations were organized in such a way as to obtain a circular structure formed by 3 types of soundscape: virtual, surreal and real [11]. This idea was applied in order to simulate an approach to the city, a tour within it and a subsequent departure to other places (fig. 9).
In this structure each performer follows a time sequence of instructions inside of wich he is free to improvise.

Performer 1:
0'00" / 3'00" - *drones* spatialization,
3'00" / 4'00" - *triggers* mode,

[2] Max/MSP patches (programmed on purpose).

4’00” / 6’00” - *sequences* mode,
6’00” / 8’00” - *triggers* mode 2 (different sounds),
8’00” / 9’00” - *drones* spatialization 2 (different sounds).

Performer 2:
0’30” / 2’00” - *drones* spatialization,
2’00” / 4’00” - *triggers* mode,
4’00” / 6’00” - *sequences* mode,
6’00” / 7’00” - *triggers* mode 2 (different sounds),
7’00” / 9’30” - *drones* spatialization 2 (different sounds).

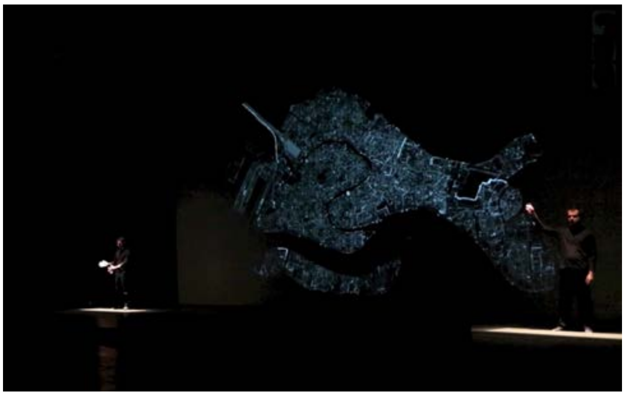

Figure 9. Performance.

7. CONCLUSIONS

Both from the technological point of view and from an aesthetic-musical perspective, the production of *Neyma* was founded on the idea of economy and that of coherence. We attempted to use the smallest possible number of technologies and focus our work on the software development of the mocap system and performance environments, aiming at maximum integration of the visual and sound media. The creation of *Neyma* demonstrates how it is possible to conceive a low cost motion capture system that is both flexible and stable even in critical situations, such as an interactive multimedia performance.

8. REFERENCES

1. D. Collinge, and S. Parkinson, “The Oculus Ranae,” in *Proceeding of the 1988 International Computer Music Conference*, San Francisco, (1988), pp. 15-19.
2. Live performance recording : http://www.youtube.com/watch?v=SRjWx7zqVsE
3. R. Murray Schafer, *The New Soundscape*. Universal Edition 1969.
4. M. Llobera, “Extending GIS-based visual analysis: the concept of visualscapes,” in *International journal of geographical information science*, London, (2003), pp. 25-48.
5. R. M. Baecker, J. Grudin, W. A. S. Buxton, and S. Greenberg, *Readings in Human-Computer Interaction: Toward the Year 2000*. Morgan-Kauffmann, 2nd edition, 1995. Part III, Chapter 7.
6. P. Depalle, S. Tassart, and M. Wanderley, ``Instruments Virtuels" *Resonance,* pp. 5-8, Sept. 1997.
7. Cycling 74 home page : http://cycling74.com
8. IanniX home page : http://www.iannix.org
9. T. Coduys, and G. Ferry, “Iannix. Aesthetical/symbolic visualisations for hypermedia composition,” in *Proceedings of the Sound and Music Computing Conference*, Paris, (2004) pp. 18-23.
10. Synapse home page : http://synapsekinect.tumblr.com
11. B. Truax, “Soundscape, acoustic communication & environmental sound composition,” in *Contemporary Music Review* 15(1), London, 1996, pp. 49-65.
12. OSC home page : http://opensoundcontrol.org