INSTRUMENTAL GESTURES AND SONIC TEXTURES

Jehan-Julien FILATRIAU

Communications Laboratory,
Université catholique de Louvain (UCL),
1348 Louvain-la-Neuve, Belgium
filatriau@tele.ucl.ac.be
and
LMA-CNRS

Daniel ARFIB

LMA-CNRS
31 Ch Joseph Aiguier
13402 Marseille Cedex 20 France
arfib@lma.cnrs-mrs.fr

ABSTRACT

The making of textures is a field largely associated to the visual domain; sonic textures are less known and explored. Instrumental gestures, especially in digital musical instruments, are often seen from a pitch-oriented point of view, and are rarely viewed as natural gestures linked to textures. Mapping between gestures and textures still remains a case per case experiment.

This article is a contribution to a framework including instrumental gestures and sonic textures. First mapping between a gesture and a sonic process is defined. The specificity of textures is then examined; algorithms implementations and sensor technology help stay on the ground of possible things with a computer and are illustrated by examples taken from the literature. A large part of the article is then dedicated to personal experiments conducted in our labs. A discussion follows, which gives an occasion to set new perspectives dealing with the notion of ecological sounds and gestures.

1. INTRODUCTION

A new field (not to say a discipline) has taken place around « gesture-controlled audio systems » [12]. Within this field, applications such as digital musical instruments have pointed out the need for a proper relationship between gesture sensing and sonic process, in so as to get the feeling of an instrument and not only the control of a process. Some of these digital musical instruments (DMI) have a strong connotation of acoustical instruments, and then some laws can be retrieved from these. Others deal with sonic textures, and look free from any convention. The goal of this article is to set a framework to study gestural audio-systems using audio textures, and to show that the specificity of these sonic textures leads to different links with gestures. In section 2 we shall see what kind of algorithms and sensors can be used, and in section 3 we shall present some realisations that can help us understand where to go in this large area. We shall finally have a discussion on the implications of these experiments and a prospective view of the possible research on gesture and texture.

2. FUNDAMENTALS ON INSTRUMENTAL GESTURES AND SONIC TEXTURES

2.1. Gesture to sound: the digital musical instruments

The matter of mapping in digital musical instruments has given rise to some formalizations of mapping possibilities between gesture sensing and sound processing (Fig. 1).

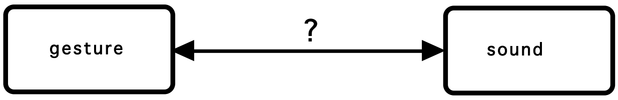

Figure 1. Gesture and sound connection

On our side, we have strongly pushed two main ideas:

2.1.1. Sound to gesture

Instead of using gestural devices and trying to connect these data to sound, which is a sonification of gesture, it may turn out musically interesting to build the inverse link: from sound to gesture. In fact many sonic processes are already known in the “non real time processing” as the result of algorithm using languages such as Csound [30] or Music V. They are generally composer oriented, which can be a drawback (they are not playable) but which also has an advantage: the sound is constructed and only the imagination sets the rules and the borders of the domain. Also in musicology one can talk about the “musical gesture” in the sense of the development of musical features apprehensible by the human brain during the listening process. So the “gesture to sound” philosophy and practise would ideally link a “gestural gesture” to a “musical gesture”. Acoustical instruments do so of course because they have been born with this following concept in mind: lutherie is the art of giving performers and composers a tool to obtain the sound they want, within limits which come either from the physics of the instrument or the ergonomics of the human gesture.

2.1.2. Intention to expression

A gesture by itself is the result of an intention. A sound is a result of an expression. So we could say that music, and especially music performance is the realisation of an expression coming from an intention. When it comes to digital musical instruments, we have to connect in a way gesture sensing parameters to sound algorithmic (synthesis or analysis synthesis) parameters. It is always a good practice to think about a way to recover intention from gesture, to map this intention to an expression and see how this expression can be translated to synthesis parameters (Fig. 2). These two intermediate levels are sometimes called "psycho acoustic parameters" [3].

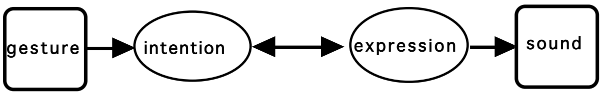

Figure 2. Intention and expression connection

2.2 Sonic textures

2.2.1. The specificity of sonic textures

Textures are sounds which grossly respond to this definition: on a short-term scale, they are composed of a succession of micro-structural elements, subject to some randomness; at a long term scale, a temporal and spectral coherence is preserved ("approximately stationary"). We will see in next subsection a more detailed typology of textures. For now what is important is to understand that the normal cues for the design of musical instruments just do not work: there is no strong temporal profile, but usually the feeling of a flux, and there is not either an harmonic structure on which one could rely for a pitch oriented direction. Thus musicians that tried to deal with textures have taken very specific views, just like a case per case situation. The feeling of "ambient music" evoked by textures makes them favourites for installations, or video music, but an instrumentalisation of textures is not a well-explored field.

2.2.2. Musical typologies of sonic textures

Electronic music has dealt a lot with sonic textures and many attempts have been done which are concerned either by the description of the sonic side, or by our perception. From this perceptual point of view, textures are good candidates for the ecological approach, which has been initiated by Gibson and applied to sound by some authors [26].

When it come to practice, we also need a more "signal processing oriented" approach, and we can find in P. Hanna's work [19] such a typology. He distinguishes

- coloured noise in which the main characteristic looks like the filtering of a stationary noise by an evolving filter;

- pseudoperiodic noise where in fact we have a texturisation of repetitive sound, especially in machines that "make noise";

- impulsive noise where the main characteristic is the successive clicks, possibly following a statistical law.

Each of these three classes corresponds to a different "musical gesture". As an example, the ocean seashore sound is belonging to the first class and the sound of rain, to the third one, even through both of them use water. The first class deals with spectral content, while the third one looks more like a transient succession, hence related to rhythm. The intermediate one is pitch oriented, and the timbre is analog to a texturalisation of a shape in the visual domain.

This points out on the fact that under the word textures, we have a wide variety of sounds, and that we should probably distinguish between them in order to use them in an audio-gestural system. If for a vocal instrument we can distinguish between the source (glottal pulse) and the resonance (articulation), here we have the distinction between the nature of a source – impulsions or noise - and its coloration. However these two, source and resonance are intermeshed, we do not usually have independence between the two.

2.3. Algorithms and implementations: analysis-synthesis techniques

In this section, we present a panel of analysis/synthesis and pure synthesis techniques dedicated to sonic textures.

2.3.1 Analysis-synthesis methods

This group of methods aims to synthesize original sonic textures from the analysis of an input texture. We can distinguish three class of analysis/synthesis techniques: first, the methods inspired of the computer graphics field for visual textures synthesis, second the methods derived from granular synthesis and finally techniques based on source-filter modelling of the sonic textures.

2.3.2. Methods inspired from visual textures synthesis

In [5] and [17], Bar-Joseph & al. proposed a method suited for both visual and sonic textures synthesis, relying on statistical learning and resampling of a tree representing the wavelet transform of an input texture. From the original tree, new random trees with the same statistical characteristics are generated and then transformed back to produce new sonic textures, statistically similar and perceptually close to the original sound. Parker & Chan [27] suggested another technique close to this approach, originally devised for visual textures synthesis [31], where the input texture is represented as a Gaussian pyramid.

2.3.3. Grain-based methods

In granular synthesis [28], original complex sounds are created in combining small audio chunks ("grains") obtained by segmenting an audio source. Granular synthesis is not well suited for analysis/synthesis process because grains are randomly sliced, which prevents from preserving the original structure of the sound. Hoskinson [20] and Lu [23] proposed similar algorithms to split up an audio source into variable sized "natural grains", relying on frame-based analysis of wavelets and mel-frequency cepstral coefficients (MFCC) respectively. Similarity and transition probability between each segment are calculated for use in the synthesis step, in which the grains are recombined into a continuous stream with following the transition probabilities to avoid audible discontinuities. Cardle [9, 10] developed an improved version of both Hoskinson's and Bar-Joseph's algorithms, by weighting the appearance of each grain in the synthesized sound to add high-level user-control over the synthesis process.

2.3.4. Source-filter approaches

This class of methods adopt a source-filter approach to model sonic textures. Analysis aims at capturing properties of both excitation and filter. Athineos and Ellis [4] suggest a method based on a double linear prediction, named synthesis by cascade time- and frequency- domain linear prediction (CTFLP). In order to render most precisely the characteristic short-term temporal structure of sonic texture, the temporal envelope of the signal is captured by a linear prediction step in the spectral domain so that the microfluctuations of the original texture are faithfully reproduced. Zhu and Wise [32] presented a extended version of the CTFLP synthesis, where sonic textures are considered as a mix of a background "din", synthesized by filtered noise, and a foreground micro events sequence following a probabilistic distribution in their occurrence.

2.4. Algorithms and implementations: pure synthesis methods

2.4.1. Noise filtering techniques

Colorization of white noise by filtering techniques allows creating sonic textures in many various ways. For instance, in the "Filtering String" instrument, we used the shape of a slow-moving string to control the gains in a filter bank with noise as sound input [2]. The coloration given by frequency resonances and fluctuations in the sound due to motion of the string enables to generate textures with complex but natural variations.

2.4.2. Functional Iteration Synthesis

Di Scipio [16] proposed an original pure synthesis method to create chaotic sonic textures. Functional Iteration Synthesis (FIS) is a derivative of the wave terrain synthesis, where wave terrains are generated by iterations of non-linear functions, which give them extremely complex – quite chaotic - relief. The resulting sonic textures present acoustic turbulences and are closed to environmental sounds like rain or thunderstorms. A digital musical instrument derived from Di Scipio's algorithm is presented in section 3.2.

2.5. Gesture and sensors typology

2.5.1. Gesture typologies

It is common sense to use a simple typology, which can be simplified as follows [11] :

- selection gestures: change a preset
- decision gestures: trigger an event
- modulation gestures: specify a curve
- accompanying gesture: do nothing

As we shall see, most gestures are combinations of different categories. For example crossing a plane can be part of a modulation gesture, but the crossing itself may be a decision gesture. Selection gestures and decision gestures are associated when it comes to hit certain zones and make a sound immediately.

Accompanying gestures, as the name indicates accompany other gestures. It has been shown on acoustic instruments [7] that such gestures are natural, that they sometimes have an acoustic meaning but that they can also only help the human performer who feels happy with them. Such gestures are very important when it comes to associate gestures and sounds: a gesture oriented towards texture must be handy, comfortable and in this sense it can provide some extraneous information that only helps the performer to feel at ease.

2.5.2. Sensor typology

Though many subdivisions can be made, the main division is between

- contact sensors,
- free sensors.

In the first category, we may have all sensors added to an acoustic instrument ("augmented instrument") or alternate devices, which are peripherals diverted from their initial purpose. This is the case for tablets, joysticks etc… which give a manual feedback or an assisted one (force feedback). Free-motion sensors allow the user/performer to produce non-constrained gestures. They usually are based on video systems. We also personally put the sensing of magnetic markers within this category, when the weight is not too heavy.

Now we give two personal realisations at LMA and UCL that show the difference between them.

2.5.3. Contact devices (tablets and joysticks)

At LMA we have been working on a project named "creative gesture in computer music" where several digital musical instruments have been designed using digital tablets and/or joysticks. The precision of such devices is great, they have been interfaced properly with Max-MSP in order to be in the core of a digital musical instrument. Constrained gestures help positioning a point, make “similii-writing” gestures (such as vibrato). The gestures are efficient, but not always beautiful or demonstrative. In most cases it is important to retrieve an intention from gesture in order to match it to an expression.

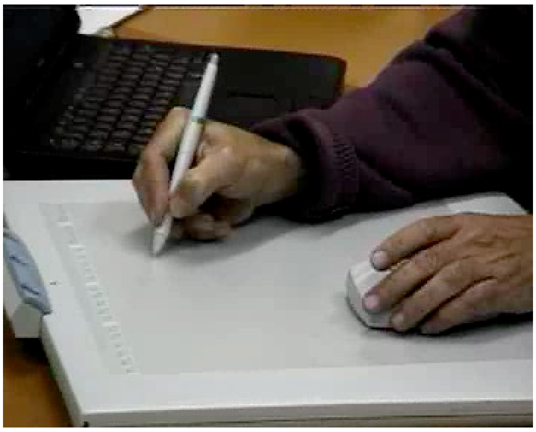

Figure 3. Vibrato gesture with the right hand.

2.5.4. Video and Gesture Recognition (UCL)

For a long time, many researches have been followed in image analysis and video segmentation fields, and notably at UCL [13]. Several systems are thus now able to extract from a video the position of hands, feet, facial characteristics and many more features (Fig. 4). These captured data are used as input parameters for several types of application, among these the gesture recognition.

Figure 4. Example of data captured by a video segmentation system.

At the Communications laboratory (UCL), the approach taken in research on gesture recognition is oriented toward recognizing the intentions of the person who performs the gestures, in order to analyze his or her global behaviour during long periods of time. The goal is about understanding the meaning and semantics of the gesture rather than simply giving a name to a recognized gesture. The treatment of position data is based on a Probabilistic Finite State Machine, modelled by a Dynamic Bayesian Network that is decomposed in many levels, each level treating a different level of abstraction, whereas the raw data form the lowest level. This gives the model the capability to deal with simple gestures made by single parts of the body, but also complex gestures performed by the whole body and over longer periods of time.

Utilization of gesture recognition knowledge in gesture-based musical creation environment already exists, especially in DIST-InfoMus lab (University of Genova, Italy) [8] where research has been conducted on the analysis of gesture expressiveness to drive musical creation. The UCL novel approach could bring a real additional value for applications in digital musical instruments driven by free gestures analysis. Indeed, it offers the possibility to build gestural control of such instruments relying more on the semantics level of the gestures rather than the syntactic one, as it is done until now.

2.6. Examples (not from the authors)

The sound synthesis techniques are quite old [30]**.** However very few has been implemented for the design of musical instruments using sonic textures. The link between texture and gesture is still a author-centered decision. Human-Computer interaction has also made a great use of “sonic icons”, some of them being textures, and it is worth seeing how they can be used. However we shall not present here a state of the art in these two domains, but merely show some samples of what is around.

2.6.1. Digital music instruments and textures

Many peripherals can drive MIDI synthesisers, and of course some presets yield sonic textures. The question there is to know which parameters can be controlled in real-time and what kind of mapping should be used. As an example, the Meta-instrument [15] has all the degrees of freedom to govern any sound, especially synthetic textures or digital audio effects giving rise to textures.

“The Hands” from Michel Waiswisz is another “classical” instrument now, at least in the hands of his inventor, and a large use of sampling techniques (with pitch shifting and time stretching or indexing) makes sounds alive, even in a theatrical sense.

An instrument really dedicated to textures is the “filtering string” designed by Couturier [2]. Here the principle is to have a graphical object, namely a string which has a dynamical behaviour (like a mushy series of masses and springs) which is on one side related to gesture (one applies forces via a 2D touch tablet) and on the other one to sound (the shape of the string is applied to an equaliser to filter a noise signal). Here we really get into a musical concept, which is even enhanced by a proper spatialisation [14].

Though many articles have been written on a possible “granulation” of sound samples, instruments really using such algorithms are few. Loic Kessous has designed such an instrument named Arpgran [22] where an excellent mapping between peripheral data and parameters for analysis-synthesis allows a musical feeling (Fig. 5).

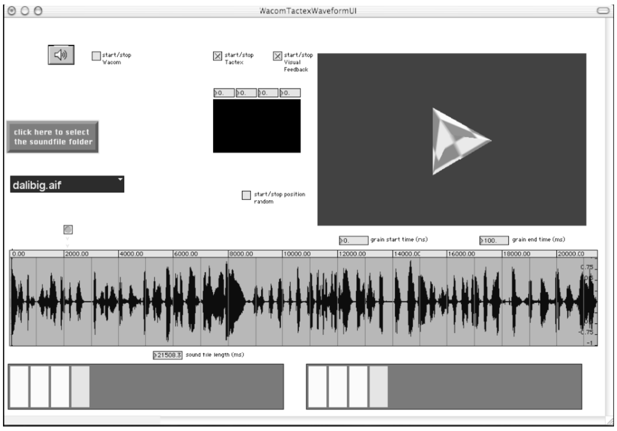

Figure 5. Graphical interface for the Arpgan instrument

Dancers equipped with sensors (or video captured) are subjected to experiments where the sonic soundscape is either synthetic or natural. Video settings can also help as devices that zoom. Travelling effects in soundscapes which are more easily rendered when images correspond to sounds (for example an artificial fire in both the image and the sound).

The simulation of DJ scratching can be considered as a texture-making [6] and devices simulating shakers of every sort linked to grainy sounds such as the one provided by the percolate toolkit (rain stick, shakers, etc…) can be considered as textural instruments.

Many other examples can be given, but the goal, as said before, is only to give a hint of the possibilities.

2.6.2 HCI and textures

Human Computer Interaction has shown the power of sound, and especially of textures in some of its applications.

The first one is the use of sonic icons. The relationship between some actions and some sound is quite straightforward: sounds are triggered by actions.

Some actions can also be “sonified” in another way. For example scales around a computer window can be sonorised, which can help blind people or users whose the vision is already attached to a task [21].

Projects like Sound Object have shown the importance of ecological relevance in the sonification of computer processes, and the sound ball can be truly considered as a musical instrument. An interesting aspect is the use of textures (Fig. 6) in collaborative environments [25]. Though music is not the intended goal, the rumour engendered by the throwing of virtual objects can be part of a new ecology of sounds.

Abb. 26: Einfache Anordnung zur Demonstration der verschiedenen Kategorien von Oberflächentexturen. Bei der Interaktion mit dem Stift auf den verschiedenen-farbigen Flächen werden unterschiedliche Textur-Kombinationen hörbar.

Figure 6. Surface scratching in MullerFelde’s thesis

Nevertheless, though the emotive part comes sometimes in account (for example for alarms) the semantic side is often bigger than the aesthetical part, and music is a by-product rather than an essential part of these HCI systems.

3. PERSONAL EXPERIMENTS

In this section we develop experiments we have personally done using instrumental gestures to interact with sonic textures. We shall see in this section the term « ecological gestures », which describes familiar gestures that people use in their daily life (though writing or using hammers are ecological gestures though they are learned). The term ecology is relevant, since it has been used in acoustic ecology, and has been defined in the context of perception by researchers such as Gibson [18]. The term « ecological gestures » means gestures properly linked to an anatomical comfort and a proper cerebral effort. A good example of « gestes écologiques » in HCI can be found in [24].

We now present some experiments the authors have been conducting at LMA using simple synthesis algorithms written in Max-MSP.

3.1. Examples (LMA I)

3.1.1. Using filtered noise

An interesting class of sounds comes from the filtering of a noise source. The reason for this is that we have at least two sorts of sounds that are easily mimicked by such source-filter algorithm: windy sounds, and whispering. Analog music has done a great use of noise generators and voltage controlled filters, so that we are really prepared for the sonic experience. But very few experiments have tried to link such sounds to gestures.

We use a Max-MSP patch, and a Max Mathews’s drum (also called radio-baton), which has the advantage of providing x,y,z in terms of Midi codes. Using the sound to gesture strategy, one has to invent gestures that can "symbolize" the sound we want to hear. Here are three different uses of the same algorithm, with different gesture strategies.

The gesture in the first instrument is a combination of a decision gesture and a modulation gesture: the initiation of the sound comes when a baton hits the surface. The x position of the hit point determines some parameters of the filters. The way the sound is generated depends upon the y position, which acts as an index in different tables, including the amplitude function. The way the sound ends depends upon the gesture. This gesture is very intuitive, because in fact we very rapidly use the "percussion-resonance" mental scheme. If we have a good combination of filtering values, one can have a musical instrument tuned to certain frequencies, for example harmonics of a drone. This is the way is has been employed in the real time version of "le Souffle du doux" (Fig.7).

Figure 7. Start and unwrapping of a filtered noise

3.1.2. Breathing gesture

Here we find another metaphor: breathing is the alternation of two windy sounds, one for inhaling and one for exhaling. One sound is linked to the right hand, and the other to the left hand. We rediscover here what are accompanying gestures: not everything is important for the control, in fact we can even trigger the sound when the y coordinate crosses a line (with a special hysteresis algorithm in order not to retrigger the sound due to some jitter of the gesture sensing) (Fig.8).

By subtly varying the coefficients of the filters, one can give the impression of a soft or deep breathing. Ideally this could be put as an additional value assigned to a controller (e.g. a foot pedal). However in the musical configuration it was meant for, a counter was incremented each time an alternation was done.

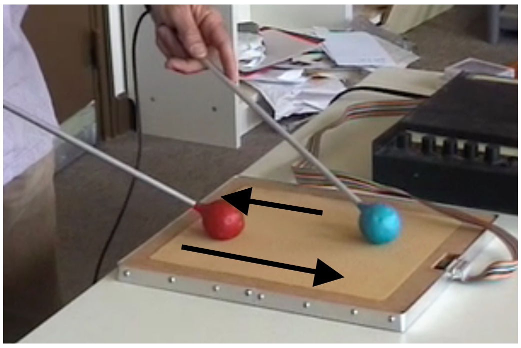

Figure 8. Alternative gestures for a breathing effect

Though the mapping may look very primitive, the gestures are very natural and one really has the impression of being part of a sonic process

3.1.3. Metaphor of the Demiurge (prince of the wind)

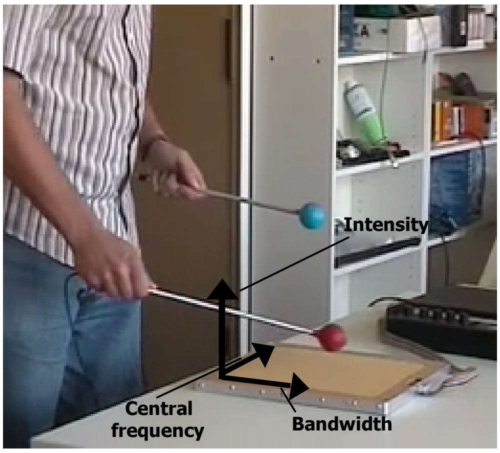

Figure 9. 3D exploration of sonic textures

Here we are in a 3D space where amplitude, central frequency and bandwidth are directly mapped to the x,y,z coordinates (Fig.9). Two sticks are used with the same algorithms. Strangely enough, this instrument immediately creates a "pedagogy" of gestures: trajectories are found that express different feelings, or expressions of sounds. One is really a "creator", hence the metaphor of the Demiurge,

3.1.4. Drone textures and stick gestures

A drone sound is created by adding three oscillators with very different values that are waveshaped in a specific way. This gives a choir effect on a simple sound, so the harmonics are beating in a kind of anarchic way. The mapping itself uses the vertical position as an index for distortion (the closer the stick, the more distorted the sound). The horizontal position of the other stick is directly linked to the frequency discrepancy between the oscillators (Fig. 10).

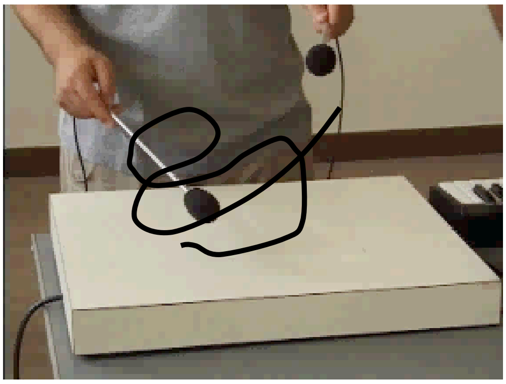

Figure 10. A 3D movement, but only one coordinate is used for each stick

It is very interesting to note that, although the vertical coordinate only is linked to the sonic process, one feels the need to play in 3D. This can be explained by the fact that a sine (or cosine) is only the projection of a circle, and that a circle may be more « ecological » than a hand oscillation when it comes to slow frequencies. Once again the sonic feedback immediately tells the performer the good regions for the two sticks. Initially devised for the proper imitation of an electroacoustic piece, the device incites to exploit the sonic material and reinvent new curves, new ways of playing. This is well known, but a musical instrument is not only a gestural control of a process, it is an « sonically output oriented » loop, the goal is to make a sound, and the rest is part of the loop.

3.2 Examples (LMA II) : Textures scratcher

In this section, we describe another digital musical instrument recently developed at the LMA. Unlike the previous instruments that were based on a noise filtering approach, the “texture scratcher” does not rely on source-filter model but rather acts directly on the source itself in order to create “chaotic textures”. Following the typology proposed by Hanna [19], sounds produced by this instrument would belong either to the “pseudo-periodic noise” or the “impulsive noise” class according the mode chosen.

This digital instrument is based on the gesturalized exploration of a visual space. It consists of a real-time adaptation of the Functional Iteration Synthesis (FIS) [16] implemented with Max/MSP driven by an advanced gestural control using a graphical tablet and a joystick. FIS is a special case of wave terrain synthesis where terrains are generated by iteration of non-linear functions. This section is divided in three parts: first part introduces the original algorithm proposed by Di Scipio to create sonic textures from fractal wave terrains. The second part describes our implementation, and especially highlights the two mapping strategies we have developed for the exploration of the terrains. Finally, considerations on musical applications and future research direction for this instrument are evoked in the last part.

3.2.1 Creation of the wave terrains by Functional Iteration Synthesis

Functional Iteration Synthesis (FIS) is a part of the wide class of wave terrain synthesis (WTS) [29], where the sound waveform corresponds to an orbit traced on a three-dimensional surface (the *wave terrain*). Characteristics of sounds produced with this type of synthesis techniques depend on both terrain properties and orbit velocity.

In Functional Iteration Synthesis, Di Scipio proposes to use fractal images as wave terrain to take advantage of their very dense and complex relief. In this intention, he suggests building terrains by iteration of non-linear functions, and takes the example, that we have followed in our instrument, of iteration of sinus function.

Given (x,y,z_n) the coordinates of the points composing the n-th wave terrain and I_x and I_y the definition domains of x and y respectively. The creation of the initial wave terrain is achieved by the following expression, where the elevation z_0 of each point is computed from its two other coordinates x and y:

$$z_0(x,y) = \sin(x*y) = f_0(x,y) \qquad (1)$$

$$\text{with } x \in I_x,\ y \in I_y \text{ and } z_0 \in [-1;1]$$

Next terrains are then calculated from (1) by an iterative process:

$$z_1(x,y) = \sin(x*z_0) = \sin(x*(\sin(x*y)) = f_1(x,y)$$
$$z_2(x,y) = \sin(x*z_1) = \sin(x*(\sin(x*(\sin(x*y))) = f_2(x,y)$$
…

until the n-th terrain, corresponding to the n-th iteration :

$$z_n(x,y) = \sin(x*z_{n-1}) = f_n(x,y) \qquad (2)$$

Sound signal s(t) is finally obtained by tracing an orbit on the n-th terrain, that is done by varying (x,y) according to time in (2) :

$$s(t) = z_n(x(t), y(t)) \qquad (3)$$

Three parameters, namely the definition domains I_x and I_y for both x and y dimensions and the number of iterations achieved *n*, are necessary to define a wave terrain. Typical values used in our instrument are $I_x=[-\pi/2, \pi/2]$, $I_y=[1,4]$, and $n<10$.

Wave terrains can be represented either in two or three dimensions (Fig. 11), the third coordinate (elevation z) being symbolized in grey level in 2D image. For the visualization of the terrain in our instrument we have chosen the 2D representation, more usable for very dense relief.

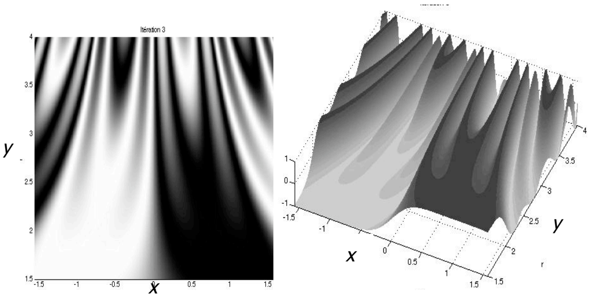

Figure 11. Two representations of the same wave terrain in 2 and 3 dimensions (x ∈ [-π/2; π/2], y ∈ [2 ; 4], n= 3).

As illustrated in Fig.12, each new iteration provides a new wave terrain with more and more complex relief. Also this technique allows creating very complex terrains, potentially worthwhile for wave terrain synthesis, by a quite simple process involving a limited number of specified parameters.

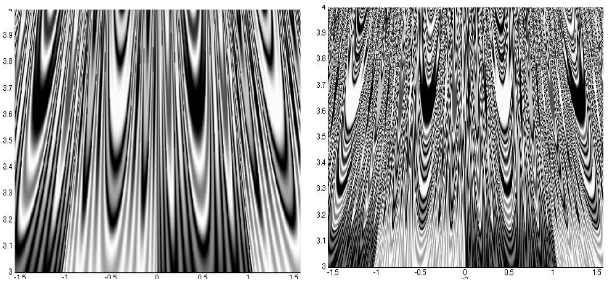

Figure 12. Representations of the 5th and 7th iteration (x ∈ [-π/2; π/2], y ∈ [3 ; 4]). The larger the number of iteration is, the more complex the terrain becomes.

In our instrument, wave terrains are computed and displayed in real-time thanks to Jitter, an additional library of objects enabling manipulation of matrix data and dedicated to image processing in Max/MSP. The terrains are displayed on an “interactive pen display”; developed by Wacom, this is an improved 15’’ graphical tablet that integrates display functionalities. Thus, we use this “tablet-screen” to make the user able to trace orbit directly on the image of terrain (Fig. 14).

3.2.2. Generation of the trajectories

From the algorithm proposed by Di Scipio, we have built a digital musical instrument by adding real-time user-interaction based on an advanced gestural control. The gestural control we built for this instrument is inspired by the metaphor of a surface scratching, and it allows the user to control the trajectories traced upon the terrain. Two modes of exploration of the terrain are possible (Fig. 13), either by means of linear trajectories (« direct mode ») or looping trajectories (« parametric mode »).

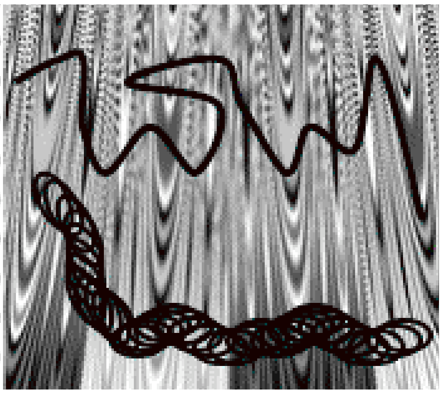

Figure 13. Examples of orbits generated by direct control (linear orbit, above) and parametric control (looping orbit, below)

a- Direct mode

In direct mode, the orbit corresponds to the actual movement drawn by the user on the tablet as if he or she were scrubbing a « sonic surface ». This mode is based on a uni-manual gestural control. Practically the user traces a trajectory with a stylet on the tablet where the 2D image of a wave terrain is displayed (Fig. 14); every 30 ms, a pair of coordinates (x,y) corresponding the position of the stylet on the tablet is captured, rescaled to the terrain dimensions, and transmitted to Max-MSP. A 44100 Hz sampled trajectory is then generated in Max/MSP by linear interpolation between two successive captured positions.

The way we generate sound waveform from orbit and terrain data is quite different from in traditional wave terrain synthesis techniques: instead of reading values in a bidimensional table associated to the terrain, we instantaneously compute each sample of the sound signal from the expression (3) for each couple (x,y) constituting the orbit. This is made possible because we know the mathematical expression defining the terrain. By this way our approach is much closer to a waveshaping synthesis method [1], than a traditional wave terrain synthesis method. In waveshaping synthesis, the output signal is the result of a function *f* applied to the result of another function *g*; we can also consider our method as a particular case of “iterative waveshaping”, and in this approach, the terrain as a graphical representation of the iterated waveshaping function. This induces an interesting property on the synthesized sound: indeed, according non-linear distorsion synthesis theory, the distorsion of a sinusoid by a k-order polynomial gives a k-order harmonic signal. In the case of an “iterative distorsion”, the signal presents after n iterations a k^n–order harmonic structure. Consequently, even after a small number of iterations, spectral structure of sounds produced by this algorithm will be characterized by a lot of foldover components. These foldover components give a very peculiar “crunchy” character to the sonic texture.

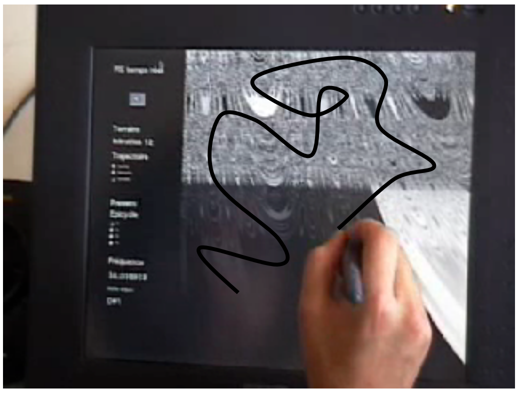

Figure 14. Unimanual gestural control used in direct mode

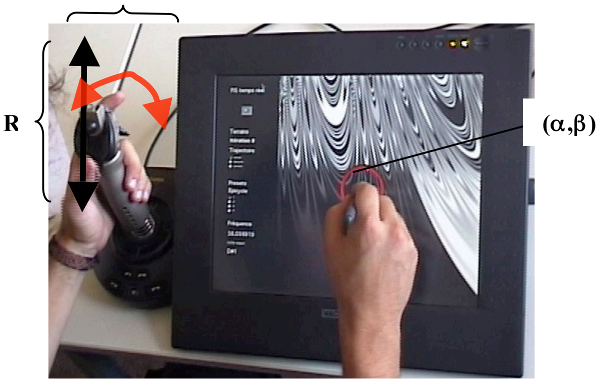

Figure 15. Bimanual gestural control used in parametric mode

In the direct mode, spectral features of the textures are directly linked to the hand motion: slow gestures will create low frequency textures with a pseudo-rhythmic structure, whereas faster gestures will enlarge the spectral content of the sound and induce lot of spectral aliasing to render chaotic textures.

b- Parametric mode

Parametric mode differs from direct mode by the fact that terrains are no longer explored by linear orbits but by rather circular/elliptic orbits that loop on themselves at a controllable frequency (Fig.13). The user governs the overall position of the circle on the terrain by moving its center and varying its radius thanks to an additional peripheral (a joystick). This mode is called parametric because the generation of the trajectories is not anymore done directly but by means of control parameters.

In this mode, coordinates (x,y) of the points composing the orbit are computed in Max/MSP from the parametric equation of a circle :

$$x=\alpha+R*\cos(\omega t) \qquad (4)$$
$$y=\beta+R*\cos(\omega t)$$

where R is the radius and (α,β) the coordinates of the center of the circle. (ωt) is governed by a saw tooth function varying between 0 et 2π to move periodically around the circle. The velocity of the orbit is also directly dependent on the frequency f of this sawtooth function. The sound signal is then obtained from the expression (3) of elevation z for each couple (x,y) calculated with circle equation (4). A similar process makes it possible to generate elliptic orbits.

A bimanual gestural control allows the user to move upon the terrain by varying each parameter used for the construction of the trajectory: the coordinates (α,β) of the center of the orbit are given by the position of the pen on the tablet, whereas the radius R and the velocity (by the way of frequency f) are increased or decreased by front/back and twist movement of a joystick respectively (Fig.15).

3.2.4. Evaluation and conclusion

The mappings associated to each mode are of different natures: in the first case, a direct link exists between the hand gesture and the position on the terrain; this mapping relies on an ecological "innate" gesture (surface scratching). The utilization of the joystick in a "assistant gesture" in the second mapping adds an intermediate layer between hand motion and the construction of the trajectory, that then becomes indirect.

Sonic textures obtained by both modes have a very peculiar "chaotic" character, especially due their many foldover components. The parametric mode allows creating different textures than the direct mode, especially pseudo-pitched textures when frequency f is in audible range. These are close to certain machine-noise sounds and well suited to render very nervous and turbulent sonic ambiances. For further musical applications, using these sounds as source in a filtering stage could give interesting results in creating "coloured chaotic textures". This is one direction we shall follow in future research.

3.3 Discussion

Our way to link gesture and sound has been the following: instead of using devices (computer peripherals) and asking what can we do with them, we prefer to think about sounds and ask ourselves: what kind of gesture would be best to produce this sound with a new digital musical instrument. This does not mean at all that we will necessarily try to dance or mimic an existing instrument, but rather that we will make a compromise between a natural gesture and an educated gesture the gesture itself will have to provide, through a specific mapping- all the values for sound definition. A natural gesture is what "comes into mind" whenever you listen to a sound. An educated gesture is a gesture you can learn, reproduce, vary around it so that the instrument becomes a "gesture-controlled audio system" and not a "gestural control of an audio-system". In terms of ergonomics and cognitive processes, this implies that the tool itself is incorporated in the physical and mental body of the performer.

Musical instruments usually rely on specific constraints that the material and the construction provide. This is the part of intervention of the "lutherie" in the sonic process. This means for example that in order to get a vibrato on a string one must make the finger oscillate. On a Theremin, vibrato is obtained by the vertical variation of height of the hand. On a digital music instrument such as the photosonic emulator, it is an oscillating scratching gesture that plays this role.

Sonic textures are however very specific in the way they can be part of such devices.

a- Sonic textures are rarely based upon the attack-sustain-release scheme of ordinary sounds. They are more like sound masses, where the attack part is diffused all over the sound and replaced by successive transients. Another specificity of textures is a loose importance of pitch, except for specific textures. More generally, control parameters are ambient rather than descriptive. Macroforms and microforms are sometimes indistinguishable. For example the sound of a river can be seen as the succession of events or as a simple granulation of a sound flow.

To see the difference between macroforms and microforms, let us consider the difference between a conductor's gesture which immediately draws a gestural sketch of the whole sound, while the way the percussionist carefully chooses the place where to strike, and the force he/she will use is typical of a meso/microlevel. Even more micro is the way one uses an analog synthesizer to control all the parts of a sound with the help of knobs. This clearly shows that there are choices to be made and they sometimes come from the way the sound is processed.

b- Sonic textures depend on algorithms which are somewhat different from classical sound. In a source-filter approximation, we can see that the source itself is a succession of different accidents and the time specification of possible transients versus noise is critical. The modelisation of a proper source is fundamental. The spectral domain is very important too as it will impart textures with their real musical meaning. It is extremely important to have a good relationship (mapping) between the gesture and the spectral control.

There is not a unique solution for the gestures that we can use. Manipulating a white noise through a filter can lead to diverse instruments because of the structure we superimpose to the sonic calculation. As an example sonic objects may either be finite short entities or a long unique event where the matter of sound is molded. Gestures associated with either will be of course different. Even with the same type of sounds, we can choose for example to link the time axis to a dimension of gesture (one unwinds the sound) or not (once an event is triggered, it will proceed to its end and one can only change other parameters)

Other choices can also lead to a bimanual gesture, where two component of the sound rendering can be associated to two concomitant gestures. This is perfectly illustrated by the "voicer" [22], a non-textural instrument, but for sure this can give some ideas about the link between a source and a filter.

c- The main specificity of sonic textures is the symbolic level it makes reference to: most of these sounds can be called ecological, and the gestures associated to them will require to be very close to natural gestures such as scratching, defence-attack movements, motions in the air. They usually have a dynamic structure which needs to be recognized and matched with specific gestures. This is also why they can be used in choreographic applications

Gestures may be redundant, and it is in this sense that they can become "ecological". As an example a free 3D gesture using a Max-drum is easier, and more sensitive, while only the height of the tip versus the surface is considered in the sonic process. But ecological gestures also have a symbolic correspondence which makes them vivid. As an example the manipulation of a sword is an ecological gesture, and it still works without any enemy (but an obstacle, real or virtual, is needed for striking sounds). When using such gestures, we come to another field, which is the one of emotion: many gestures are connoted with an emotion. Rocking a cradle is not an innocent gesture. In fact we enter a domain where other arts have set markers: theatre and dance use gestures charged with emotion, be they symbolic or not. "Gracious gestures", whatever the style, are indeed ecological gesture because usually they minimize some jerking, so as to have a smooth side. Contemporary dance sometimes tries to get out from this but for sure gestures remain gesture and it is the overall structure that can become anarchic, not the gestures themselves.

4. CONCLUSION AND PERSPECTIVES

Gesture is not gesticulation. Gesture is constrained by two things: the first one is the feasibility of this gesture, and this constraint is linked with ergonomy research. The second one is that the gesture is linked with the sonic result, and it is aesthetically constrained. The definition of a domain and trajectories inside the domain is a quite unexplored field and it has much to do with the double meaning of a musical gesture: it is an action and perception movement.

Although many musical experiments use gestures and textures, there is no real state of art of this powerful combination or alliance. While it has brought up some experiments done in this field, this article is not a state for art either, and this is a good perspective for the future.

Finally, the evaluation of such links between gesture and texture must not let us forget an important thing: music is an intense process, not limited to algorithms and mappings; there is an intense implication of emotion, and one always has to remember that textures are not merely "soup music" but can be an awakening of the senses. This means that a musical point of view always must be the guardian of computer sonic research.

5. ACKNOWLEDGEMENTS

Authors would like to thank Kosta Gaitanis and Pedro Correa, PhD students at Communications laboratory in UCL, for their help on the writings of gesture recognition part. Many of the ideas concerning digital musical instruments would not have risen without the help of Jean-Michel Couturier, Loic Kessous and Vincent Verfaille who successfully passed their PhD at LMA in the two last years. Many thanks to all the ConGAS delegates of COST287 action for their helpful conversations and comments.

6. REFERENCES

[1] Arfib D. "Digital synthesis of complex spectra by means of multiplication of non-linear distorted sine-waves", *Journal of the AES*, 27 (10), 1979.

[2] Arfib, D., Couturier, J.M., Kessous, L. "Gestural Strategies for specific filtering processes", Proceedings of *5th International Conference on Digital Audio Effects DAFx 02*, pp. 1-6, 2002, Hamburg, Germany.

[3] Arfib D., Couturier J.M., Kessous L., Verfaille V., "Mapping strategies between gesture control parameters and synthesis models parameters using perceptual spaces", *Organised Sound 7(2)*, Cambridge University Press, pp. 135-152 .

[4] Athineos M., Ellis D. P. W., "Sound texture modelling with linear prediction in both time and frequency domains", *Proceedings of the International Conference on Acoustics Speech, and Signal Processing,* Hong Kong, 2003.

[5] Bar-Joseph Z, Lischinski D., Werman M.,Dubnov S., El Yaniv R., "Granular synthesis of sounf textures using statistical learning", *Proceedings of the International Computer Music Conference*, Beijing,1999.

[6] Bresin R., Hansen K.F, Högskolan K.T., "Complex gestural audio control: the case of scratching", *SOB Book*, freely available at http://www.soundobject.org/

[7] http://www.music.mcgill.ca/musictech/MT_Staff/Wanderley/Trends/P.CadWan.pdf

[8] Camurri A., Coletta P.,MassariA., Mazzarino B., Peri M.., Ricchetti M., Ricci A., Volpe G., "Toward real-time multimodal processing: EyesWeb 4.0", in *Proceedings AISB 2004 Convention: Motion, Emotion and Cognition*, Leeds,UK, March 2004.

[9] Cardle M., Brooks S. "Directed sound synthesis with natural grains", *Proceedings of the Cambridge Music Processing*, Cambridge University, 2003

[10] Cardle M., Brooks S., Bar-Joseph Z., Robinson P., "Sound-by-numbers: motion driven sound synthesis", *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer animation*, 2003

[11] Cadoz,C. and. Wanderley M. Gesture-Music. "Gestures in Music" In M. Wanderley and M. Battier (eds.) Trends in Gestural Control of Music". Paris, Fr: IRCAM - Centre Pompidou, 2000 available at http://www.music.mcgill.ca/musictech/MT_Staff/Wanderley/Trends/P.CadWan.pdf

[12] ConGAS, COST action n°287 "Gesture-controlled audio systems": http://www.cost287.org

[13] Correa Hernandez P., Czyz J., Umeda T., Marqués F., Marichal X., Macq B., "Silhouette Based 2D Motion Capture for Real-Time Applications", *IEEE International Conference in Image Processing*, Genova, 2005.

[14] Couturier, J.M., "Etheraction: Playing a Musical Piece Using Graphical interfaces", *Actes de la de la conférence internationale "Sound and Music Computing" (SMC 04)*, pp. 213-218 (6 pages), Paris, octobre 2004.

[15] De Laubier, S., "Meta-instrument description" http://www.lagrandefabrique.com/pages/meta_instrument.html.

[16] Di Scipio A., "Synthesis of environmental sounf textures by iterated non linear functions and its ecological relevance to perceptual modeling", *Journal of New Music Research*, Vol. 31 p. 109-117, 2002.

[17] Dubnov S., Bar-Joseph Z., Yaniv R.E, Lischinski D, Werman M. 'Synthesizing sound textures through wavelet tree learning", *IEEE Computer Graphics and Applications,* vol.22, no. 4, pp.38-48, Jul/Aug 2002.

[18] Gibson J-J., "The ecological approach to visual perception, Boston: Hougton Mifflin, 1979.

[19] Hanna P., Louis N., Desainte-Catherine M., Benois-Pineau J., "Audio features for noisy sound segmentation", *Proceedings of International Conference on Music Information Retrieval (ISMIR)*, Barcelona, 2004

[20] Hoskinson R., Pai D, ”Manipulation and resynthesis with natural grains”, *Proceedings of the International Computer Music Conference*, Havana, 2001.

[21] IEEE Multimedia special issue on Interactive Sonification

[22] Kessous L. and Arfib D., “Bimanuality in alternate musical instruments”., *Proceedings of NIM03*, pp 140--145. Wanderley~M. (ed.) (ed.), Mc Gill University, 2003.

[23] Lie Lu, Liu Wenyin, Hong-Jiang Zhang, “Audio textures: theory and applications”, *IEEE transactions on speech and audio processing*, vol.12, no. 2, pp. 156-167, March 2004.

[24] Mertz E., Vinot J-L., Etienne D., “Entre manipulation directe et reconnaissance de l’écriture : les gestees écologiques”, http://www.tls.cena.fr/divisions/PII/Rapports/ergoihm2000.pdf, 2000.

[25] MüllerFelde Christian, Sounds@Work “Akustische Repräsentationen für die Mensch-Computer Interaktion in kooperativen und hybriden Arbeitsumgebungen”, available at http://elib.tu-darmstadt.de/diss/000313/

[26] Oliveira, André L. G. and Oliveira, Luis F. “Toward an ecological conception of timbre”, *Proceedings of Auditory Perception Cognition and Action Meeting, Kansas* City, 2002.

[27] Parker J.R, Chan S., ”Sound synthesis for the web, games, and virtual reality”, *Proceedings of the SIGGRAPH conference on Sketches & applications*, 2003.

[28] Roads C., ”Automated granular sound synthesis”, *Computer Music Journal 2(2)*, pp. 61-62

[29] Roads C., ”The Computer Music Tutorial”, *MIT Press* Cambridge, Massachusetts, 1996.

[30] “The Csound Book Perspectives in Software Synthesis, Sound Design, Signal Processing,and Programming”, *Edited by Richard Boulanger, MIT Press*, 2000.

[31] Wei L.Y., Levoy M., “Fast texture synthesis using tree structure vector quantization”, *Proceedings of ACM SIGGRAPH Conference*, New Orleans, LA, 2000

[32] Zhu X., Wyse L., (2004) ”Sound texture modelling and time-frequency LPC”, *Proceedings of the 7th International Conference on Digital Audio Effects DAFx 04*, Naples, Italy, October 2004.