INTERACTIVE DIGITAL AUDIO ENVIRONMENTS: GESTURE AS A MUSICAL PARAMETER

Sylviane Sapir

Via Cimate, 92
03018 Paliano (FR) - Italy
syl@rtmol.it

ABSTRACT

This paper presents some possible relationships between gesture and sound that may be built with an interactive digital audio environment.

In a traditional musical situation gesture usually produces sound. The relationship between gesture and sound is unique, it is a cause to effect link. In computer music, the possibility of uncoupling gesture from sound is due to the fact that computer can carry out all the aspects of sound production from composition up to interpretation and performance.

Real time computing technology and development of human gesture tracking systems may enable gesture to be introduced again into the practice of computer music, but with a completely renewed approach. There is no more need to create direct cause to effect relationships for sound production, and gesture may be seen as another musical parameter to play with in the context of interactive musical performances.

1. INTRODUCTION

The study and the analysis of human gesture in music area for the control of digital instruments, along with standardization requirements coming from the industry have led to the birth of MIDI devices and their communication protocol. This protocol has often been judged insufficient and too much limited to the pianistic gesture reproduction [1]. This is even more true today as the relationship between composer and technology is no more exclusive but is extended to other forms of artistic projects in which multimedia interact with other forms of expression such as dance, theatre or video art - predilected fields of body communication.

The development of interactive audio environments which capture and analyze movements of one or more performers in real time to arbitrary control sound processing systems allow the traditional aspects of composition and sound synthesis in computer music to be connected with those less codified of the body expression. In this way it'll be possible to establish new gestural interactions in the perspective of an extended musical creative process which involves the spectacular, affective and emotive aspects of the human gesture.

2. SOUND AND GESTURE

The possibility of creating music material in all its details by physically describing the sound event and by taking into account the limits of our perception, has freed us from the mechanical constraints of traditional musical instruments.

Furthermore, the possibility of producing music in a different place and at a different moment in which it may be listened has incredibly powered up the possibilities of artistic and creative processes.

Finally, the possibility of detaching the control of sound from its production either in space or time opens new horizons to composers and performers.

2.1. Computing sound

The progress of digital audio technology of these last 20 years is so impressive that we may wonder how it will continue and what its impact could be on music production and other forms of creativity.

J. A. Moorer in an AES lecture [2] foresees that the main problem facing digital audio engineers will not be how to perform a particular manipulation of sound, but how the amount of power that will be available can possibly be controlled.

In effect, the growth of processor speed, the development of new processors architecture and networking capacity, the reduction of memory access time, the increasing volume of hard-disk storage, etc. leads Moorer to assume that in 20 years we should be able to process 9000 channels of sound in real time, doing million-point FFTs and that processes that were considered unfeasible because they were too complex will become matter of fact.

But what will happen if a mixing console has 2000 or more controls on it , or if a digital audio workstation has two million virtual sliders and buttons? The only viable solution, according to Moorer, will be the rise of "intelligent assistants".

2.2. Control and automation

Intelligent assistance is the crucial point of computer music, and it is fundamental for the control of real time performances.

Nowadays it is possible to interact with a digital system through its user interface which can convey information through many communication channels such as hearing, sight and touch. But as seen before, the number of parameters to be managed simultaneously may be very big which makes real time control a complicate task for performers.

To overcome this problem more solutions have been found such as grouping parameters or by introducing automation in performances. A software layer is always necessary between gesture acquisition and sound production: gesture management software and performance software.

Performance software includes sequencers, arrangers, score followers and other rule-based applications which allow computers to be charged with specific tasks in order to relieve performers from some low level controls. Dividing control between computer and performers allows one to go beyond traditional performance modalities

When gestural data follows the MIDI format, it is easy to find some program that can manage it, like MIDI sequencers and more general development environment such as Max [3]. But other environments like PLAY [4], HMSL [5], or ARCTIC [6] have been specifically developed for interactive music performance systems. These environments support dedicated languages that allow one to define and handle musical control flow, i.e. a sequence of instructions that manage musical and gestural events in a formal way.

2.3. New meanings for gesture

Musical instruments are the most natural and traditional interfaces used by musicians to perform music. But technology has brought to light new instruments that can transduce human motion into electrical signals, and then pilot any kind of sound processing.

As described by T. Winckler [7], pioneering work on gestural input devices was carried out during the analog electronic music period. The invention of the Theremin anticipated the development of new expressive instruments. It is noticeable that the first musical instrument, a simple oscillator, performed without physical touch could produce very subtle and varied sounds because the sound reflected more the expressive quality of human movement than its own timbre quality.

A main question in using new input devices is whether to use controllers based on existing instruments or innovative ones. The main advantage of using the former is that musicians are already familiar with them and can easily exploit their performance skills learnt over years of practice. But gesture is no more linked to sound production and traditional expressions of virtuosism hardly find place in today's music.

On the other hand, new models may be simpler or more intuitive to use for non skilled performers and allow the introduction of new gestures which better fit control features of digital sound processing - mixing for example.

Ergonomics, mechanical and temporal precision, flexibility and programmability are the characteristics more often required for new controllers. In the mean time innovative input devices make performances more spectacular and may extend expressive possibilities of computer-based interactive music systems.

Finally the selection of the gestural input device, and consequently those of the category of gesture to be used for the performance belongs more to a musical or a compositional choice than to a technological one - gesture becomes part of the new musical parameters.

3. GESTURE TRACKING TECHNOLOGY

The explosion of new devices that enables capturing and tracking movement is symptomatic of a great deal of interest from many fields. The development of new interactive systems based on gesture is concerned with both industrial and artistic applications just like virtual reality and multimedia applications.

3.1. Requirements for a tracking system

As described by A. Mulder, human movement tracking systems can be classified as inside-in, inside-out and outside-in systems [8].

Inside-in and out systems employ sensors on the body while outside-in systems employ external sensors. Both types have strengths and weaknesses. For example, the former are generally considered more obtrusive than the latter.

But it's the category of gesture that has to be captured and will determine the kind of system to use [9]. In the case of instrumental gesture, it is possible to directly capture the gesture through sensors. MIDI devices belong to this category. It is also possible to indirectly capture gesture by analyzing the sound produced by an instrument and deducing information about the gestures applied by a performer to generate it. In this case, it is necessary to provide very sophisticated programs that extract significant gestural/audio information in real time.

The following figure describes a very generic overall tracking system:

Figure 1. *Scheme for a human gesture tracking system*

3.2. Some hardware technology

Any technology that can transduce human gesture or movement into electrical signal is available to build sensors. Regardless of sensor type, it is important to recognize not only what is being measured, but how it is being measured: resolution and sampling rate are fundamental aspects to take into account. The commonly used technologies include infra red, ultra sonic, hall effect, electromagnetic and video.

Keyboard, switch, pushbutton, slider, joystick, mouse, graphic tablet, touch-sensitive pad, etc. belong to the computer hardware. However, although such devices have been intensively used as gestural input devices for musical purpose they hardly fit musical gestuality requirements.

On the other hand, musical keyboards, foot pedals, drum pads, ribbon controllers, breath controllers and other instrument-like controllers belong to the MIDI world. They have been exclusively developed for musical aim and capture musical gesture on the scheme of the MIDI Protocol, but they are sometimes used for other purposes, lighting controls for example.

Many other technologies have been developed to sense and track a performer's movements in space, and not only for musical performances. On-body sensors measure angle, force, or position of various parts of the body. For example, piezo-resistive and piezo-electric technologies are very common for tracking small body parts such as fingers through the use of gloves. Spatial sensors detect location in space by proximity to hardware sensors or by location within a projected grid. The Radio Drum, for example, measures the location of two batons in three dimensions in relation to a radio receiver [10]. Videocamera-based technologies are used to track larger marked body parts and are often used to integrate music and movement languages such as dance or theater [11].

Force feedback devices developed at ACROE for interactive systems using physical modeling synthesis are very interesting attempts to offer tactile-kinesthetic feedback to performers [12]. By feeding back the sensation of effort, it is possible for the performer to have a finest control on its movements.

Many musicians have built their own input devices as prototypes by using microphones, accelerometers and other types of sensors combined to electronic circuitry. But all this hardware has sense if it is connected to the computer and managed by some software - the performance software.

3.3. What about software?

As described by C. Roads in its Computer Music Tutorial [13], at the core of every performance program is a software representation of the music. By working with DSPs to process audio data we have to deal with a lot of different types of parameters at a low level.

Some parameters have discrete representation, some other are continuous, some have time constraints, some not, etc. On the other hand gestural data also has its own internal representation. Performance software has to support all these representations, and to map sensor data onto algorithm parameters in a meaningful musical way, i.e. adapted to the styles of music and movement.

Timing is another focal point of performance software, as for any kind of reactive computing application. These applications require reliable, real-time performance compatible with intimate human interaction. Software research for the integration of robust temporal scheduling in operating systems, drivers, programming languages and environments is fundamental to provide good software tools that allow musical control flow description and interactive music applications without latency problems.

4. THE MAPPING CONCEPT

The second step after having captured some gesture is to decide how the gestural data will be related to sound processing. The mapping problem is at the center of the effectiveness of an interactive environment. To a certain extent, as described by T. Winkler, we can consider human movement in space as a musical instrument if we allow the physicality of movement to impact on musical material and processes, and if we are able to map physical parameters of movement on high level musical parameters down to low level sound parameters [7].

Adding some musical intelligence inside the control system implies thinking of the range of operations to deal with, the conversion mechanisms involved, the temporal independence of parameters, grouping or splitting gestural channels, adding causal or random transformations on gestural data, and adding design flexibility and programming facilities to the interactive environments.

Gesture may be used to produce immediate sound by triggering it or by directly controlling synthesis or other sound processing algorithms. But gesture may be used to control higher level processes that automatically generate variables for sound processing, introducing the concept of performance variations.

Gestural data may also be tracked in order to compare the actual performance with an ideal one stored into a score. The objective of this job is to follow and interpret in order to be able to produce some performance dependent accompaniment with the digital audio system;

Gesture may also have a conducting meaning, which help the control system to take some decisions and transform the state of the running processes during performances. This includes compositional processes and musical structure management.

4.1. Gesture analysis

First of all, human gesture has to be noise filtered, calibrated and quantized in order to have a normalized representation. Such data can then be analyzed in order to extract pertinent information.

Human beings do not reproduce their gesture with precision, and there is no general formalisation of movement that may help us to recognize its meaning. Gesture allows human to interact with their environment to modify it, to get information and to communicate. Gesture applied to a musical context may follow the same objectives: modifying musical or sound material, getting information on the performance and communicating information to other processes.

Segmentation, statistical analysis, pattern matching, neural networks, information retrieval are all examples of analysis techniques which can be used to extract meaningful information on gesture or on movement of various parts of the body. These techniques have to be implemented in real time and have to produce stable and reliable outputs on stage for the specific musical purpose of the performance.

But as for other domains, such as sound analysis-synthesis, it is maybe into a dynamic relationship between gesture analysis and some kind of synthesis feedback - auditive, tactile, or visual - that we can imagine better exploring and studying gesture with a particular focus on signification and emotional content.

4.2. Mapping strategies

Interactive musical performances require the simultaneous manipulation of inter-dependent parameters. A good mapping metaphor will help performers and the audience understand the effects of gesture on sound. A good mapping strategy will fulfill all the requirements of the interactive environment:

- Hardware and software robustness with reliable timing response;
- Data reduction and significant perceptive results according to movement changes;
- Practice facilities for the performers to learn the idiosyncrasies of the system;
- Quick prototyping facilities for rehearsals;
- Different abstraction levels of data representation and mapping for visualizing and editing both gestural and musical parameters.

We can consider five different ways to map or connect gestural data to sound parameters.

One-to-one connections are typically used for selecting formatted or preset sound parameter values, or for dynamically updating sound parameters. In the former case, the current value of the gestural parameter is used as a branching condition to select materials. In the latter, gestural parameter values need to be converted in order to be adapted to the range of sound parameters extension. This can be done in an absolute mode, as for the amplitude, or in a relative one, as for a vibrato variation.

Convergent connections are used to get more information on movement or to add control dimensions to a higher level sound parameter. Pitch, for example, may depend on a fundamental frequency driven by a key pressure, by some frequency deviations coming from some gliding device and by the intensity which may be controlled through some other device. In this case too, conversion rules are necessary to fit gestural parameter range to those of sound.

Divergent connections are used to reduce the number of input devices and to optimize the control process. This mapping strategy is often used for commercial digital music instruments. A typical example is the scaling mechanism on the keyboard extension to manipulate envelope profiles, to balance amplitudes or to adjust spectrum bandwidths of generated sounds according to the selected pitch.

Metaphors may be a good approach to solve multidimensional control aspects of a sound. A. Mulder demonstrates the validity of introducing some virtual objects or surfaces, like spheres or sheets, manipulated with hands through instrumented gloves that measure hand shapes [14]. The metaphor used in this case is sound sculpting and it has been studied for helping sound designers in their difficult job of adjusting and editing sound algorithm parameters. The implementation of such mapping is very computer expensive and requires a sophisticated programming environment - in this case MAX/FTS has been chosen.

Rules which link gesture to sound or musical parameters may also be chosen arbitrarily. They may be considered as part of the compositional material of an music piece.

4.3. Data synchronization

Real time processing does not mean updating sound parameters at the moment gestural data is available. On the contrary, in a musical system performers actions may introduce delayed reactions. They also may be analyzed over longer spans of time to extract information on the direction of the movement, on its repetition, and on other kinds of patterns.

Immediate actions are used to trigger preset parameters or to dynamically update parameter values using the one-to-one mapping strategy.

Delayed actions may be used as a compositional or musical choice, such as delayed sounds in live electronics applications. Other delayed actions may be necessary when waiting for some other events to introduce synchronization mechanisms between processes.

5. MUSIC V, 4I and MARS

5.1. Computer score performance

During the author's working period at the Centro di Sonologia Computazionale (C.S.C.) of the Padua University, the problems that derived from using gesture to control the real time performace of a computer musical score, such as Music V ones have been studied [15]. This led to the description of a theoretical frame in which gesture and sounds may be organized into two independent and structured spaces that may interact according to arbitrary relationships. A set of primitive performance controls has been defined to allow composers to describe the way gesture interacts with sound [16].

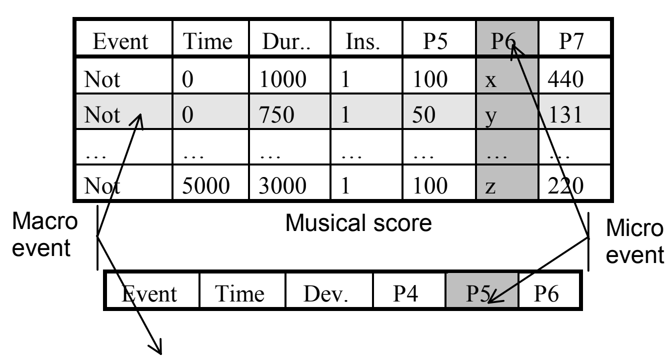
The problem is not to play with an instrument, but to interpret a score with different levels of interaction from the complete automatic performance (press a button and the sequence runs by itself) up to improvisation (gesture built musical phrases).

Something has to be said about the significance of the time data: gesture time (start, end or duration) may influence notes or silence timings and vice-versa. Real-time - loading a value or a set of sound parameters at the moment a gesture occurs - is only one possibility among others.

By considering the computer sound score (notes, timings, parameters) and the gesture score (gesture, timings, parameters) as two grids of events as in Figure 2, it is possible to mix elements of these grids in order to produce a performance which is musically significant.

A system that allows score performance has been designed and implemented for the digital sound processor 4i. This processor was physically controlled by only two sliders and the host computer keyboard - a DEC PDP-11 with the RTI4I proprietary real time operating system.

Even with such a simple prototype, numerous musical pieces have been produced, such as the digital part of the Prometeo by L. Nono [17].



Event	Time	Dev.	P4	P5	P6
Gest	0	sensor	100	a	30
Gest	0	sensor	80	b	60
…	…	…	…	…	…
Gest	3000	sensor	100	c	25

Gestural score

Figure 2. *Musical and gestural scores as grids of events*

5.2. Real time control of a music workstation

This work has been proceeded at the IRIS research center with the MARS workstation. This musical workstation is a totally programmable digital machine dedicated to real time digital sound processing. In its latest version, this workstation includes a sound board, NERGAL, to be plugged into the host computer and the graphic user interface ARES which runs under the Windows platform [18].

ARES is a set of integrated graphic editors which allow the interactive development of hierarchized sound objects. It is possible to design algorithms for sound processing and to control them by associating any type of MIDI device to their parameters with arbitrary relationships. The real time control is done by the embedded RT20 operating system of NERGAL [19].

MARS allows users to develop single sound objects called tones. Tones are specific configurations of all algorithm input variables such as parameters, envelopes or LFOs. The MARS operating system and real time control is MIDI oriented, thus all references to note events, continuous controls and other messages conform to the standard MIDI protocol. Tones are related to their algorithms and are associated to a MIDI channel. Tones may be dynamically loaded with Program Change messages.

Parameters have a symbolic name, a default representation unit (such as dB, Hz or ms) and a state that defines the way their values will be updated:

- Static parameters are initialized at the beginning of a note event by a constant value.
- Dynamic parameters are initialized at the beginning of a note and are continuously updated during the life-time of the note, according to a conversion formula.

The setup of a dynamic parameter is made in two phases:

1. Formula selection - a set of 17 predefined formulas is available as combinations of a maximum of three functional blocks called M. In the Figure 3 the formula associated to the pitch parameter is:

   Pitch = M1*[1+ M2]

2. Blocks definition - Each M block is defined by a conversion function (an array), a MIDI index, and a scaling factor for output values. For example:

   M1=HZTEMP(Key)*1.0

   M2=BEND12HT(Pitch-bend)*1.0

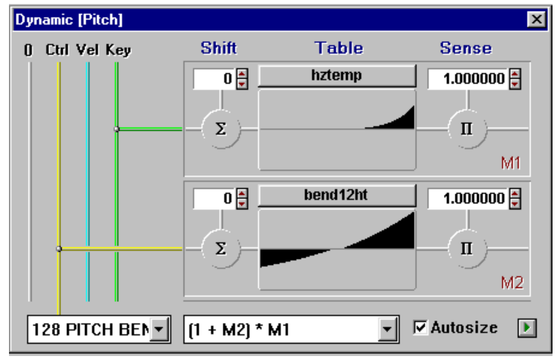


Figure 3. The MARS dynamic parameter editor

The mapping features provided by the MARS workstation allows different kinds of mapping which fit different control situations.

One to one mapping strategies have been intensively used for DSP research and quick sound prototyping.

Convergent mapping strategies have been used to add performance possibilities and allow subtle live control mechanisms, as illustrated in the previous example on pitch.

Divergent mapping has been commonly used to simulate commercial instrument behaviors and to optimize the ergonomic aspects of their control part.

5.3. Live electronic performance

Live electronic pieces may be considered a challenge for creative gesture to control sound processing. That's why interactions between standard sound synthesis, sound transformations and traditional music performance often fascinate composers.

By dealing with existing sound and recognizable instrumental gesture, we have to think how sound processing and other type of gesture may be musically congruent to the musical performance. This is done without having any direct control on the sound emission, but by transforming audio data according to external events which are not always explicit on stage but which depend on states of the performance itself.

Live electronics is often combined with sound projection for spatialization and/or sound reinforcement. Thus very complex situations on stage and at the mixing console have to be managed including audio and MIDI setups, computers network with eventual backup systems in case of crash, and different communication channels between performers. Interactive environments must be designed for the final musical performance situation but also for the critical rehearsal phase.

With such complex setups, it's generally more comfortable to work with cue lists of pre-defined musical material such as sets of sound or transformation parameters, or audio samples. This strategy allows to lower risk levels and is well suited when live electronics have been encoded on the score. Performers at computer, or performers on the stage will only have to trigger the cues at a given time through keyboards, pedals, buttons or other devices.

When sound processing depends on the musical situation, logic conditions on audio or gesture signals has to be tested thanks to modules such as threshold, multiplexer or noise gates.

Parameter extraction from sound analysis with pitch, envelope or peak followers, may be used to trigger effects or to pilot automatic score following.

Finally, sound processing algorithms may be considered as instruments to play. This situation requires immediate dynamic control which can be done on stage with remote controllers or by some hidden performer close to the mixing console.

The 4i system, and the MARS workstation combined with the MAX environment have both been used for live electronic pieces. Their interfaces provide four important features related to the sound - gesture relationship which usually have fit the interactive digital environments requirements.

1. Gestural data conversion facilities;
2. Gestural data grouping rules to define "meta-gesture";
3. Sound parameters extraction and logic facilities on audio signal;
4. Time control and synchronization features for parameter updating.

6. CONCLUSIONS

Human gesture has always been at the center of the musical creation process either with the writing form of a score or for instrumental or conducting performance. On the contrary, listening music often leads to emotions which reverberate at body level; a phenomenon which is at the basis of dance and choreography.

Music and gesture are closely linked into a dynamic relationship between acting and listening, cause and effect. Today's technology allows one to imagine and build in an arbitrary way new relationships between gesture and sound. Besides new hints for creativity these new possibilities allow the development of innovations for education, help for the handicapped and overall communication.

Composers and performers have an important role to assume for helping developing interactive systems with meaningful relationships between music and human gesture.

7. REFERENCES

[1] Moore, F. R., "The dysfunction of MIDI", Proc. International Computer Music Conference, San Francisco, USA, pp. 256-263, 1987.

[2] Moorer, J. A., "Audio in the new millennium", R.C. Heyser memorial lecture series, AES, 2000.

[3] Puckette, M., "Combining event and signal processing in the MAX graphical programming environment", Computer Music Journal 15(3), pp. 68-77, 1991.

[4] Chadabe, J., Meyers, R., "An introduction to the PLAY program", *Computer Music Journal* 2(1), pp. 12-18, 1978.

[5] Polansky, J, Rosenboom, D., Burk, P., "HMSL: overview and notes on intelligent instrument design", Proc. International Computer Music Conference, pp. 256-263, 1987.

[6] Dannenberg, R., McAvinney, Rubine, D., "Arctic: functional language for real-time systems", *Computer Music Journal* 10(4), pp. 67-78, 1986.

[7] Winkler, T., "Making motion musical: gestural mapping strategies for interactive computer music", Proc. International Computer Music Conference, pp. 261-264, 1995.

[8] Mulder, A., "Human movement tracking technology", tech. rep., Simon Fraser University, Canada, 1994.

[9] Depalle, P., Tassart, S., Wanderley, M., "Instruments Virtuels", *Résonance*, IRCAM, pp.5-8, 1997.

[10] Boie, R, Mathews, M., Schloss, A., "The radio drum as a synthesizer controller", Proc. International Computer Music Conference, San Francisco, USA, pp 42-45, 1989.

[11] Camurri, A., "Interactive dance/music systems", Proc. International Computer Music Conference, pp. 245-252, 1995.

[12] Cadoz, C., Luciani, A., Florens, J.L., "Responsive input devices and sound synthesis by simulation of instrumental mechanisms: The CORDIS system, *Computer Music Journal*, 8(3), pp.60-73, 1984.

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

[14] Mulder, A., Sidney Fels, S., Mase, K., "Empty-handed gesture analysis in Max/FTS", Proc. Kansei - The technology of Emotion Workshop, Genova, Italy, 1997.

[15] Azzolini, F., Sapir, S., "Score and/or gesture: a real time system control for the digital sound processor 4i", Proc. International Computer Music Conference, Paris, France, pp 25-34, 1984.

[16] Sapir, S., "Informatique musicale et temps réel: geste - structure - matériaux", actes du colloque international "Structures Musicales et Assistance Informatique", Marseille, France, pp. 5-19, 1988.

[17] Sapir, S., Vidolin, A., "Interazioni fra tempo e gesto - Note tecniche sulla realizzazione informatica del Prometeo", *Quaderno 5 del LIMB*, pp.25-33, La Biennale di Venezia, 1985.

[18] Andrenacci, E. Favreau, N. Larosa, A. Prestigiacomo, C. Rosati, S. Sapir, "MARS: RT20M/EDIT20 - Development tools and graphical user interface for a sound generation board", ICMC proc., pp 340-343, San Jose 1992.

[19] Andrenacci P., Armani, F., Paladin, A., Prestigiacomo, A., Rosati, C., Sapir, S., Vetuschi, M., "MARS = NERGAL + ARES", JIM Proc., pp 145-152, Lyon, 1997.