Issues for Designing a flexible expressive audiovisual system for real-time performance & composition

Enrique Franco
Interaction Design Centre
Dept. Of CSIS
University of Limerick, Ireland
+353 61 202798
enrique.franco@ul.ie

Niall J L Griffith
CCMC
Dept. Of CSIS
University of Limerick, Ireland
+353 61 202785
niall.griffith@ul.ie

Mikael Fernström
Interaction Design Centre
Dept. Of CSIS
University of Limerick, Ireland
+353 61 202606
mikael.fernstrom@ul.ie

ABSTRACT

This paper begins by evaluating various systems in terms of factors for building interactive audiovisual environments. The main issues for flexibility and expressiveness in the generation of dynamic sounds and images are then isolated. The design and development of an audiovisual system prototype is described at the end.

Keywords

Audiovisual, composition, performance, gesture, image, representation, mapping, expressiveness.

1. INTRODUCTION

The use of visual representations to specify music has a long history. The advent of new interfaces offers the opportunity to go beyond the traditional static time-line representation of the score. Most existing audiovisual systems (AVS) (FMOL, Hyperscore, Loom, etc) emphasize either the graphics display or the sounds for mapping flexibility and control. Therefore, the composer has to adjust to what a system's designer has decided beforehand about how the form and characteristics of images or diagrams represent the identity and quality of sounds and vice versa.

An interesting aspect of various existing audiovisual environments is the use of a dynamic image as a representation of the sound. Some of these systems take advantage of the implicit temporality of gestural 'mark-making' to add expressivity to the visual and sonic composition. However, there is an important challenge to the visual representation of the evolution of events in time when you are dealing with a dynamic image (animation). While the composer is drawing, the time line unfolds. However, once the drawing is complete this dynamic is lost and unless the image is re-animated the representation of the music is obscured by our inability to see the time line in the static image. If you are using an image to specify a sequence there is an advantage in making a timeline through the image. However, if you have several animated audiovisual sequences, a global view with a common time reference, is useful for both understanding and controlling the temporal relations between them.

Existing systems lack this kind of visual (temporal) representation. This is understandable in the context of real-time performance systems that don't have editing capabilities. We think this is an important issue that is open to improvement in order to have more control over the composition process. However one could ask: Is it worth editing a real-time performance of experimental music? How far can you take the level of sophistication and flexibility of a real-time performance audiovisual system? This issue is directly relevant to the actuation of sound within performance. Specification, whether it be represented in Tibetan neumes or western notation always involves an interpretive gap that is to filled by the performer. Thus the definition of an interface involves crucially how the determinate and expressive aspects of an instrument are to be realized.

In order to explore these issues we will discuss some of the most interesting interactive audiovisual environments taking into account a variety of requirements and their strengths and weaknesses.

2. FACTORS IN INTERACTIVE AVS

As a starting point we have examined various systems, in terms of the following properties. These we consider essential for building an interactive AVS.

- ***Real-time (improvisatory) performance capabilities for the creation of images and sound****:* This point is essential if the aim is to build a system that can take advantage of the nuances and variety of the user's gestures and expand and augment them in an audiovisual environment. Also, it should have the capacity to create images and sounds from scratch and not just interpret stored ones.
- ***Compositional structures: events organization and modification****:* Most of the systems that allow the creation of sound and image in real-time don't have the capability for organizing events at a global level. This is however, required if the aim is to allow the composition of a piece that involves feedback from events sonic and visual, in the construction of interactive audiovisual compositions.
- ***Expressiveness****: detailed gestural control over visual and sonic parameters.* In order to have "unlimited" expressivity it is necessary to arrive at a balance between the degrees of freedom and the number of parameters of control. Also, a controller that can capture a wide range of nuances from the user's performance is needed.
- ***Mapping flexibility between image and sound****:* There is no "objective" mapping from sounds to image or vice versa. Therefore, mapping flexibility between the aural and visual dimensions is necessary for the user to feel comfortable with the audiovisual feedback from the gestures. This is related to the form of synthesis used. Arguably, some mappings are less arbitrary than others in the case of physical models of synthesis and can be implemented in the system as default

interactions. However, mappings to spectral synthesis methods are intrinsically arbitrary.

- ***Modifiers, effects and filtering for audio and image****:* As a part of a complete tool for audiovisual performance and composition, it is necessary to implement modifiers, effects and filters, that can be applied at different organizational levels.
- ***Learnability****:* It would be ideal to have a system that is easy to learn and powerful all at once. Therefore it is desirable to have a system that offers different learning curves. A very cryptic functioning environment would dishearten the user but at the same time he/she will rapidly get bored with one that is too simple.

3. CONTROL METAPHORS IN AVS

There are four principal metaphors for sound-image relationships in the field of visually-orchestrated computer music, the first three were described by Golan Levin in [7]:

- *Timelines and diagrams* (Performer, Cubase, Protools, Cakewalk, etc): these systems offer different views of musical information such as standard music notation, digitized sound waveforms, MIDI notes displayed on a "piano roll" among others.
- *Control-Panel Displays* (Reaktor, Audiomulch, ReBirth, etc): these systems often mimic the controls (knobs, dials, sliders, buttons) afforded by analog synthesizers.
- *Reactive widgets* (FMOL, Aurora, etc.): virtual objects, which can be manipulated, stretched, etc. by a performer in order to control and modify sounds.
- *Drawings & free-form images* (Yellowtail, Loom, Hyperscore, etc.): these permit the generation and control of sound and/or music by gestural mark-making.

These systems discussed below all make use of one or more of these metaphors and are either specialized in performance or composition. They all make use of gestural input and synthetic image for controlling sound. These systems are *Yellowtail, Loom, Warbo, Aurora, Floo, Hyperscore, Metasynth, Videodelic, Music Sketcher and FMOL.* One of the most interesting collections of imagistic gestural interfaces is that of Golan Levin.

3.1 Golan Levin's Image Systems

Golan Levin approach is interesting because it recognizes the essential variety of possible interfaces a user may relate to. This is for real-time and simultaneous performance of dynamic imagery and sound. The painterly interface metaphor is exemplified by five interactive audiovisual synthesis systems: *Yellowtail, Loom, Warbo, Aurora and Floo.* Videos and images of these systems can be found at [8] and detailed information about the design in "Painterly Interfaces for Audiovisual Performance" at [7].

- *Loom:* In this application every visual element is associated with a corresponding sound-event. It wraps an animated score around the spine of a user's mark. As the marks are perpetually redrawn, they are sonified by a curvature-sensitive FM synthesizer.
- *Aurora:* Permits the creation and manipulation of a shimmering, nebulous cloud of color and sound. This glowing formlessness evolves, dissolves and disperses as it follows and responds to the user's movements.
- *Floo:* disperses and deflects soft-edged tendrils in response to user movements. Sound granules in a circular pitch-space create chorused drones as the tendrils grow.

The diversity of Levin's representational views is associated with forms of synthesis and the sound qualities envisaged. Strokes, clouds, blobs, tendrils are transformed into sonic analogies; linear, diffuse, discrete sonic events. The diversity of views reflects the desirability of diverse interfaces as it reflects the diversity of sonic forms envisaged.

3.2 FMOL

Another system that uses close feedback in its interface is Sergi Jorda's FMOL. This presents a closed feedback loop between the sound and the graphics: *the same GUI works both as the input for sound control and as an output that intuitively displays all the sound and music activity* [2].

However, FMOL is designed to be a playable instrument, not a compositional environment and for that reason users cannot edit performances or trigger pre-recorded sequences while improvising. It is therefore hard to play a fixed sequence of pitches or a precise rhythm, as *the interface is good for large-scale or statistical control but poorer for detailed specification* [2]. However the large amount of sound synthesis algorithms (more than 100) makes FMOL a very flexible system in terms of sound generation.

3.3 Comparison of AV systems

The examination in Table 1, shows that none of the current systems fulfill all the conditions that our "ideal system" should posses.

Table 1. Comparison of different AV systems

System	Real-time performance		Composition capabilities		Expressivity		Modifiers, FX & filters		Modul-arity	Learnability	
	Sound	Image	Sound	Image	Sound	Image	Sound	Image		Easy	Difficult
Yellowtail	X	X			X	X		X		X	
Loom	X	X			X	X					X
Warbo	X	X			X						X
Aurora	X	X								X	
Floo	X	X				X					X
FMOL	X	X			X		X		X	X	
Hyperscore		X	X				X			X	
Music Sketcher			X				X			X	
MetaSynth		X	X	X		X	X	X	X		X
Videodelic		X		X				X	X	X	
IDEAL SYSTEM	X	X	X	X	X	X	X	X	X	X	

However, some of them are close to match this "ideal system", e.g. *Metasynth* [9]. The biggest drawback of this system, from the point of view of this research, is the lack of real-time performance of sound, a property that is present in both Levin's and Jorda's systems. Also, Levin and Jorda's systems incorporate dynamic visual feedback. Other systems, which are much closer to, while being more permissive than sequencers or score based specification, are *Hyperscore* [5] and *Music Sketcher* [10]. These both make use of the *timelines and diagrams* metaphor but present alternative ways of control and generation of audiovisual material such as drawing strokes that are mapped to structural elements in the music in *Hyperscore;* or insertion of small blocks of musical content in *Music Sketcher*. However the audiovisual outcome of these last two systems has a reduced expressivity. Other systems such as *Floo* and *Aurora,* although they allow the

generation of interesting and variable images and sounds, most aspects of the interaction are predetermined by the system.

Table 2. Controllers and techniques for generating image and sound.

System	Image		Sound		Controllers
	Static	Dynamic	Synthesis Technique	MIDI	
Yellowtail		X	Spectrogram-based additive synthesis oscillators.		Mouse, keyboard
Loom		X	Frequency Modulation (FM)		Mouse, Wacom tablet
Warbo		X	Waveshaping (Chebyshev polynomials)		Mouse and Wacom tablet
Aurora		X	Granular		Mouse
Floo		X	Granular. Shepard tones.		Mouse, keyboard
FMOL		X	More than 100 synthesis algorithms		Mouse, keyboard
Hyperscore	X			X	Mouse, keyboard
Music Sketcher	X			X	Mouse, keyboard
MetaSynth	X		Spectogram-based FM and Wave table		Mouse, keyboard
Videodelic		X		X	Mouse, keyboard, audio input, any MIDI controller

Table 2 summarizes the different techniques used for sound synthesis and shows there are several algorithms used in this kind of system that reflect the advantage of having timbre as a core element to achieve expressivity and sonic variety. This way one might think in terms of having a sound design environment rather than a music composition tool.

Another core issue in audiovisual systems design is the way the user controls synthesis. Table 2 compares the different systems and it shows that the mouse is the most common controller. Though the mouse is available everywhere, it doesn't permit the capture of a wide range of movements executed by the user. Only *Loom, Warbo* and *Videodelic* [12] allow the use of different kind of controllers like drawing tablets and MIDI controllers.

4. FLEXIBLE & EXPRESSIVE AVS

The main issues that we have isolated through evaluating various audiovisual environments for flexibility and expressiveness in the generation of dynamic sounds and images are: *mapping flexibility, gestural control, dynamic visual feedback, sound synthesis algorithms, and timelines*.

4.1 Mapping Flexibility

Because there is no "objective" mapping from sounds to image or vice versa, the flexibility of the mapping between the aural and visual dimensions is crucial if the user is to feel comfortable with the audiovisual feedback from the gestural input.

A visual shape generated by mark-making using an electronic drawing device can sound like anything. We are using abstract synthetic sounds (timbres) and therefore the images or animations are abstract.

Also, our wish is to have different kinds of sounds and images playing at the same time. We have implemented a variety of graphics generation algorithms (paint tools) the composer can explore to match specific sound generators. In this way the performer/composer can decide what is associated in a personal, perceptually motivated way. This is achieved through a user controlled mapping switching mechanism.

4.2 Gestural Control & Dynamic feedback

In this way we are exploring the possibility of alternative representations that can give the player/composer indications of how the sound synthesis parameters vary over time by having a dynamic visual feedback. For instance, if the thickness of a mark is mapped to the intensity of the sound, you can both see and hear clearly that thinner marks sound softer and become louder as they grow thicker. In this way the visualization is not just a static representation of the control interface, but is an active element of the composition itself with an aesthetic value.

Within this approach the graphics are not symbolic notations to be read by the users, but a representation and control input for the sound generated by the synthesis algorithms. This reflects the active status of such representation in the actuation of the sound. In normal notation the gap between specification and performance has to be filled by the player.

Dynamic image generation adds expressivity to the visual representation. It takes advantage of the implicit temporality of gestural mark-making and that's why a high-resolution physical interface is needed in order to capture the nuances of the user's movements. The possibilities of expression when you use an electronic drawing device such as the Wacom tablet are endless. You can draw or paint whatever you want and can make as many marks as you like, you are free to create any two-dimensional image. The visual expressivity can be achieved by the "brushwork" (nature of the marks, shape, texture, sensitivity of the brush) and the use of colour (brightness, intensity) as a result of the signal analysis used to extract relevant information from the raw temporal and spatial data.

Also, as the painterly schema proposed by Golan Levin, "*the visual material is not situated along a set of coordinate axes like in the score-based systems, but rather in the free-form visual structure of a dynamic abstraction*"[7]. One important question that arises from this schema is: Can the visual output of the system be read as a painting or a score, or both? As we have stated above the ideal is that it should be both.

4.3 Sound synthesis and Musical Aspects

If we link free form gestures to control sound synthesis parameters we can create an infinite range of timbres and control their evolution over time by calculating geometric or statistical properties of the marks, or by creating representations in the frequency domain.

The sequences of dynamic images and their iteration, spatial and temporal accents can create rhythm patterns linked to the creation of sounds in real-time. The development of colours, textures and shapes is linked to the development of timbre. The kind of pieces you can compose or play with such a system is something that combines electroacoustic (timbral), electronic and visual music.

4.4 Timelines

How can we edit and organize dynamic audiovisual events?

If you make timelines through animated marks, you cannot have a clear idea and detailed control of what is going on in terms of the temporal relation between them unless you also have a frame by

frame representation, such as used in video editing. An alternative solution to this is a two-dimensional representation of the evolution of the visual and sonic parameters in time. In this way we have a multitrack-like display (e.g. Protools, Cubase) with a common clock and can make variations in different characteristics by modifying diagrams.

5. IMPLEMENTATION

The systems that we use as a starting point for our design are *Loom* for the visual and control concepts and *FMOL* for the multiple sound synthesis algorithms. We have a visual representation similar to *Loom* where every visual element is associated with a corresponding sound-event and a timeline is "wrapped around" the user's marks, linked to various graphics and sound generators through a mapping switching mechanism.

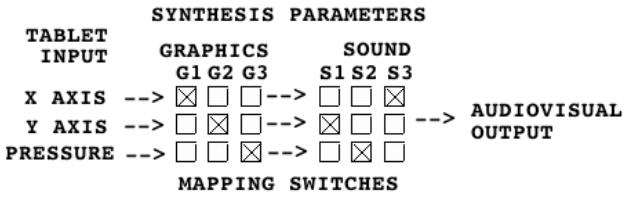

Figure 1: *Miró's* mapping mechanism.

We are currently finalizing a prototype called *Miró* using *PD+GEM* [11][3]. In our prototype the gestural input comes through a *Wacom* tablet that allows the modification of the visual window by playing directly on it. The continuous representation of the gestures is recorded and mapped to some aspects of the user's marks such as the local velocity, pressure, X and Y coordinates. So far the system has three different sound generators (two variations of FM synthesis and Phase Aligned Formant) and three different graphics' generators or tools (paintbrush, spray and fountain) that you can assign to any track.

In this way we have dynamic visual feedback in the first stage, i.e., when the user is interacting with the system by manipulating the stylus for the generation of audiovisual sequences. The gestures' nuances in terms of trajectory, force and temporal variations (modulations) add expressiveness to the audiovisual substance that is fed back in real-time. Then, in a second stage we can playback and organize this dynamic visualization. The outcome is a multiple dynamic visual representation of sounds synthesis generators in the same view.

It is possible to stretch the duration of each track and select sections within them. Also it is possible to playback the sequences in two ways a) synchronizing them to a common metro that triggers each section according to its position on a timeline and b) according to its own period by creating loops. The second method of playback allows the user to improvise by changing "on the fly" different controls of the track panels such as slot selection, stretch factor, visual depth, colour, playback direction and the graphics and sound synthesizers. Obviously, these modifications can also be done off-line. For this purpose separated control panels for each track or sequence and a set of general controls have been implemented.

As a result of the designing process we have realized that there is a contradiction in modifying audiovisual characteristics of data recorded from gestural input; if we want to keep the audiovisual specification, realization and expressiveness close linked. To change a section we just "re-draw" it rather than go through every recorded millisecond. Therefore, our priorities are the mapping switching mechanism between gesture-image-sound and the temporal organization for playing back the audiovisual sequences.

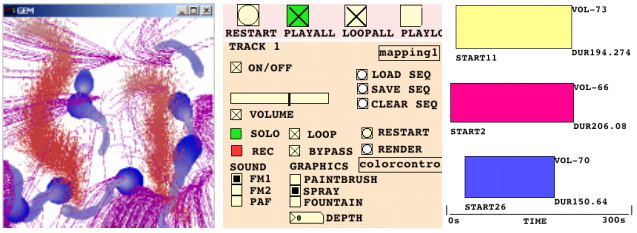

Figure 2. A screen shot of *Miró* prototype in action: a) graphics window b) Main control panel section c) Timelines

6. CONCLUSION

The integration of diverse visual representations and sonic forms controlled by a gestural input, offers an interesting alternative to current systems to allow expressivity and flexible control in software-based music. More associations between sound, image and gesture are waiting to be discovered and experimented as an important area for further research.

7. ACKNOWLEDGMENTS

We would like to thank the members of the Interaction Design Centre for their continuous feedback and suggestions.

8. REFERENCES

[1] Abbado, A. *Perceptual Correspondences of Abstract Animation and Synthetic Sound*.1988. http://www.abbado.com/thesis/corpo.html

[2] *FMOL* homepage: http://www.iua.upf.es/~sergi/FMOL

[3] *GEM* homepage: http://gem.iem.at/

[4] Girling, L. *Granulator*. Software prototype developed at Interval Research Corporation, Palo Alto, 1997-1998

[5] *Hyperscore*: http://web.media.mit.edu/~mary/hyperscore/

[6] Jorda, S. *Sonigraphical Instruments: From FMOL to the reacTable**. Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03). (Montreal, Canada, May 22-24, 2003) http://www.music.mcgill.ca/musictech/nime

[7] Levin, G. *Painterly Interfaces for Audiovisual Performance*. Master Thesis, Massachusets Institute of Technology, 2000. http://acg.media.mit.edu/people/golan/thesis/

[8] Levin, G. *Audiovisual Environment Suite.* http://acg.media.mit.edu/people/golan/aves/

[9] *Metasynth* homepage: http://www.uisoftware.com

[10] *Music Sketcher*: http://www.research.ibm.com

[11] *PD*: http://www-crca.ucsd.edu/~msp/software.html

[12] *Videodelic* homepage: http://www.uisoftware.com

[13] Willats, J. *The Syntax of Mark and Gesture*. 2002. www.lboro.ac.uk/departments/ac/tracey/somag/willats.html

[14] Winkler, T. Composing Interactive Music. Cambridge, MA: MIT Press, 1998.