A Poetics of Simulation for Audiovisual Performance

Randy Jones

Department of Computer Science
University of Victoria, Canada

Abstract

Audiovisual performance is a fertile area for creative expression, an intersection of experimental cinema and computer music that has seen a groundswell of interest in recent years. To create works in this emerging medium, a complex network of relationships between sounds, images and sensor input must be organized. This complexity poses major technical and aesthetic challenges which a systematic approach can help address.

This paper presents an analysis of audiovisual performance as two parts: a real time simulation which produces dynamic form, and a visualization by which that form is aestheticized. This approach to a systematic study, or poetics, of the medium is drawn from the study of successful works as well as from film theory and cognitive psychology. Recent audiovisual work by the author is discussed, and technical details are presented. Approaching audiovisual performance as real time simulation provides a practical framework for collaboration between artists and researchers in aesthetic visualization.

Categories and Subject Descriptors (according to ACM CCS): J.5 [Arts and Humanities]: Performing Arts; Keywords: audiovisual performance, aesthetic visualization, sound-image relationships.

1. Introduction

The term "visual music" refers to a broad range of artistic pratices united by a common thread: that abstract moving imagery can have the dynamics and structure of musical sound. The past decade has seen a great deal of activity in the area, including a flurry of creative work and a major museum exhibition focused on the history of the idea [Mat05]. Artists since the 1700's have approached visual music by either using the most suitable tools available or inventing new ones—the idea has been realized in technologies from mechanical light projection machines or "color organs," to film and video, to the computer.

Though some visual music is silent, an exploration of moving light as an unaccompanied music for the eye, many works combine light and sound. The computer has been instrumental in making possible a performative approach to visual music. Though there exist precedents for audiovisual performance which predate computer technology, activity in the medium has grown greatly in recent years based in part on the availability of affordable, high-quality graphics hardware. An overview of the history of visual music as it leads to current audiovisual performance can be found in Golan Levin's Masters thesis "Painterly Interfaces for Audiovisual Performance" [Lev00].

One result of visual music's influence on popular culture has been the birth of the VJ. The VJ (short for video jockey, as a parallel with the club DJ or disc jockey) uses computer and video hardware to create a real time accompaniment to a DJ or live performance. Most VJs work by triggering and mixing prerecorded video segments. Though effects applied in live mixing can make these clips all but unrecognizable, the dominant mode of VJ work relies heavily on source material for its style. Software tools for VJs such as Motion Dive, Resolume and VJamm allow the user to mix of two or more channels of live video and apply a variety of effects.

Audiovisual performance tends to differ from VJ work in that no one formal structure characterizes the medium. A wide variety of methods have been applied to the simultaneous organization of audio and live motion graphics. The flexibility of the software required to enable a plurality of approaches to this problem has meant that the only successful tools either are programming languages, or share the generality and steep learning curves of programming languages.

Though the ability to program in one form or another is certainly a valuable conceptual tool, it is a skill which takes significant time to acquire—not all media artists want to learn to code. The requirement to learn programming, or a tool with a comparably steep learning curve, can be a significant hurdle to the artist with a specific audiovisual idea in mind. This represents an opportunity for both collaboration and the creation of new tools. If there is any rhyme or reason behind audiovisual performance, a poetics, or analysis of the functioning of the medium, can identify the organizational principles which underlie successful works. These principles can in turn inform the creation of tools for the medium. Any tool easier to learn than a general-purpose programming environment will also be more specific, facilitating a specific approach to organization. One successful approach, presented here, is the use of simulation as an abstraction which mediates the connections between performance sensors and audiovisuals.

2. Tools for Audiovisual Performance

The difficulty of learning existing software tools for visual music, and audiovisual performance in particular, is the primary motivation for the analysis presented here. As both a software developer and practitioner in the area, it has been my experience that composers of audiovisual performance typically spend more time learning and building software tools than composing.

The list of available tools for audiovisual performance is fairly small. From 2001–2004 I worked on designing and implementing one such tool: Jitter, an extension to the Max/MSP environment. Max/MSP/Jitter and the other few commonly used tools such as Pd/GEM, Isadora and vvvv can all be categorized as visual programming or *patching* environments. A *patch* in these environments is the equivalent of a program in a textual language. Composed of objects and connections between them, it can also contain a hierarchy of subpatches. The connections define the flow of data between objects; in Max/MSP/Jitter this flow is implemented using a messaging system. A simple video generator patch is shown in Figure 1.

The top of the window shows some of the GUI objects available in the environment. The boxes in the window connected with lines are named objects which pass messages. Aside from the top "metro" object, which acts as a metronome, and the "jit.noise" object, which generates a matrix of random values, all of the objects shown here perform transformations on incoming matrix data, the final result of which is seen at the bottom of the patch. Flexible patching environments have an inherent complexity resulting from the large library of objects required. Though the designers of Max/MSP/Jitter have taken pains to provide easy access to and ample documentation for their library, its sheer size— over 600 objects—represents a significant hurdle for newcomers.
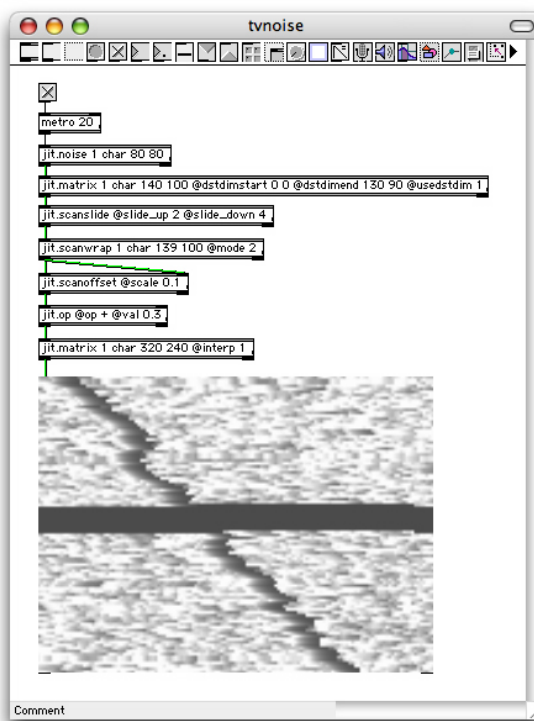

Figure 1: *A small Max/MSP/Jitter patch.*

Enabling a general approach to audiovisual performance, as both programming languages and patching environments do, requires complex tools. The process of designing one's own structural approach to visual music and grappling with these tools can be intellectually rewarding. But a desire common among composers working in this area, myself included, is to spend less time toolmaking and more time composing. Furthermore, many of these practitioners have worked along similar lines in developing their own tools. If one is prepared to give up some generality, tools for visual music can be made which do not require the composer to reinvent the wheel.

3. Poetics in *Six Axioms*

The term "poetics" has been applied to the study of various media. In his writing on film, David Bordwell defines poetics as "a kind of middle-level theorizing-an attempt to describe or explain particular craft practices." [Bor93] Through an analysis of successful works, approaches to poetics for audiovisual performance can be developed which have the potential to inform the creation of more specific tools for the medium. I will discuss how a simulation-based approach has supported various compositional goals in my recent work *Six Axioms.*

Six Axioms is an audiovisual composition, approximately

30 minutes long, which immerses both performer and audience in fluid, synaesthetic imagery. The piece has three main sections. In the opening section, a series of simulations is controlled by the performer; time-varying fields direct the motion of particle clouds which are both visualized and sonified to create tightly linked audiovisuals. In the middle section, synthesized percussion and full-screen images taken from prints of plants create a kind of stately, ancient procession. The closing section recapitulates the first, drawing the clouds out into a vast skyscape. Figure 2 shows a representative frame from each section.

3.1. Natural Mappings

My primary compositional goal, one shared with most composers, has been to create musical meaning. In order to create meaning a ground must be established, a set of correspondences which are understood as normative within the context of the piece. These provide states of rest. Moving towards and away from them shapes the work at various levels of structure. Consonance in tonal music is an example of this kind of correspondence.

The simultaneous organization of audio and visuals has often been approached through *mappings,* transformations which convert a sampled quantity from one sensory domain into another. Mappings can flow from aspects of sound to image or vice versa, or from control data such as gestures to both domains. The composer typically seeks to create meaning by arranging stimuli over time. If meaning is to be communicated through the use of mappings, correspondences should be chosen which are qualitatively similar to the various people who might constitute a work's audience.

Synaesthesia is a well-documented condition which occurs in approximately one in 100,000 individuals. To a person with synaesthesia, stimuli in one sensory modality evoke sensations in multiple modalities. For example, musical notes can create the perception of certain colors: the note C may always evoke the sensation of green to a synaesthete, while E flat 'appears' red. [RH03] Because these associations are consistent, synaesthesia might seem to be a good basis for mappings in audiovisual performance. But the logic of synaesthesia is subjective: the mappings that could be derived by interviewing a given synaesthete are not meaningful to another, let alone to a person without synaesthesia.

Certain mappings do exist which can be perceived as natural by many individuals. [JN05] A proportional mapping from sound amplitude to image brightness, for example, can be considered natural because amplitude and brightness are both measures of the same concept, stimulus intensity, in different sensory domains. Likewise, when linking a frequency to the size of an object, mappings from small sizes to high frequencies and large sizes to low frequencies are more natural than the reverse. Due to the physics of vibrating objects, this scaling is consistent with our experience of the world.

Natural mappings, like harmony in tonal music, allow the composer to communicate meaning by moving towards and away from deeply and intuitively understood states of rest.

### 3.2. Mapping Through Simulation

In mapping between a large number of simultaneous parameters, strategies are needed to manage complexity. Simulation, or modeling of Nature, is an effective way of creating a large number of coordinated mappings which also have an intuitive correctness. Through simulation as a layer of abstraction, the general problem faced in audiovisual performance, that of constructing a set of mappings from sensors to stimuli, can be reduced in complexity.

In Six Axioms I faced the particular problem of mapping attributes of a large number of audiovisual particles to performance data from the Radio Drum. The Radio Drum is a sensor for live performance created in the 1980's at Bell Labs by Max Matthews and Bob Boie [MS89], and recently refined in ongoing work at the University of Victoria by Andrew Schloss, Ben Nevile and others [NDS03], [SD01]. Using very low frequency radio waves on the order of 10 kHz, it senses the position of two drumsticks over an antenna/percussion pad at a high sampling rate, enabling its use as both a 3D spatial and percussive controller.

Six continuous values are sampled from the drum: the x, y and z positions of the end of each stick. These values are the input to a simulation which produces a dynamic vector field as its output. Stirring motions of the sticks produce overall rotations of the field; bringing the sticks towards and away from each other produces a global "breathing" in which all vectors are oriented towards or away from the origin. The simulation's output can also be mixed with stored vector fields from a library of prerendered forms. These create turbulent flow and other phenomena which are not feasible to compute in real time. Notes played by striking the Radio Drum sticks against the pad create dynamic form, by instantiating collections of particles which are advected through the field. In this way, desired forms are not painted directly from performance input, but coaxed into being.

The simulation is a layer of abstraction between sensor data and the audiovisuals, an effective tool in managing the complexity presented by the large number of possible mappings. By specifying position and orientation attributes for each particle globally as by-products of the simulation, the decisions made in mapping are reduced to a tractable number.

In addition, the simulation divides the work into two conceptual parts: a dynamic form, which can be considered more or less devoid of aesthetic qualities, and an aesthetic rendering of the form. The choices made in the form's visualization and sonification are largely responsible for the perceptual qualities of the output, what Kirschenbaum [Kir] calls a "radical aestheticization." This structure offers
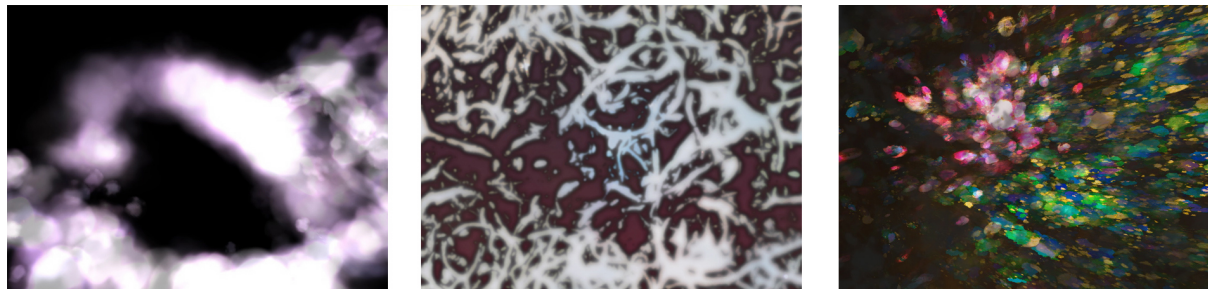
**Figure 2:** *A representative frame from each section of* Six Axioms.

a promising framework for creative collabrations between composers of audiovisuals and researchers in aesthetic visualization.

3.3. Visualization / Sonification

In visualizing the simulation, I made a number of technical choices based on aesthetic goals. One goal was to remove any visual signature particular to the underlying hardware. This is an attempt to relate the piece to its influences from film and painting, rather than works which are concerned with the graphics technology of the computer itself. To further this goal, the aspects of the OpenGL rendering pipeline which impart a visual style particular to current technology were not used. OpenGL's default lighting calculations produce a distinctive appearance, especially in interpolation across polygons. Polygon edges and the intersections of z-buffered polygons produce aliasing. These artifacts can be minimized by techniques such as adding vertices, adding custom lighting models, and enabling full-screen antialiasing, but currently only at the cost of real time performance. The rendering of non viewplane oriented polygons also imparts a particular visual signature, both in texture filtering and perspective transformation.

To address these issues, each particle is drawn with a viewplane-oriented texture mapped polygon, or impostor. When applied to rendering the dymanic form, these impostors to create large-scale stimuli with an interesting microstructure. This is a direct visual analog to granular sound synthesis as described in depth by Roads [Roa01]. The textures are taken from a collection of thousands of small prints, about 3 by 3 cm in size, made from plant material found on the Olympic Penninsula of Washington State in the Cascadian bioregion. The prints are divided into categories, each made with a different plant and printing gesture. Examples of prints from one category are shown in Figure 3. Once a category is chosen for a particle, a random texture image within that category can be assigned either for the entire life of the particle, or for each frame of the video. The ability to show a different image for each frame is inspired by the hand-painted films of Stan Brakhage. The flickering of the

spatially incoherent motion recalls ephemeral phenomena like the motion of light on water, or the phosphene activity which is an aspect of what Brakhage [BS76] called "closed-eye vision."
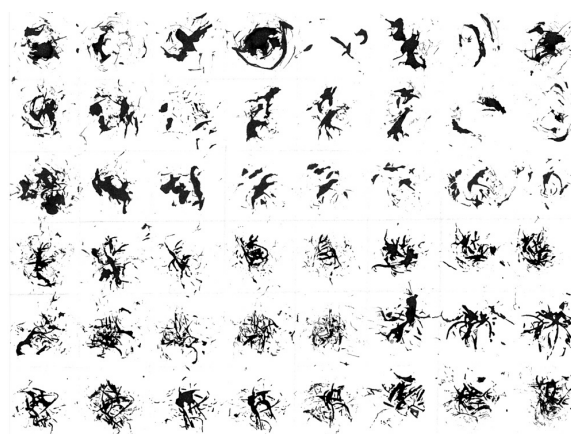
Figure 3: *Multiple prints from "Old Man's Beard" lichen.*

The textured images on the particles can either maintain a fixed rotation, or rotate to match their direction of motion as projected onto the viewplane. When the texture is kept constant for each particle, the latter option paints a coherent visualization of flow direction through the vector field. Another setting allows the particles to elongate based on their velocity, again projected onto the viewplane. This gives rise to an expressive effect which Scott Snibbe calls "the cartoon notion of motion blur [SL00]." Other parameters which can be varied for each grain include color, transparency and blend mode.

Particles moving through the simulation are sonified by spatializing individual sine waves. Each wave is panned to match the location of a particle in the virtual space, volume and dry/reverberant balance are varied based on distance, and Doppler shift of the wave's frequency is calculated according to the particle's motion. These mappings are motivated by the natural world, yet allow the changing of

parameters to create unrealistic sonic spaces. For example, lowering the speed of sound in the simulation maintains the phenomena created by Doppler shift but makes them first more obvious, then cartoonish in effect.

Since the parallel to granular synthesis of audio was mentioned here in the context of visualization, a few words on why this technique was not used for the sound seem germane. A main goal in aestheticizing the simulation was to make the audio and graphics match on as fine-grained a level as possible. This has resulted in an additive synthesis implementation which plays thousands of partials simultaneously. Playing this many audio grains is beyond the limitations of the current hardware used. I chose to embrace this limitation as part of the work's style. Reducing the timbral complexity of the material which is spatialized allows the listener to focus on the spatialization itself.

Just Intonation is the basis for the musical scales in which the tonal parts of *Six Axioms* are written. The pure harmonic ratios produce striking shimmering effects when partials of the notes are detuned relative to one another by the spatialization. Learning to listen in Just Intonation allows one to tune in to found melodies which come from harmonic resonance. In particular, I was given one melody when riding on a British Columbia car ferry, in the cavernous hold of the boat, a very busy sonic environment full of interesting drones and thrummings. One of the many vibrating pieces of machinery was changing slowly between different resonant modes; then out of the blue it played the melodic figure which begins the last section of the work.

Sonification and visualization are always tightly linked. Both are created from the motion and distribution of the particles; parameters which affect the rendering have been chosen to maintain this correspondence. The correlation between audio and visuals is a fixed ground for the piece which is reinforced as other aesthetic attributes change, resulting in the perception of a coherent space which surrounds both performer and audience.

3.4. Nonverbal Communication

Another compositional goal of mine is to make the medium an effective form of nonverbal communication. A gem-like anecdote that explains what I mean by nonverbal communication is given by Richard Feynman. In his autobiography *What Do You Care What Other People Think* [FL01], Feynman recalls a conversation he had at the age of eleven or twelve, with a young friend named Bernie:

> [I said,] "But thinking is nothing but talking to yourself inside."
> "Oh yeah?" Bernie said. "Do you know the crazy shape of the crankshaft in a car?"
> "Yeah, what of it?"
> "Good. Now, tell me how did you describe it when you were talking to yourself?"

So I learned from Bernie that thoughts can be visual as well as verbal.

By using simulation to create dynamic forms with physically plausible motion, I evoke nonobjective yet familiar spaces which these forms either inhabit or enclose. A catalog of such spaces is the vocabulary in my nonverbal language. Gaston Bachelard's *The Poetics of Space* [Bac69] is the classic text on how intimate qualities of place shape our experience. Bachelard's phenomenology is a valuable tool for our analysis of audiovisual performance, for, as he states, "only phenomenology makes it a principle to examine and test the psychological being of an image, before any reduction is undertaken." In *The Poetics of Space,* Bachelard discusses in great detail our experience of different kinds of spaces such as houses, nests, shells and corners. Bachelard's analysis is based on examples from written poetry. But by classifying qualities of experience evoked by different types of spaces in poetic imagery, he points the way to a language of direct audiovisual perception.

The types of spaces presented in *Six Axioms* are determined by both the simulated forms and their aesthetic rendering into sound and visuals. For example, a rigid round envelope of filaments which surrounds performer and audience, possibly creaking gently, puts us both in a nest. As humans we can not literally enter a bird's nest, but nevertheless we understand intuitively, bodily, the ramifications of nesting. This shared experience gives the poetic image a meaning which transcends the intellectual. Composing with oneiric spaces allows me to approach visual music as wordless poetry.

3.5. Implementation

Six Axioms was realized using Max/MSP/Jitter. External objects were written in C to perform the simulation and sound spatialization.

To record the work I developed a Jitter patch, render_node. Using this patch, performance data is captured and used to make a high-quality rendering offline with greater resolution than is possible in real time. Adding temporal aliasing or motion blur as well as other effects is also possible. Keeping the gestural information intact from performance gives the finished video a meaning as documentary and, I feel, an accompanying vitality. A surround DVD of *Six Axioms* is in progress. Render_node is available to Jitter users as an open-source project [Jon07].

4. Conclusion

In building software for and finally completing *Six Axioms,* simulation proved an effective tool for managing complexity and generating natural correspondences between audio and visuals. Factoring the work into two parts using simulation as interface has been both an approach to poetics and a

practical engineering step. The simulation creates a dynamic form which is then aestheticized; this aestheticization offers a clear opportunity for applying research in visualization to the creation of new works. I hope that this approach to a poetics for audiovisual performance, as well as my description of certain compositional goals and their technical realization, will encourage the development of new tools for the medium.

References

[Bac69]   BACHELARD G.: *The Poetics of Space*. Beacon Press, 1969.

[Bor93]   BORDWELL D.: *The Cinema of Eisenstein*. Harvard University Press, 1993.

[BS76]   BRAKHAGE S., SITNEY P.: *Metaphors on Vision*. Film Culture, 1976.

[FL01]   FEYNMAN R., LEIGHTON R.: *" What do you care what other people think?": Further Adventures of a Curious Character*. Norton, 2001.

[JN05]   JONES R., NEVILE B.: Creating visual music in Jitter: approaches and techniques. *Computer Music Journal 29*, 4 (2005), 55–70.

[Jon07]   JONES R.: 2up Technologies, http://2uptech.com., Retreived May 2007.

[Kir]   KIRSCHENBAUM M.: A White Paper On Information. http://www.iath.virginia.edu/~mgk3k/white.

[Lev00]   LEVIN G.: *Painterly interfaces for audiovisual performance*. Master's thesis, Massachusetts Institute of Technology, Program in Media Arts and Sciences, 2000.

[Mat05]   MATTIS O.: *Visual Music*. London: Thames and Hudson., 2005.

[MS89]   MATHEWS M., SCHLOSS W.: The radio drum as a synthesizer controller. In *Proceedings of the International Computer Music Conference* (1989).

[NDS03]   NEVILE B., DRIESSEN P., SCHLOSS W.: A new control paradigm: software-based gesture analysis for music. *Communications, Computers and signal Processing, 2003. PACRIM. 2003 IEEE Pacific Rim Conference on 1* (2003).

[RH03]   RAMACHANDRAN V., HUBBARD E.: The phenomenology of synaesthesia. *Journal of Consciousness Studies 10*, 8 (2003), 49–57.

[Roa01]   ROADS C.: *Microsound*. MIT Press, 2001.

[SD01]   SCHLOSS W., DRIESSEN P.: New Algorithms and Technology for Analyzing Gestural Data.

[SL00]   SNIBBE S., LEVIN G.: Interactive dynamic abstraction. *Proceedings of NPAR 2000, Symposium on Non-Photorealistic Animation and Rendering (Annecy, France* (2000), 21–29.