Sawtooth: Interactive Clarity and Aesthetic Complexity

Christopher Burns
University of Wisconsin-Milwaukee
Department of Music
P.O. Box 413
Milwaukee, WI, 53201 USA
cburns@uwm.edu

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
Sawtooth (2009) is an artwork integrating performance, sound, and animation. This paper describes the design of *Sawtooth*, with particular reference to the balances it strikes between control, clarity, and complexity.

Keywords
Motion capture, electroacoustic music, animation, improvisation

ACM Classification Keywords
H5.5. Sound and music computing: Systems.

General Terms
Design

Introduction
Sawtooth is an improvised multimedia performance. A performer's gestures are captured by a video camera, and translated into both music and animation. The size, location, and frequency of gestures correlate to the complexity and intensity of sound and image. Motion capture, gesture analysis, and animation for *Sawtooth* are implemented in Processing; Open Sound Control messages transmit gesture data to Pd, which provides audio rendering [3, 4, 5].

The basic audiovisual materials provided by the *Sawtooth* environment are relatively simple, as are the physical gestures which instantiate those materials in performance. Artistic complexity arises primarily from skillful improvisation, as many instances of these materials are combined and encouraged to interact with one another. For the performer, *Sawtooth* is designed to strike a balance between visceral, accurate, and expressively rewarding control of the system, and complex interactions or autonomous behaviors which foster a feeling of improvisational dialogue between the performer and instrument, and encourage exploration of texture and form in performance. From the audience perspective, this balance plays out along a slightly different axis. Clearly perceptible relationships between the performance and the multimedia result clarify the interactive relationships afforded by the system. At the same time, a performance should be more than a demonstration of interactivity; aesthetic richness and complexity are paramount.

Audiovisual materials

Sawtooth incorporates four main types of audiovisual materials, referred to here as "small tiles," "spinners," "sweepers," and "large tiles." All four types are activated and controlled via motion capture; the two-dimensional plane of the animation parallels the two-dimensional plane of the video camera feed used for motion analysis. The four material types have a variety of dependencies upon and interactions with one another; each is explained in turn below.

Small tiles

The small tiles are laid out in the motion capture and animation planes as a grid of 64 elements horizontally and 48 elements vertically. Activations of each small tile are represented in the animation field as a square divided into a pair of solid-color triangles. The color relationship between the triangles describes the current state of the tile (its receptivity to additional activations, which might lead to the instantiation of spinners). Each activation is represented sonically by a short-duration waveshaped sine. Pitch and timbre are governed by spatial location in the grid, and can vary over time through performer interaction.

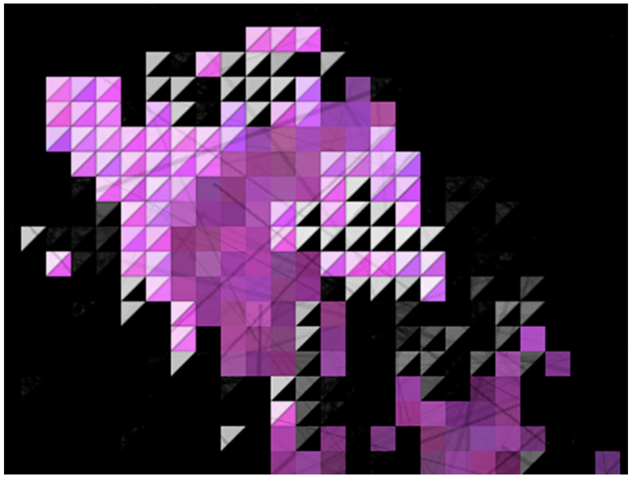

Figure 1. Detail showing small tiles in various states.

Spinners

When a small tile is activated three times in a short timeframe, it begins to launch spinners in a periodic rhythm (based on the duration between the second and third activation of the host tile). Spinners are represented visually by a moving triangle (both translating and rotating), and sonically by a brief frequency modulation (FM) tone, with a pitch derived from the location of the small tile that instantiated the spinner. Once in motion, spinners are attracted (via a simple gravity simulation) to the most significant movement in the motion-capture plane; the performer can direct and redirect them using movement.

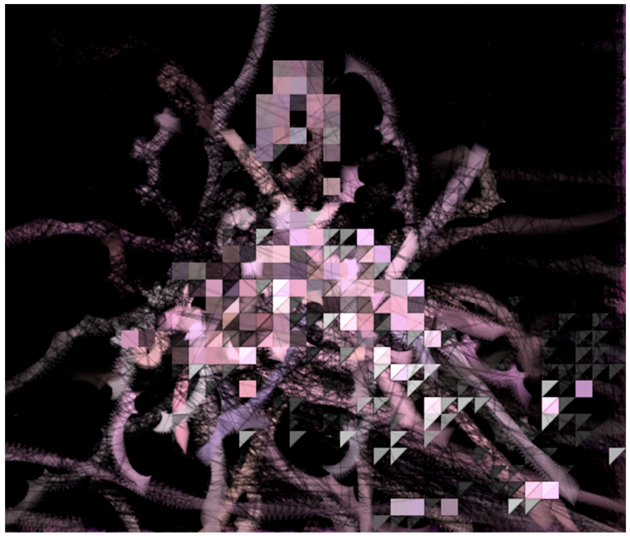

Figure 2. Detail featuring small tiles launching spinners.

Sweepers
Sweepers are instantiated from collections of adjacent active small tiles. They appear visually as outlined triangles with a gradually lengthening linear extension, and are represented sonically with a sustained FM sonority. As with spinners, sweepers are attracted by motion, and can be guided around the animation plane by the performer.

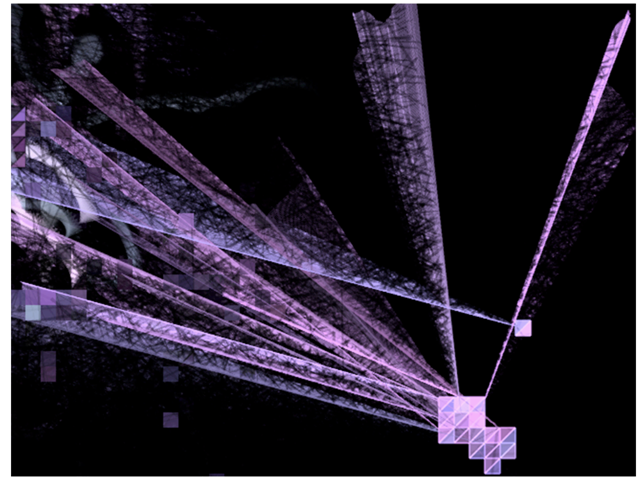

Figure 3. Detail featuring sweepers.

Large tiles
Large tiles are visually identical to small tiles except in scale; they form a spatial and visual grid 8 elements horizontally by 6 elements vertically. Large tiles produce no audio on their first activation; multiple activations within a short time span establish an audio feedback network between adjacent active large tiles. Each row of tiles in the feedback network applies a different type of signal processing to the audio in the feedback network: tremolo, vibrato, bandpass filtering, waveshaping, harmonization, or echo.

Large tiles can interact with sweepers and spinners. When sweepers enter the visual space of an active large tile, the tile changes color, and the sweeper sound changes to a much more inharmonic FM sonority. When a spinner enters the visual space of an active large tile, the tile flashes with the spinner's color, and the system produces an FM synthesis event. Each column of large tiles is associated with a different FM timbre, ranging from relatively harmonic to very noisy; pitch is derived from the spinner's activation pitch. These synthesis events are also input into the audio feedback network associated with the large tile and any adjacencies.

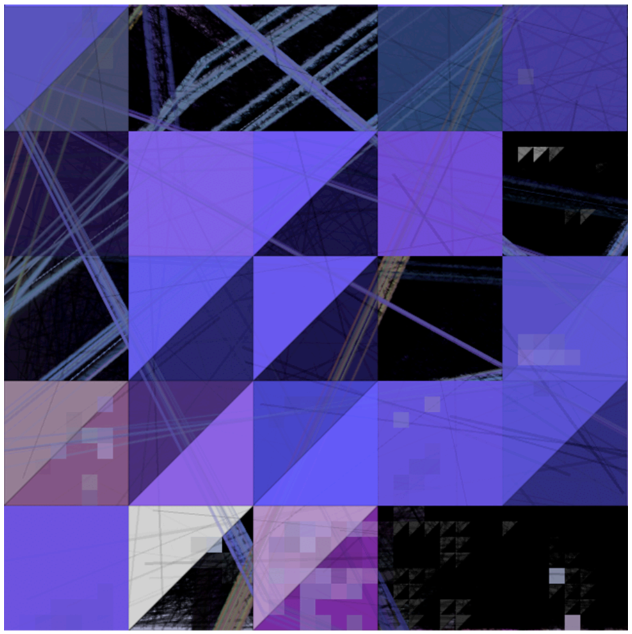

Figure 4. Detail featuring large tiles.

Interaction design

The performance of *Sawtooth* is conducted almost entirely through motion capture. The system captures video in 64 X 48 pixels of resolution at 30 frames per second. The basic motion analysis technique employed is frame differencing – searching pixel-by-pixel for color change between adjacent frames [2].

The initial analysis locates motion throughout the video frame, and associates that motion with small and large tile activations, and with sweeper and spinner attractions. A 20% or greater color change in a given video pixel activates the small tile associated with that pixel. (Because of this threshold requirement, very gradual motions are not detected. This enables the performer to reposition in the motion-capture plane without triggering animation or sound events, by moving slowly). If more than 20 adjacent small tiles are activated within the span of a few frames of video, the enclosing large tile is activated (and additional small tile activations in that region are suppressed). The persistence of small and large tile activations is correlated with the intensity of the color change that instantiates them. Sweepers and spinners are attracted (via motion derived from a gravity simulation) to the pixel with the largest color change value between the last two frames.

The second level of analysis concerns repeated motions through a given tile location. Motion through an activated tile starts a timer. If a third motion through the tile occurs between 800 and 3200 milliseconds after the timer begins, an additional behavior is established. Small tiles begin to launch spinners in a periodic rhythm; large tiles join their audio local feedback network, spiking and then reducing the feedback coefficient, again in a periodic rhythm.

A third level of analysis is provided by comparing previous and current moving averages of the small tile activations over various timescales:

- a change of more than +/-1400 activations across two 5 second windows sets a new scale type for changes to the grid of pitches associated with small tiles
- a change of fewer than +/-100 activations across two 5 second windows changes one column of the pitch grid to the current scale
- an increase of more than 100 activations across two five-second windows decreases the brightness of small tile and spinner synthesis events
- any decrease, or an increase of less than 10 activations across two 5 second windows increases the brightness of small tile and spinner synthesis events
- a change of more than +/-100 activations over two 10 second windows changes the color gamut for new small tile and spinner activations
- an increase of more than 1800 activations across two 10 second windows, and fewer than 6 large tiles activated in the last 10 seconds, causes a search for adjacent active small tiles to instantiate as sweepers

In summary, small motions activate small tiles. Large motions activate large tiles, and suppress small tiles. Repeated motions through a given space activate periodic behaviors. Large quantities of small-scale motion activate sweepers. And any detected motion guides sweepers and spinners around the animation plane, to avoid or create interactions with large tiles.

Balancing control and complexity

The first level of motion analysis results in control that is viscerally satisfying – the performer's movements immediately result in image and sound. This level of control is also easy to learn and control. The fixed correspondence between the motion capture plane and the animation plane is intuitive, the catalog of audiovisual materials and their associated gestures is small, and the animation component of the work serves as an information display for the performer [1].

However, the performer's task is complicated by the multipoint nature of the interface. It is not only possible but aesthetically desirable to activate many audiovisual materials simultaneously, and to manipulate many areas of the animation plane at once. Complexity is also increased by the overloading of gestures: movements activate materials, redirect moving spinners and sweepers, and feed the moving average analyses, all at the same time. These issues are manageable with experience, and exciting to grapple with in performance. Personal criteria for a convincing improvisation with *Sawtooth* include effective use of physical gesture, subtle deployment of the multipoint interface and overloaded gestures, and the creation of convincing temporal form.

The interactions which result from the moving average analyses are controllable, but significantly less precise. The performer is "encouraging" behaviors like the activation of sweepers and changes to spatial mappings of pitch, rather than controlling them. The work also includes behaviors which are entirely autonomous or unpredictable: color evolutions in the animation, and audio outputs from the feedback networks associated with the large tiles. These elements of the work in particular change the relationship between performer and system. The sense of precise control recedes, replaced by a feeling of dialogue with the system. The performer guides the animation and sound over time, but also reacts to and accounts for unexpected results from the software. This sensibility discourages the facility of improvisational clichés and rehearsed gambits, and encourages a genuine sense of exploration and spontaneity in performance.

Balancing clarity and complexity

If *Sawtooth*'s interface is designed to balance performer control with complex, unpredictable, or autonomous behavior from the software, the intention is to strike a similar balance for the audience. Since the audience doesn't experience the interactivity directly, the correlation between physical gesture and multimedia result must be clear and convincing. At the same time, if the work consists only of one-to-one mappings between gesture and result, it is unlikely to rise from the level of technology demonstration to become a convincing aesthetic experience.

Sawtooth's approach to this challenge is to provide an initial sense of clarity. The direct correlation between the motion capture and animation planes, the tight connection between movement and small tile activations, and the sizable physical gestures required to activate large tiles and sweepers all help to clarify the interactive relationships for the audience. The flashes of color which occur as large tiles interact with spinners and sweepers also have an explanatory function.

Beyond that initial threshold, the work prioritizes aesthetic richness over interactive clarity. The more difficult-to-discern behaviors (especially those resulting from the moving average analyses, and from scenarios with many materials interacting) facilitate surprise, complexity, and continuing audience engagement with the performance. My performances of *Sawtooth* typically begin with very small movements, showcasing the easily perceived relationships between motion, sound, and image. Once the basic interactive mappings are established, the performance can continue to larger scales and greater complexity of gestures. The result is both a "tour" of the various capabilities of the performance environment, and a genuinely improvised response to some of the autonomous and unpredictable behaviors of the system. For many viewers, the performer recedes into the background, and the amplification of the performer through sound and animation becomes the foreground experience.

Citations

[1] Burns, C. Sawtooth: Interface as visualization. In *Proc. ICMC 2009*, International Computer Music Association, 549-552.

[2] Levin, G. Computer vision for artists and designers: pedagogic tools and techniques for novice programmers. *AI & Society* 20, 4 (2006), 462-482.

[3] Puckette, M. Pure data. In *Proc. ICMC 1996*, International Computer Music Association, 269-272.

[4] Reas, C., and Fry, B. *Processing*. MIT Press, Cambridge, MA, USA, 2007.

[5] Wright, M. Open Sound Control: An enabling technology for musical networking. *Organised Sound* 10, 3 (2005), 193-200.