Mapping Sound to Image in Interactive Multimedia Art

Zune Lee
CCRMA, Stanford University
The Knoll, 660 Lomita
Stanford, CA, 94305-8180, USA
(650) 723-4971
zune@ccrma.stanford.edu

Jonathan Berger
CCRMA, Stanford University
The Knoll, 660 Lomita
Stanford, CA, 94305-8180, USA
(650) 723-4971
brg@ccrma.stanford.edu

Woon Seung Yeo
CCRMA, Stanford University
The Knoll, 660 Lomita
Stanford, CA, 94305-8180, USA
(650) 723-4971
woony@ccrma.stanford.edu

ABSTRACT
We describe an approach and framework for integrating synthesis and processing of digital image and sound. The approach is based on creating an analogy between sound and image by introducing the notion of the *soxel*, a representation analogous to the pixel. We describe some simple mappings between two domains that map pitch, time, spatial coordinates and timbre to bitmap mode, gray-scale mode, RGB color mode, and layer mode for images. The framework described, sonART, is a powerful multimedia application for integration of image and sound processing with flexible network communication.

Keywords
sonification, music, spatializaion, timbre, image, mapping, visual composition, musical composition

1. INTEGRATING SOUND AND IMAGE

1.1 Spatial Representation of Sound
Consider the piano keyboard, a musical interface in which pitch is arranged in a two-dimensional spatial mapping arbitrarily placing lower pitches to the left of higher pitches (Figure1). Alternate mappings exist in traditional musical instruments. The 'cello', for instance, maps ascending pitch to descending hand position on a single string. The Theremin maps ascending pitch to a leftwards motion. These mappings are easily learned and, in fact, musicians are capable of learning alternative mappings with moderate effort. Other spatial representations of sound exist. These include various timbral representations such as the MDS model of Grey [1], the tristimulus model of Pollard and Janson [2], and the cardinal vowel chart. These later mappings are multi-dimensional. Some (the cardinal vowel chart and the tristimulus model) can describe dynamic temporal timbral change by tracing a trajectory through the coordinate space. A performance oriented multi-dimensional map may be considered by extending the 2-dimensional keyboard map to 3D space (Figure2).

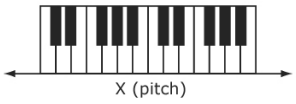

Figure1: Sound as a 1d space (with pitch).

1.2 Sonic Representation of Image
Conversely, the inherent spatial representation of images can be mapped to a multi-dimensional representation of sound. Although the idea of simple mappings of coordinate space to sound is not new (see for example Spiegel's 1981 program, *MusicMouse* [3]), establishing compelling multi-dimensional mappings that express natural time-variant gesture remains elusive.

Our recent work on a framework for exploring multi-dimensional audio-visual mappings has resulted in a layer-oriented framework that offers new flexibility for multi-media and collaborative art. Moreover, the simple three-dimensional mapping suggested in (Figure 2) provides a means of representing gesture in time. The dimensional mappings can represent a broad range of musical dimensions including pitch, loudness, virtual space, etc), or can integrate attributes of a single dimension such as timbre.

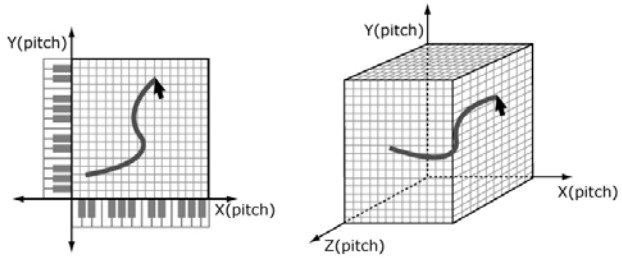

Figure2: Sound as a 2d or a 3 space (with pitch).

1.3 Mapping Auditory Factors to Visual Factors

1.3.1 Sound Element: Soxel
If sound can be expressed as a spatial object with a position in a 2d space, we can simply associate it with a digital image because it is defined as a representation of a two dimensional image as a finite set of pixels [4]. To map sound to image in a 2d space, we introduce the concept of the *soxel,* which is analogous to the pixel. The soxel is defined as the smallest discrete component in a given representation of multi-dimensional musical or auditory space. Just as a pixel can be described by its dimensions (RGB, intensity, etc), the soxel can be described by whatever auditory parameters are employed in a given mapping.

1.3.2 A Simple Mapping
A simple example of mapping between soxel and pixel might entail mapping pitch and loudness to visual color and intensity. The visual representation is dependent on selecting a particular mode such as bitmap mode (black-white), gray-scale mode, RGB color mode, or layer mode.

1.3.2.1 Mapping in Bitmap Mode
In the bitmap mode, we can map the loudness or pitch of soxel to the 1bit intensity of pixel. In Figure 3-1, pitch control is mapped to X and Y coordinates, and the loudness is linked to the 1bit

intensity of bitmap mode. As x and y value increase from 0 to 255 in the XY plane by dragging a mouse, we control the pitch. But the loudness is on-off according to black (0) and white value (1). Assuming that a mouse or other pointing device moves from A to B, the sound volume is 0 outside the white area but 1 inside it while the pitch changes inside the white square. Thus the square functions as an on-off switch rather than a graduated volume dynamic control. In Figure 3-2, the mapping is inversed. Figure3-1 and 3-2 show that the resulted sound can be different depending on the mapping between soxel and pixel under the same image structure.

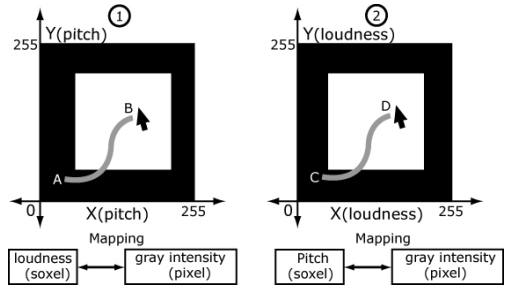

Figure3: Mapping between soxel and pixel in bitmap mode.

1.3.2.2 Mapping in Gray Scale Mode

More dynamic control can be achieved by mapping loudness to gray-scale. In Figure 4-1 the loudness gradually changes inside the gray-scale gradation area by dragging the mouse. However, the pitch change is not heard outside the gradation area because the loudness is zero (black). Figure 4-2 represents a more complex mapping of loudness.

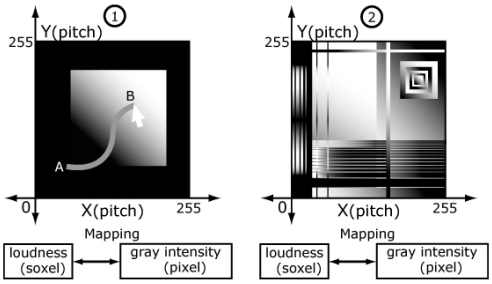

Figure4: Mapping between soxel and pixel in gray scale.

1.3.2.3 Mapping in RGB Color Mode

A still more complex mapping of pitch and loudness can be achieved by using RGB color values. In this example frequency of the soxel is mapped to both the RGB values and the XY coordinates so that dragging a mouse on the square, for example, can change the frequency while doing the corresponding color of Figure5-1 to that of Figure5-2. The four initial colors in Figure5-1 have their own different initial pitches (For instance, red is arbitrarily 440Hz, green 530Hz, blue 680Hz, and yellow 550Hz). The initial colors shift to different colors by dragging the mouse (Figure 5-1 and 5-2). Additionally, each area of the colors is mapped to each loudness of the pitches so that it can vary as the each color area does (Figure5-3). This mapping demonstrates that RGB color composition can interact with sound composition.

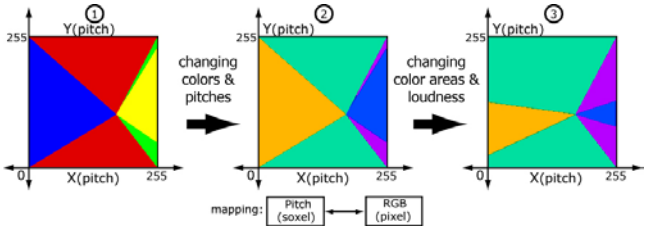

Figure5: Mapping between soxel and pixel in RGB color.

1.3.2.4 Mapping in Layer Mode

An image layer with opacity and blending mode can be mapped to diverse audio parameters [6]. The above-mentioned color mode and soxel concepts can be applied to each layer and also the layer can be merged into one. Mouse interaction on XY coordinates can manipulate the layer opacity and blending amount along with linked audio factors to the layer. Controlling opacity of two layers can create image transition effect between the layers. It can be used for image fade-in and out analogous to sound fades.

2. SONART

The examples described above were explored using SonART, a flexible, multi-purpose multimedia environment that allows for networked collaborative interaction with applications for art, science and industry. SonART supports an arbitrary number of layered canvases, each with independent control of opacity, RGB values, etc., and can transmit or receive data using Open Sound Control, thereby providing a framework for image and data sonification as well as a powerful tool for multimedia performance over a network and facilitates real-time interactive creation, manipulation and exploration of audio and images. The program is currently implemented as a Cocoa-based OS X application.

2.1 Core Features of SonART

Originally created for sound driven data exploration and diagnostic purposes [7], the software has evolved to facilitate real-time multimedia projects and in particular, networked collaborative multi-media art.

- Image display and processing capabilities
    - Intuitive GUI features of audio mixers and of both audio and graphic editing software are preserved.
    - Analogies between audio and visual processes are utilized.
    - An arbitrary number of images can be simultaneously layered.
    - User controlled parameters (i.e., opacity, compositing operations, visibility, etc.) for each layer can be controlled within the program's interface.
- OSC messaging
    - Transmitting and receiving data for sound and image generation and processing.
    - Visual parameters can be both transmitted and received to and from an arbitrary number of networked connections using OSC.
    - Data can be used for sonification using any OSC-supporting synthesis or audio processing engine, or

image processing on another host running SonART separately.

- Data sonification
  - Sonification can be done directly from images of from data linked to image data. For example, the RGB values of a given layer may be used as sonification parameters, or an image pixel might be a reference to a vector of data in a given dataset.
  - The most recent version supports data-stripe mode, in which numeric data could be stored as a line on image to provide the dataset as well as its visualized information at the same time.

In addition we consider color histogram information for visual mapping. Example work 1 requires the information of overall color distribution for mapping background image to a single pitch; this made us realize the need for SonART to provide the color histogram of an image.

3. EXAMPLES

We next demonstrate two examples of soxel-pixel mapping using SonART with synthesis and audio processing done in MAX/MSP via OSC. In these examples the user interfaces with SonART using standard mouse operations of point, click and drag.

3.1 Example Work 1

The example mainly focuses on mapping sound loudness to image size in a linear perspective space. The entire mapping is below:

- Mapping
  - Foreground image object size: foreground sound loudness.
  - Background RGB color (pixel): background pitch (soxel).
  - Background RGB Color area: background sound loudness.
  - X coordinates: background sound pitch & RGB value control.
  - Y coordinates: foreground sound loudness & image size control.

Figure6: Example work 1.

Foreground: The person in the foreground image is mapped to a particular sound. The size of the person is mapped to sound intensity. By dragging the mouse on the Y axis (Figure 6) the loudness of the representative 'person' sound dynamically changes in intensity.

Background: The initial four colors of the background image mapped to the initial pitches and initial volumes as big as each color area. Each pitch and its corresponding RGB value varies with mouse drag along the X axis (Figur6).

3.2 Example Work 2

The next example interposes three image-sound transitions using SonART's layer mode. The mapping is as follow:

- Mapping
  - 1st layer opacity: 1st layer sound loudness and sampling rate.
  - 2nd layer opacity: 2nd layer sound loudness and sampling rate.
  - X coordinates: 1st layer opacity, sound loudness, and sampling rate control.
  - Y coordinates: 2nd layer opacity, sound loudness, and sampling rate control.

1st layer: The 1st layer image is linked to a preloaded sound file. By moving the mouse along the X axis, the image opacity, sound volume, and loudness are simultaneously changed.

2nd layer: The 2nd layer image is also connected to a sound file. The mouse movement along the Y axis simultaneously vary image opacity, sound volume, and sampling rate .

3rd layer: The 3rd layer image is linked to a sound file. As a background image and sound, its opacity and other all sound properties are fixed to their initial values.

Figure7-4 shows a result of the layered mapping of sound to image.

Figure7: Example work 2.

4. REFERENCES

[1] Grey, J. M. "Multidimensional Perceptual Scaling of Musical Timbre." *Journal of the Acoustical Society of America*, Vol. 61, pp. 1270-1277, 1977.

[2] Pollard, H. and Janson, E. V. A tristimulus method for the specification of musical timbre. Acustica 51, 162 – 171, 1982.

[3] Spiegel, L. *MusicMouse.* http://retiary.org/ls/

[4] http://www.wordiq.com/definition/Image

[5] http://www.wordiq.com/definition/Digital_image

[6] Lee, Z. *Painting Music and Playing Visual Arts: The Meta-Synthesis, and Communication in Technology Art*. The Journal of Design Culture and Criticism(JDCC), 4th issue, Ahn Graphics Publishing Company, Seoul, Korea, 2001.

[7] Yeo, W., Berger, J., Lee, Z. *SonART: A framework for data sonication, visualization and networked multimedia applications*. ICMC, 2004.

[8] Cook, P. *Music, Cognition, and Computerized Sound: an Introduction to Psychoacoustics*. MIT Press, Cambridge, MA, 1999.

[9] Rossing, T. *The Science of Sound.* 2nd ed., Addison-Wesley, 1990.

[10] Zelanski, P., Fisher, M.P. *Design Principles and Problems.* 2nd ed., Harcourt Brace College Publishers, Orlando, PL, 1996.

[11] Kandinsky, W. *Point and Line to Plane.* Dover Publication, New York, NY, 1979.