Navigating Landscape Representations as Sound: An Introduction

Eric Kabisch
Arts, Computation and Engineering
University of California, Irvine
ekabisch@uci.edu

ABSTRACT

This paper describes ongoing work that explores human perception of physical space as it relates to the cross-sensory phenomena related to modalities of vision and hearing. Specifically, it is concerned with the use of abstracted landscape data as formal structures for real-time embodied interactive experiences. Using physical space and projected image, users engage in navigation of a representative landscape whose corresponding data structures are simultaneously realized as sound. A description of background motivations is given, followed by details on technical implementations. As this is a preliminary discussion of work in progress, I lay out areas for continued exploration of the themes addressed.

INTRODUCTION

There are two main areas of inquiry motivating this work. The first set of goals involves developing compositional techniques for real-time interactive sound environments. The history of formalism in music composition reflects interest in the application of formal structures to generate innovative and variable works of art from an abstracted analytical aesthetic [Rowe, Cope, Kabisch 2004]. These aesthetic or formal structures have included liturgical chants, the rolling of dice, and serial pitch sequences among others.

Of interest to the author is the use of natural and manmade landscapes as compositional source material. Countless composers such as Debussy and Ravel incorporated ideas from nature in their work, and in 1768 the form of composition known as tone painting was lauded by philosopher Jean-Jacques Rousseau saying it "paints every picture, renders every object, submits the whole of nature to its ingenious imitations." [Rousseau] Most commonly, composers have attempted to mimic sounds of nature in their compositions as opposed to representing visual perception of space. While the *music concrète* of Pierre Shaeffer and GRM was not meant specifically to represent the physical world, his early work has been termed as "directly expressive" whose characteristics include "the relatively primitive nature of the material." [Shaeffer] The *World Soundscape Project* was a study initiated by R. Murray Shafer in the late 1960's, whose main activity was the collection of recorded "found" sounds as a sort of sonic mapping of the physical environment. [Shafer] Their approach was largely to gain perspective on the physical world and social change by becoming an objective listener to naturally occurring sounds (including the less natural forms of noise pollution).

It takes a greater leap to imagine the rendering of visual landscape representations into sound. This is not to say that it has not been attempted. Many philosophers, composers, poets and others have approached the conceptual bridging of distinct sensory phenomena. Take Baudelaire's *Correspondences* [Baudelaire] where he poetically illustrates that "perfumes, colors and sounds correspond." Many writings, including the compilation of essays [Morton] discuss the evolving understanding of relationships between music and painting, or sight and sound. An observation commonly made is that the influence of sound or "musicality" in painting is not matched by equal or compelling reflection of painting in music [Morton, Bosseur]. While daunting, this is an area that holds great potential for experimentation.

A second area of inquiry concerns the ways in which humans perceive, understand, and represent space and landscape. The social and collaborative understandings of physical-auditory space have been preliminarily explored in previous work by the author [Kabisch 2005, Williams]. This work focuses more on the presentation and representation of an abstracted space through the user's navigation of that abstraction. Through representation, the goal is to develop new ways to communicate and understand the complex physical and ephemeral qualities of inhabited space.

The junctions between these stated goals provide vision for long-term trajectories. Along the way, many other issues present themselves. This paper seeks to explicate reflections – in process – upon current explorations of these ideas. No concrete answers or user studies are provided, merely new perspectives and techniques that are informing the author's current research trajectory.

DATA STRUCTURES AND EXTRACTION

Potential methodologies for extracting landscape data are diverse and myriad. The use of three-dimensional geo-referenced data is desirable on many levels and is being pursued concurrently with the methodology described here. However, the aesthetic qualities of computer-generated 3D graphics were determined to be unsuitable to the project's goals at this time. As the current aesthetic goal is to represent natural landscape, it was deemed inappropriate to visualize that landscape through such artificial renderings. In addition, the main thrust at this stage of development is to experiment with various mappings of data to sound, so the method of data derivation is not the paramount concern.

In these explorations, data is extracted from landscape images, which serve as an abstraction of the physical landscape. The images are selected to encompass a range of landscape types in order to gain varied data sets for sound mapping. There exists research on the use of digital images as a data source for sound generation [Meijer, von den Doel, Kamel, Zhao]. I look to their results to find both successful techniques and those which are not appropriate for my goals. It should be noted that the four papers referenced are intended for the mapping of data to sound with no visual component. I will illustrate that I wish to combine the two sensory modalities, and thus have slightly different concerns than the work we reference above.

In [Meijer] and [van den Doel] sound is mapped directly from pixel information. By scanning the image, pixel information such as brightness, hue or saturation determines various parameters of the sound. This allows for a relatively simple mapping, however it also incurs a large amount of processing resources, as the entire image is scanned. Concerns with the abundance of pixel information led the designers to keep their sounds mappings relatively simple in order to limit information overload for the end users.

I propose that by limiting redundant information in the image, we can create more complex and interesting mappings without losing important image data. When looking at landscape images, it can be noted that the defining characteristics are contained within the contours of the terrain. On a micro level, the pixel-to-pixel gradations do contain texture information, however for this project the concern is with larger landscape contours. With this in mind I have experimented with edge detection algorithms in the extraction of contour data from imagery. Various algorithms have been tested and a simple Sobel algorithm was determined to give the best results across a range of images. All image processing is currently done in the program Jitter, which is advantageous in that it shares the same Max/MSP environment used for audio and MIDI processing. Figure 1 shows an example of an unprocessed image, and one processed using the jit.sobel edge detection algorithm.

Figure 1. Image before and after applying the Sobel edge detection algorithm.

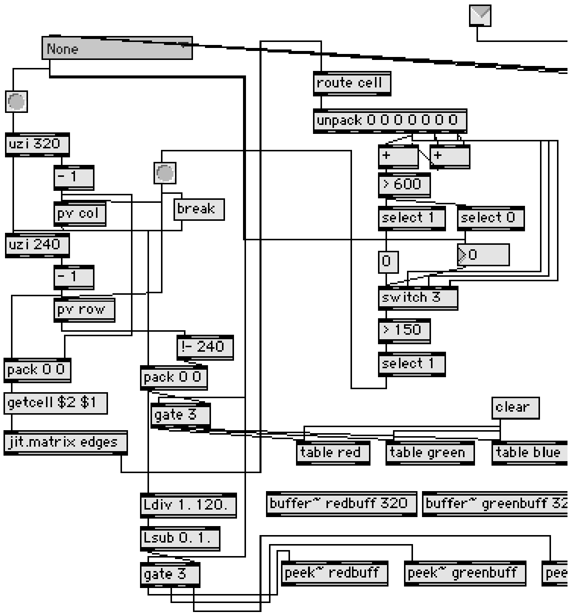

Figure 2. Populating tables and buffers with image data from the RGB planes.

The result of using edge detection on the input image matrix is that we are left with a simplified matrix that has much of the unnecessary data removed. When navigating and mapping the new image matrix black pixels can be ignored, focusing only on the pixels whose color value is over a set threshold.

In addition to directly utilizing data from the edge detection image matrix, other Max tables are populated with even simpler one-dimensional arrays. For instance, by taking the highest (in location) pixel row above the brightness threshold for each column, we end up with an array that shows the contour at the highest point in the image – typically the skyline, ridgeline or horizon. To gain more diverse data sets, this can be done for each plane within the matrix so that we have an array for each color channel (RGB). (see Figure 2)

PRESENTATION AND NAVIGATION OF DATA SPACE

There is an inherent difference in the way that we perceive sound vs. the way we perceive visual stimulus. Standing in one location, a person can see only a subset of the 360° panorama. And even within the range of their peripheral vision they are actively focused on only a small portion at any given time. The way that the human auditory system, and sound waves in general operate, we hear sounds from all directions at the same time. While rotating the head might enact slight changes in timbre, azimuth angle, and other perceptual qualities, in general the focusing on specific sounds within a field of sound is not overtly physical (as the rotating of the head or movement of the eyes) but seems to happen at a cognitive level.

While this is true, trying to listen to all of the visual content contained in an image at once would be a very difficult task. The temporal nature of sound allows us to make sense of sound as a function of time, not as discrete snapshots. This principle informs our design, and the design of most similar work. In [Meijer, von den Doel, Kamel, Zhao] they use one point, or one scan line to determine the area of the image that is being sonified at any given moment. [Meijer] uses a vertical scan line that sweeps across the image at a given rate, then issues a noticeable click indicating to the listener that a new frame is being scanned. The listener becomes a passive receiver of the experience with no control over the exploration of the data. For their purposes this is ideal, but in our scenario we wish to give the user more control. [von den Doel] places control of the image space in the hands of the user by allowing them to move a pointing device over the image surface. However instead of using vertical scan lines it renders sound on a pixel-by-pixel basis.

My experiments use a combination of these two approaches. The combination of user control with the use of vertical scan lines enables the user to become immersed in a virtual space that they can navigate on their own terms. Given the distant, panoramic nature of the imagery used, the vertical scan line works quite well to navigate the image in relation to the user. The user is not controlling a pointing device, but is inside a circular projection of the panoramic image. Physical movement around the circumference of the circle is sensed, and the vertical scan line follows the user around the image, enacting the mapped sounds of the landscape's data space. In order to reinforce the relationship of the user to the projected image, pixel blocks are highlight as their relevant data is mapped to sound. This creates an understanding of the space informed by the synergy of the visual display, the user's physical location within the installation, and the spatialized soundscape.

MAPPING OF DATA TO SOUND

Decisions made in the mapping of data to sound are as much artistic and philosophical as they are technical. Much of the literature on data sonification aspires to a clean one-to-one mapping of the data to sound in order to maintain the integrity of the source data. This is especially true and necessary in [Meijer] where the goal is to represent images to the visually impaired. However, representations in the arts have the luxury, and even a compulsion, to interpret data is a way that communicates more than the source material itself. Metaphors of cartography are perfect for this project, which seeks to represent landscape and space. Jorge Luis Borges tells the tale of the Empire whose cartography became so advanced that they created a full scale map of the entire Empire and placed it right on top of the land, point corresponding to point [Borges]. It seems that in map making, as in other forms of representation, it is not only what is included that matters but what is left out.

In these experiments however, the goal of a simplified mapping is not thrown out entirely, but seen as a starting point. The hope is that the underlying image data, when left to simplistic devices, will speak for itself and allow for interesting contrasts between landscapes. Once grounded in more simple mappings, the system can become more complex, incorporating higher levels of intervention and representation.

Mapping on the Formal (Macro) Level

Data extracted from the images can be used to generate musical structures either on the macro level, as applied to elements such as rhythm and form; or it can be applied on the micro level through sound synthesis and timbre. I outline experiments with both.

As the user navigates the installation space, the vertical scan lines read edge pixels at a rate than maps fairly well to the triggering of individual notes. This idea is used for several examples of varying complexity. *(Implementation notes: the physical sensing and installation is under development and discussions of embodied user interaction are at this time theoretical. All other techniques discussed have been fully implemented, and user interaction is currently simulated by mouse and keyboard input. All of the following examples are programmed in Max/MSP using internal sound processing and also sending MIDI data to synthesizers in Propellerheads Reason software)*

The mapping that seems to be most common, used in [Meijer] and [Zhao] among others, is a mapping from vertical pixel location to pitch. Their tests have shown that this mapping is fairly intuitive to the average user. This technique is used for the first mapping in the system outlined here. The image matrix obtained from the edge detection algorithm is scanned as the user's position moves horizontally. Locations around the circumference of the installation space are mapped to a horizontal column of the image. As the user's movement enacts each column, the entire column is scanned for pixels whose combined red, green, and blue values are above a certain threshold. When pixels are found, their row number is sent out as control data. This control data is remapped to relative frequencies within the audible range and sent to a sine wave oscillator. The resulting sound mapping allows you to trace the contour of the image, with pitch increasing as the height of the image contour increases. In addition I have made the sound pan from left to right with the movement of user navigation.

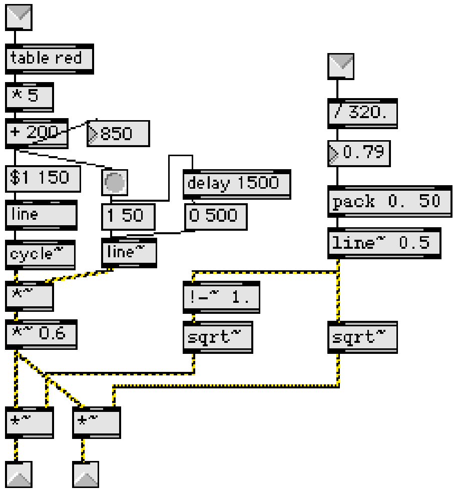

Figure 3. Control data mapped to a panning oscillator.

The next experiment involves the use of all three matrix planes to control separate oscillators. This is a more interesting mapping as it illustrates the differences within the color planes. With some images, the planar array data is quite similar, however with the proper image and edge detection settings the system arrives at data that results in contrapuntal movement between the various planes. The implementation shown in Figure 4 displays three sine wave oscillators panned and set in octave intervals from one another. Other interesting relationships can be set in motion when changing the types of synthesis and intervals, but I have kept it simple for this illustration.

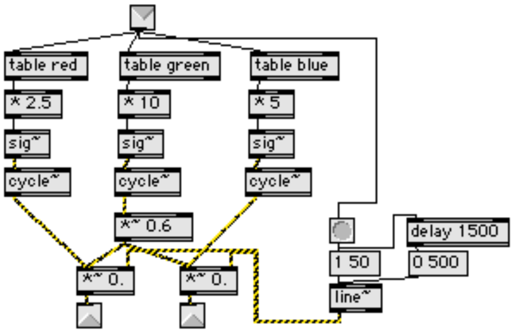

Figure 4. Image color planes mapped to separate oscillators.

A fourth mapping begins to explore a more subjective way of representing the image data. In this example, the user or composer draws in pitch contours to integrate with the navigation of the image data. These pitch contours could also be derived from other sources in similar ways that we have extracted data from the images. The interesting component of this technique is the imposition of multiple pitch contours onto the same navigational framework. I have also built in the capacity to record and playback loops of recorded control data. This concept could become very interesting within the context of the interaction if employed properly.

Mapping on the Synthesis (Micro) Level

Not only are these extracted data structures useful for control data within a larger formal context, but they can also be used as information for the synthesis of sound on the micro level. One way to do this is to use our existing data as lookup tables for waveshaping or additive synthesis [Rhoads]. If familiar with viewing Cartesian representations of audio waveforms, one can almost imagine a skyline or ridgeline as one big waveform waiting to be realized (that was actually one of the initial inspirations for this work). By reading through the buffers filled by our image analysis shown in Figure 2, we have done just that. However, the sound produced by these waveforms, while periodic, tends to be perceived as noise. These waveforms can reflect their own character in a more pleasing way when used as waveshaping lookup tables. Figure 5 shows a simple way in which we can use the same data to control both the formal structures (pitch and rhythm) and the timbral structure by shaping a frequency-controlled sine wave with two of our data buffers.

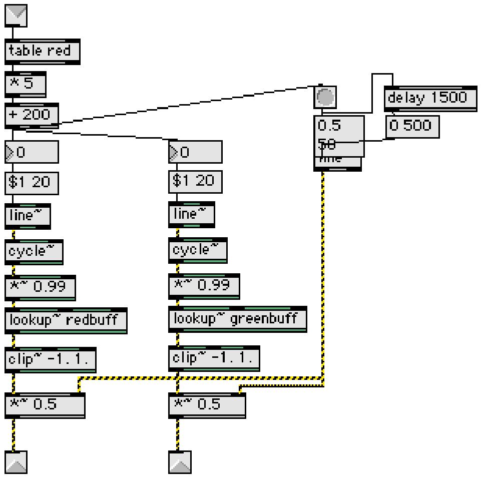

Figure 5. Wavetable synthesis using the image data from both frequency control and a waveshaping buffer.

CONCLUSIONS AND FUTURE DIRECTIONS

The research and implementation described here characterize a starting path toward the representation of landscape through abstracted formal sound structures. While the second goal outlined in the introduction – that which relates to the phenomenological experience of space and landscape – is not directly addressed in these more technical exercises, it has informed decision making in this experimentation as described.

Having developed methodologies for deriving and structuring landscape image data, there is now a sound base on which to further explore mapping of those structures to sound. These techniques only scratch the surface of that which is yet to be discovered, but they ground my research in an understanding of simple mappings from which to branch out. In addition to basic mappings, techniques were explored that hold great potential.

The use of multiple color planes as data structures will allow for many possibilities. The inherent cohesiveness of the coordinated but contrapuntal data obtained through this method suggests a structuralist outlook toward object representation.

The mapping of external structures onto the abstracted data also holds promise. This line of experimentation leads to a level of artistic and compositional representation that is ultimately desired for the use of these techniques in communication. The philosophical give-and-take of asserting control over vs. relinquishing control to the system will be an ongoing concern.

Perhaps the most interesting mapping, musically and ontologically speaking, was achieved by using the same data structures for both macro and micro control of the sound. This reflexivity in sound generation can be viewed as a kind of meta-composition, which has the potential for tremendous congruence and cohesion.

In addition to technical explorations, research into the history of cartography and the collection of geo-referenced data is being conducted. Of particular interest are the socio-political climates under which mapping technologies and the associated data spaces are and have historically been constructed. The ultimate artistic goal is to represent space not only topographically but also to illustrate social, political, ecological and other spatial data through these mappings. By grounding the navigation of these data sets in visual, physical, topographical models, the representation of more phenomenological, temporal and ethereal data can be more readily constructed.

REFERENCES

Baudelaire, C. Correspondences. *Les Fleurs du Mal* (1857).

Borges, J.L. Of Exactitude in Science. *Collected Fictions.* Viking Penguin (1998).

Bosseur, J. *Music: Passion for an Art.* Rizzoli (1991).

Cope, D. *Virtual music*. The MIT Press (2001).

Kabisch, E. *Generative Aesthetics: The Pursuit of Musical Behavior.* Working paper (under review) (2004).

Kabisch, E., Williams, A., and Dourish, P. Symbolic Objects in a Networked Gestural Sound Interface. *Ext. Abstracts CHI 2005,* ACM Press (2005).

Kamel, H.M., Roth, P., and Sinha, R.R. GUESS: Providing Interrelational Representation of Objects in a Non-visual Environment. *Proc. International Conference on Auditory Display.* (2001).

Meijer, B.L. An Experimental System for Auditory Image Representations. *IEE Transactions on Biomedical Engineering, Vol. 39, No. 2, (1992), 112-121.*

Morton, M.L. From the Other Side. From *The Arts Intertwined: Music and Painting in the Nineteenth Century.* Garland Publishing (2000).

Rhoads, C. *The Computer Music Tutorial.* MIT Press (1996).

Rousseau, *Dictionaire de Musique.* Paris (1768). (quoted in Morton)

Rowe, R. *Interactive Music Systems*. The MIT Press (1993).

Schaeffer, Pierre (ed.). *Vers une musique expérimentale: La revue musicale 236*. Richard-Masse (1957).

Shafer, R.M. *The Soundscape: Our Sonic Environment and the Tuning of the World.* Destiny Books (1977).

Stalbaum, B. *The Landscape and Culture: Data as mediator, modes of engagement.* (2003) http://www.c5corp.com/research/landscapeculture.shtml Last accessed on March 19, 2005.

van den Doel, K. SoundView: Sensing Color Images by Kinesthetic Audio. *Proc. International Conference on Auditory Display.* (2003).

Williams, A., Kabisch, E., and Dourish, P. From Interaction to Participation: Configuring Space through Embodied Interaction. Working paper (under review) (2005).

Zhao, H., Plaisant, C., Shneiderman, B., and Duraiswami, R. Sonification of geo-referenced data for auditory information-seeking: Design principle and pilot study. *International Conference on Auditory Display* (2004).