APPLICATION OF IMAGE SONIFICATION METHODS TO MUSIC

Woon Seung Yeo and Jonathan Berger
Stanford University
Center for Computer Research in Music and Acoustics
woony@ccrma.stanford.edu

ABSTRACT

To utilize visual information for musical purpose, inevitable time-based nature of sound should be understood and considered. Time is the principle dimension within which all other auditory parameters are placed, and this poses a particular challenge to effective sonification of time independent images and their applications to music.

In this paper we present two concepts of time mapping, scanning and probing, to provide a framework for conceptualizing mappings of static data to the time domain. We then consider the geometric characteristics of images to define meaningful references in time. Finally we discuss combination of scanning and probing methods in relation to human image perception model, and proceed to suggest its musical applications and implementation with SonART.

1. INTRODUCTION

Due to its inevitable time dependence, data representation in the auditory domain should involve the problem of organization over time. In the area of auditory display, time is not just a parameter, but the principle dimension within which all other auditory parameters are placed. Although this hardly becomes an issue for time ordered information, it plays a crucial role in the problem of sonification of static data such as still images, which are neither organized in time nor containing any time-relevant information inside. Therefore, it requires a mapping to time that is not arbitrarily oriented towards a left-right scan. So far, however, the role of time as the principle in the auditory display has not been paid enough attention compared to its significance for designing and analyzing methods of image sonification. Even the most widely used and effective methods such as inverse spectrogram mapping have not been categorized in terms of time mapping.

Another important factor for image sonification is its geometric characteristics. A still image is defined on a two-dimensional space, and each of its pixels can have three (RGB) or four (RGBA, or CMYK) different color values. This multi-layered two-dimensional property makes it possible to define different types of reference pointers by which current dataset for sonification is located.

In this paper we present two major concepts of time mapping for image sonification, *scanning* and *probing*, to construct a meaningful mapping of static data to time domain. Together with this concept of time, we also suggest the idea of *pointers* as an essential components of time reference based on the geometric framework of images, and discuss the problem of defining their *paths*. Examples of mappings are also presented.

These concepts of time and geometry provide us a soild theoretical background for understanding the mappings of image sonification. Furthermore, they suggest the musical perspective of image sonification: scanning and probing methods can be understood by anaolgy with two different playing styles of musical performances - strictly following scores vs. freely improvising. We discuss this in 4.2 together with the issue of human perception of images, and present *SonART*[1] - a sonification framework which offers a number of powerful features for processing multi-layered image data.

2. CLASSIFICATION OF IMAGE SONIFICATION MAPPINGS

For categorization of image sonification mappings, both *time* and the *geometric reference to time* play crucial roles.

2.1. Construction of time: scanning vs. probing

Methods for organizaing time-independent data for auditory display can be generally classified into two major categories: whether they are pre-scheduled and fixed, or arbitrary and freely adjustable.

2.1.1. Scanning

The term *scanning* refers to the case in which data is scheduled to be sonified in a fixed, non-modifiable order. Figure 1 illustrates the mapping of inverse spectrogram method, which is by far the most popular method for sonifying images. We refer to this as the inverse spectrogram in that the sonification is analogous to the reconstruction of a sound from its spectrogram.

The speed of scanning is usually fixed, and not allowed to be changed arbitrarily throughout the process of sonification. More detailed description of inverse spectrogram mapping is in 4.1. Scanning, however, is not necessarily to be performed along a continuous path. Furthermore, it does not have to cover the whole image area. This issue of *path for scanning* is discussed in 2.2.2

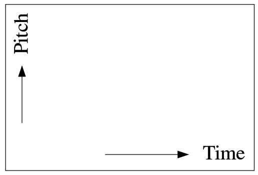

Figure 1. Inverse spectrogram scanning.

2.1.2. *Probing*

If the speed and the location of sonification "pointer" can be varied by an operator arbitrarily during sonification, we classify this as *probing*. In this case, the time domain is sequential only insofar as one may arbitrarily probe anywhere and anytime within an image.

Compared with the scanning process whose result is constructed in a pre-determined and fixed manner, probing delivers much more flexibility for determining the location/path, as well as the speed, of sonification. Any pixel or collection of pixels (i.e., line, area, etc.) can be selected for sonification by navigating around the visual image, with the time domain being sequential.

One of the major characteristics of probing is that it usually requires the image to be scanned (or analyzed) prior to its sonification. This "combination" of scanning and probing processes is discussed in 4.

Freedom in time and location is certainly the most important criterion for sonification methods. However, this by itself still falls short of providing sufficient background to analyze, categorize, and explain every type of sonifications. In the following section we discuss the geometric charcteristics of images and introduce *pointers* and *paths*, which provide us a new perspective.

2.2. Reference to time: pointers and paths

In terms of data sturcture, a still image can be respresented as a three-dimensional matrix. This multi-dimensional property makes it both essential and complicated to define different types of reference by which current dataset for sonification is pointed.

2.2.1. *Pointer*

A pointer means a data element, or a set of data elements, that is mapped to auditory domain *at the same time*. Pointers in different geometric shapes are illustrated in figure 2.

Under our definition, a pointer can be mapped to more than one auditory parameters at the same time. However, it is not supposed to be utilized for multiple events over time: instead, we suggest that this be considered as a *path*.

2.2.2. *Path*

Pointers defined in the previous section can be maneuvered along a number of different *paths* over time. They can be pre-defined (scanning), or freely improvised (probing). Figure 3 shows typical examples of paths, including (a) - a "single point" path, which becomes the case when the same pixel value is repeated to be mapped, or different color values (i.e., RGB or CYMK) are sent in order.

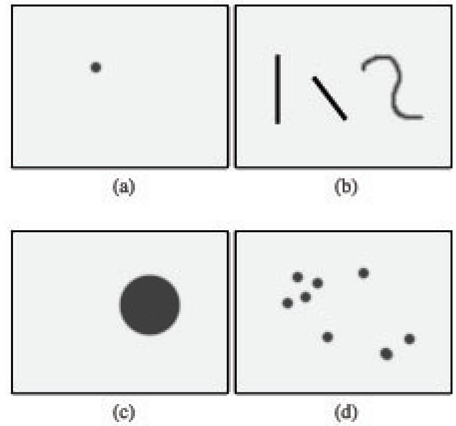

Figure 2. Pointers in different shapes: (a) single point, (b) line/curve, (c) area, and (d) set of distributed points.

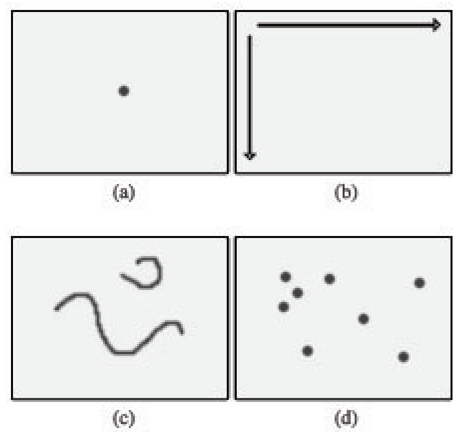

Figure 3. Examples of paths: (a) stationary point (no geometric movement), (b) straight lines along axes, (c) arbitrary curves, and (d) set of distributed points.

3. EXAMPLES OF SONIFICATION MAPPINGS

In this section, we provide five mapping examples of image sonification and classify them according to the aforementioned concepts to show the effectiveness of our model.

3.1. Inverse spectrogram mapping

The general mapping of inverse spectrogram is shown in figure 1. By the terms we defined, this can be classified as a scanning process with a line-shaped pointer moving linearly along the x-axis.

As mentioned before, this is mostly performed as a "scanning" process. However it can also become a probing, which means that its pointer does not have to move

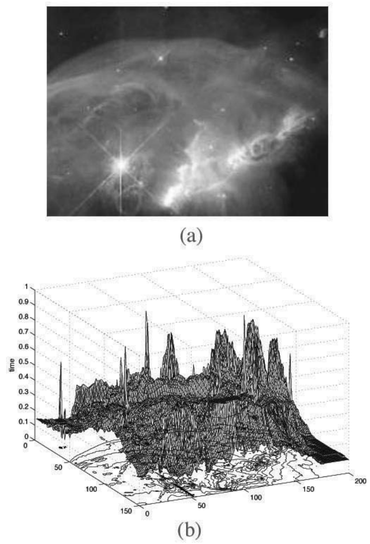

Figure 4. 3D representation of color/brightness. (a) original image, and (b) its 3D representation.

according to the pre-defined schedule; in fact, its position on the x-axis can be freely changed during sonification process. This suggests us an important point: *any set of pointer and path can be used both for scanning and probing*. Although convenient and familiar, this mapping is rather restrictive in that it fails to account for attention grabbing features that may demand prominence in effective auditory display.

3.2. Inverse raster scanning mapping

Raster scanning is a well-known technology for display devices (i.e., TV, CRT monitors) in which two dimensional information is represented as a serial data stream. Consequently, this could serve as a method for sonification in which a path of visual exploration is sonified as a melodic stream. Obviously this is an example of a scanning process with a point pointer, which scans through the whole image horizontally.

3.3. Virtual paths

To show that paths for pointers can be taken in various ways - especially neither x nor y axis, we present two new mappings with “virtually” created paths.

3.3.1. Path on color depth

With this mapping, horizontal and vertical position of each pixel is mapped to spatial parameter and pitch, respectively. Path for time pointer is not mapped to either axis: instead, the color/brightness value of each pixel determines the duration of corresponding sound component. This can be defined as a scanning process, with an areal pointer which takes up the whole image area, and has a path that lies on the third dimension of image data (i.e., brightness/color).

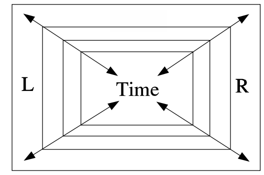

Figure 5. Time on “perpendicular” axis.

3.3.2. Path on an imaginary axis

This model was created under the inspiration of one of the illustrations in ’The Pedagogical Sketchbook’ by Paul Klee [2]. Although being a result of optical illusion, the perpendicular direction that goes into (or out of) the paper/screen provides us a new dimension to which any sonic properties, including pointer paths, can be assigned.

Figure 5 illustrates a model of sonic space in which we are listening to sounds from walls, floor, and ceiling while going “into” the paper/screen. This can be categorized as a scanning process, with a rectangular pointer of time-varying size moving along a perpendicular path.

3.4. Pointers on arbitrary paths

In this section we present the *Viewer.app*[5], one of our sonification and data analysis applications. It displays a visual representation of complex dataset (or an image itself) on screen, and users can probe over the image to send out the data values of visibly salient areas over network.

Unlike the examples shown previously, Viewer.app can be identified as a combination of probing and local scanning: once a visually salient area is detected, user can either probe around that point, or set a linear path over the area to perform a scanning sonification. Therefore, Viewer.app can be described as a probing process combined with local linear scanning, with a point pointer which could move both on an arbitrary path (probing) and a linear path (scanning).

Combinations of scanning and probing methods, together with the issue of preprocessing, will be discussed more in the following section.

4. MUSICAL APPLICATION OF IMAGE SONIFICATION: COMBINATION OF SCANNING AND PROBING METHODS

Numerous composers and artists have used sonification for musical purposes[3]. However, in addition to the simple fact that results of sonifications are sounds, the aforementioned concept regarding time provides us a new perspective of image sonification as a musical device. We now discuss the *combination of scanning and probing methods*, starting from the question of modeling *the way human perceive images*.

4.1. Human image perception model: a new metaphor for music

A variety of scanning methods and mappings provide numerous means of auditory represention of a static image. The primary drawback of these approaches lies in the inability to sequentially focus on particular details or features of an image. Meanwhile, probing method enables us to focus on points of interest and examine its details. However, it lacks the feature of overviewing the whole image in a contextual sense. In addition, unlike the scanning methods which depend on pre-defined trajectory of timeline, paths of probing will vary from user to user. Therefore, it is quite natural to design an alternate approach as a combined form of scanning and probing.

The process of human image perception is a perfect example of this: when viewing a static image, the brain engages in separate processes of object feature finding and spatial adjustment as the visual system combines foveal vision with peripheral vision. The *scanpath* theory [4] proposes that eye movement is driven in a top-down process in which a path of discovery is created between attention grabbing features. In fact, this scanpath combines fixation with saccadic eye movements during image viewing. Therefore, a simple model of these processes can be constructed with local probing and global scanning methods to integrate identification with localization. This not only fits well into our theoretical framework, but also serves as a metaphor for “music inspired by images.”

4.2. Analogy to performance styles: image sonification as musical instrument

Scanning and probing methods can be related with the ideas of musical performance: scanning is like strict following of scores as opposed to the free, improvisatiorial characteristic of probing. Or, this can be stretched even further to the relationship between pre-recorded audio and real-time live performance. In either case, we are interested in not only their combination but also the improvisatorial nature of probing method.

4.3. Application of SonART for Combined Methods

Combination of scanning and probing methods can be well realized with *SonART* [1] - a flexible, multi-purpose multimedia environment that allows for networked collaborative interaction with applications for art, science and industry.

An arbitrary number of layered canvases, each with independent control of visual parameters, can transmit or receive data using *Open Sound Control*. Data from images can be used for synthesis or audio signal processing and vice versa. Therefore, SonART provides an open ended framework for integration of powerful image and audio processing methds with a flexible network communications protocol. Among SonART’s core features, the followings are most essential to our model of combined methods:

- layered image support. At the core of SonART’s image display and processing capability is the canvas of layered images: eacn layer can be thought of an audio channel, and its intuitive GUI with sliders and toggle switches draws an analogy between itself and the audio mixers.

- real-time data transfer. Data such as pixel location on selected layers, opacity settings, RGB values on a per pixel and per layer basis, etc., can be both transmitted and received to and from an arbitrary number of network connections instantaneously. Furthermore, data can be sent to any OSC-enabled synthesis or audio processing engine.

These features, when combined together, make it possible to use SonART as an interface to musical instruments or sound sources that provides users a sonification environment incorporating both scanning and probing support.

5. CONCLUSION

The new framework for image sonification analysis, based on the concepts of time and its reference, provides us a complete and clear description and definition of sonification mappings. Combination of scanning and probing methods serves as a new model for musical applications. And for this, SonART offers an ideal environment.

Future research should be focused on issues regarding implementations of various scanning methods, which will then be followed by the comparison of those models in terms of effectiveness when used with probing method.

6. REFERENCES

[1] Woon Seung Yeo, Jonathan Berger, and Zune Lee, “SonART: A framework for data sonification, visualization and networked multimedia applications” in *Proceedings of the Internaional Computer Music Conference*, Miami, FL, USA, November 2004.

[2] P. Klee, *Pedagogical Sketchbook*, Faber and Faber, Boston, 1968.

[3] Jonathan Berger and Oded B. Tal, “De natura sonoris” in *Leonardo Music Journal*, 37(3), 2004.

[4] L. W. Stark, and S. R. Ellis, *Scanpaths Revisited: Cognitive Models Direct Active Looking*, 1981.

[5] Woon Seung Yeo, Jonathan Berger, and R. Scott Wilson, “A Flexible Framework For Real-time Sonification with SonART” in *Proceedings of the Internaional Conference on Auditory Display*, Sydney, Australia, July 2004.