A FRAMEWORK FOR DESIGNING IMAGE SONIFICATION METHODS

Woon Seung Yeo and Jonathan Berger

Center for Computer Research in Music and Acoustics,
Department of Music, Stanford University,
Stanford, California, 94305 USA.
woony@ccrma.stanford.edu

ABSTRACT

Time is not just a parameter of auditory disply, but is rather the principle dimension within which all other auditory parameters are placed. While this is a feature of time ordered information, it poses a particular challenge to effective sonification of time-independent data, such as images. 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 proceed to suggest a new approach of modeling human image perception for image sonification, which can be designed upon our framework.

1. INTRODUCTION

Due to its inevitable time dependence, representation of data 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. Since a still image is defined on a two-dimensional space and each pixel can have up to four different color values, it can be represented by a three-dimensional matrix. This multi-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 the time domain. Together with this time concept, 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*. Based on these concepts, we present and analyze examples of mappings, and then proceed to suggest a new approach for image sonification.

2. CONSTRUCTION OF TIME DOMAIN: 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. 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.

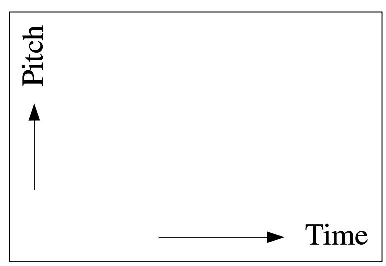

Figure 1: *Inverse spectrogram scanning.*

In this system x-axis is mapped to time, and each timestep the corresponding vertical line is sonified. This, at the same time, is also similar to the time vs. pitch structure of traditional musical scores.

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 3.3.

2.2. 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 analyzed prior to its sonification. This *preprocessing* of images, together with the issue of combination of scanning and probing processes, is discussed in 4.2.

In case of scanning, an auditory display of the general periphery may be informative about context, while probing might be informative about features of specific points. Also in musical terms, probing is an *improvisational* action as opposed to the strict *score following* characteristic of scanning.

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.

3. 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.

3.1. 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.

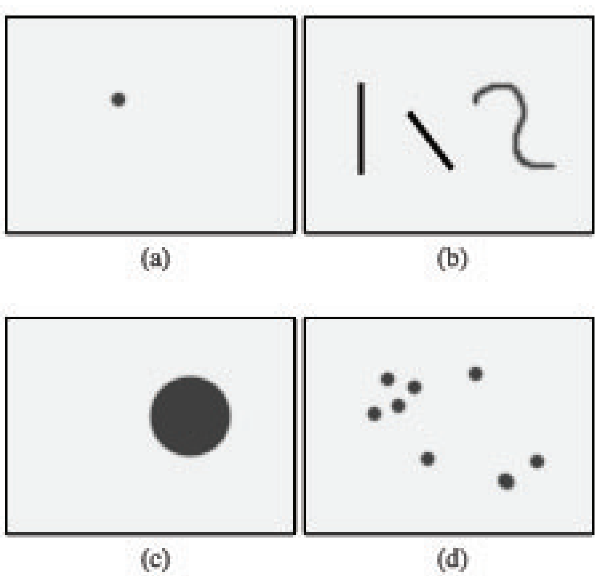

Figure 2: *Pointers in different shapes: (a) single point, (b) line/curve, (c) area, and (d) set of distributed points.*

A single point on image shown in figure 2 (a), however, does not necessarily correspond to one data element: in case of images with colors, up to three or four values could be contained within a point. This possible multi-dimensionality should also be considered for other types of pointer shapes.

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*.

3.2. 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 examples of typical paths.

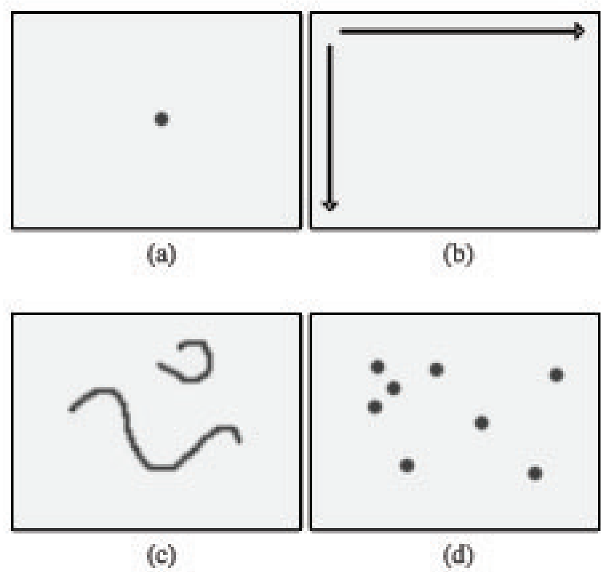

Figure 3: *Examples of paths: (a) stationary point (no geometric movement), (b) straight lines along axes, (c) arbitrary curves, and (d) set of distributed points.*

Figure 3 (a) shows the example of a “single point” path: this 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,

over time.

4. ANALYSIS OF SONIFICATION MAPPINGS: EXAMPLES

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.

4.1. 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 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.

4.2. Raster Scanning Mapping

Raster scanning is a well-known technique 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.

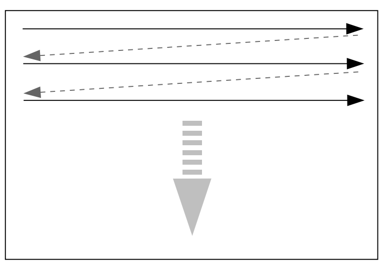

Figure 4: *Raster scanning method.*

And this is an example of

- a scanning process,
- with a point pointer, which scans through the whole image horizontally.

One advantage of this model is its nature of repetition, which could serve as a guide to regular tempo; in musical terms, the contents (pixels) of image could be thought of note events, while both ends of image function as bar lines. Also, in a much shorter time scale, size of repetition can function as the length of a delay line, thereby determining the frequency of resulting sound.

For sounds with more than one channel, we can apply the idea of interlaced raster scanning.

4.3. 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 “virtual” paths.

4.3.1. Path on Color Depth

In this mapping, horizontal and vertical position of each pixel is mapped to spatial parameter and pitch, respectively. Path for time

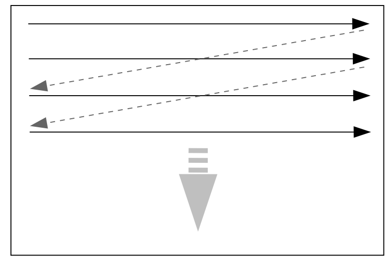

Figure 5: *Interlaced Raster scanning.*

pointer is not mapped to either axis: instead, the color/brightness value of each pixel determines the duration of corresponding sound component (i.e., brighter pixels get longer durations).

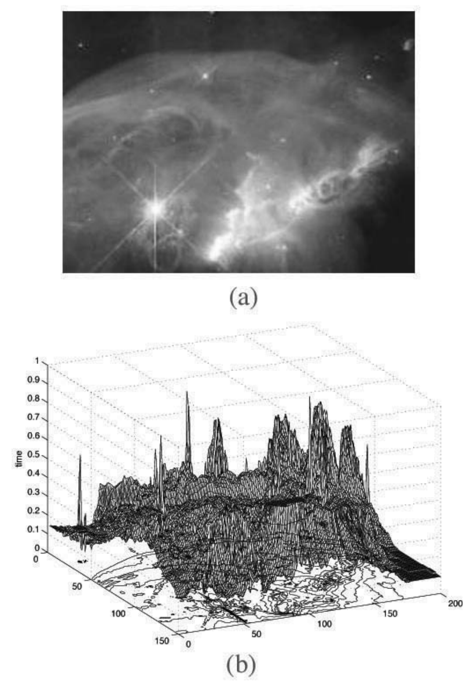

Figure 6: *3D representation of color/brightness. (a) original image, and (b) its 3D representation.*

And it can be defined as

- a scanning process,
- with an areal pointer which takes up the whole image area, whose path lies on the third dimension of image data (that is, color values of pixels)

This mapping can provide an effective means for implementation of various detailed decay patterns of sound not only over frequency but also over panning space. For colored images, color components can be arranged in various sequences to produce different distributions of sound particles over time.

4.3.2. *Path on 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 can be assigned. In this case, it is utilized as the dimension for pointer paths.

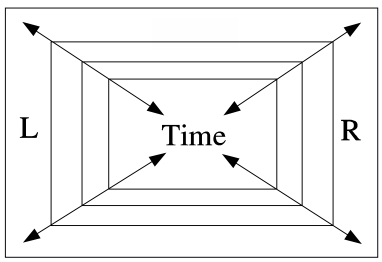

Figure 7: *Time on "perpendicular" axis.*

This is a model of sonic space in which we are listening to sounds from walls, floor, and ceiling while going "into" the paper/screen. And this can be categorized as

- a scanning process,
- with a rectangular pointer, moving along a perpendicular path. This makes the size of the pointer vary.

4.4. Pointers on Arbitrary Paths

In this section we present two of our sonification tools. Unlike the previous examples, the followings can be identified as a probing process, and a combination of probing and local scanning, respectively.

4.4.1. *SonART*

SonART [3] is a flexible, multi-purpose multimedia environment that allows for networked collaborative interaction with applications for art, science and industry. Figure 8 is a screenshot of SonART.

It supports an arbitrary number of layered canvases, and each canvas can display a unlimited number of images in layers. By dragging mouse over any canvas, the opacity and RGB values of pointed pixel in every layer can be transmitted through network for sonification. Detailed description about SonART can be found in [3].

- a probing process,
- with a point pointer on an arbitrarily controllable path.

4.4.2. *Viewer.app*

Viewer.app [4] is a data analysis application which works quite similar to SonART: a visual representation of complex dataset (or an image itself) is displayed on a canvas, and users can probe over the image to send out the data values of visibly salient areas over network. Figure 9 shows its screenshot, with a hyperspectral image of a colon cell displayed on canvas.

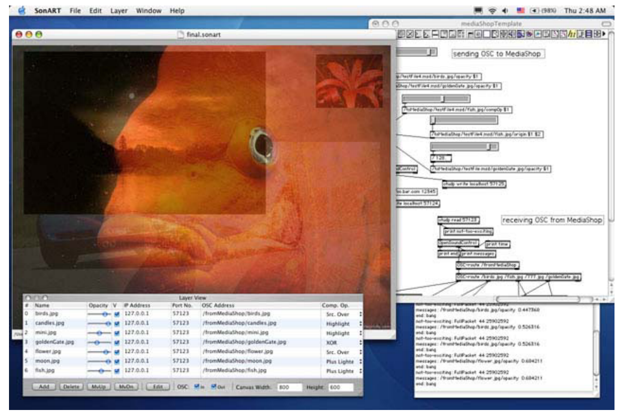

Figure 8: *SonART: a Mac OS X application for multi-purpose, networked multimedia application.*

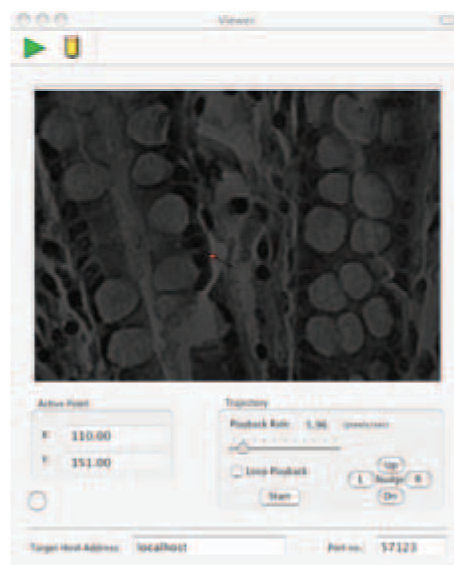

Figure 9: *Viewer.app: a Mac OS X application for probing images of hyperspectral data.*

Unlike SonART, however, Viewer.app provides a local scanning feature: 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).

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

5. IMAGE ANALYSIS PRIOR TO SONIFICATION: PREPROCESSING

5.1. Combination of Scanning and Probing

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. Viewer.app, previously discussed in 4.4.2., serves as an example of global probing supported by local scanning.

In general, however, two methods are utilized together as a combination of global scanning followed by local probing. In those cases, the foregoing scanning process is referred to as *preprocessing* of the sonified image. By nature, most probing methods involve preprocessing of data, which could be categorized into two groups: preprocessing in visual domain, and in auditory domain.

5.2. Preprocessing in Visual Domain

Auditory probes can be initiated and guided by vision, and an attention-grabbing area may prompt placing a pointer over a particular point to hear the sonified output. This process of locating the position to probe - whether being performed with enough "attention" or not - can be interpreted as a processing/analysis of data, which occurrs prior to the actual sonification of data.

Pre-processing of images could also be done automatically by scanning the image with threshold rules, and probing any point that crossed the threshold.

5.3. Preprocessing in Auditory Domain

In certain areas of auditory display, preprocessing in visual domain is not available as an option; For example, synthetic vision through auditory representations such as the *vOICe* [5], and other types of auditory aids for people already concentrating on different visual representations, cannot utilize the image itself for presenting information; instead, it must be scanned in the auditory (or some non-visual) domain for further analysis on salient points.

6. POSSIBLE EXTENSION: HUMAN IMAGE PERCEPTION AS A SONIFICATION MODEL

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 [1] 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 thought of in terms of probing and scanning to integrate identification with localization, and this fits well into our framework. We believe this will result in a new meaningful method providing comprehensive and informative auditory feedback - both in global context and local feature - of a static image.

7. CONCLUSION

In this paper we presented a framework for image sonification analysis based on scanning and probing copcents. It provides us a complete and clear description and definition of sonification mappings. We believe this will help us develop a new, effective model of human image perception as a method for image sonification.

8. REFERENCES

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

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

[3] 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.

[4] 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.

[5] Peter Meijer, *The vOICe: Visual Prosthesis for the Totally Blind*, http://www.visualprosthesis.com/voice.htm.