Raster Scanning: A New Approach to Image Sonification, Sound Visualization, Sound Analysis And Synthesis

Woon Seung Yeo and Jonathan Berger
CCRMA, Department of Music, Stanford University
woony@ccrma.stanford.edu

Abstract

Raster scanning is a technique for generating or recording a video image by means of a line-by-line sweep, tantamount to a data mapping scheme between one and two dimensional spaces. While this geometric structure has been widely used on many data transmission and storage systems as well as most video displaying and capturing devices, its application to audio related research or art is rare.
In this paper, a data mapping mechanism of raster scanning is proposed as a framework for both image sonification and sound visualization. This mechanism is simple, and produces compelling results when used for sonifying image texture and visualizing sound timbre. In addition to its potential as a cross modal representation, its complementary and analogous property can be applied sequentially to create a chain of sonifications and visualizations using digital filters, thus suggesting a useful creative method of audio processing.
Special attention is paid to the rastrogram - raster visualization of sound - as an intuitive visual interface to audio data. In addition to being an efficient means of sound representation that provides meaningful display of significant auditory features, the rastrogram is applied to the area of sound analysis by visualizing characteristics of loop filters used for a Karplus-Strong model. Construction of new sound synthesis systems based on texture analysis / synthesis of the rastrogram is also discussed.

1 Introduction

Data conversion between the visual and audio domain has been an active area of scientific research and various multimedia arts. Examples include waveforms, spectrograms, and numerous audio visualization plug-ins, as well as visual composition and image sonification software such as Metasynth (U&I Software) and Audiosculpt (IRCAM).

Since these conversions essentially represent data mappings, it is crucial to understand and utilize the nature of datasets in both audio and visual domains to design an effective mapping scheme. The temporal nature of a sound and the time-independent, two-dimensional nature of an image requires that data mappings between the two media address these fundamental differences.

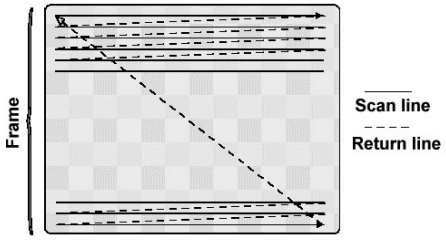


Figure 1: Raster scanning.

1.1 Raster Scanning as a Data Mapping

Raster scanning is a technique for generating or recording the elements of a display image by sweeping the screen in a line-by-line manner. More specifically, it scans the whole area, generally from left to right, while progressing from top to bottom of the imaging sensor or the display monitor, as shown in figure 1.

In addition to being the core mechanism behind most video display and capturing devices, geometric framework of raster scanning - a mapping between one- and two-dimensional data spaces - can be found in other places such as communication and storage systems of two-dimensional datasets. This is, in fact, the property that receives primary attention in our choice of raster scanning as a new mapping framework between image and sound.

Raster scanning provides an intuitive, easy to understand mapping scheme between one- and two-dimensional data spaces. This simple, one-to-one mapping also makes itself a totally reversible process: data converted into one representation can be reconstructed without any loss of information.

2 Review of Comparable Works

Together with almost every introduction of devices which could record and store audio and/or visual media information, numerous attempts have been made to convert data from one domain into the other. Among other works in the early days, special attention is paid to the technique of *animated sound* developed by McLaren (McLaren and Jordan 1953), in which he "drew" lines and curves on the audio portion of his films to create the sound for his motion pictures. In spite of being possibly subjective and unclear, his method has a huge implication in that, for each picture frame, it was a mapping from two-dimensional stationary images to one-dimensional time dependent sounds.

Computer technology and digital media formats opened up a new world for multimedia art, pioneered by Whitney (Whitney 1980) and others. By constructing a highly organized set of data mappings between musical events and visual motion patterns/colors, Whitney combined both domains to create "an inseparable whole that is much greater than its parts:" instead of being satisfied with only unidirectional conversions, he had the vision of interchangeability between audio and visual domains, and emphasized the advantage and power of digital media that enabled it.

In terms of geometric framework, applications of raster mapping to sound-image conversions can rarely be found. However, there are some comparable works based on different types of scanning methods.

2.1 Sound Scanner

In (Kock 1971), Kock used a "sound scanner" to make sound visible: he put a small microphone to a long motorized arm which swept out a raster-like arc pattern. Attached to the microphone was a small neon light driven by an amplifier connected to the microphone. Sound received by the microphone would then light up the lamp, which was photographed in darkness with a long exposure. The resulting picture, therefore, depicted the sound level as bright patterns.

While the result of our raster visualization is time dependent, Kock's work was spatial rather than temporal; although the light source swept the space over a period of time, its results were standing waves. In terms of loudness to brightness conversion, however, they are based on a similar mapping rule.

2.2 Spiral Visualization of Phase Portrait

To visualize period-to-period difference of a sound, Chafe (Chafe 1995) projected phase portrait of a sound onto a time spiral, thereby separating each period. In this mapping, time begins at the perimeter and spirals inward, one orbit corresponding to one full period of the waveform. Also, trace can be colored according to specific spectral qualities of the sound.

Spiral drawing path, together with the use of spectral transform, sets this technique apart from raster visualization. However, it should be noted that this was proposed as an analysis tool for designing physical synthesis models, especially with pitch-synchronicity in mind. Application of the raster mapping method to a similar problem will be discussed in §5.

2.3 Wave Terrain Synthesis

In wave terrain synthesis (Bischoff, Gold, and Horton 1978), a "wave surface" is scanned in an "orbit" (a closed path), and movement of the orbit causes variations in the generated sound. Obviously this is a mapping from two- to one-dimensional data space, which is applicable to image sonification.

Techniques that scan a wave terrain have been explored by several researchers, including (Borgonovo and Haus 1984). None of them, however, are similar to the raster scanning path we propose.

2.4 Scanned Synthesis

Scanned synthesis (Verplank, Mathews, and Shaw 2000) is a sound synthesis technique which "scans" a closed path in a data space periodically to create a sound. Due to its emphasis on the performer's control of timbre, data to be scanned is usually generated by a slow dynamic system whose frequencies of vibration are below about 15 [Hz], whereas the pitch is determined by the speed of the scanning function. The system is directly manipulated by motions of the performer, therefore can be looked upon as a dynamic wavetable control.

While scanned synthesis can be characterized by various scanning patterns and data controllability, raster sonification features a fixed geometric framework dedicated to converting rectangular images to sound.

2.5 Research on Mapping Geometry

In (Yeo 2001), Yeo proposed several new image sonification mappings whose scanning paths and color mappings are different from those of the inverse spectrogram method that is most widely used. Examples include vertical scanning, and scanning along a virtual "perpendicular" axis, with horizontal panning.

This research for mapping geometry has been further refined in (Yeo and Berger 2005) to provide the concept of *pointer - path* pair, which serves as the basis of a general framework for mapping classification.

2.6 Significance and Contributions

In summary, the following can be proposed as possible contributions of this research in relation to existing works.

- By adopting raster scanning method, we can construct a simple and reversible mapping framework between image and sound with complementarity. This enables us not only to utilize images as sound libraries (or sounds as image libraries), but also to edit sounds by modifying corresponding images (or vice versa), in a highly predictable way.
- Raster sonification creates a sound which evokes the visual texture of the original image in detail. Although it lacks the freedom of control provided by scanned synthesis, its geometric framework proves to be highly effective with two-dimensional images.
- Raster visualization is also proposed as an intuitive tool for timbre visualization, sound analysis, and filter design for digital waveguide synthesis.
- Moreover, combination of raster sonification and visualization not only suggests a new concept of sound analysis and synthesis based on image processing techniques, but also has strong implications for artistic applications, including cross-modal mapping and collaborative paradigm.

3 Image Sonification

Currently, raster mapping for sonification is defined as follows:

- Brightness values of grayscale image pixels, ranging from 0 to 255 (8-bit) or 65535 (16-bit), are linearly scaled to fit into the range of audio sample values from -1.0 to 1.0.
- One image pixel corresponds to one audio sample.

Figure 2 illustrates these rules.

3.1 Basic Properties

Because of the natural periodicity found in the majority of images, the sonified result of an image sounds "pitched:" the width of an image determines the period (thereby the pitch) of its sonified sound. Also, by the one-to-one sample mapping, area of an image corresponds to the duration of its sonified sound.

In addition to width, pattern changes in the vertical direction are represented as similar changes in timbre over time.

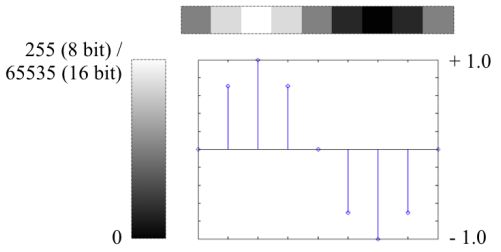
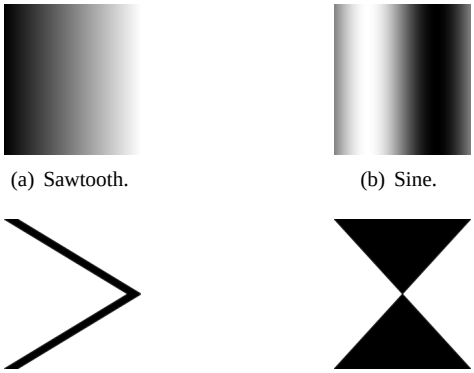


Figure 2: Rules of raster mapping.



(a) Sawtooth. (b) Sine.

(c) Pulse frequency modulation. (d) Pulse width modulation

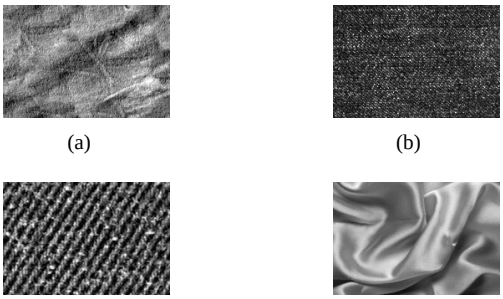
Figure 3: Images with constant and changing vertical patterns. Sonified results are specified individually.

This is depicted in figure 3: images in 3(a) and 3(b) sonifies into sounds with constant sound quality, whereas 3(c) and 3(d) are exaples of time-varying timbre.

3.2 Texture Sonification

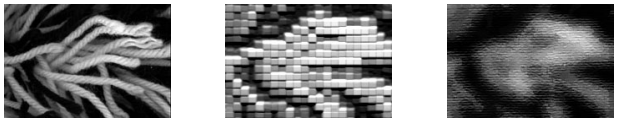
Raster mapping proves to be highly effective when used for sonifying the fine "texture" of an image. Sonified sounds preserve the feeling of the original images quite similarly in the auditory domain, and are quite useful for discriminating relative differences between various image textures. Figure 4 illustrates four images with contrasting textures: a number of tests have shown that most people could correctly match the original images with their sonified sounds when they were given all of them simultaneously.

Providing an absolute auditory reference to a particular visual texture, however, is a challenging task. To this end, construction of a large set of image-sound pairs as a mapping library, together with their classification and training, is desirable.



(a) (b) (c) (d)

Figure 4: Images used to compare textures by sound.



(a) Original. (b) "Patchwork". (c) "Grain".

Figure 5: Comparison of visually filtered images.

3.3 Effects of Visual Filters

Raster mapping also produces interesting and impressive results when used for sonifying the effects of various visual filters. Figure 5 contains an image together with its filtered results. Sonified results of these images are very convincing. Compared with the sound generated from the original image of figure 5(a), sound of figure 5(b) produces the auditory image of small bumpy texture, while that of figure 5(c) feels more noisy and grainy.

Sonification of visual filter effects could be further developed as the concept of "sound manipulation using image processing techniques" when combined with corresponding raster visualization.

4 Sound Visualization: Rastrogram

Visualization of sound with raster mapping is an inverse process of raster sonification.

- Values of an audio samples (from -1.0 to 1.0) are linearly scaled to fit into the range of brightness values of image pixels (within the range from 0 to 255).
- One audio sample corresponds to one image pixel.

A new term *rastrogram* is proposed as its name.

Compared with ordinary waveform displays and/or spectrograms, rastrogram can be considered to be highly space-efficient - equally as much as spectrogram. For example, to display a waveform of a one-second long sound recorded at the sampling rate of 48 [kHz] *without any loss of sample*, an image with the width of 48,000 pixels is needed. Same sound, however, can fit into a rastrogram with an *area* containing the same number of pixels (i.e., 240×200), viewable within most computer display. Naturally, sound duration contributes to the image area (and height).

4.1 Width And Pitch Estimation

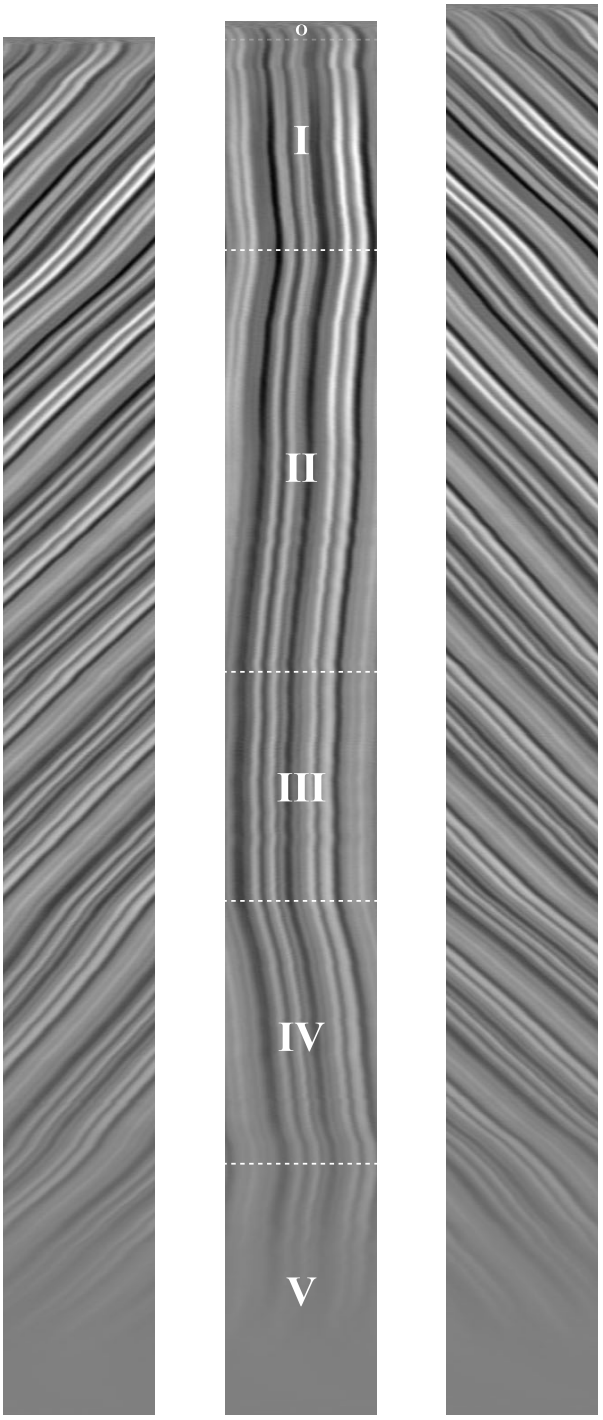
Rastrogram is basically a representation of short segments of audio samples stacked from top to bottom over time. Since it shows the phase shift between each of those segments, rastrogram can be a useful tool for visualizing changes of pitch over time.

To make this effective, its width should be properly chosen to match the length of one period of sound as closely as possible, thereby being *pitch-synchronous*. In case of an exact match, every "stripe" of rastrogram should align on a perfectly vertical direction. Due to the limited precision of frequency values obtained by integer-only image widths, however, unwanted "drifts" can be introduced by round-off errors: stripes will usually slope in either way, depending on the instantaneous pitch value.

Figure 6 shows three rastrograms that are generated from the same violin sound, but of different width. Obviously 6(b) is most well-synchronized to its pitch, which is supposed to be around $44,100/170 = 259.4$[Hz]. From a closer inspection of this, we see the followings:

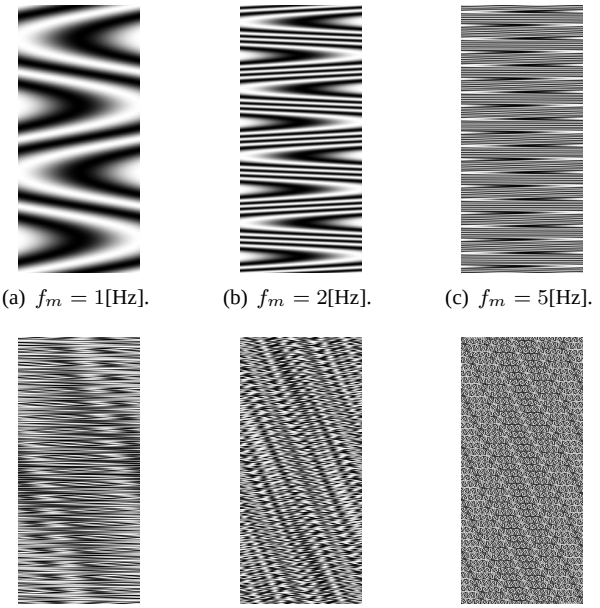
- Inclination at a certain point provides the amount of relative delay to the width of rastrogram, which makes it possible to derive more precise pitch. Figure 6(b) shows that the original sound could be largely segmented into five different pitch sections, with roughly estimated pitch values of 259.3 [Hz] (I), 259.5 [Hz] (II), 259.4 [Hz] (III), 259.2 [Hz] (IV), and 259.6 [Hz] (V), respectively.
- In addition, it is clearly shown that there are a number of short, subtle changes of pitch throughout the duration of sound. These include the relatively big one in the beginning of the sound (O), which introduces temporary pitch shift down to about 257.9 [Hz].

It should be noted that rastrogram visualizes fine details of pitch variations with extremely high precision *in both time and frequency*, which can hardly be achieved by spectrogram alone.



(a) 171 pixels. (b) 170 pixels. (c) 169 pixels.

Figure 6: Comparison of rastrograms from the same violin sound, but of different widths (as specified). Sound sample was obtained from the *Musical Instrument Samples* of *Electronic Music Studios, University of Iowa*.



(a) $f_m = 1$[Hz]. (b) $f_m = 2$[Hz]. (c) $f_m = 5$[Hz].

(d) $f_m = 10$[Hz]. (e) $f_m = 20$[Hz]. (f) $f_m = 50$[Hz].

Figure 7: Rastrograms of frequency modulated sounds, $y(t) = \cos\left[2\pi\{f_c t + I \cdot f_m \cdot \sin(2\pi f_m t)\}\right]$, with $f_c = 220.5$ [Hz], $I = 1$, but different f_m values (as specified).

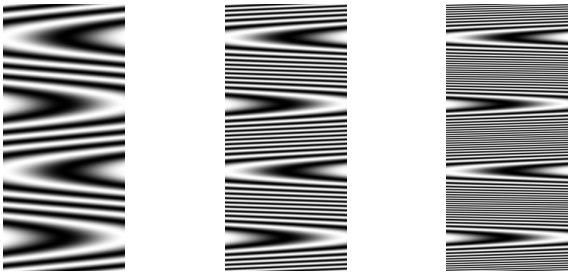
4.2 Frequency Modulation

Rastrograms of frequency-modulated sounds with different values of modulation frequency are depicted in figure 7. With the same modulation index, threshold of image "clarity" lies around 10[Hz] of modulation frequency, which roughly matches the perceptual characteristics in the auditory domain.

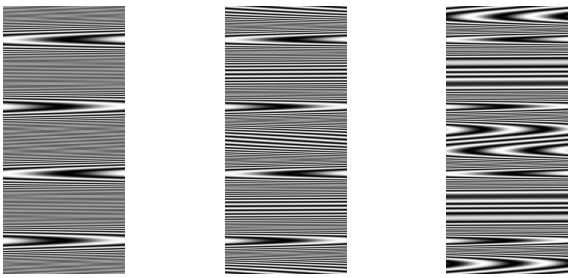
Figure 8 shows the results of various modulation indices. Compared with the previous case of modulation frequency, image patterns remain relatively clear for higher values of modulation index.

Rastrograms of FM sounds with higher modulation frequency and index values are generally complex, and require further research to be analyzed and fully understood. On the contrary, sounds with relatively low modulation frequency and moderate modulation index produce quite simple rastrograms, from which both parameters could be derived. Also, they show a wood-like texture.

This, in turn, means that selected woodgrain images as in figure 9(a) could be raster sonified to synthesize FM-like sounds. Figure 9(b) shows a simplified "synthetic" rastrogram, which is generated from an FM synthesizer whose parameters were designed to emulate this natural woodgrain. Except for the fine details of the natural woodgrain side, sounds of both figures are quite similar to each other in terms of pitch.



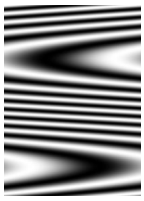
(a) $I = 2$. (b) $I = 5$. (c) $I = 10$.



(d) $I = 20$. (e) $I = 30$. (f) $I = 40$.

Figure 8: Rastrograms of FM sounds from the same equation as in figure 7 . Each sound is generated with $f_c = 220.5$ [Hz] and $f_m = 1$ [Hz], but with different I (as specified). Note that $I = 1$ would be the same as figure 7(a)

(a) Natural woodgrain image.



(b) "Synthetic" woodgrain created from an FM sound with $f_c = 220.5$[Hz], $f_m = 0.7$[Hz], and $I = 0.6$.

Figure 9: Woodgrain images.

(a) Original. (b) Filtered by $y[n] = \frac{1}{10}\sum_{k=0}^{9} x[n-k]$. (c) With noise.

Figure 10: Rastrograms of sounds modified by different audio processes.

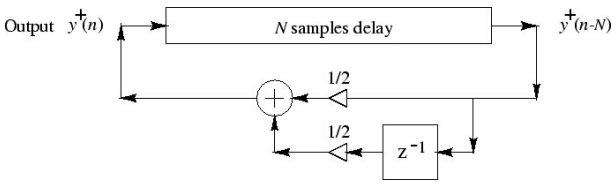


Figure 11: System diagram of Karplus-Strong algorithm, with a simple loop filter.

4.3 Effects of Audio Filters

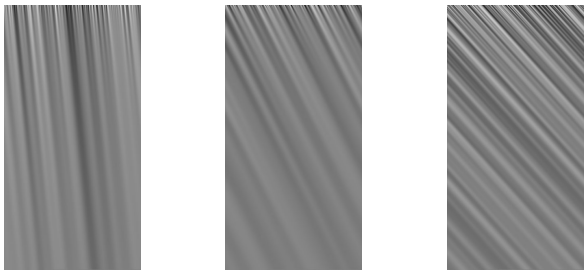
We suggest the idea of "visualizing the effect of audio filters" as a dual to its counterpart in the other domain. In figure 10, sound generated from 10(a) is lowpass-filtered and visualized back to create the rastrogram in 10(b), while 10(c) is converted from the same sound of 10(a) with added white noise.

Naturally, this could develop further into the idea of "image manipulation with audio filters, which will be discussed in §7.

5 Visualization of Karplus-Strong Model

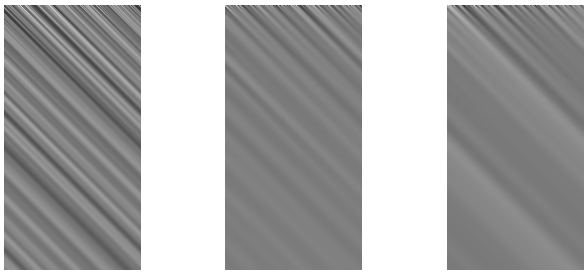
As mentioned in 4.1, rastrogram shows short segments of audio samples in order of time. For digital waveguide sound synthesis models, it becomes a visual record of delay line status at every cycle and depicts particular characteristics of the system. This, therefore, becomes an ideal method to visualize sounds generated from Karplus-Strong algorithm (Karplus and Strong 1983) (Jaffe and Smith 1983), whose diagram with a simple (or, possibly the simplest) filter is shown in figure 11.

Figure 12 illustrates rastrograms generated from a Karplus-Strong model using different filters in the feedback path. In this comparison, inclined lines in each rastrogram receive primary attention. Inclination in the rastrogram of a waveguide represents an additional delay whose length can be determined by degree, thereby showing instantaneous pitch change, as discussed in §4.1. The loop filters used here can be considered as *fractional delays* using linear interpolation: delay



(a) $0.9 + 0.1z^{-1}$. (b) $0.5 + 0.5z^{-1}$. (c) $0.1 + 0.9z^{-1}$.

Figure 12: Rastrograms of plucked-string sounds generated from Karplus-Strong algorithm, with different loop filter $H(z)$s as individually specified. Delay line length N is 337, which corresponds to about 130.9 [Hz].



(a) $0.05 + 0.9z^{-1} + 0.05z^{-2}$. (b) $\frac{1}{3} + \frac{1}{3}z^{-1} + \frac{1}{3}z^{-2}$. (c) $0.45 + 0.1z^{-1} + 0.45z^{-2}$.

Figure 13: Rastrograms of the same Karplus-Strong model as figure 4.1, with different second-order loop filter $H(z)$s that introduce one sample delay.

sizes of the filters used in 12(a), 12(b), and 12(c) are 0.1, 0.5, and 0.9, respectively. From inspection of figure 12, we see that each of these coincides the inclination of corresponding rastrogram expressed as $\Delta x/\Delta y$.

In Figure 13, Karplus-Strong rastrograms with second-order loop filters are depicted. Each of these can be considered as a second-order interpolation that is equivalent to one sample delay, thereby having the same degree of inclination. Differences between their amplitude response characteristics, however, are clearly visualized: obviously, the result of 13(a) is not much affected, while 13(c) is the most lowpass-filtered.

From the results presented in this section, we can see that rastrogram produces an effective visualization of particular features of the loop filter used in Karplus-Strong algorithm, including the length of interpolated delay. Although rastrogram does not provide any precise measurement of filter parameters by itself, it has a strong potential as a visual interface of a filter analysis/design tool.

6 Image Processing Techniques for Sound Synthesis

When used together, raster sonfication and visualization make it possible to modify/synthesize a sound using various image processing techniques.

6.1 Basic Editing and Filtering

From the aforementioned geometric properties of raster mapping, it is obvious that *time scale modification* and *frequency shifting* of a sound can be performed by resizing its rastrogram. More precise control of frequency would be possible by parallelogram-shaped *skew* transforms to consider the roundoff drift. Other geometric modifications, such as rotation, can change the timbre and pitch: as a special case, rotation by 180° produces a time-reversed sound.

In addition to these processes, visual filters can be applied not only to replace corresponding audio filters but also to create unique variations of the original sound, as suggested by examples in §3.3.

6.2 Texture Analysis and Synthesis

Numerous algorithms for analysis and synthesis of visual texture have been developed in the filed of computer graphics, as summarized in (Wei 1999). We believe that these techniques can be used for analyzing the rastrogram of a sound to create a new synthesized one, which then can be raster-sonified to produce a *visually synthesized* version of the original sound. Advantages of using texture synthesis techniques for sound include the followings:

- While preserving the quality of the original image, synthesized textures can be made of any size, providing full control over the pitch and duration of raster sonified sound.
- Texture synthesis can also produce *tileable* images by properly handling the boundary conditions: this enables us to eliminate any unwanted noise components introduced by abrupt jumps at the edge.
- Potential applications of image texture synthesis include de-noising, occlusion fill-in, and compression. Therefore, techniques for these could be applied to similar problems in audio domain.

Future research on this topic will be focused on constructing a set of *visual eigenfunctions* which can span various rastrograms with "auditory significance". We believe this will provide a simpler and more effective algorithm for analysis and resynthesis of complex tones.

6.3 Hybrid Synthesis Model

A carefully designed image analysis algorithm might be able to divide a rastrogram into multiple sections depending on the existence of particular visual patterns. This, with a set of proper mappings between image patterns and sound synthesis algorithms, leads to the idea of *hybrid sound synthesis algorithm*: a sound could be created as a series of multiple segments, each of which is generated from a specific synthesis method chosen by the visual pattern of equivalent image section.

7 Chained Audio/Visual Conversions

In §4.3, we have seen the use of raster sonification and visualization for converting the effect of audio filters. This, when combined with the ideas in §6, constitutes a framework of chained conversions between audio and visual domains: audio and visual processes can be connected to each other through raster sonification and visualization, thereby forming a long chain of cross modal conversions.

It should be noted that the reversible, one-to-one nature of raster mappings makes these conversions more than a random transformation: data converted into the other domain still contains its original "meaning", not only artistically but also mathematically.

Also, to fully understand how filters in one domain affect signals in the other, relationship between one and two dimensional signal processing techniques is to be investigated. The *helical coordinate system* (Claerbout 1998) would serve as a framework which enables us to analyze two-dimensional data with methods for one-dimensional space.

8 Online Examples

Examples presented in this paper, together with sound files, are available at (Yeo 2006).

9 Conclusion

We have here proposed raster scanning method as a new mapping framework for image sonification and sound visualization. Raster sonification proves to be a powerful method for creating a sound which contains the "feeling" of the original image, and becomes the elementary background for sonifying the effects of visual filters. Rastrogram, on the other hand, has a strong potential as a visual interface to audio: in addition to being a space-efficient audio data display, it can intuitively visualize the timbre and some fundamental auditory properties of a sound, and filter characteristics as well. Together, both sides of raster mappings form a complete circle of conversion between audio and visual data, thereby making it possible to utilize image processing methods for sound analysis and synthesis.

Future works will also include artistic applications of raster mappings. In addition to the simple idea of using images as sound libraries, sonification of a painting will be proposed as an auditory clue for its visual patterns. Also, the chain of audio-visual conversion will be developed into a new framework of collaborative art.

References

Bischoff, J., R. Gold, and J. Horton (1978). Music for an interactive network of microcomputers. *Computer Music Journal 2*(3), 24–29.

Borgonovo, A. and G. Haus (1984). Musical sound synthesis by means of two-variable functions: Experimental criteria and results. In *Proceedings of the International Computer Music Conference*, pp. 35–42. ICMA.

Chafe, C. (1995, September). Adding vortex noise to wind instrument physical models. In *Proceedings of the International Computer Music Conference*, pp. 57–60. ICMA.

Claerbout, J. (1998). Multidimensional recursive filters via a helix. *Geophysics 63*, 1532–1541.

IRCAM. Audiosculpt. *http://forumnet.ircam.fr/*.

Jaffe, D. A. and J. O. Smith (1983). Extensions of the karplus-strong plucked string algorithm. *Computer Music Journal 7*(2), 56–69.

Karplus, K. and A. Strong (1983). Digital synthesis of plucked string and drum timbres. *Computer Music Journal 7*(2), 43–45.

Kock, W. (1971). *Seeing Sound*. New York: Wiley-Interscience.

McLaren, N. and W. Jordan (1953, Spring). Notes on animated sound. *The Quarterly of Film, Radio, and Television 7*(3), 223–229.

U&I Software. Metasynth 4. *http://uisoftware.com/MetaSynth/*.

Verplank, B., M. Mathews, and R. Shaw (2000, September). Scanned synthesis. In *Proceedings of the International Computer Music Conference*. ICMA.

Wei, L. (1999). Deterministic texture analysis and synthesis using tree structure vector quantization. In *Proceedings of The SIGGRAPH*. ACM.

Whitney, J. (1980). *Digital Harmony : on The Complimentarity of Music and Visual Art*. Peterborough, NH: McGraw Hill.

Yeo, W. S. (2001). Image sonification: Image to sound. *http://www.mat.ucsb.edu/∼woony/research/winter01/mat310/index.html*.

Yeo, W. S. (2006). Raster scanning. *http://ccrma.stanford.edu/∼woony/works/raster/*.

Yeo, W. S. and J. Berger (2005, September). Application of image sonification methods to music. In *Proceedings of The Internaional Computer Music Conference*. ICMA.