Digitally Extending the Optical Soundtrack

Alexander Dupuis
Brown University
Alexander.P.Dupuis@gmail.com

Carlos Dominguez
Dartmouth College
cdominguez87@gmail.com

ABSTRACT

The optical soundtrack has a long history in experimental film as a means of image sonification. The technique translates image luminance into amplitude along the vertical axis, enabling the sonification of a wide variety of filmed patterns. While the technical challenges of working with film preclude casual exploration of the technique, digital implementation of optical image sonification allows interested users with skill sets outside of film to access this process as a means of sonifying video input.

This paper presents an overview of the workings of optical image sonification, as well as a basic implementation in the Max/MSP/Jitter environment. The benefits and drawbacks of the technique in its digital and analog forms are discussed, as well as the greatly improved control over the sonification process afforded by digital image processing. An example of these expanded possibilities in the context of audiovisual composition is presented, and supplementary code is given to provide a basis for users to test and apply the technique.

1. INTRODUCTION

Numerous artists working in experimental music and film share a common interest in the study and production of integrated audiovisual content. Though the tools and approaches used between the two fields often differ, this diversity provides a wealth of perspectives and processes that can be employed in the creation of co-related sound and light.

One such process found in film is the repurposing of the optical soundtrack as a means of image sonification. Designed to provide a simple means of storing synchronized images and sounds on the same film strip, the optical soundtrack has also proven to be a useful method for turning certain spatial periodicities into pitches and rhythms, creating a distinctive sound which complements the visual features in carefully selected images and patterns. While the nature of the process has largely limited its appeal to those skilled at working with film, programs for digital sound and image synthesis like Max/MSP/Jitter provide an accessible means of adapting the technique for an entirely new set of users. Implementing the process in a digital context also greatly improves the flexibility of the technique, allowing it to be applied more specifically to visual features of the artist's choosing.

This paper presents a brief background on optical sound and its strengths and weaknesses as an image sonification method. A basic digital implementation of the process and its differences from the analog equivalent are described, followed by examples of some of the enhanced sonification possibilities afforded by the digital version. These examples, combined with the collection of documentation patches available at alexanderdupuis.com/code/opticalsound, provide an initial demonstration of the potential uses of this sonification method as a means of generating audiovisual material for a variety of sources and contexts.

2. BACKGROUND

2.1 The 16mm optical soundtrack

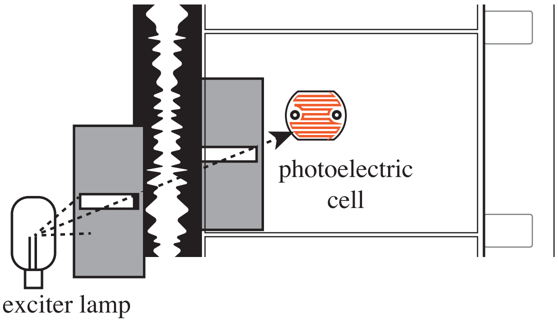


Figure 1. Diagram of the optical sound head

The invention of the optical soundtrack in 1929 provided filmmakers with an elegant solution for synchronizing recorded audio with its accompanying film [1, 2]. While several alternatives existed such as the magnetic track, optical sound was the dominant film sound format until the development of digital sound-on-film in the 1990s.

The variety of film format sizes and their different modes of dissemination led to a number of differing optical soundtracks, with some larger formats carrying stereo or even four-channel tracks. For the purposes of this project, we will be using the 16mm optical soundtrack as our model. The majority of the optical sound experiments pertinent to this investigation were created using 16mm, and as such it offers the best opportunities to test the effectiveness of the digital processes.

Optical soundtracks are printed onto film rolls as fluctuating patterns of light and dark, occupying a narrow strip next to the images. In the case of 16mm film, the small size requires the use of single-sprocket film, with the optical soundtrack taking the place of the second set of sprocket holes. The projector sonifies this soundtrack by means of an optical sound head, shown in Figure 1. An exciter lamp shines through the film onto a photocell, filtered by narrow horizontal slits on either side of the film. As the film passes across this thin band of light, it produces a fluctuating voltage which is processed and output as the audio signal [3]. Due to the need for continuous film speed when producing sound, as opposed to the stopping and starting required when projecting images, the optical sound pickup in a 16mm projector is placed 26 frames ahead of the lens. Thus, assuming a playback rate of 24 frames per second, the audio on any point of an optical soundtrack will be heard a little over a second before its adjacent image is seen.

The use of horizontal slits and a single photocell within the optical pickup means that a soundtrack can be represented on film in a variety of ways, provided that the average lightness along the horizontal axis at any point in time is equivalent. This flexibility has given rise to a number of optical soundtrack formats and applications, several of which are shown in Figure 2. The conventional kinds are variable area (2a) and variable density (2b) soundtracks, with the more common variable area representing its information by fluctuating the width of a white waveform on a black background, and variable density translating its values to differing shades of gray. The other two examples show possible applications of the optical soundtrack for image sonification: the first (2c) shows how whole frames might be sonified, though they must be horizontally scaled to fit into the smaller area and offset by 26 frames if they are to be in sync with original images. The second example (2d) makes use of a 16mm widescreen format known as Super16, which extends the image into the area occupied by the optical soundtrack. While this approach allows for the sonification of only a small part of the image and produces a 26 frame offset between each frame and its sonified output, several experimental films, such as Roger Beebe's *TB TX DANCE*, have exploited these idiosyncrasies to great effect.

2.2 Historical experiments in optical sound

The visual depiction of sound on the physical medium of film opened up a variety of new sound editing and synthesis possibilities. Many of the earliest experiments with optical sound revolved around the manipulation of recorded sounds using new editing techniques afforded by the medium, which had already been developed for the creation of motion pictures. Sounds could now be easily studied and modified in a number of ways such as cutting, splicing, and overlaying, all of which would be used years later by pioneering electronic musicians working with tape [1].

Animators quickly realized the potential of the optical soundtrack as a means of applying their skills to the creation of novel sounds. Early animated sound experiments in the 1930s included the research at Leningrad's Scientific Experimental Film Institute, as well as Oskar Fischinger's work documenting audiovisual links between the aural and visual aspects of optical sound [1]. Filmmakers found that by varying the positioning, shape, and exposure of sequenced abstract patterns, they could predictably control the pitches and amplitudes produced as well as effect changes in the resulting timbres [2]. By the 1970s, Scottish-Canadian animator Norman McLaren elevated the practice of animating sound to new technical and artistic heights, developing a set of templates for rapid production of several waveforms at different pitches and a variety of masks which functioned as envelopes [4]. McLaren's 1971 piece *Synchromy* highlights the cross-modal nature of the process, juxtaposing the optical patterns with their sonic results to form a psychedelic audiovisual spectacle.

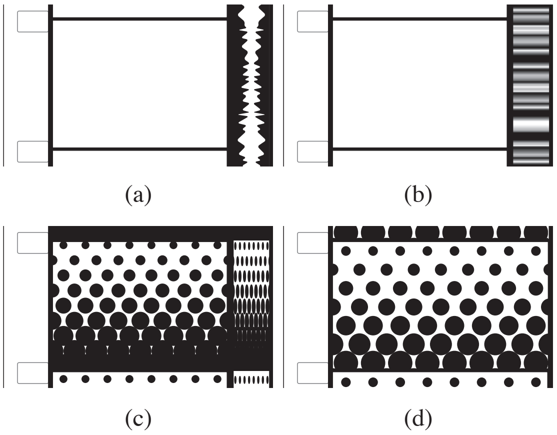


Figure 2. Examples of optical soundtracks: (a) variable area, (b) variable density, (c) soundtrack made from camera images, (d) Super16 images extending onto soundtrack area

While *Synchromy* hints at the transmodal possibilities of film and optical sound, filmmakers such as Guy Sherwin and Lis Rhodes pushed the process to its limits by using the same source material to create the image and sound. Their works demonstrated and exploited the fact that anything put on film could be sonified if placed on the optical soundtrack, from the gritty images in Sherwin's *Musical Stairs* to the morphing abstract animations in Rhodes' *Dresden Dynamo*. Their work also reveals the limits of the process: all images can be sonified, but not all information contained in an image is communicated equally. Rhodes' piece *Dresden Dynamo* from 1971 is a particularly powerful exposition of the possibilities and limits of this technique, with her morphing abstract patterns allowing us to see gradual changes in timbre, pitch, and amplitude. As the patterns evolve, we also encounter the boundaries of the sonification process: the same pattern that produces a steady pitch at one angle fades to nothingness as it rotates, only to gradually reemerge as it comes back into alignment.

2.3 Relationship to Electronic Music

Many sound-on-film experiments paralleled and in some cases predated similar technical developments in electronic music. The early film sound montages naturally evoke comparisons to the approaches later found in tape music and *musique concrète* [1], while the animated sound exemplified by McLaren aptly fits within the broader field of graphical sound synthesis [5]. Optical sound systems were also directly utilized in the creation of sound synthesis instruments and systems, including the ANS synthesizer [6], Daphne Oram's Oramics [7], and the Whitney brothers' system of pendulums used in the creation of their *Five Film Exercises* [8].

The specific audiovisual transduction that occurs in optical soundtrack image sonification lacks a direct parallel outside of film, though some approaches bear similarities to the process and its goals. Yeo and Berger's raster scanning technique, inspired by analog television image rendering, is particularly pertinent in its translation of spatial patterns to audio waveforms [9]. Raster scanning renders every pixel of an image as a single audio sample by traversing each pixel row, preserving all the information in a way which can be decoded back into the original waveform.

In a video context, however, the amount of data preserved by raster scanning becomes a possible difficulty. When working with video, the number of pixels in each frame multiplied by the framerate can be hundreds or thousands of times larger than the audio sampling rate, requiring some means of reducing the information [10]. Optical sound achieves this by discarding a great deal of the image's information through its averaging of luminance values, limiting the number of samples per second to the matrix height multiplied by the framerate. This technique is therefore more closely described as *scanned synthesis*, another method defined by Yeo and Berger [11]. Drawing on their terminology, the optical sound head essentially acts as a pointer which reads one horizontal line at a time, scanning along a vertical path.

3. MOTIVATIONS

The sonification limitations of analog optical sound have served, in many ways, to inspire filmmakers working with the technique. By eliminating almost all control over the transduction process itself, filmmakers are forced to explore the far reaches of these constraints, both through the careful selection and shooting of subject matter as well as physical manipulation of the film and projection. Indeed, subversions of the technique often emphatically remind us of the physical nature of film rather than transcending it. Guy Sherwin's *Railings*, for instance, manages to sonify the horizontal luminance in the image rather than the vertical, but only by rotating the projector itself by ninety degrees [3].

Freeing the optical sonification process from these constraints opens up the technique to an entirely new group of users and applications. Digital implementation makes the basic technique available to artists and composers lacking the requisite knowledge of film shooting and manipulation necessary to produce these pieces. It also bypasses the trial and error of the production and development stages that can delay work, effectively giving instant feedback on the sounds that will be produced for a given image as well as providing a means of real-time sonification. Finally, the flexibility of digital video processing allows the technique to be applied for much more precise sonification of features, allowing artists to more effectively tailor the method to their chosen visual input.

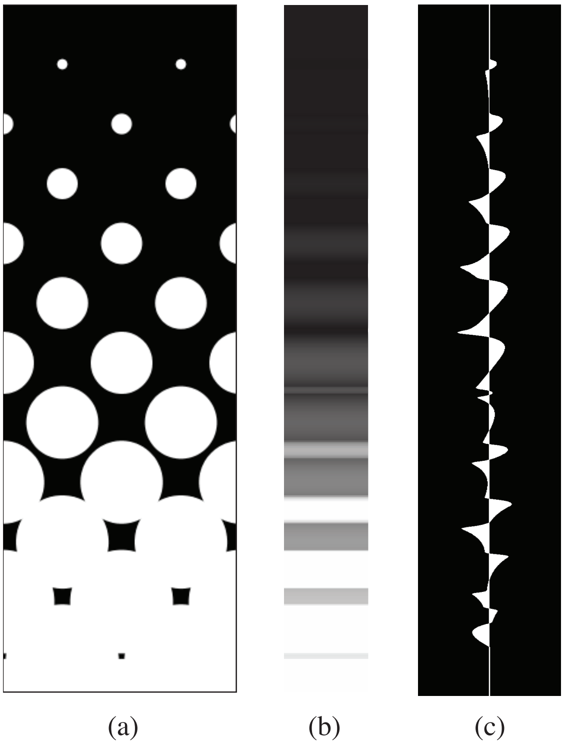


Figure 3. Depiction of the optical sonification process: the lightness values of the source frame (a) are averaged to return a column of values (b), which are then filtered and multiplied by a scalar to produce the waveform (c)

4. IMPLEMENTATION

All software was created in Max/MSP/Jitter due to its ease of interchangeability between audio and video data, as well as its modularity and popularity amongst audiovisual artists. The sonification process takes a single-plane matrix of arbitrary size as its visual input, and stores the resulting audio sample sequence as single-row matrix for playback and sound manipulation. A number of image processing techniques and their effect on the sonification process are discussed, as well as the expanded playback options afforded by this approach.

4.1 A basic digital optical sound head

The core of the processing is the digital equivalent to the 16mm optical sound head. This implementation is not an attempt to model the specific output of any particular analog optical sound circuitry, which would require introducing numerous distortions to the signal representing noise from the film, the projector, and the optical sound head itself [12]. Rather, the process mimics an ideal optical sound head reading a film with discrete pixel values.

Figure 3 depicts the steps of deriving a waveform from an input grayscale matrix. Each row of the incoming matrix is averaged using matrix multiplication, returning a single column of values. The column is transposed to become a row, and resampled with interpolation to the desired playback length. The resulting matrix, representing the waveform for the frame, can then be high-pass filtered to remove DC offset, along with any other desired filtering. This filtering can alternately be done during audio playback, especially if the playback rate is manipulated in real-time. The processed waveforms are stored sequentially in a larger matrix, available for retrieval during playback (particularly useful for replicating the 26 frame offset in film) and for saving to an audio file when all the frames have been rendered. While an image's waveform would ordinarily be read through from left to right over the frame's duration, the waveform can also be manipulated using buffer playback techniques, providing greater control over the resulting pitches and rhythms, or even be used for more complex audiovisual synthesis as in Shawn Greenlee's *graphic waveshaping* technique [13].

4.2 Digital expansion

The translation from grayscale input matrix to audio waveform is largely similar to the original analog process. The decision to preserve the major sonification aspects produces limitations similar to those affecting filmmakers, with the resulting sound highly dependent on the input frame's spatial periodicities and their orientation. Unlike the analog process, however, the digital version affords a vast increase in the number and flexibility of image manipulations prior to passing the optical sound head. Artists can now focus on any of a number of desired spatial features, through manual manipulation as well as analysis and processing.

The simplest operation on a three-plane RGB image would be to convert it to a single-plane matrix of luminance values and send the result directly to the virtual optical sound head, giving us the result most similar to that of a projector and its film. Processing the image before it reaches the virtual optical sound head gives a variety of results that can vary considerably depending on the source material, though some generalizations can be made. Zooming in or out along the y-axis will change the pitch of the output sound, while zooming along the x-axis will alter the timbre. Rotating the image changes the axis along which spatial periodicities are sonified, boosting some while diminishing others. These processes and others, such as color keying, convolution, and cropping, provide a means to greatly improve (or destroy) the cleanliness of the visual signal before it is heard, and allow for innumerable opportunities to explore and establish audiovisual relationships.

5. TESTING: *MUSICAL STAIRS*

Although faithful modeling of the idiosyncrasies of the analog optical sound process was not a goal of this work, a brief comparison to verify the basic similarities between the analog and digital approaches seemed prudent. A Max/MSP patch implementing the aforementioned virtual optical sound head was used to re-render the optical soundtrack to Guy Sherwin's *Musical Stairs*, working from a digitized copy of the film. This piece was chosen in part due to the good quality of the image and the soundtrack on the capture, as well as Sherwin's documentation verifying the methods used to produce the piece [3].

The patch loads the digitized version of the film, which has a resolution of 720 by 576 pixels and a framerate of 25 frames per second. The original film almost certainly runs at 24 fps, but was sped up four percent to conform to the PAL standard. The soundtrack is extracted to a separate buffer to provide access to the actual samples. Frames are read one at a time, with the image of the frame displayed next to its corresponding waveform, and the image is processed using the virtual optical sound head. The rendered sound is displayed next to the original, and both waveforms can be synchronously looped for comparison.

Despite a host of distorting factors including the relatively low input resolution, the film capturing process, and the lack of any modeling of the frequency response of the optical sound head, the synthesized and analog waveforms usually exhibit highly similar morphologies, as seen in Figure 4. Minor additions to the process improved the results even further, beginning with the addition of zero padding to the height of the input video. The capture process of the video seems to have eliminated some of the vertical information of the film, perhaps by cropping, and while this information cannot be reintroduced, padding the height of the video and windowing the edge of the frame preserves the original frequencies in the rendered waveform. The synthesized waveform was resampled to a length of 1920 samples (the length of one frame at a framerate of 25 fps and a sample rate of 48000 Hz), interpolating to minimize the artifacts of the low image resolution. Finally, an temporal offset variable for the analog waveform was introduced to allow for manual correction to the film's drift.

After making these adjustments, a synthesized audio file for the full piece was rendered. A short-time Fourier transform (STFT) was taken for the analog and synthesized versions to compare the frequency content of each version over time. As seen in the excerpt in Figure 5, the two versions exhibit similarly timed attacks, as well as an evolution from lower-level noise to the sustained lower frequencies beginning at roughly ten seconds. Significant differences can also be found: the original recording is considerably noisier up until roughly 10,000 Hz, followed by a sharp reduction in amplitude. This behavior is not unexpected, given the noisy nature of the 16mm optical sound process, as well as the frequency-range limits of film and typical equalization to roll off high frequencies [14]. Other observable discrepancies include the tendency of the digital rendering to produce sustained tones, and the analog process' lack of sensitivity to changes in bright values.

These issues will no doubt be of great concern in the production of a faithful optical sound emulator, and some adjustments, such as a more accurate equalization curve, should be simple to implement. For the purposes of this project, however, the comparison between the analog and synthesized outputs reveals that the digital application of optical sound principles produces a sufficiently similar result to be considered the same process, even under adverse

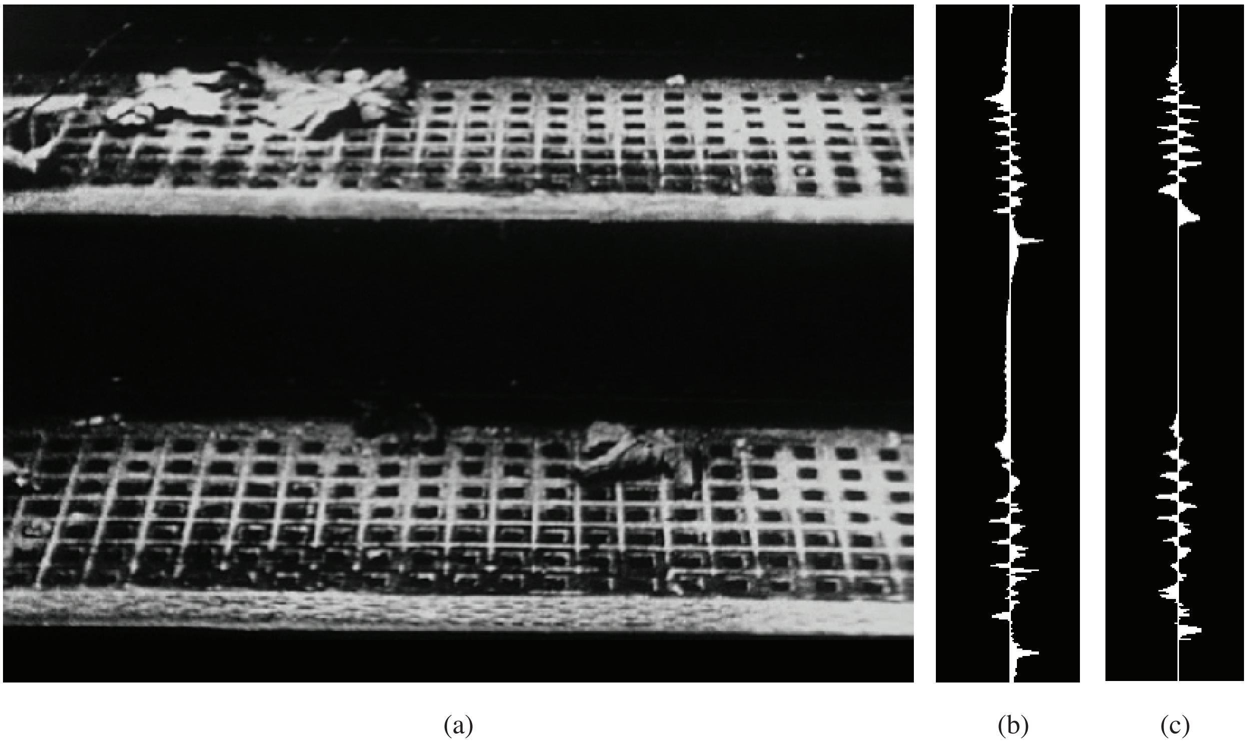


Figure 4. Original (b) and synthesized (c) waveforms corresponding to frame 1117 (a) in *Musical Stairs*

rendering conditions. The digital process retains much of the characteristic sound of the analog version, with the noise and drift of film playback available as possible aesthetic additions rather than practical necessities.

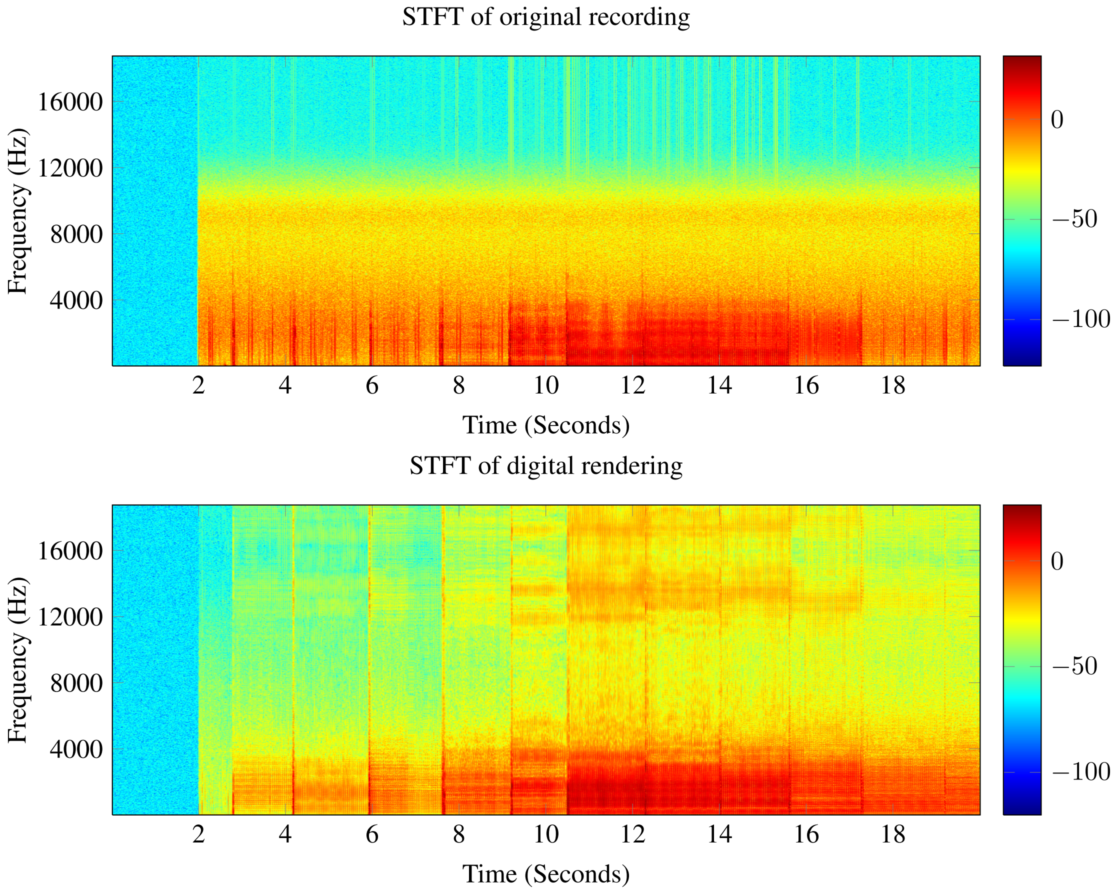
6. APPLICATION: *NO-INPUT PIXELS*

This virtual optical sound processing was used extensively in the creation of the author's piece *No-Input Pixels*, providing an example of some of the extended possibilities of the digital version [15]. The video source material for the piece was generated using a digital video synthesis feedback system, rendering evolving kaleidoscopic patterns with spatial qualities almost ideally suited for optical sonification. The regular spacing between elements creates areas of geometric periodicity, a feature which optical sonification can translate into pitched sounds. The visual structures are also consistently oriented vertically or horizontally, avoiding rotations which could obscure the rendering of the waveform.

However, the video possesses other qualities which would dilute or nullify the potential sonified effects in an analog environment. The regularly spaced patterns are composed from a palette of eight colors, with overlapping patterns of multiple colors sometimes occupying the frame at the same time. Flattening an entire frame down to its luminosity values would significantly obscure the differences between some colors and lead to a noisier result. There would also be no way to tailor the sonified result of each color to reflect its unique role and impact in the piece at different points in time. Furthermore, although the local spacing between the elements of each color is apt for sonification, the kaleidoscopic arrangement of elements can serve to blur horizontal and vertical periodicities, nullifying both.

While these issues would be extraordinarily difficult if not impossible to solve in an analog environment, digital image processing allows for the isolation of relevant information before the sonification takes place. A Max/MSP workflow, shown in Figure 6, was designed to produce individually controllable sonifications for each color based on specified areas of the source. The input frame (6a) is filtered into eight black and white maps (6b) corresponding to the eight colors of the source. The maps are then analyzed using horizontal and vertical edge detection (6c) to find the busiest lines along the x and y axes: that is, the lines containing the greatest number of changes from light to dark. The selected rows and columns (6e) are output from the corresponding color masks (6d) and sonified using the virtual optical sound head, giving sixteen unique audio waveforms per frame. The amplitude and playback pitch of each channel is then altered using feature data extracted from its corresponding each color map, including the rate of change between successive frames and the overall overall quantity of the color.

The ability to isolate relevant information and areas within each frame enabled a far more specific transduction of geometry to audio within the piece. Without the digital ability to filter, analyze, and crop the input image, many of the periodicities and visual events which are readily apparent to a viewer would have been muted or absent. Other source materials would, naturally, invite different approaches to tailor the sonifications to the artist's choosing, and this video itself might have been subjected to any number of alternative processes to establish compelling audiovisual relationships.


STFT of original recording

STFT of digital rendering

Figure 5. Comparison of STFT plots on a dB scale between the original and re-rendered versions of *Musical Stairs*, using an FFT size of 4096 and a hop size of 1024

7. CONCLUSIONS AND FUTURE WORK

The digital implementation of optical image sonification vastly expands the capabilities of the original technique. The principles of the analog processing have been preserved, but can now be applied more easily and specifically to a greater range of source materials. Max/MSP examples demonstrating the technique and its applications can be found at alexanderdupuis.com/code/opticalsound, providing a basis from which interested parties can study and adapt the process.

While several specific applications of the technique have been detailed here, different source materials and pieces will call for their own unique processing. However, the core of the sonification, the virtual optical sound head, is used in any implementation of this process, and improvements to this area would potentially benefit future applications. Future work will therefore concentrate on improving the quality and accuracy of the sonification process, especially a version which more accurately models the idiosyncrasies of the analog sound and can provide filmmakers with a means of testing experimental optical soundtracks before committing them to film.

8. REFERENCES

[1] R. S. James, “Avant-garde sound-on-film techniques and their relationship to electro-acoustic music,” *The Musical Quarterly*, vol. 72, no. 1, pp. 74–89, Jan. 1986.

[2] A. Smirnov, *Sound in Z: Experiments in Sound and Electronic Music in Early 20th Century Russia*. London: Koenig Books, 2013.

[3] G. K. Sherwin and S. Hegarty, *Optical sound films 1971-2007*, B. Cook, Ed. London: LUX, 2007.

[4] R. Russett and C. Starr, *Experimental animation: an illustrated anthology*. Van Nostran Reinhold Co., 1976.

[5] C. Roads, *The Computer Music Tutorial*. MIT Press, Jan. 1996.

[6] S. Kreichi, “The ANS synthesizer: Composing on a photoelectronic instrument,” *Leonardo*, vol. 28, no. 1, pp. 59–62, Jan. 1995.

[7] P. Manning, “The oramics machine: From vision to reality,” *Organised Sound*, vol. 17, no. 02, pp. 137–147, 2012.

[8] J. Whitney, *Digital harmony: on the complementarity of music and visual art*. Peterborough, USA: Byte Books, 1990.

[9] W. S. Yeo and J. Berger, "Raster scanning: A new approach to image sonification, sound visualization, sound analysis and synthesis," in *Proceedings of the International Computer Music Conference*. ICMA, 2006.

[10] J.-M. Pelletier, "La sonification de séquences d'images à des fins musicales," in *Actes des JIM'09*, Grenoble, France, Apr. 2009.

[11] W. S. Yeo and J. Berger, "Application of image sonification methods to music," in *Proceedings of the International Computer Music Conference*. Barcelona, Spain: ICMA, 2005.

[12] E. W. Kellogg, "History of sound motion pictures," *SMPTE Motion Imaging Journal*, vol. 64, no. 8, pp. 422–437, Aug. 1955.

[13] S. Greenlee, "Graphic waveshaping," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Daejeon, Korea, 2013.

[14] I. Allen, "The x-curve: Its origins and history electro-acoustic characteristics in the cinema and the mix-room, the large room and the small," *SMPTE Motion Imaging Journal*, vol. 115, no. 7-8, pp. 264–275, Jul. 2006.

[15] A. Dupuis, "No-Input Pixels," 2013.

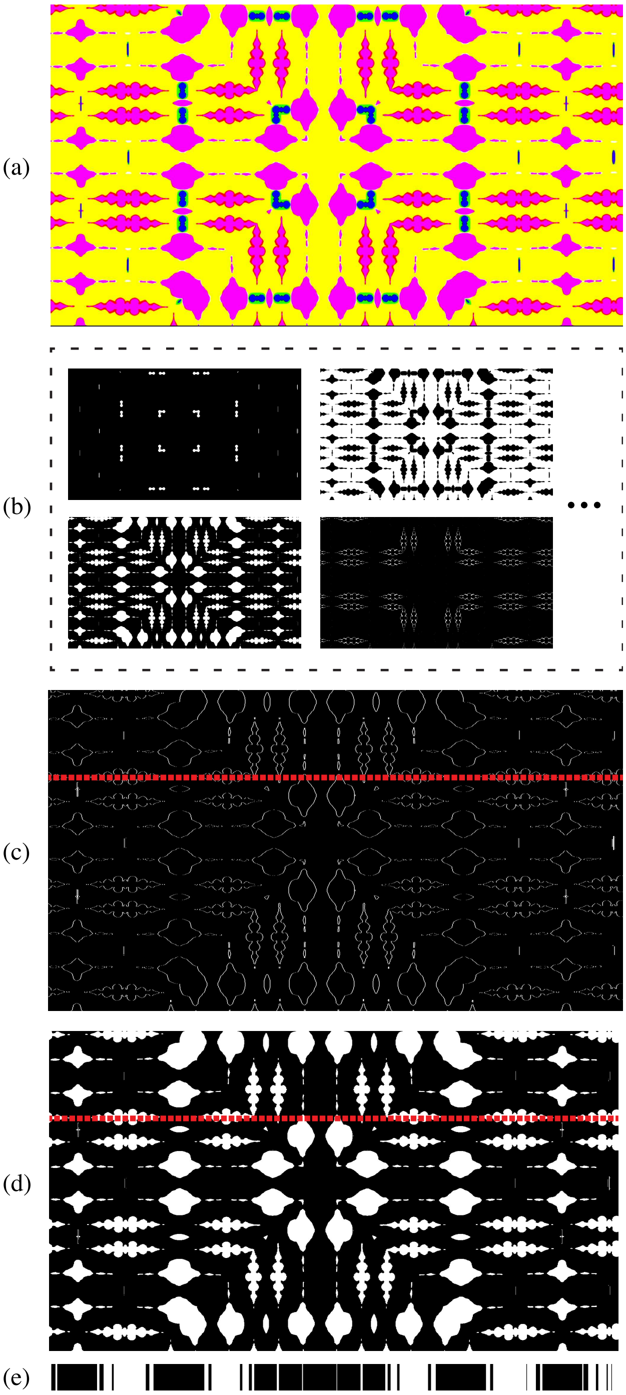


Figure 6. Image processing stages in *No-Input Pixels*. The frame (a) is filtered into eight color maps (b). Each map is analyzed using horizontal and vertical edge detection (c), and the busiest row and column (represented by the red line) are selected from the original map (d). The selected rows and columns are output (e) and sent to the virtual optical sound head.