Sound Synthesis from Video, Wearable Lights, and "The Watercourse Way"

Roger B. Dannenberg[1], Barbara Bernstein[2], Garth Zeglin[1], and Tom Neuendorffer[3]
roger.dannenberg@cs.cmu.edu, bbernste@ringling.edu, garthz@cs.cmu.edu, tneuendorffer@yahoo.com
http://www.cs.cmu.edu/~rbd

[1]School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
phone: 412-268-3827
fax: 412-268-5576

[2]Department of Fine Arts
Ringling School of Art and Design
2700 Tamiami Trail
Sarasota, FL 34234

[3]130 Pineview Drive
Wexford, PA 15090

Abstract. "The Watercourse Way" is a mostly-music interactive multimedia performance for violin, cello, percussion, and dancer. The work uses a computer to process sounds from the performers, to synthesize sound, and to generate computer animation. A novel synthesis technique is introduced in which the sound spectrum is controlled in real time by images of light reflected from a shallow pool of water. In addition, performers wear computer-controlled lights that respond to video and sound input, using a wireless radio link to the computer. This work explores connections between the senses using technology to both sense and generate images and sounds.

Introduction

We are interested in the creation of art that connects multiple media and multiple senses. All of us have some connection to both music and the visual arts and want to express ideas both visually and aurally. We are also drawn to technology and particularly to computers because computation itself is seen as a medium for artistic expression. "The Watercourse Way" is primarily a music composition, but it integrates dance, sculpture, instrument making, and lighting. The work also features interactive electronics to create a multidimensional world of interconnected sound and light.

"Water Course Way" is a literal translation of the words "Tao Te Ching", Lao Tzu's 6th century B.C.E. treatise describing a particular attitude and perception as a code of life and livelihood. Its essence asks for acceptance, allowance and awareness of things as they are, without struggle, conflict or resistance. In the text, water is the metaphor used to illustrate this freedom and fluidity of thought, word and deed. Subsequent action is based on an immediate but clear, unencumbered response in the present moment, with full awareness of an eternal "now." Interdependency is a key element in the text as effects of awareness ripple and change the immediate environment as well as the distant future, but initially and most importantly, the perception of ourselves. The revealing of truth is both reflective and active, an ongoing process.

This piece, 'The Watercourse Way," invites the removal of rigid boundaries and definitions of sound, light, and visual art while investigating simultaneity, synaesthetics, and the interdependent relationships of time, space, movement and form. Just as water ripples to change and shift the shape of water, we ask, "How can perception itself be changed? Can we see sound? Can light be heard?"

This collaborative piece investigates these questions with the attitude that the space within an empty cup is what makes the cup useful and necessary, that "letting go" and "letting in" are the same and that without "trying" we can begin to be. "The Watercourse Way" uses technology as a 21st Century means to access and dialog with the 6th Century B.C.E. text, collapsing and stretching time with both literal and metaphorical formats.

Description of the Work

As a composition, "The Watercourse Way" focuses on natural but unconventional sounds of instruments created by scraping, bowing, plucking, and tapping. Small sounds are essentially magnified through real-time digital audio processing to create rich sonic textures. It is hoped that these sounds will evoke the feeling of natural settings of

Published as: Roger B. Dannenberg, Barbara Bernstein, Garth Zeglin, and Tom Neuendorffer, "Sound Synthesis from Video, Wearable Lights, and 'The Watercourse Way.'" In *Proceedings: The Ninth Biennieal Symposium on Arts and Technology. New London, CT: Ammerman Center for Arts and Technology, (2003), pp. 38-44.*

water, plants, wind, and soil. Complementing the instrumental performance is the sound of water, especially water drops which are amplified and processed to augment the instrumental textures.

In contrast to the natural sounds, a new synthesis technique has been created for this work, using a camera to sense changing patterns of light caused by reflections from the surface of a pool of water. This process allows a dancer to literally create sound by touching water. The water waves generate rich flowing light images which are immediately transformed into sound.

"The Watercourse Way" is conceived as a visual and theatrical setting as well as a concert performance. The theme of the Tao Te Ching led to the incorporation of a pool of water, and water sounds are incorporated into the piece. Video input allows the dancer to interact with the water setting in a way that can be translated directly to sound, thus the dancer is also a musician, and the moving images created by light reflections serve both as a lighting effect and a musical instrument. It is our hope that the connections between media will create many paths for the mind and imagination of the listener.

On the technological side, the main innovations are the use of images to synthesize sound and lighting systems worn by the dancer and musicians. The synthesis technique is inspired by "terrain synthesis" and "scanned synthesis" in which a terrain or surface is scanned at audio rates, such that variations in the terrain cause audio-rate fluctuation, i.e. sound. After some experimentation, we determined that interpreting the video image as a time-varying spectrum was more interesting than using the image as time-domain data.

The wearable lighting systems used in "The Watercourse Way" have evolved through many variations. A very interesting precursor was a system that used optical fibers driven by high-intensity LEDs to create points of lights at the ends of the fibers, literally suspended in space. The current generation of lights uses a microcomputer to control red, green, and blue LEDs, which are covered by translucent cloth to diffuse the light.

A real-time computer system coordinates the performance and allows translation between media. The computer captures live video of light patterns reflected from the water and synthesizes sound accordingly. In addition, the computer processes sound from live musicians. Based on video and audio input, the computer generates animation and also outputs MIDI commands to control the LED lighting systems worn by performers. Our goal is to create an atmosphere where media are interconnected, evoking a sense of synaesthesia. The animation is inspired by the Barbara Bernstien's work (see Figure 1.)

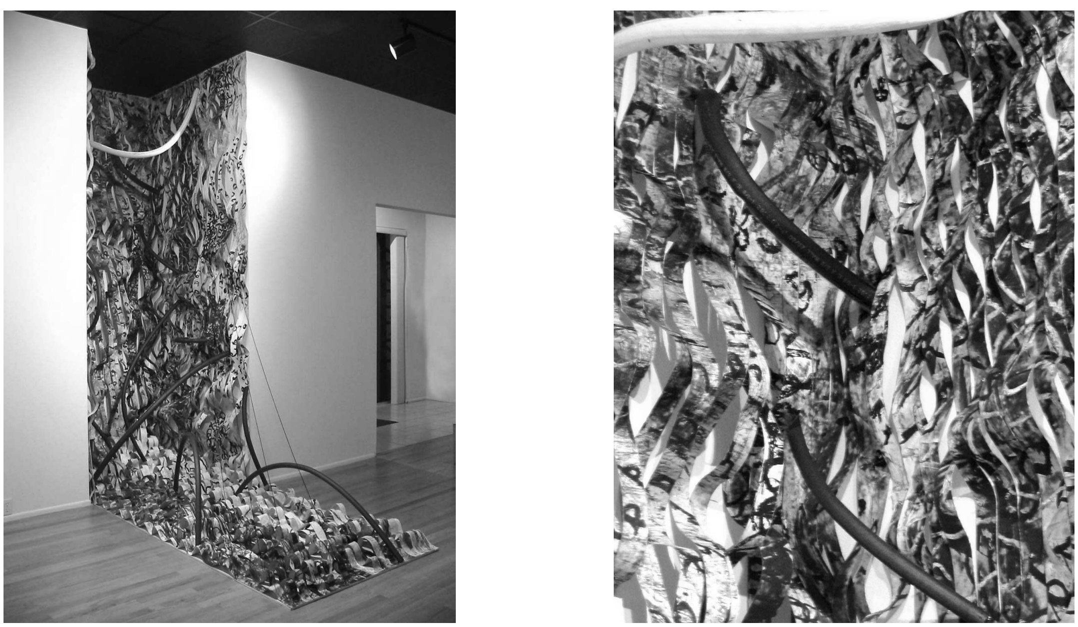

Figure 1. Installation (left) and detail (right) by Barbara Bernstein, Selby Gallery, Sarasota, Florida, 2002.

Sound Synthesis from Video

A central goal of this work, at least for Dannenberg, has been to incorporate live video as an input to an interactive sound synthesis system, creating a dramatic connection between the visual and sonic aspects. This idea evolved into a specific technique in which video images modulate the spectrum of synthesized sound. Our approach is related to terrain synthesis and scanned synthesis.

In terrain synthesis, as described and implemented by Rich Gold (Bischoff, Gold, & Horton, 1978), a real or imaginary terrain is scanned in a closed path, and movement of the path causes variations in the generated waveform. Borgonovo and Haus describe a similar system in which the terrain and path are generated by various mathematical formulas. (Borgonovo & Haus, 1985) Scanned synthesis (Boulanger, Smaragdis, & Ffitch, 2000; Verplank, Mathews, & Shaw, 2000) allows the terrain (sometimes in just one dimension) to vary or vibrate at a very low rate, causing interesting variations in the audio waveform. It seemed a natural extension to use video intensity as a source of "terrain" for synthesis. Imagine moving a photocell in a circle over the video image. If the circular path is completed at audio rates (faster than 20 cycles per second), then a quasi-periodic waveform will be generated by the photocell, giving rise to a perceptible pitch. As the image changes, the generated waveform will also change. Some sort of time-domain interpolation might be used to avoid clicks and pops when a new video image arrives. Of course, we want to simulate this system digitally rather than physically scanning a real image.

To experiment with these ideas, we constructed a small pool of water using a wooden frame lined with plastic. After some experimentation, we determined that interesting images could be obtained by shining a light on the water at a shallow angle of incidence. Reflections from the pool can be observed by placing a screen behind the pool (see Figure 2). This is essentially the configuration we use for the performance.

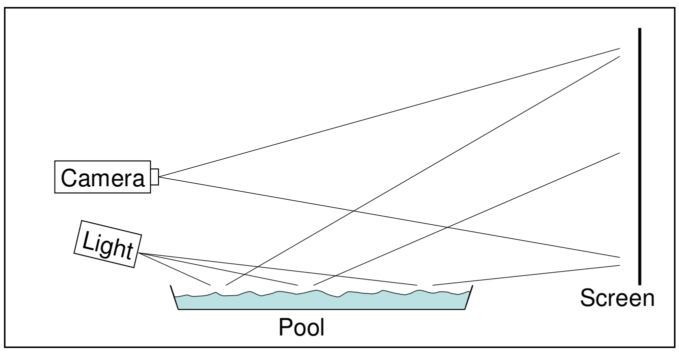

Figure 2. Light reflected from water forms moving images on a screen. These are visible to the audience and captured by a video camera interfaced to a computer.

Using video captured from an experimental setup, we determined that time-domain interpolation was not as interesting as we had hoped. At least with water images as an input, the rather sudden changes in intensity from one location to the next create corresponding jumps in the waveform, giving the sound a fairly "buzzy" sound full of high frequencies and lacking in low harmonics. It was thought that as the water waves dissipated, there would be a corresponding decrease in high-frequency sound energy. In practice, water in our tests did not calm very quickly. Instead, the high frequencies, represented by small ripples, continue to reflect back and forth for many seconds. Even though there was obvious wave motion in the image, the generated sound was relatively static, and the connection to the image was not at all obvious. We decided to start again from first principles.

Captured video shows that the main image movement is in the form of horizontal bands of light moving vertically. (See Figure 3.) The orientation is due to the position of the light. It occurred to us that time variations along the vertical axis might make an interesting time-varying spectrum. We take a vertical strip of the video image and apply smoothing to reduce high-frequency data in the image. The smoothing reduces the video to 32 values, which become spectral data. Although it might be interesting to synthesize using an Inverse Fourier Transform, there are well-known problems associated with discontinuities between successive frames. We decided to use the video data to determine a waveform and then use spectral interpolation synthesis to generate sound.

Spectral interpolation synthesis (Dannenberg, Serra, & Rubine, 1990) was originally developed to model the time-varying spectra of acoustic instruments (Dannenberg & Derenyi, 1998). It works by interpolating between two waveform tables that store single periods of the waveform. Interpolation results in a smoothly varying spectrum at a low computation cost. In practice, the interpolation factor changes linearly from zero (all table 1) to one (all table 2) over a fraction of a second. When the factor is one, table 1 is replaced by a new target table, and the interpolation factor is linearly ramped back to zero. At that point, table 2 is changed to yet another target, and the interpolation cycle repeats. Figure 4 illustrates the spectral interpolation algorithm. Notice that only one phase is computed for both tables. If the phases of corresponding harmonics in the two tables are equal, then linear interpolation of the table data is equivalent to an interpolation of their spectra. Notice also that the fundamental frequency is independent of the spectrum, allowing pitch decisions to be made independently from the spectral control.

Figure 3. Light reflected from water, captured by digital video.

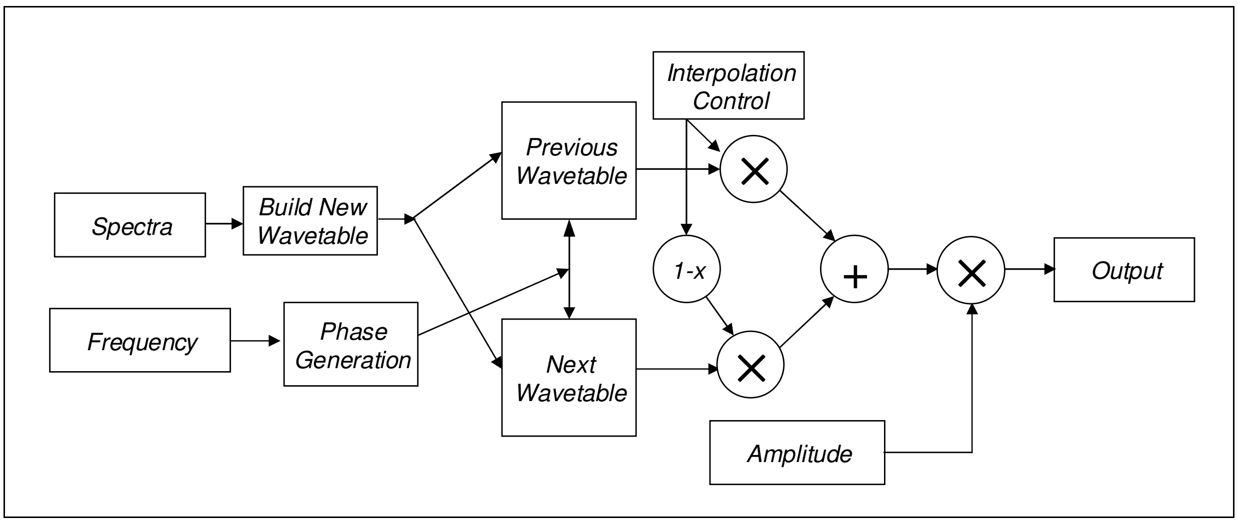

Figure 4. Spectral Interpolation Synthesis algorithm.

In the implementation for "Watercourse Way," we compute 512-sample tables containing 32 harmonics, which are derived from the video as described above. A naïve approach would simply wait until the next wave table is needed, then compute the table from the video data. This would cause a large computational demand (512 samples times 32 harmonics is 16K samples) that is undesirable in a real-time system. To avoid this, we spread the computation over time, using a third wave table to hold the intermediate data until it is ready. In our system, audio is computed in blocks of 32 samples, so every time we compute 32 samples using spectral interpolation, we also add one harmonic to the third wave table. After 32 blocks of 32 samples (1024 samples in all), the new table is ready. Even when the new table is ready, we still have to wait until the interpolation factor ramps to its endpoint at zero or one. Then we swap tables and begin interpolating with the new one. This leaves the previous table free to use in constructing the next wave table.

The result is a very interesting sound. As water waves travel up and down the screen, the listener hears a sweep through the harmonics, somewhat like a swept filter or a flanging effect. With multiple waves, this gives the sound a very animated character that has a very strong correlation with the moving water image, which was our main goal.

Wearable, Computer Controlled Lighting Systems

Several of the performers wear costumes with internal lighting that changes and modulates to complement the sound environment. The overall effect is a soft illumination from within that appears to flow across the body. The coloration follows the water theme with gently rippling shades of blue and green. During the course of the piece, the light slowly moves from ankle to shoulder, metaphorically filling the body with water and continuing the gesture of filling the on-stage pool. The body of the costume references the visual theme of waves by using curved strips of gauzy material that diffuse the light in a volume around the body.

The pattern of light is subtle; it is intended to accompany the music, not distract the audience. The color and tempo of the lighting tracks the sections of the music, but is not keyed to specific sounds. The costume conceptually closes a cycle represented in the piece; just as the ripples of water modulate light to create sound, the sound and score are processed to create ripples of light.

Technical Description

The light is generated by a number of small custom lighting modules sewn into the costume. Each unit incorporates a microprocessor, high-intensity light-emitting diodes (LEDs), and RS-485 data communication (see Figure 5). These units are interconnected by modular wiring that carries power and data from unit to unit within the costume. This network communicates with the on-stage controller either via a tether or radio data link. This allows the on-stage central controller to algorithmically generate the lighting pattern for several costumes in close coordination with the score and sound generation.

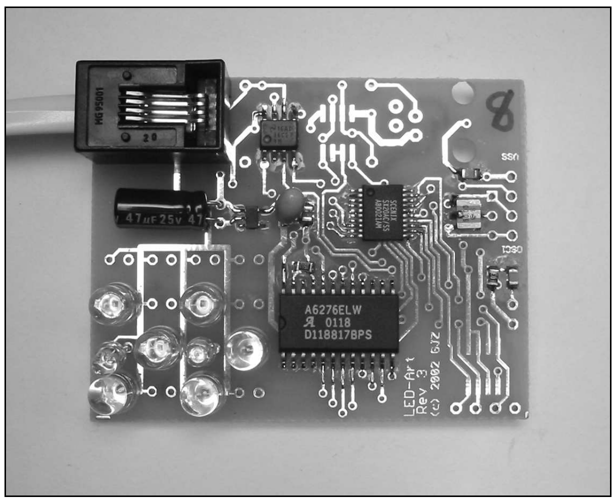

Figure 5. Electronics for wearable lighting module, approximately actual size. Connector (upper left) carries power and control. Red, green, and blue LEDs are at the lower left.

Each costume has a modest number of lighting units (less than ten), each of which has red, green and blue LEDs that can be modulated to control the overall color within RGB color space. Each unit is individually controlled, and the intensity and color of each source can be changed instantaneously with precision. This allows many effects: slow fades, color changes, fast pulses and rhythms, cross-fading between units in different locations, etc. The color is modulated by pulse-width-modulating the LED current; once the light is blended by the costume fabric the effect is a fairly uniform region of color. The LEDs can be pulsed at a rate invisible to the human eye, although fast body motions can make a slight strobing effect visible. For this application, this strobing is a less-than-desirable consequence of this practical modulation scheme. However, it does demonstrate that light has a temporal quality often overlooked in discussions of light and color.

This wearable lighting system has evolved through several variations. One precursor used optical fibers driven by high-intensity LEDs to create points of lights at the ends of the fibers, literally suspended in space. In that version, each costume was self-contained, with the light pattern algorithmically generated on the lighting microprocessor based on motion of the body as measured by an accelerometer. The optical fibers also created a visual effect as they swayed in response to body motion, causing the points of light to follow dynamic paths about the body.

The technical goals for the current hardware were to increase modularity and ease of prototyping, and to allow networking and central control. Development is generally easier on a central controller, plus the lighting control can be directly coupled to the sound generation process. As with the previous system, part of the lighting algorithm uses a simulated physical system, i.e., a system of differential equations modeled after a physical system. This is a rich source of behavior with "natural" qualities, since the rhythms and cycles follow patterns familiar to our eyes from the real world.

Future versions may once again incorporate sensing onto the body; it may be possible to include microphones that would make the costumes respond to the sound field precisely at the performer. This would follow through on the overall theme of making the invisible manifest; it would allow the audience to visually perceive a bit of the performer's experience in the context of the whole musical environment.

System Integration

Dealing with video, audio, MIDI, and computer graphics poses a challenging software design problem. We are using the Aura system (Dannenberg & Rubine, 1995) as a foundation for our work. Aura is designed to support interactive multimedia systems, with a special emphasis on interactive audio processing (Dannenberg & Brandt, 1996). To achieve low-latency audio processing, Aura runs multiple threads at different priorities. Audio processing is performed in the highest-priority thread so that it can preempt other computation that might otherwise lead to audio buffer underflow. A "control" thread handles performance information including MIDI, compositional algorithms, and other low-latency control. Video and graphics, which have the lowest frame rates and least critical timing, run at the lowest priority.

Aura implements a message-passing and scheduling system so that objects running in different threads can communicate efficiently. For example, the video object reads video frames, processes them to obtain spectra, and sends the spectra via an Aura message to the spectral interpolation objects running in the high priority thread.

A relatively new real-time scripting language, Serpent (Dannenberg, 2002), is integrated with Aura and provides a simplified way to build audio instruments and control them. Serpent commands can be entered and evaluated at run-time, which provides a powerful aid to testing and debugging. Using serpent, Aura objects can be instantiated, connected to inputs and outputs, and controlled. In addition, Serpent programs can be loaded from files. Much of "The Watercourse Way" is implemented as Serpent programs. Only time-critical, low-level objects such as the spectral interpolation synthesis object are implemented in C++, the implementation language of Aura.

Aura depends upon an operating system to provide threads and priority-based scheduling. We use Linux as a platform because of its good real-time performance. Linux is capable of dependably running the highest-priority thread within about 1ms when the thread becomes ready to run. Since Aura computes audio in blocks of 32 samples, we normally run the high-priority thread every block period of 726 microseconds. Several blocks must be buffered to avoid buffer underflow due to operating system latency, but the total latency can be kept down to a few milliseconds.

Unfortunately, Linux, as distributed, does not have good real-time performance, so it is necessary to patch the kernel. We use the PlanetCCRMA resources (Lopez-Lezcano, 2002), which provide pre-compiled real-time Linux kernels, device drivers for audio and MIDI, and extremely helpful documentation to make the system-installation process less painful.

Conclusions

"The Watercourse Way" is a new work resulting from the collaboration of the authors. Our goal has been to produce a musical work with visual and theatrical elements. We are attracted to technology for many reasons, including its ability to deal with audio, control, images, and video, and to transfer information from one medium to another. We hope that some of our synthaesthetic inclinations and experience are expressed through this work.

Technology and interactive art offers many opportunities to explore. In this work, we introduce a new synthesis method based on mapping real-time video data to time-varying spectra. We also present newly developed wearable lighting systems controlled interactively via wireless MIDI. All of this is operated by a single computer running Linux and using the Aura software platform.

A longer version of "The Watercourse Way" is planned for summer 2003 where it will be performed by the Pittsburgh New Music Ensemble under the direction of Kevin Noe. We hope to use the same platform to create new works and explore other possibilities, particularly the use of live video with real-time computer graphics and the use of multichannel audio to interact with many musicians and to work with specialized sound output.

Acknowledgements

We wish to thank Fernando Lopez-Lezcano at Stanford University's CCRMA for the fantastic job he has done creating PlanetCCRMA. Without his support, we and many others might still be struggling to get audio, real-time, MIDI, and graphics cooperating under Linux. Our computer was built by Tempest Computers, Inc. This work is partially supported by the Pennsylvania Partners in the Arts program of the Pennsylvania Council on the Arts, a state agency, the National Science Foundation Award #0085945, and by a commission from the Pittsburgh New Music Ensemble.

References

Bischoff, J., Gold, R., & Horton, J. (1978). "Music for an Interactive Network of Microcomputers." *Computer Music Journal, 2*(3), 24-29.

Borgonovo, A., & Haus, G. (1985). "Musical Sound Synthesis by Means of Two-Variable Functions: Experimental Criteria and Results." *Proceedings of the International Computer Music Conference 1984*. San Francisco: International Computer Music Association, pp. 35-42.

Boulanger, R., Smaragdis, P., & Ffitch, J. (2000). "Scanned Synthesis: An Introduction and Demonstration of a New Synthesis and Signal Processing Technique." *Proceedings of the 2000 International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 372-375.

Dannenberg, R. B. (2002). "A Language for Interactive Audio Applications." *Proceedings of the 2002 International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 509-515.

Dannenberg, R. B., & Brandt, E. (1996). "A Flexible Real-Time Software Synthesis System." *Proceedings of the 1996 International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 270-273.

Dannenberg, R. B., & Derenyi, I. (1998). "Combining Instrument and Performance Models for High-Quality Music Synthesis." *Journal of New Music Research, 27*(3), 211-238.

Dannenberg, R. B., & Rubine, D. (1995). "Toward Modular, Portable, Real-Time Software." *Proceedings of the 1995 International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 65-72.

Dannenberg, R. B., Serra, M.-H., & Rubine, D. (1990). "Analysis and Synthesis of Tones by Spectral Interpolation." *Journal of the Audio Engineering Society, 38*(3), 111-128.

Lopez-Lezcano, F. (2002). "The Planet CCRMA Software Collection." *Proceedings of the 2002 International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 138-141.

Verplank, B., Mathews, M., & Shaw, R. (2000). "Scanned Synthesis." *Proceedings of the 2000 International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 368-371.