Sonic Panoramas: Experiments with Interactive Landscape Image Sonification

Eric Kabisch[1], Falko Kuester[1,2], Simon Penny[1]
[1]Arts Computation Engineering, [2]Calit2 Center of Gravity
University of California, Irvine; Irvine, CA 92697, USA
{ekabisch, fkuester, penny}@uci.edu

Abstract

This paper describes research into the sonification of large-scale landscape imagery. Techniques for image sonification are coupled with users' physical navigation of the image data and a visual representation of the correspondence between the image and sound. One application of this research, an interactive art installation titled *Sonic Panoramas*, is discussed. Through the use of vision-based motion tracking, immersive projection, image analysis and real-time spatialized sound, a multi-modal representation of landscape image data is deployed.

Keywords: Visual and auditory display, motion tracking, image sonification, immersive projection, interactive art, interactive sound, multi-modal interface.

1. INTRODUCTION AND MOTIVATION

This work was motivated largely by two areas of inquiry. The first is in developing compositional techniques for real-time interactive sound environments, such as those required in immersive art and VR experiences. A second area of investigation in this work concerns the ways in which humans perceive, understand, and represent physical landscapes. The objective is to enrich a participant's experience of space through sonic interpretations of visual landscapes, providing a multi-modal interface for data exploration. The user's physical movement through the immersive projection space is tracked in real-time and used to generate a position-specific visual and auditory representation.

The history of formalism in music composition reflects interest in the application of formal structures to generate innovative and variable works of art from an abstracted analytical aesthetic [Rowe, Cope, Kabisch 2004]. These aesthetic and formal structures have included liturgical chants, the rolling of dice, and serial pitch sequences among others.

The use of natural and manmade landscapes as compositional source material is explored in this work. Countless composers such as Debussy and Ravel have incorporated inspiration from nature in their work. Most commonly, composers have attempted to mimic sounds of nature in their compositions as opposed to representing visual perception of space. The *World Soundscape Project* was a study initiated by Shafer in the late 1960's, whose main activity was the collection of recorded "found" sounds as a sonic mapping of the physical environment [Shafer]. Their approach was largely to gain perspective on the physical world and social change by becoming an objective listener to ambient sounds, both natural and manmade.

The sonification of visual landscape representations poses a unique set of challenges. An observation commonly made is that the influence of sound or "musicality" in painting is not matched by equal or compelling reflection of painting in music [Morton, Bosseur]. The various color organ experiments of the early 20th Century and Oskar Fischinger's Lumigraph [Fischinger] are often pointed to as examples of experiments in synaesthesia. However, these examples demonstrate the desire to create visual displays that are "musical" in nature or influenced by accompanying music. This paper explores the reverse: creating sound from visual source material.

Our data representation employs the mapping of visual information into sound, coupled with a visual representation of that sound's relationship to the source material. In addition, social and collaborative understandings of physical-auditory space are important considerations [Kabisch, Williams]. This paper focuses on the representation of an abstracted space through the user's navigation of that abstraction, a panoramic digital image. The goal is to develop new techniques to communicate and understand the complex physical and ephemeral qualities of landscape through the synergy of motion, sound and vision.

2. RELATED WORK

The presented research extends algorithms and techniques from the fields of real-time vision-based user tracking, data visualization, and image sonification.

2.1 Camera-based Motion Tracking

Reliable, fast and accurate tracking of subjects in the environment is critical for proper fusion of position-dependant visuals and sound. For interactive and collaborative spaces that allow multiple users to interact with the presented data concurrently, it is furthermore important that position information can be obtained without constraining the user. Camera-based motion tracking techniques have proven particularly efficient and many algorithms and APIs are now available, including SoftVNS [Rokeby], Cyclops [Singer], and Isadora [Coniglio].

While vision-based motion tracking has become more tractable in the last ten years, much pioneering work occurred earlier in the context of interdisciplinary media arts. Myron Krueger (1977) employed live video in computer driven interactive installations such as Video Place, establishing a paradigm for camera-based interaction. David Rokeby built Very Nervous System during the mid 1980's, a piece in which music was created from the motions of the user. Simon Penny has built custom machine vision systems for user tracking and real time interaction in immersive interactive environments including Fugitive (1996) and Traces (1998) [Penny].

In the project “living-room,” an augmented reality tracking toolkit ARToolKit available as opensource library from the HITLab (University of Washington) is used for two-camera motion tracking [Galantay]. EyeCon is a motion tracking software made for creating interactive works of art [Wechsler]. This software uses optical flow analysis on a video image to determine the movement of users. One can draw shapes and lines onto the top of the video image which act as controllers for sound and video parameters. When movement is detected in the same pixels as the drawn shapes, programmable actions are enacted. Livingstone and Miranda (2004) use a system of motion tracking, capturing and recording to allow for an environment that learns and responds to various user gestures and actions.

Some of the packaged software solutions noted are very general purpose and are meant to allow for tracking in a wide range of scenarios, while software developed for specific projects often employs solutions that are very specialized. For the creation of Sonic Panoramas a custom solution to the tracking problem was developed to eliminate unnecessary overhead while gaining experience with fundamental challenges that will inform future projects.

2.2 Image Sonification

There exists research on the use of digital images as a data source for sound generation, including Meijer (1992), von den Doel (2003), Kamel (2001) and Zhao (2004). It should be noted that the four papers referenced are intended for the mapping of data to sound with no visual component. This paper illustrates a desire to combine the two sensory modalities, and thus has slightly different concerns than the work referenced above. Further elaboration on the above work is included in the image analysis and sonification sections of this paper.

3. IMPLEMENTATION

The main installation is contained within a 15’ diameter circular space, demarcated by a three-foot high cylindrical screen. The screen is suspended from the ceiling and extends from four feet to seven feet from the floor. Users access the space by ducking under the screen. A 360-degree panoramic landscape is projected onto the screen, which is made of a material suitable for rear projection. As the participants walk into and around the space, their movements are tracked and position information is projected visually onto the projection and aurally through a four-channel speaker system. In addition, there are two other speakers that play ambient sound. Multiple participants are tracked concurrently and can interact with the environment individually. If users move close to each other, or if the number of participants exceeds three, they are tracked as one entity with merging visual and auditory trails. There is also a fairly simple control panel in the center of the space, used to change the landscape and other parameters.

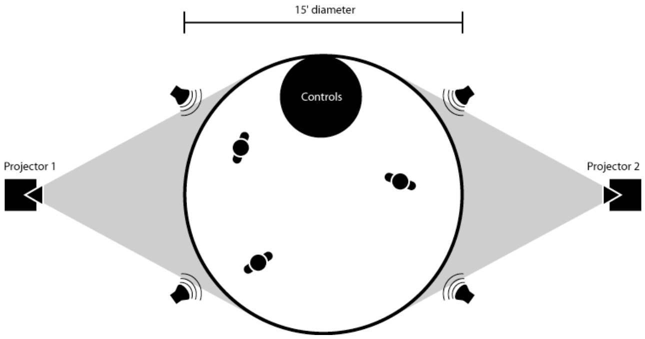

Figure 1. Top View of Installation showing position of projection screen, projectors and sound system.

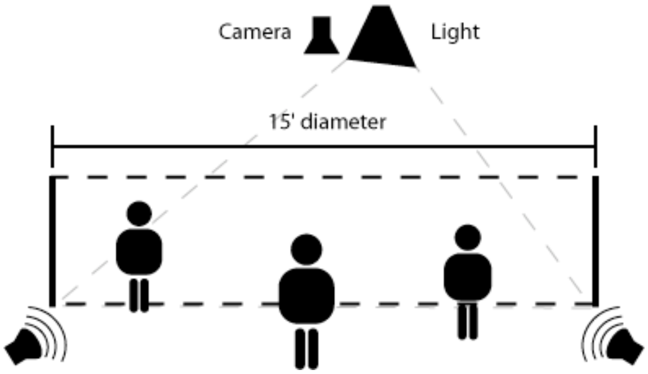

Figure 2. Side View of Installation showing layout of projection screen, vision system and sound system.

The participants are inside a circular projection of the panoramic image. Physical movement around the circle is sensed, and scan lines follow the users around the image, enacting the mapped sounds of the landscape’s data space. In order to reinforce the relationship of the user to the projected image, pixel blocks are highlighted as their relevant data is mapped to sound. This creates an understanding of the space informed by the synergy of the visual display, the user's physical location within the installation, and the spatialized soundscape.

Figures 1 and 2 show diagrams of the installation setup. Figure 3 is a block diagram of the data flow. The following topic headings proceed in the order of data flow shown in Figure 3.

3.1 Motion Tracking System

Development of the motion tracking system proved to be an exercise in trial and error. Many techniques were experimented with before devising a functional system.

Sharp infrared ranging sensors were employed initially to determine location around the circular interaction space. Using various sensor array configurations, viable tracking results were obtained. However none of them produced the type of control information desired. One problem with these sensors is that their functional range was not long enough. Though the sensors were advertised at 60", their true range seemed to be about half of that. Another issue was that they were not adequate for sensing multiple users, because one user could easily obscure the sensing of another user by reflecting the infrared signature before it reached the second user. The only way devised for these sensors to work was to place a large number at the center pointing outward and using them as binary sensors.

A video-based approach was subsequently explored using an overhead camera to track motion. The first trial was to break the video image down into many component image matrices, performing Horn-Schunk optical flow algorithms [Horn] on each component matrix to detect motion within that area. This allowed the tracking of multiple discrete areas of movement. However, successfully determining continuity of movement from one matrix to the next was quite difficult.

The optical flow algorithm was then employed on the entire video image, taking the output matrix of the optical flow and converting it to a binary image based upon an optimal threshold. Agglomerative hierarchical clustering algorithms were used to find three centroids of the moving mass. This technique proved to be quite effective for a single user. However, multiple users were more problematic because as one person came to rest, they would essentially disappear and the centroids would jump toward the moving objects.

The next experiment entailed converting the source video stream directly to a binary stream without optical flow. This required substantial contrast adjustment to the image due to the low light conditions necessary for a rear-projected environment. But this did allow for the user to still be seen by the system even when they were at rest.

The main problem with the motion tracking system was difficulty in tracking individuals in dark clothes. They

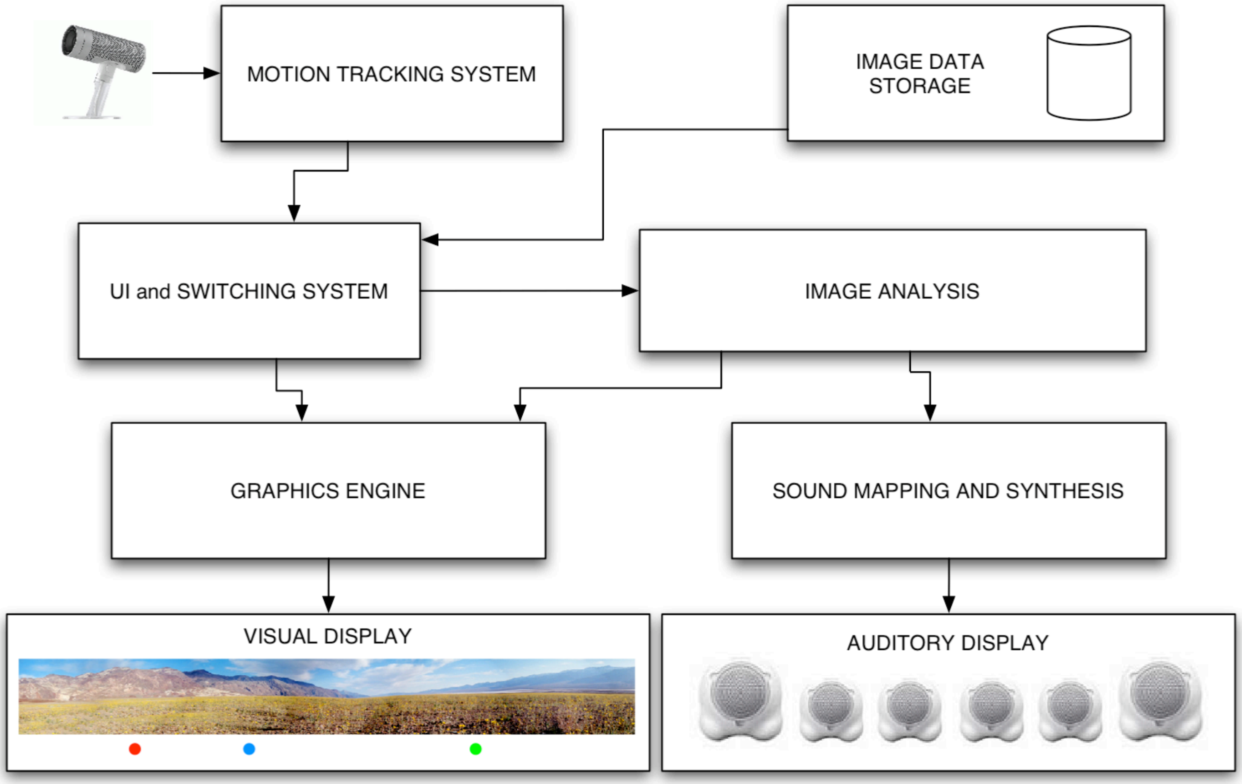

Figure 3. Processing Pipeline

were often nearly invisible to the camera. This was solved by supplying white hats that users could wear or carry if they were not being tracked sufficiently.

The motion tracking system was run on one dedicated computer. This computer performs the required image acquisition and processing, obtains the x and y coordinates and intensity of three centroids, and transmits this information at approximately 20 frames per second to the audio/video server over an Ethernet connection using the UDP protocol.

3.2 Image and Sound Acquisition

The source images and location sounds were acquired at various locations in Southern California. The images were shot using a Sony digital camera at three megapixels per image. After finding an optimal location and time for each shot, the camera was placed upon a tripod to maintain a consistent focal point. The camera was rotated for each successive image to allow for approximately a 20% overlap per image. The number of images for each panorama varied since a different focal length was chosen based on optimal framing of the scene.

After the images were taken they were stitched together, blended, warped and cropped to produce a continuous rectilinear image that could be wrapped onto a cylinder. The panoramas were stitched and edited using Arcsoft Panorama Maker, a commercial software package for Windows. The images were cropped and scaled to a resolution of 2048 by 150 pixels, providing a one-to-one pixel mapping from the panoramic image to the combined horizontal resolution of the projectors.

In addition to photography, ambient location sound was recorded at each location. At least ten minutes of ambient sound was recorded to allow for an adequate sample. In some locations, baffling was necessary to keep wind from overpowering the ambient sound.

3.3 Image Analysis

In these explorations, data is extracted from landscape images, which serve as an abstraction of the physical landscape. The images are selected to encompass a range of landscape types in order to gain varied data sets for sound mapping.

Meijer (1992) and van den Doel (2003) map sound directly from pixel information. By scanning the image, pixel information such as brightness, hue or saturation determines various parameters of the sound. This allows for a relatively simple mapping, however it also incurs a large amount of processing resources, as the entire image is scanned. Concerns with the abundance of pixel information led the designers to keep their sounds mappings relatively simple in order to limit information overload for the end users.

Here we propose that by limiting redundant information in the image, more complex and interesting mappings can be created without losing important image data.

Figure 4. Image before and after applying the Sobel edge detection algorithm.

When looking at landscape images, it can be noted that the defining characteristics are contained within the contours of the terrain. On a micro level, the pixel-to-pixel gradations do contain texture information, however for this project the concern is with larger landscape contours. This inspired experimentation with edge detection algorithms in the extraction of contour data from imagery. Various algorithms were tested and a simple Sobel algorithm was determined to give the best results across a range of images. Figure 4 shows an example of an unprocessed image, and one processed using the Sobel edge detection algorithm.

The result of using edge detection on the input image matrix is that we are left with a simplified matrix that has much of the unnecessary data removed. When navigating and mapping the new image matrix black pixels can be ignored, focusing only on the pixels whose color value is over a set threshold.

In addition to directly utilizing data from the edge detection image matrix, other data structures are populated with even simpler one-dimensional arrays. For instance, by taking the highest (in location) pixel row above the brightness threshold for each column, we end up with an array that shows the contour at the highest point in the image – typically the skyline, ridgeline or horizon. To gain more diverse data sets, this can be done for each plane within the matrix such that an array for each color channel (RGB) is obtained.

3.4 Presentation and Navigation of Data

There is an inherent difference in the way that we perceive sound versus the way we perceive visual stimulus. Standing in one location, a person can see only a subset of the 360° panorama, and even within the range of their peripheral vision they are actively focused on only a small portion at any given time. The way that the human auditory system, and sound propagation in general operates, we hear sounds from all directions at the same time. While rotating the head might enact slight changes in timbre, azimuth angle, and other perceptual qualities, in general the focusing on specific sounds within a field of sound is not overtly physical (as the rotating of the head or movement of the eyes) but seems to happen at a cognitive level.

While this is true, trying to listen to all of the visual content contained in an image at once would be a very difficult task. The temporal nature of sound allows us to make sense of sound as a function of time, not as discrete snapshots. This principle informs our design, and the design of most similar work. In Meijer, von den Doel, Kamel, and Zhao they use one point, or one scan line to determine the area of the image that is being sonified at any given moment. Meijer uses a vertical scan line that sweeps across the image at a given rate, then issues a noticeable click indicating to the listener that a new frame is being scanned. The listener becomes a passive receiver of the experience with no control over the exploration of the data. For their purposes this is ideal, but in our scenario we wish to give the user more control. von den Doel places control of the image space in the hands of the user by allowing them to move a pointing device over the image surface. However instead of using vertical scan lines it renders sound on a pixel-by-pixel basis.

Sonic Panoramas uses a combination of these two approaches. The combination of user control with the use of vertical scan lines enables the user to become immersed in a virtual space that they can navigate on their own terms. Given the distant, panoramic nature of the imagery used, the vertical scan line works quite well to navigate the image in relation to the user. The user is not controlling a pointing device, but is inside a cylindrical projection of the panoramic image. Physical movement around the circle is sensed, and the vertical scan line follows the user around the image, enacting mapped sounds of the landscape's data space. In order to reinforce the relationship of the user to the projected image, pixel blocks are highlight as their relevant data is mapped to sound. This creates an understanding of the space informed by the synergy of the visual display, the user's physical location within the installation, and the spatialized soundscape.

3.5 Visualization

Visualization and sonification were done within the Max/MSP/Jitter [Cycling] programming framework. A switching and timing program was developed to change between the various panoramic images, datasets, sound mappings, and sound engines.

A 2048 pixel by 150-pixel image matrix was populated with the appropriate imagery and an additional image matrix of 2048 by 50 pixels was used as feedback to the user indicating where the motion tracking centroids were located (see Figure 3). As motion tracking information was received the Cartesian coordinates were converted to polar coordinates of radial angle, amplitude and intensity. The angle was mapped to the appropriate pixel location in the rectilinear image matrix, and circles were drawn using the angle for radial position and the amplitude for circle diameter. Each of the three motion tracking centroids was colored differently, red, blue and green.

Image analysis data was stored in arrays with the index corresponding to the horizontal pixel index of the image matrix. This data was then mapped to visualization and sonification engines according to the movement of users. As each pixel was mapped to sound, a pixel the color of the tracking centroid was drawn in the corresponding location on top of the image in the main image matrix.

3.6 Sonification

Decisions made in the mapping of data to sound are as much artistic and philosophical as they are technical. Much of the literature on data sonification aspires to a clean one-to-one mapping of the data to sound in order to maintain the integrity of the source data. This is especially true and necessary in Meijer where the goal is to represent images to the visually impaired. However, representations in the arts have the luxury, and even a compulsion, to interpret data is a way that communicates more than the source material itself. Metaphors of cartography are ideal for this project, which seeks to represent landscape and space. Borges (1998) tells the tale of the Empire whose cartography became so advanced that they created a full scale map of the entire Empire and placed it right on top of the land, point corresponding to point. It seems that in map making, as in other forms of representation, it is not only what is included that matters but what is left out.

3.6.1 Previous Techniques

We first outline some of the techniques that served as a foundation on which the sonification for Sonic Panoramas was built.

3.6.1.1 Mapping on the Formal (Macro) Level

Data extracted from the images can be used to generate musical structures either on the macro level, as applied to elements such as rhythm and form, or it can be applied on the micro level through sound synthesis and timbre. The following outlines experiments with both.

As the user navigates the installation space, the vertical scan lines read edge pixels at a rate that maps fairly well to the triggering of individual notes. This idea is used for several examples of varying complexity.

The mapping that seems to be most common, used by Meijer and Zhao among others, is a mapping from vertical pixel location to pitch. Their tests have shown that this mapping is fairly intuitive to the average user. This technique informs the first mapping outlined here. The image matrix obtained from the edge detection algorithm is scanned as the user's position moves horizontally. Locations around the circumference of the installation space are mapped to a horizontal column of the image. As the user's movement enacts each column, the entire column is scanned for pixels whose combined red, green, and blue values are above a certain threshold. When pixels are found, their row number is sent out as control data. This control data is remapped to relative frequencies within the audible range and sent to a sine wave oscillator. The resulting sound mapping allows the user to trace the contour of the image, with pitch increasing as the height of the image contour increases.

In addition the sound pans around the spatialized audio system with the movement of user navigation.

The next experiment involves the use of all three matrix planes to control separate oscillators. This is a more interesting mapping as it illustrates the differences within the color planes. With some images, the planar array data is quite similar, however with the proper image and edge detection settings the system arrives at data that results in contrapuntal movement between the various planes. One implementation employs three sine wave oscillators panned and set in octave intervals from one another. Other interesting sounds can be achieved when changing the types of synthesis and harmonic intervals, but they are kept simple for this illustration.

A fourth mapping begins to explore a more subjective way of representing the image data. In this example, the user or composer draws in pitch contours to integrate with the navigation of the image data. These pitch contours can also be derived from image analysis instead of user input. The interesting component of this technique is the imposition of multiple pitch contours onto the same navigational framework. This system also has the capacity to record and playback loops of recorded control data. This concept could become very interesting within the context of the interaction if employed properly.

3.6.1.2 Mapping on the Synthesis (Micro) Level

Not only are the extracted data structures useful as control data within a larger formal context, but they can also be used as information for the synthesis of sound on the micro level. One way to do this is to use our existing data as lookup tables for waveshaping or additive synthesis. If familiar with viewing Cartesian representations of audio waveforms, one can almost imagine a skyline or ridgeline as one big waveform waiting to be realized (which was actually one of the initial inspirations for this work). By reading through buffers filled with image analysis results we accomplish just that. However, the sound produced by these waveforms, while periodic, tends to be perceived as noise. These waveforms can reflect their own character in a more pleasing way when used as waveshaping lookup tables. We can easily use the same image data to control both the formal structures (pitch and rhythm) and the timbral structure by shaping a frequency-controlled sine wave with two of our data buffers.

3.6.2 Extensions of Previous Techniques

In previous experiments the goal was that the underlying image data, when left to simplistic devices, would speak for itself and allow for interesting contrasts between landscapes. Now, grounded in those simple mappings, the system has become more complex, incorporating higher levels of intervention and representation.

Many of the same "macro-level" approaches outlined above are still employed. The pixel height of selected edge pixels is used to control enacting of pitch in many of the sound mappings. These pitches are then filtered in both the temporal and frequency domains. This results in a decrease of note density in the time domain to avoid sensory overload and it allows for scalar pitch mapping within the relative placement of the data source.

On the micro level, more complex synthesized tones were used to evoke musical timbres. The choice of musical voicings was informed by previous "objective" experiments, but was also based upon subjective associations with the source imagery.

3.7 Installation of Auditory and Visual Displays

A custom projection screen and sound system were configured specifically for the deployment of Sonic Panoramas.

Figure 5. Photo of Installation from a low angle outside of the screen looking into the interaction space

3.7.1 Screen Construction

The main concerns in designing the projection screen were finding a material that was both cost effective and also exhibited good rear projection properties. A thin vinyl taffeta material from a local plastics supplier was acquired, by purchasing a rectangular piece of material that was 52 feet long by three feet wide and having it welded with three-inch seems in both the top and the bottom. This allowed the structural support to be threaded through the seams.

Again concerned with cost and ease of fabrication PVC plumbing pipe was used to construct the structure. By gluing five ten-foot lengths of PVC together and then bending and attaching it end-to-end, the pipe naturally conformed to a near-perfect circle. After experimentation, 1" thick pipe was found to be ideal for the top ring of the structure and ¾" pipe worked for the bottom ring, in order to stretch the material. Five pipes were connected together for both rings, and then threaded through the hems of the material. The ends of both pipes were then glued together to form the circles, and the material was stretched out around the structure.

Six 1/8" steel braided cabled suspended the screen from the ceiling. Clamps were attached to the ends of the cable and clamped onto the screen.

Figure 5 shows the fully constructed and hanging projection screen with images projected from both sides. The low angle of the photograph shows the depth of the installation and multiple participants interacting with the scene concurrently.

3.7.2 Projection

Only two projectors were used for this installation due to resource constraints. This was not ideal, as the pixelation was fairly evident when projecting 1024 pixels to a width of 15 feet, making the pixels almost ¼" wide. The projectors were placed on opposite sides of the screen at a height equal to the mid-point of the screen.

In addition to the pixelation, the image was not perfect due to challenges of projecting onto a curved surface. As the screen distance increased on the sides, some image warping, blurring, and increased pixel size resulted. However, the artistic nature of the installation made this acceptable and the image was not pre-processed before being sent to the graphics card. All in all, the image distortion was not as bad as predicted.

Another challenge came up due to the size of the installation space. There was not enough space to get enough projector throw distance for the full 15-foot width on both sides. A mirror was used to increase the throw for one of the projectors. The projector was placed near the floor pointing directly up at a mirror. The mirror was suspended from the ceiling with steel cable, hung at the same height as the other projector, and placed at a 45-degree angle to the floor and projection surface. This increased the throw distance just enough to fill the projection screen with the width of the image.

3.7.3 Camera and Lighting Setup

An Apple iSight Firewire video camera was used for image capture in order to track motion in the installation space. The camera was placed overhead at the exact center of the circular installation space. It was hung from the ceiling at approximately 15 feet above the floor. In order to increase the viewable area a wide-angle camera lens was attached to the iSight. The camera was hung with a flexible gooseneck mount in order to allow for easy adjustment.

The challenge of using visible light for motion tracking in this installation was that too much light tended to wash out the projection image, while too little light made tracking difficult. A solution to explore in the future would be to use another spectrum of light such as infrared. In order to light the space without lighting the projection screen an impromptu lampshade was constructed, specially designed to aim light all the way to the bottom of the screen without casting light upon it. The light had to be offset from the center of the space to be out of the way of the camera. Therefore, the lampshade was angled such that the side projecting a longer distance was shorter, allowing light to be cast further.

3.7.4 Speaker Setup

There were six channels of audio employed for this installation. Two powered speakers were placed at one side of the installation space, underneath the control table. Four other speakers were placed on the floor at points equidistant around the circle. They were all angled toward the center of the circle and angled up from the floor to propagate sound to the ear height of an average standing user.

A six-channel firewire audio interface was used. The ambient location sound was directed to the stereo pair underneath the table. Spatialized audio was fed to the four surround speakers, with sounds being localized to the area where detected movement was enacting sound.

4. CONCLUSIONS

Much was learned in the development and construction of this installation. Techniques for motion tracking were explored, with both challenges and successes. Ultimately an approach was achieved which minimized processing overhead and effectively tracked multiple users in a large (180 ft.^{2}) interaction space with a single camera. The use of a portion of the visual display to reinforce participants' effects on the tracking system was a significant success. A cylindrical projection screen was constructed quickly and inexpensively from readily available materials. This technique will inform future prototyping of large, atypically shaped screens. In addition to construction, problems associated with projecting on a curved surface were explored.

Existing research into the representation of images as sound was extended, utilizing edge detection and image analysis to extract data structures. Continued development and experimentation with the mapping of data to sound was pursued with good results. Perhaps the greatest achievement was the integration of various real-time processing techniques for motion tracking, image analysis, user interactive, sonification and visualization into a cohesive user experience. By forming a coherent link between the sonification and visualization, a greater understanding of the underlying data was produced through synergy of vision, sound and movement.

On a technical level this project was a success for the reasons listed above. As an art project there were also accomplishments. The interaction design turned out to be quite successful. Users were readily able to interact with the piece and understand their affect on the system with little or no instruction. The piece was largely received as an aesthetically pleasing experience, and it engaged users for periods up to fifteen or twenty minutes, a long time for a gallery setting.

5. ACKNOWLEDGEMENTS

We would like to thank Elizabeth Chattin, Christopher Dobrian, Robert Nideffer and Kevin Ponto for their consultation, support and input through the course of this project.

REFERENCES

J. L. Borges, "Of exactitude in science," in *Collected Fictions*, Viking Penguin, 1998.

J. Bosseur, *Music: Passion for an Art*. Rizzoli, 1991.

M. Coniglio. "Isadora." http://www.troikatronix.com/isadora.html last accessed September 2005.

D. Cope, *Virtual Music*. The MIT Press, 2001.

Cycling '74. "Max/MSP/Jitter." http://www.cycling74.com/ last accessed September 2005.

K. van den Doel, "Soundview: Sensing color images by kinesthetic audio," in *Proceedings of the International Conference on Auditory Display*, pp. 303-306, 2003.

E. Fischinger. "Writing Light." http://www.centerforvisualmusic.org/WritingLight.htm last accessed September 2005.

T. Galantay, J. Torpus, and M. Engeli, "living-room: Interactive, space-oriented augmented reality," in *Proceedings of the 12th annual ACM international conference on Multimedia*, pp. 64-71, 2004.

B. K. P. Horn and B. G. Schunk. "Determining optical flow." *Artificial Intelligence*, 17:185-203, 1981.

E. Kabisch, "Generative aesthetics: The pursuit of musical behavior." http://e.fluxt.com/papers/ last accessed September 2005, 2004.

E. Kabisch, A. Williams, and P. Dourish, "Symbolic objects in a networked gestural sound interface," in *Proceedings of the Conference on Human Factors in Computing (CHI)*, pp. 1513-1516, 2005.

H. Kamel, P. Roth, and R. Sinha, "Guess: Providing interrelational representation of objects in a non-visual environment," in *Proceedings of the International Conference on Auditory Display*, pp. 261-266, 2001.

M.W. Krueger. "Responsive Environments," in *Proceedings of the AFIPS 46 National Computer Conference*, pp. 423-33, 1977.

D. Livingstone and E. Miranda, "Composition for ubiquitous responsive sound environments," in *Proceedings of the International Computer Music Conference*, pp. 321-324, 2004.

B. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112–121, 1992.

M. Morton, "From the other side," in *The Arts Intertwined: Music and Painting in the Nineteenth Century*, Garland Publishing, 2000.

S. Penny, J. Smith, P. Sengers, A. Bernhardt, and J. Schulte. "Traces: Embodied Immersive Interaction with Semi-Autonomous Avatars," *Convergence*, vol. 7, no. 2, 2001.

D. Rokeby. "SoftVNS." http://homepage.mac.com/davidrokeby/softVNS.html last accessed September 2005.

R. Rowe, *Interactive Music Systems*. The MIT Press, 1993.

R. Shafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books, 1977.

E. Singer. "Cyclops." http://www.ericsinger.com/cyclopsmax.html last accessed September 2005.

R. Wechsler, F. Weiß, and P. Dowling, "Eyecon – a motion sensing tool for creating interactive dance, music and video projections," in *Proceedings of the Society for the Study of Artificial Intelligence and the Simulation of Behavior (SSAISB)'s convention: Motion, Emotion and Cognition*, 2004.

A. Williams, E. Kabisch, and P. Dourish, "From interaction to participation: Configuring space through embodied interaction," in *Proceedings of the Conference on Ubiquitous Computing (Ubicomp)*, pp. 287-304, 2005.

H. Zhao, C. Plaisant, B. Shneiderman, and R. Duraiswami, "Sonification of geo-referenced data for auditory information-seeking: Design principle and pilot study," in *Proceedings of the International Conference on Auditory Display*, 2004.