Bringing to Life the Musical Properties of the Eyes

Anthony Hornof
Computer and Info. Science
University of Oregon
Eugene, OR 97403 USA
hornof@cs.uoregon.edu

Troy Rogers
Department of Music
University of Virginia
Charlottesville, VA 22903
tsr9f@virginia.edu

Jeffrey Stolet
School of Music
University of Oregon
Eugene, OR 97403 USA
stolet@uoregon.edu

Tim Halverson
Computer and Info. Science
University of Oregon
Eugene, OR 97403 USA
thalvers@cs.uoregon.edu

ABSTRACT
Art, music, and science come together as computer musicians and computer scientists collaborate to create EyeMusic, a system for composing and performing eye-controlled musical and multimedia compositions. In EyeMusic, an eye-tracking device communicates with a musical and multimedia authoring environment to enable composers and artists to sonify and visualize eye movement data. During a live performance, the performer moves his or her eyes to interact with visual stimuli and, in doing so, triggers a range of musical and visual elements. Two major challenges in designing eye-controlled multimedia software include (a) composing pieces that use the noise and error in eye movements and eye trackers for an expressive and artistic outcome and (b) finding a balance between having the performer directly control that outcome and having the composition proceed on its own while responding in meaningful and interesting ways to eye movements. EyeMusic compositions have been selected by multiple juries for live performance at major computer music conferences.

Author Keywords
Computer music, eye tracking, new media art, performance.

ACM Classification Keywords
Categories and Subject Descriptors: A.m [General Literature]: Miscellaneous; H.4.3 [Information Systems Applications]: Communications Applications; H.5.2 [Information Interfaces and Presentation]: User Interfaces - input devices and strategies, interaction styles; J.5 [Arts and Humanities]: Fine arts, performing arts. General Terms: Design, Experimentation, Human Factors.

INTRODUCTION
Eye tracking continues to hold great promise, not yet fully realized, for human-computer interaction, both to analyze and understanding how people interact with visual displays, and to provide an alternative means of interacting with a computer in real time. Eye movement data are useful to a variety of disciplines. Cognitive psychologists study eye movement data to understand human information processing capabilities. Human factors practitioners employ eye tracking to understand how people interact with devices and to improve usability. Accessibility researchers write software to enable physically disabled people to communicate by controlling the computer with their eye movements. Jacob and Karn [4] provide a good overview of eye tracking research in human-computer interaction.

Computers have fundamentally transformed how music is produced and consumed. Apple's iPod, for example, has radically extended the concept of music portability. Computers have profoundly changed our notions of art [13] and music [9]. What is really exciting, though, is that given the timeline of human creative expression, computers have by now only barely touched art and music. Massive exploration of computer-mediated art, music, and performance is yet to come. This article discusses what may be the current "state of the art" in using eye movements for real-time computer-mediated creative expression.

Collaborators for this project include (a) computer music composers and performers and (b) eye tracking and cognitive modeling researchers. This collaboration has been active for four years. Eye tracking researchers teamed with domain experts to explore new eye-interaction techniques that support the users' tasks (composing, performing, watching, and listening). This paper documents what did and did not work. Direct objective assessment by domain experts is critical for the development of new computer-mediated creative expression. Recent developments presented in this paper pertain to live performance—system enhancements that make live performances possible, compositions that emphasize live performance, and pieces that have been performed live. This project builds on previous work [3] by exploring the interface design, human factors, and musical challenges of creating compositions that can be performed live using an eye tracker in front of an audience.

There are a number of reasons that eye movements are useful and interesting for musical composition: (a) A performer could alternate between a higher-level visual task such as studying a painting, in which eye movements are to some extent programmed subconsciously, and another task in which he or she deliberately controls the music with his or her eye movements. (b) There is an inherent musical or at least rhythmic quality to eye movements that lends itself to composition. (c) People with physical disabilities who already interact with their computer by moving their eyes could enjoy new opportunities for musical expression. (d) From the perspective of scientists who analyze eye movement data to understand patterns of human visual processing, there may be patterns in the data that become most salient when the data are sonified rather than

University of Oregon Department of Computer and Information Science Technical Report 08-05.

visualized. This project focuses on creating compelling artistic and musical performances that contribute to contemporary computer music in a meaningful way, while exploring the interactive design issues of creating eye-controlled musical instruments, compositions and performances.

This work is of particular interest to human-computer interface design because (a) it traces the development of eye-controlled interactive systems for creative expression and provides a model for other scientists and artists/ musicians who wish to collaborate and (b) it describes the architecture, interface, and evaluation of an innovative new eye-controlled interactive system.

Previous Eye-Controlled Music

We are aware of only one previous body of work in which eye movements direct musical compositions, work by the digital artist Andrea Polli [6]. Her musical composition with eye tracking entitled *Intuitive Ocusonics* has been performed internationally. Excerpts are online at (www.andreapolli.com). In this work, the eyes directly control aspects of the composition as it is performed. The pieces are striking, filled with haunting electronic sounds and digital samples of the human voice, sometimes singing and sometimes screeching. The compositions tend to be sparse, with just a few instruments or voices playing at a time.

EyeMusic differs from the work of Polli in that EyeMusic benefits from decades of scientific research on eye tracking that is embedded within a commercial eye tracking system, the LC Technologies Eyegaze System, which provides accurate gazepoint data using the standard pupil-center corneal-reflection technique. Polli's compositions respond to video images of the eye—not specifically the pupil center or corneal reflection—which are parsed and processed twelve times per second using STEIM's BigEye software (www.steim.org). EyeMusic enjoys the benefit of specialized algorithms for translating the video images of the eye into screen coordinates, with spatial and temporal accuracy that permits specific, deliberate musical composition. Polli's live performances include a large video image of the eye as it moves. EyeMusic performances, on the other hand, include a video image of the eye to make it clear how the piece is being played, but the primary visual presentation is of the gaze moving through a visual scene.

GETTING THE HUMAN DATA TO THE MUSICIANS

This section introduces the technical components of EyeMusic system, and explains how the components interact. In that the EyeMusic system architecture interacts directly with the human physical "architecture," a few terms pertaining to eye movements and eye tracking must be defined. The *gaze* is the vector that goes from a person's eye to the gazepoint, which is the point in a scene where he or she is looking. The eyes (and thus the gaze) examine a scene with a series of quick jumps called *saccades*, each of which lasts roughly 30 ms. Between saccades, the eyes (and the gazepoint) stay at the same location (with a slight tremor) for a *fixation* that lasts roughly 100 to 400 ms. People acquire visual information primarily during fixations, not saccades. The eyes move so that people can put items of interest into the high resolution vision at the center of the gaze.

Individual eye movements are typically made subconsciously in the service of a higher-level strategy to accomplish a visual task, such as reading. However, people can also make deliberate, conscious decisions to move their eyes to a specific location, such as to fixate this letter *X* for two seconds and to then pass the control back to the higher-level reading strategy. For a more detailed explanation of how and why the eyes move, see Rosenbaum's *Human Motor Control* [10].

Eye movements are monitored using an *eye tracker*, which typically incorporates a camera that sits below a computer video monitor and is focused on the eyes of the person using that computer. The video images are transformed, via software algorithms, into the (*x, y*) coordinates of where the person is looking on the screen.

Figure 1 shows the major software and hardware components in the EyeMusic system. Arrows indicate the flow of data. A *scene generator* displays a visual image on a video display. A person, the *eye performer* or *oculist*, moves his or her gaze around the scene. A video camera captures an image of the eyes and, in the *eye tracking computer*, converts it to the corresponding (*x, y*) coordinates of the video display. EyeMusic currently uses the L. C. Technologies Eyegaze System (www.eyegaze.com), which monitors the change in the spatial relationship between the pupil-center and a corneal-reflection as the gaze moves across the screen, and reports the gaze position sixty times per second, once every 16.67 ms. In our experience working with this eye tracker, once a good calibration is acquired for a performer, the accuracy is within 1° of visual angle. The sampling rate and the accuracy of the eye tracker are more than adequate for capturing the relevant physiological phenomena as well as the unique personality and characteristics of eye movements. The temporal and spatial accuracy are adequate to clearly see and hear the eye movements translated into the intended music.

The system is designed based on an understanding of how computer musicians work, which was gained via interviews, observations made at computer music concerts and performances, and participant observation. One of the most widely used software applications for computer music composition is Cycling 74's *Max*, a graphical environment for creating music and multimedia (www.cycling74.com) named after Max Mathews, a pioneer in computer music. Max is typically used in conjunction with the Max Signal Processor (*MSP*) and thus referred to as *Max/MSP.* It is often used with an object library called *Jitter* that supports real-time manipulation of graphics and video, and is typically used on an Apple Macintosh computer.

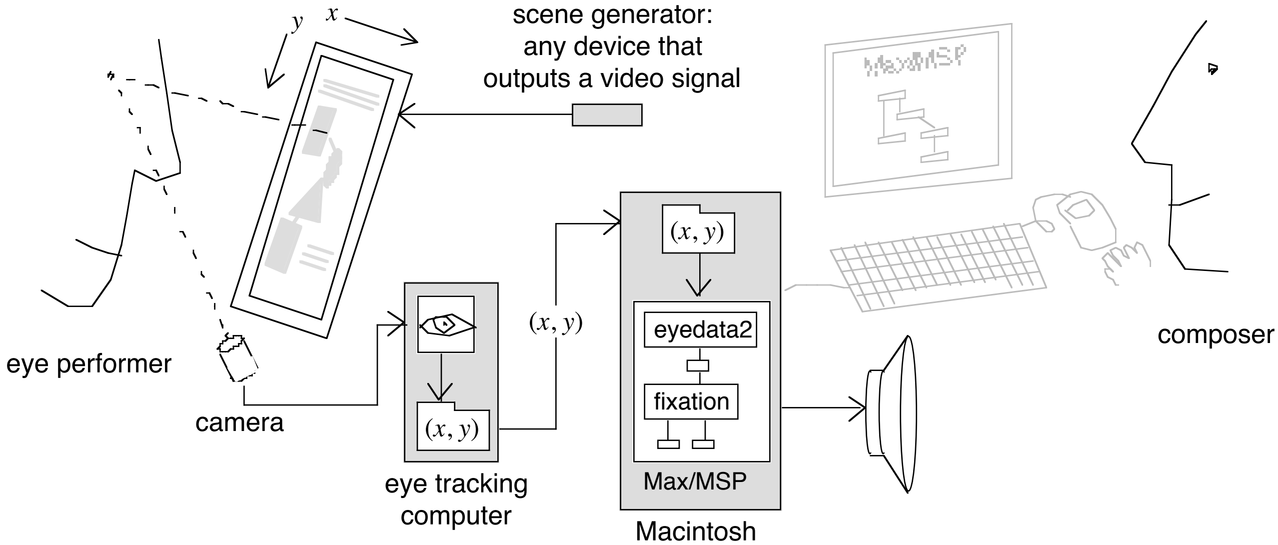

Figure 1. The EyeMusic system architecture is designed to get the human eye movement data to the musician in a context and format that are most useful for composing and performing. Arrows indicate the flow of data.

To enable computer musicians to make the best use of eye movement data in the musicians' native environment, the eye movement data are made available in Max/MSP by means of *external objects*, written in the C programming language, which insulate the composer from the technical details of eye tracking. The data appear in Max/MSP by flowing out of an *eyedata* external object at a rate of 60 Hz and, optionally, are parsed into fixation data by means of a *fixation* external object. Both external objects were created for this project and are discussed next. Two versions of eyedata are discussed, version 2 which reads prerecorded eye movement data from a file, and version 5 which connects to an eye tracker and delivers the gaze position in real time.

Playing Back Prerecorded Eye Movement Data

In EyeMusic, eye movement data are read from disk into the Max environment by means of an external object called eyedata (version 2). Figure 2 shows the eyedata object as it would appear in the Max/MSP authoring environment. This is where the eye tracker connects to Max/MSP. The gaze position (and the pupil radius) flows out of the eyedata object at a rate of sixty samples per second, to any Max object that is connected to the outputs at the bottom of eyedata.

Table 1 shows sample eye movement data output from the external. *Sample #* increments 60 times a second. *Eye found?* indicates whether the eye was tracked for that sample (1 = yes, and 0 = no). The *x* and *y* are the screen coordinates of the gaze, in pixels, with (0, 0) at the top left of the screen. *Pupil radius* is reported in mm. The data shown in Table 1 were collected while a person was reading. On Sample #390, the eyes finish a fixation at roughly (553, 112). On Sample #391, the eyes make a horizontal saccade to the right. On Sample #392, the eyes start a fixation at roughly (652, 109).

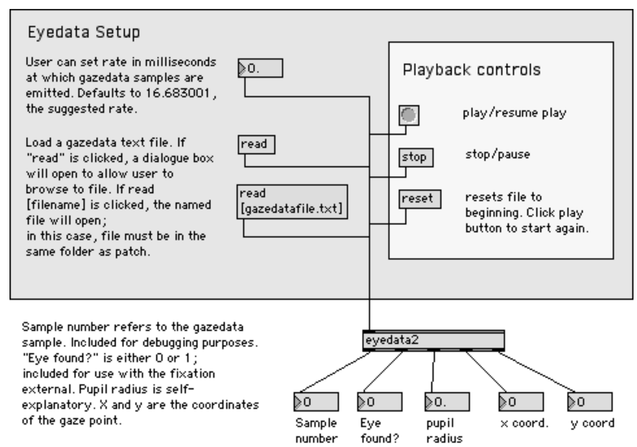

Figure 2. Composers can read eye movement data files from disk using the *eyedata2* object in the Max/MSP multimedia authoring environment. Data flows down the *patch cords.*

Table 1. Sample data from the eyedata object.

Sample #	Eye found?	*x*	*y*	Pupil radius
...	...	...	...	...
388	1	553	112	1.31
389	1	552	112	1.32
390	1	554	112	1.31
391	1	576	111	1.32
392	1	634	108	1.3
393	1	663	108	1.31
394	1	659	111	1.3
...	...	...	...	...

Parsing the Eye Movement Data into Fixations

The fixation external object converts the gaze samples into fixations. The samples flow out of the eyedata object at a rate of sixty samples per second. The rate is an artifact of the eye tracker used, and different trackers have different rates. Where and when the performer makes a fixation, either deliberate or subconscious, is a more relevant human physiological phenomenon than the sample rate of the eye tracker. The fixation object parses the gaze position samples and identifies when and where a fixations occurs. This way, a composer can work directly with the eye movement data that eye tracking specialists believe to be most relevant. Figure 3 shows the fixation object as it would be used by a person creating a multimedia composition in Max/MSP. The fixation object uses an established dispersion-based algorithm [12]. The object is further detailed in previous work [3].

Playing Eye Movement Data in Real Time

For live performance, gaze data needs to flow from the eye tracker to Max/MSP system in real time. This is made possible with an updated version of eyedata, version 5, which connects Max/MSP to the eye tracker with a network connection. The physical connection is typically made by connecting the two computers directly with a single ethernet cable, but the connection could also be made across the Internet. The object permits the musician to specify a remote eye tracker within Max/MSP, such that the data could be retrieved from an eye tracker anywhere on the Internet. We have not yet explored the possibilities of remote collaborative performances, though, in part because we expect that this would introduce delays in data stream.

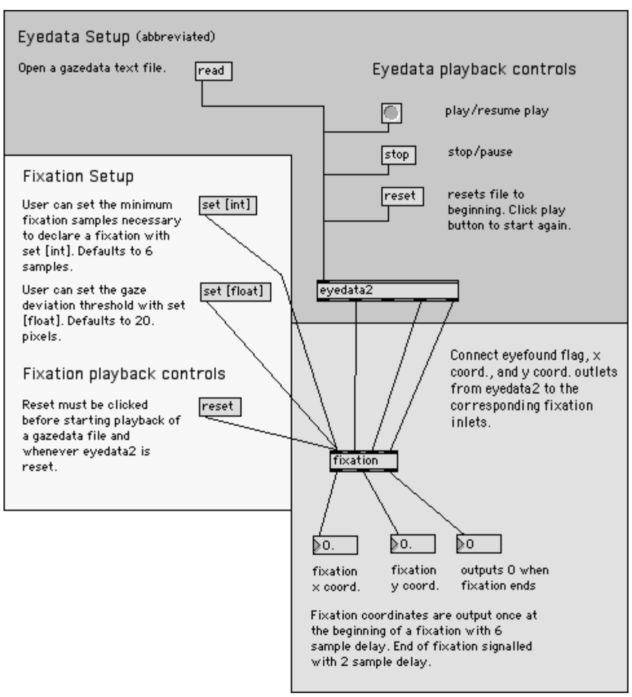

Figure 3. To encourage the musicians to compose to the physiological phenomenon of gaze fixations, the computer scientists created a second *fixation* object that parses the eye movement data using established algorithms.

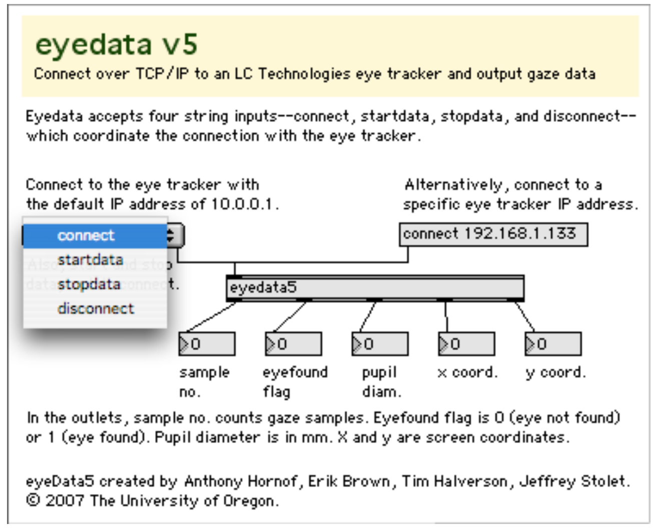

Figure 4. The most recent version of the *eyedata* object (v5) connects directly to an eye tracker for real-time performance.

Figure 4 shows the eyedata version 5 object as it would appear in a Max program. It is similar to eyedata version 2 discussed earlier except that the performer now has a series of commands for connecting to the eye tracker. Before connecting, the performer must activate software on the eye tracker (called *EGServer*) that sends eye movements to other computers. The eyedata5 object was designed to insulate the composer and performer as best as possible from the technical details of making the connection. Numerous technical issues are addressed within the C code of the external that enable the composer and performer to focus on the music. For example, the external keeps internal track of its current state, such that if a performer tries to connect again after a connection is already established, the request is ignored, thus avoiding a software crash.

Recent commercial technology developments make it much more feasible to do live performances of EyeMusic compositions. Until recently, performances required not only the eye tracking computer, cameras, displays, audio converter, and many cables, but also a 20 kg Macintosh PowerPC (Dual G5) tower computer, the only Macintosh with the processing speed required to run a Max/MSP composition with full audio resolution and a Jitter video composition at 30 fps, as needed to simulate smooth motion.

The recent boost in processing power of Macintosh laptop computers that resulted from Apple moving from the PowerPC to the Intel Core 2 Duo central processor, and our subsequent revising of the eyedata5 external to a Macintosh “Universal Binary” format so that the external can run on the new laptops, means that a 2 kg laptop computer can now replace the 20 kg tower. Portable hardware has finally caught up to the processing requirements of EyeMusic. This is an important human factors issue because computer musicians are continually trying to reduce the amount of equipment that they must travel with, even down to replacing 2 m cables with 0.2 m cables because they are smaller and lighter. The size and weight of equipment constrain where and when a computer musician can perform live. It is not much easier to transport EyeMusic to remote venues.

The EyeMusic externals, documentation, eye movement data, and audio-visual recordings can be downloaded at <www.cs.uoregon.edu/research/cm-hci/EyeMusic/>. An eye tracker is not needed to use eyedata2.

COMPOSITIONS THAT SONIFY EYE MOVEMENT DATA

The process of composing music based on eye movements will sometimes focus on how the different gaze positions and timings can be mapped to sounds, rather than the live interactive potential of "playing" an instrument or composition with the eyes. For this sonic development, it is easiest for the composer to be able to play prerecorded eye movement data that are read from disk during the "performance." Further, the development of a musical instrument or composition designed for *live* performance with eye movements introduces fundamentally different challenges, such as classic human factors issues pertaining to controls and feedback. Initial EyeMusic compositions instead focused on sonification, and read eye movement data from disk using the eyedata2 external discussed in the previous section.

We created several compositions using the EyeMusic system reading data from disk. In this mode, the system reads from a file that contains the (*x, y*) coordinates of the gaze positions recorded in an earlier session in which the composer viewed a scene, perhaps "preperforming" the piece. To compose a piece in which the music responds to where the eyes were looking in the scene, a visual recording of how the eyes moved through a visual scene is also needed. Without such a record, it is not possible for the composer to plan or the audience to observe a correlation between the eye movements and the sounds created by those movements. For example, the composition could be designed to play a bass note when the gaze lands on a blue blob, but both the composer and the audience must be able to see the gaze land on the blue blob. To solve this problem, when collecting the (*x, y*) data, we also record a video that captures the contents of the display that the performer is looking at, and superimpose the gazepoint on the scene using a hardware video mixer and the video output from the eye tracker (not shown in Figure 1). The video with the superimposed gazepoint can be played back within Max/MSP by synchronizing it with the corresponding (*x, y*) data.

The *Reading/Typing* Composition

One early compositional sketch sonifies the eye movements that a person makes when reading text and is entitled *Reading/Typing.* We recorded the eye movements of a performer as he read a passage of text displayed on the screen. To create a certain self-referential aspect to the piece, the text passage is from Rayner and Pollatsek's *A Psychology of Reading* [7] and describes the characteristics of eye movements that people make, without realizing it, when reading. In addition to collecting the eye movements, a video was also made that superimposed the gazepoint on the reading material on the page.

Figure 5 shows a screenshot from *Reading/Typing.* The eye movement data collected for Reading/Typing are sonified as follows. Every time a fixation occurs, a typewriter keypunch sound is played. Every time the eyes move back to the start of a new line, an old-fashioned typewriter carriage return and bell sound is played. The fixations sound remarkably like typing because the fixations occur at the same rate that a slow typist would press the keys, about four per second, and because of the slightly varied rhythm resembles that of typing. The sounds are played at the same time that the videotape is played back, and you can both see and "hear" the performer read the text. Reading/Typing is a rhythmic sonification of reading sonified as typing. An audio-video recording of "Reading/Typing" is available on the EyeMusic web site.

The *Viewing Kandinsky* Composition

We next set out to produce a more musical playback of eye movement data, and produced *Viewing Kandinsky*, also known as *EyeMusic v.0.9b.* For this composition, the composer watched a series of Kandinsky paintings that were slowly zoomed and panned across the video display. Data collected included the (*x, y*) coordinates of the gaze samples, and a video of the fixation point superimposed on the Kandinsky paintings.

The sonification of gaze data in *EyeMusic v.0.9b* are as follows. The eye movement data are output from the eyedata object at the usual rate of 60Hz, and also selectively sampled at slower rates within Max/MSP. The selective sample rate is roughly once every 500 ms, but for musical purposes is varied slightly during the piece and ritards at the end. *EyeMusic v.0.9b* produces a primary melody and a pointillistic counterpoint. The primary

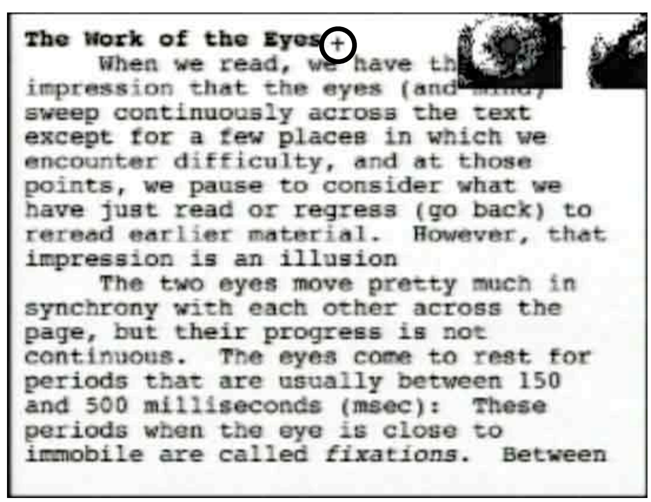

Figure 5. A screenshot from the *Reading/Typing* composition, with the current gaze position circled.

melody is derived from the horizontal dimension of the gaze location, with the pitch of the melody increasing as the gaze moves from left to right, as on a piano keyboard. The counterpoint is derived from the vertical dimension, with the slowed-down samples cycling through eight different MIDI (Musical Instrument Digital Interface) instruments, and each sound triggered roughly once every four seconds. The vertical dimension is also mapped to pitch, with higher gazepoints triggering higher pitches. The resulting music, influenced in important ways by the Max processing, displays clear parallels with the original data and produces a mysterious and lyrical ambiance.

EyeMusic v.0.9b is a little under six minutes long, and was presented at the Future Music Oregon concert on November 15, 2003, in Eugene, Oregon. During the presentation, the video of the gazepoint superimposed on the paintings was shown, and Max/MSP processed the eye movement data in real time during the presentation. An audio-video recording of the piece can be viewed on the EyeMusic web site.

COMPOSITIONS EMPHASIZING LIVE PERFORMANCE

The most compelling way to have eye movement data influence a musical performance is not through prerecorded eye movement data, as discussed in the previous section, but instead by enabling the performer to directly control the music in real time with their eyes. This would have the greatest potential of bringing to life the musical properties of the eyes. Once some prerecorded sonifications were completed, we explored how the eye tracker might be used for real-time eye-interactive multimedia performance. This requires both the creation and the performance of the multimedia work. It also involves an interactive process of determining how the eyes can interact with the multimedia environment, a process akin to designing a new musical instrument.

Figure 6 shows how the EyeMusic system is configured for live performance. The diagram is derived from Figure 1, but emphasizing the control and feedback loop. The video image that the performer watches during the performance is no longer a prerendered video, but is instead created by the Max/MSP multimedia system, which is now extended with the Jitter object library for real-time manipulation of graphics and video. A feedback and control loop has been created, such that the performer can now visually interact with objects on the screen, such as by looking at objects and having the objects change based on the gaze. The auditory and visual feedback loop creates an experience analogous to playing a physical instrument. The performer's video monitor is also shown to the audience (though flipped horizontally) so that the audience can observe the visual elements of the performance, and see how the performer is interacting with the display. The image needs to be flipped horizontally so that the performer's physical gaze corresponds to gazepoint projected on the multimedia video display shown to the audience; that is, when the gaze is to the performer's left, it interacts with visual elements on the audience's right. The output from the camera that is used by the eye tracker is run through a video splitter so that the audience can view, in a separate video monitor, the same image of the eye that is used by the eye tracker. This provides a close up view of what the eye is doing, and helps to communicate the performer's physical movements and gestures which are much smaller than, say, that of a violinist.

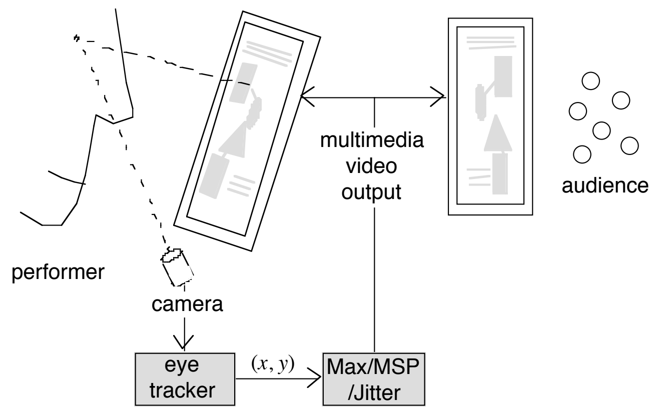

Figure 6. The EyeMusic system configured for live performance, with the performer directly interacting with visual objects to trigger visual and musical responses.

The Max/MSP/Jitter composition is created in advance of the performance, and this preparation is analogous to both the construction of a musical instrument and the composition of a musical score. The next subsections discuss compositions created to explore the real-time expressive potential of the EyeMusic system, and our discoveries of how eye movements can be readily used for real-time creative multimedia expression. Our design exploration included the creation of multiple case studies, including an eye-controlled piano.

Eye-Controlled Piano

To explore one possible means of playing music with the eyes, we created an eye-controlled piano. The eye piano functions analogously to eye typing, a well-established communication technique in which a key is pressed by holding an eye-controlled cursor on that key [5]. In our eye piano, a piano keyboard replaces the typing keyboard. Figure 7 shows part of the Max/MSP/Jitter interactive composition that we wrote to create an eye piano. When actually played, however, the octave of piano keys filled the entire computer display to give the performer larger targets and thus greater control and ability to land on the desired key with a single eye movement. Small dots were also placed on the keys as peripherally-salient targets for eye movements. The eye piano played a note, sonically and visually, when a fixation was detected on the key, using the fixation external.

An eye-controlled piano based on a dispersion-based fixation-detection does not work well. The second author, an experienced analog and computer musician, practiced playing the fixation-based eye-controlled piano for roughly an hour a day for two weeks, and intermittently for another month after that, and reported that (a) he was only slightly

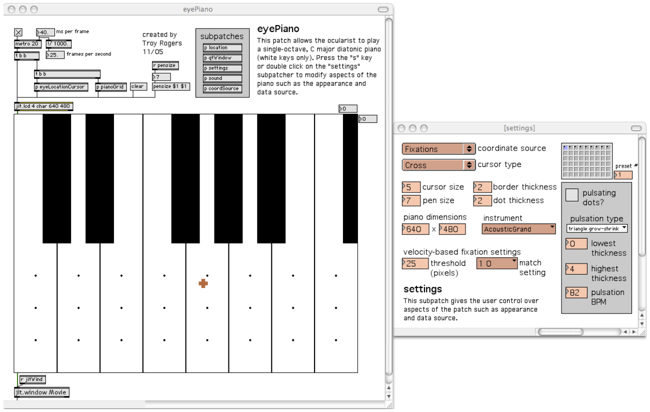

Figure 7. The front end of an eye-controlled piano in Max/MSP.

able to improve his ability to move to the intended piano keys and (b) he was not at all able to improve his rhythmic accuracy. He specifically worked on trying to improve his and the instrument's ability to play a steady beat, such as by practicing alternating between two keys, but with no improvement.

It might seem, in hindsight, obvious that an eye-controlled piano is not a good idea. But it is important to report that an it did not work. Numerous interface designers, computer musicians, and researchers (familiar with the success of eye typing) suggested that we create an eye-controlled piano. Based on the characteristics of eye movements (roughly four per second) and the fact that there is only one gazepoint as opposed to many fingers, we did not expect the piano to work very well, but we did expect it to help us to better understand the capabilities and limitations of the eyes in a musical context. This exploration contributed to the design and discovery of the new idiomatic interface that was ultimately constructed.

Rhythmic Control

Since rhythmic control is critical in musical performance, and since it could not be achieved with an eye piano using a dispersion-based fixation-detection, subsequent instrument design was made working almost exclusively with the "raw" gazepoint data that arrives in Max/MSP/Jitter every 17ms. This created an interesting tension between the eye tracking researchers and the electronic musicians collaborating on the project. At least one of the eye tracking researchers insisted that fixations (when the gaze stays at the same place for roughly 100 to 400 ms) are the relevant psychophysical phenomena, and so the composition should be designed around the fixations, not the raw data from the eye tracker, which is merely an artifact of the instrumentation. The computer musicians, accustomed to working with the noise and idiosyncrasies associated with a wide range of sensors (accelerometers, infrared, etc.), preferred to work with the data from the sensor in its most native form, in this case the (*x, y*) gazepoint samples supplied every 17 ms. Over time, the musicians successfully demonstrated how using this raw eye movement data not only gave the performer more direct control and expressive opportunity with their instrument and the performance, but also permitted more interesting and varied compositions.

By starting with the raw sample data rather than the fixation data, we further determined that, despite its initial failure, the notion of an eye-controlled piano should not be entirely dismissed. There were two ways that it did work somewhat well. First, when the piano was modified so that a key would play when a single sample landed on it, rhythmic control improved, though at the expense of melody accuracy since every movement between piano keys produced a glissando. Second, a discovery we made is that an eye-controlled piano that uses a velocity-based (as opposed to dispersion-based) fixation-detection algorithm works better for rhythmic control with the eyes. When we tried this, rhythms could be performed better. This makes perfect sense in retrospect because the velocity-based algorithm can more quickly indicate the start of a fixation. The discovery came after the following compositions were created, however, and so the remaining compositions work primarily with the raw gazepoint data, which is what the computer music composers prefer to work with anyway.

Visually Interacting with Dynamic Objects

We started our exploration of gaze-based control with the established paradigm in which the eyes look at a fixed location to actuate a control or, in other words, you just look at a big button to click it. This is basically trying to use the eyes to control devices that are designed for the fingers. Our exploration eventually turned to where the gazepoint was considered to be just one of many objects in the display, such that each object has its own physical and dynamic characteristics, as does the gazepoint. But that the gazepoint was a special object in that other objects would interact with it differently. This explores other expressive control capabilities that might be better suited for the eyes. Now the performer moves an eye-controlled cursor around on the screen, and makes the cursor come into direct visual contact with other visual objects on the screen, producing a visual and sonic reaction. We also sonified the eye cursor in a variety of ways so that it continually emits sound based on its position and movement. These interactions produced a variety of feedback loops, as with analog musical instruments, and a loop that is tighter and more responsive than the eye-controlled piano.

The *Swarming Dots* Composition

The most recent major milestone of the project has been numerous live performances of the *Swarming Dots* or *EyeMusic v1.0* composition, which is an eye-controlled musical and interactive visual composition in which the performer visually interacts with dynamic objects. Storyboarding was used to help design the content and flow of time-based, multimodal, musical compositions. Storyboarding is an established means for describing human-computer interactive behavior [11]. Though this design technique was new to the computer music composers, they quickly saw its value in collaborating on the design of an interactive composition. Figure 8 shows an excerpt from the storyboard, which was augmented with a vocal track that sonically sketched out some basic ideas for what the various interactions might sound like. Perhaps the most important point of the storyboards is that an eye-controlled multimedia composition needs to be designed as an interactive experience, and not as instrument that waits idly for the performer to take control and play an independently scripted (such as memorized) composition. In other words, until storyboarding was introduced, the composition stood somewhat still, with a single eye-controlled interaction. Storyboarding pushed the composition through a series of interaction sequences, very roughly analogous to pages of sheet music.

One of the goals of the composition and staging was to ensure that the audience could easily and directly understand that the composition was performed with eye movements. This was accomplished by using extremely clear and deliberate visual effects, and with great consideration of the physical staging. Figure 9 shows a rough approximation of the assembly of stage elements. Showing the video image of the eye that was used by the eye tracker, for example, helped a lot. This was inspired by what we learned in our accessibility research, in which the video image helps users stay in range of the camera [2]. Directly asking the audience what they just saw following an early performance of EyeMusic 1.0 demonstrated that

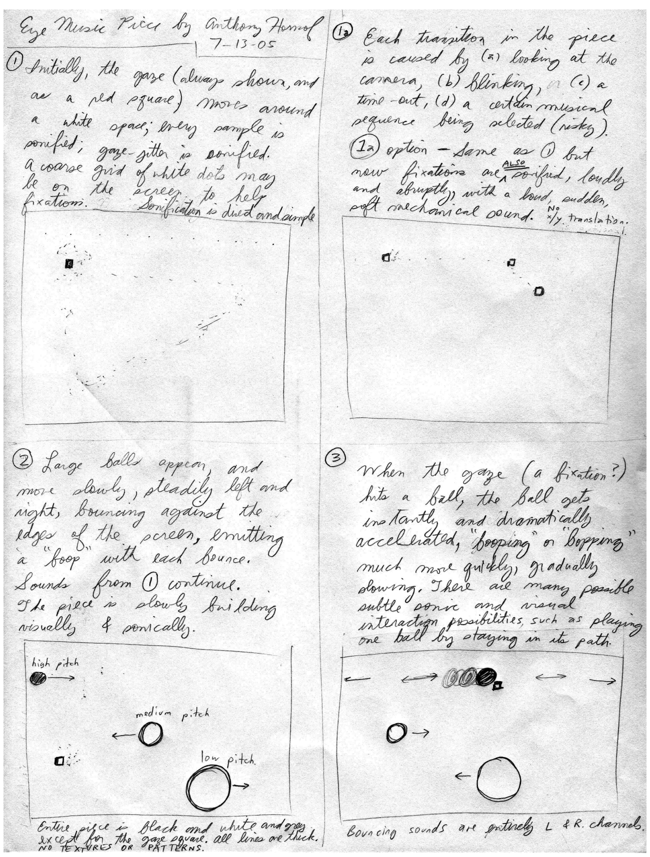

Figure 8. This storyboard is an example of how even rough sketches can successfully convey the time-based elements of an interactive experience. An off-the-cuff recording of vocal sound effects accompanied the drawings, further providing a concrete starting point for interactive design.

the audience was clearly able to understand that the eyes were controlling the performance.

EyeMusic v1.0 is roughly five minutes long, and starts with the opening of the eyes which fills the small monitor with the eye image captured by the camera, and the appearance of a single gazepoint eye cursor appearing on the rear-projected screen. The eye cursor is a large white dot, drawn on a black background. Throughout the performance, quiet, scratchy, jittery noises play near the gazepoint, using stereo panning to move the sound from left to right as the eyes moved left and right. The performer moves through segments of the composition by blinking his or her eyes. Each blink is sonified as sudden, loud, slamming sound. Sometimes the eyes stay closed, during which the eye-closing sound decays, and the rear-projection is blank.

As the composition progresses, red circles appear on the display. The performer moves the eye cursor close to, but does not touch, the circles. The near misses create some tension. Eventually, the eye cursor touches a circle, which activates the circle to bounce between the left and right edges of the display, slowly decelerating over time. The performer "plays" the circles like an instrument, first tentatively and in a somewhat subdued manner with occasional glances, and eventually in a frenetic, over-energized manner such as by keeping the gaze directly in the left-right path of a circle, insuring continual re-activation. Each bounce produces a rubbery bouncing sound, with smaller circles creating higher-pitched sounds.

A loud blink replaces the bouncing balls with a cartoonish eye in the center of the screen. The eye mimics, with a simple graphic, the image of the eye that is used by the eye tracker and shown to the audience. The eye cursor moves around the screen and plays with a swarm of red dots that

Figure 9. The first author practicing *EyeMusic v1.0* before performing the piece at a concert at NIME 2007.

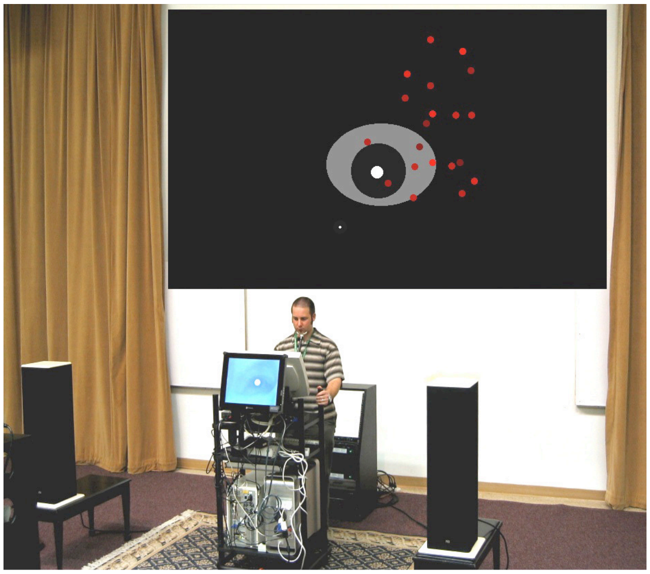

Figure 10. The second author practicing EyeMusic v1.0 before presenting the piece at a concert at SEAMUS 2006. The performer's video is also projected above and behind him during the performance, shown here with a photo inset.

follow the *boids* model of bird flocking [8]. The screenshot in Figure 10 illustrates the cartoonish eye and swarm of dots. The swarm initially avoids but is eventually attracted to the eye cursor, like moths around a light bulb. A small blip sound is played when the eye cursor touches a red dot. To the performer, it feels a little as if the swarm of dots is tickling the eye. The entire piece is punctuated with the loud, sudden blinks, which dramatically increase in frequency near the end of the performance.

EyeMusic v1.0 was selected by juries and performed live at the annual meeting of SEAMUS 2006 (the Society for Electro Acoustic Music in the United States) and NIME 2007 (New Interfaces for Musical Expression). A video of EyeMusic v1.0 is available on the EyeMusic website.

LESSONS LEARNED

The project was immediately compelling to the lead computer musician collaborator who, in his creative practice, is continually looking for interesting new data streams and sensors, because eye tracking provides the opportunity to extend something so essential to the human condition that could be at least interesting, and at best very expressive and fascinating. Working together, the collaboration provides an opportunity to create hybrid compositions that don't exist in a nonhybrid world. Much of the collaboration time was spent translating between disciplines. For example, we had numerous extensive discussions regarding how eye tracking is used for analytic purposes in the sciences and how fixations are the first-order data of interest to the scientists. Though the computer musicians were eager to assist the scientists in developing sonifications of eye movement data that would be useful for scientific analysis, a consensus developed among all that the compositions developed for live performance ultimately needed to prioritize aesthetic issues over scientific goals, at least for the pieces to contribute to and be taken seriously in the world of computer music. It is difficult to serve two masters, and thus far EyeMusic has leaned towards the aesthetic.

The roles of performer, composer, musician, artist, and scientist became blurred. Several hundred hours of computer programming were done in Max/MSP, mostly by a computer musician. The NIME 2007 performance of EyeMusic v1.0 was done by one of the computer scientists, with coaching from an experienced performer. Important advice included, as in the musical tradition, to practice, practice, practice, though in this case the practice also included the unpacking, assembly, disassembly, and packing of the instrument. The distinction between composition and instrument also becomes blurred in computer music.

Two of the problems of eye tracking that continue to keep it from its great promise are issues pertaining to noise and control. Working with computer musicians who routinely deal with these issues with a wide range of sensor technology provides a refreshing perspective. The musicians readily embrace the noise and sometimes even monopolize it as a feature, as did Jimi Hendrix with feedback and distortion on the electric guitar. When working with eye tracking data, rather than trying to remove all the noise, whether it comes from the eye tremor or the eye tracker, the musicians embrace and want to work with the noise.

Another extreme form of noise that sometimes occurs is the temporary loss of the gaze position on the screen. Future compositions could even incorporate this "noise" by means of corrective control measures built into the composition. For example, when the gaze is lost during a performance, such as when the eye cursor jumps to seemingly random locations on the screen, a performer familiar with the functioning of the eye tracker might want to see the eye tracker system status output to try to address error in real time such as with small head movement. System feedback that might used to fix the problem in real time could be incorporated into the composition itself, such that both the performer *and* audience see the eye tracker system status while the adjustments are made. This would be somewhat akin to seeing a performer make a small tuning adjustment or even a small hand movement to fine tune a string or note while an instrument is being played. The composisition's transition to this adjustment mode could even be made automatically, triggered by the composition itself detecting the error.

Regarding control, the musicians are comfortable with creating a composition that cannot necessarily be played the same way every time. In his closing plenary talk at CHI 2005, Michel Waisvisz, the director of the Studio for Electro-Instrumental Music (STEIM), explained that, when designing electronic and computer musical instruments, the goal is not always control. Some materials such as a violin string have a life of their own, and the goal is to bring those materials to life.

CONCLUSION

This paper discusses the exploration of new interaction techniques for real-time eye-controlled music and multimedia performances. Though designing any human-computer interface requires attention to the intended context of use, the user's expectations, and the intended outcome of the interaction, these requirements are perhaps even greater for eye-controlled interfaces. The basic input-output and human-computer control mechanisms that work at the 100 to 500 ms grain size must be redesigned for each task. For example, the fixation-detection algorithm generally used for eye-typing does not work well for eye-piano-playing because it does not provide good rhythmic control. Further, some people with severe motor impairments have deliberate control of their gaze but not their blinks, and so for these performers the blink control in EyeMusic v1.0 would need to be replaced, such as with an off-screen glance. EyeMusic continues as an exploration into bringing to life the musical properties of the eyes.

EyeMusic will continue to be developed on a number of different fronts. We will continue to explore possibilities for live performances, to see how a musician can control a musical passage with his or her eyes. We will also explore the sonification of eye movements for data analysis purposes. Eye movement analysis for scientific purposes is difficult and time-consuming. Sonification of eye movement data will likely enhance current techniques. There are characteristics of human audition, such as a slower decay in working memory, that may make sonification of eye movements useful for displaying certain data trends. Lastly, we will explore opportunities for EyeMusic to open doors to musical composition and performance for people with severe mobility impairments, who interact with the world via eye tracking. We have already developed software that enables disabled children to draw with their eyes [2]. Making music with the eyes would follow nicely.

ACKNOWLEDGMENTS

Ying Tan provided invaluable aesthetic critique. Erik Brown, Rachel Nehmer, and Linda Sato programmed the Max/MSP externals. The project was supported in part by the National Science Foundation (NSF) through Grants IIS-0308244 and IIS-0553117. The eye tracking lab used for these projects is supported by the Office of Naval Research (ONR) through Grants N00014-02-10440 and N00014-06-10054. All grants were made to the University of Oregon with Anthony Hornof as the principal investigator. Any opinions, findings, conclusions, or recommendations expressed in these materials are those of the authors and do not necessarily reflect the views of NSF or ONR.

REFERENCES

1. Findlay, J. M. and I. D. Gilchrist. *Active Vision: The Psychology of Looking and Seeing*. Oxford University Press, New York, 2003.
2. Hornof, A. J. and Cavender, A. EyeDraw: Enabling children with severe motor impairments to draw with their eyes. In *Proc. of CHI 2005*, ACM Press (2005), 161-170.
3. Hornof, A., & Sato, L. EyeMusic: Making Music with the Eyes. In *Proc. NIME04*, Shizuoka University of Art and Culture (2004), 185-188.
4. Jacob, R. J. K. and Karn, K. S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyona, R. Radach, & H. Deubel (Eds.) *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*. Elsevier Science, Amsterdam, 573-605, 2003.
5. Majaranta, P. & Räihä, K-J. Twenty years of eye typing: Systems and design issues. In *Proc. ETRA 2002*, ACM Press (2002), 15–22.
6. Polli, A. Active vision: Controlling sound with eye movements. *Leonardo* 32, 5 (1999), 405-411.
7. Rayner, K., & Pollatsek, A. *The Psychology of Reading*. Prentice Hall, Upper Saddle River, NJ, 1989.
8. Reynolds, C. Flocks, herds and schools: A distributed behavioural model. *Computer Graphics* 21, 4 (1987), 25-34.
9. Roads, C. *The Computer Music Tutorial.* MIT Press, Cambridge, MA, 1996.
10. Rosenbaum, D. A. *Human Motor Control*. Academic Press, New York, 1991.
11. Rosson, M. B., & Carroll, J. M. *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*. Morgan Kaufmann, San Francisco, 2002.
12. Salvucci, D. D. and Goldberg, J. H. Identifying fixations and saccades in eye-tracking protocol. In *Proc. of ETRA 2000*, ACM Press (2000), 71-78.
13. Wardrip-Fruin, N. & Montfort, N., Eds.. *The New Media Reader*. MIT Press, Cambridge, MA, 2003.