EyeMusic: Making Music with the Eyes

(Submitted to New Interfaces for Musical Expression, NIME'04. Please do not cite or distribute.)

Anthony Hornof and Linda Sato
Department of Computer and Information Science
University of Oregon
Eugene, OR 97403 USA
{hornof, lsato}@cs.uoregon.edu

ABSTRACT
Though musical performers routinely use eye movements to communicate with each other during musical performances, very few performers or composers have utilized eye tracking devices as a means of influencing musical compositions and performances. EyeMusic is a system that uses eye movements as an input to electronic music compositions. The eye movements can control the music, or the music can respond to the eyes moving around a visual scene. EyeMusic is implemented so that any composer using established composition software can incorporate prerecorded eye movement data into their musical compositions.

KEYWORDS
Electronic music composition, eye movements, eye tracking, human-computer interaction, Max/MSP.

1. INTRODUCTION
An eye tracker is a device that reports where a person is looking. Eye trackers usually incorporate a camera that sits next to a computer screen and is focused on the eyes of the person using the computer. The video images are transformed, via software algorithms, into the (x, y) coordinates of the location on the screen where the user is looking. Eye trackers generally report eye movements within 100 ms.

Eye movement data are useful to a variety of disciplines. Cognitive psychologists study eye movement data to understand basic human physiology. Human factors practitioners employ eye tracking to understand how people interact with specific devices and to thus improve the usability of those devices. Accessibility researchers write software that enables severely disabled people to communicate by controlling the computer and typing with their eye movements. Interface designers explore how computers can indirectly assist a user based on their eye movements, such as by determining when they are interruptible. Jacob and Karn [1] provide a good overview eye tracking research.

This report discusses EyeMusic, a system that transforms eye movement data from an eye tracker into musical compositions and data sonifications. The system will enable musicians, composers, and performers to directly control a composition as it is being performed—by just moving their eyes.

2. HOW THE EYES MOVE
To understand how EyeMusic works, a few terms pertaining to eye movements and eye tracking must be defined. The *gaze* is the vector that goes from the eye to the *gazepoint*, which is the point in a scene where a person is looking. The eyes (and thus the gaze) move around a static 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. A *dwell* is a long fixation. People acquire visual information during the fixations, not saccades. The reason that the eyes move, in short, is so that people can put items of interest into the high resolution vision which is at the center of their gaze.

Individual eye movements are usually decided based on a higher-level strategy that is used to accomplish a visual task, such as reading, and the eye movements are programmed subconsciously. However, people can also make deliberate, conscious decisions to shift from higher-level strategy execution to program individual eye movements that will move their eyes to or from a specific location, such as to fixate this *X* for two seconds and to then pass the control back to the higher-level reading strategy.

3. WHY MAKE MUSIC WITH EYE MOVEMENTS
There are a number of reasons that eye movements are useful and interesting for musical composition: (a) Eye movements are made based on a combination of conscious and deliberate programming, creating an interesting opportunity to make music while engaged in a different task, while sometimes taking direct control. (b) There is an inherent musical or at least rhythmic quality to eye movements that lends itself to composition. (c) Eye movements offer a new modality for interacting with a musical composition, both for regular nondisabled musicians and performers, as well as for disabled users who already interact with the world primarily by just moving their eyes; the project opens the door to new musical control, and to musical expression by people with severe disabilities. (d) From the perspective of scientists who analyze eye movement data to understand patterns of human visual processing, there may be some obscure patterns in the data that become most salient when the data are sonified rather than just visualized.

4. PREVIOUS MUSIC WITH THE EYES

We are aware of only one body of previous work in which eye movements direct musical compositions, work by the digital artist Andrea Polli [2]. Her musical composition with eye tracking entitled *Intuitive Ocusonics* has been performed internationally. Excerpts are online at www.andreapolli.com. She has also recorded a CD entitled *Active Vision: Musical Improvisation Using Eye Movement to Create Complex Sound Patterns*, available online at www.zucasa.com/rfzc/albums/activevision.html.

In Polli's work, the eye directly controls aspects of the composition as it unfolds. The musical compositions are striking, filled with haunting electronic sounds as well as digital samples of the human voice, sometimes singing and sometimes screeching. Sometimes the sounds ease in over time; other times they strike suddenly. The compositions tend to be somewhat sparse, with just a few instruments, sounds, or voices playing at a time.

There are three ways that our project differs from the work of Polli. First, our system uses more accurate eye movement data, resulting in better control of composition and performance. Polli's compositions respond to the video image of the eye rather than the screen coordinates provided by a commercial eye tracker. Her compositions use Steim's BigEye software to parse and process video frames of the eye at a rate of 12 frames per second, without the benefit of specialized algorithms for converting video images into screen coordinates. EyeMusic currently uses the L.C. Technologies Eyegaze System (eyegaze.com), the core technology of which is a patented algorithm that accurately converts video images of the eyes into screen coordinates sixty times a second.

The second way that our approach differs from that of Polli is that EyeMusic identifies fixations, which is the relevant physiological phenomenon. Resulting compositions and data sonifications will relate more clearly to the human processes that are at work.

The third departure is that EyeMusic puts more emphasis on the scene that is viewed during the performance, and how the eyes move around the scene, and less emphasis on the image of the eye moving. Polli's live performances include a large video image of the eye as it moves. Our performances include a video image of the gazepoint superimposed on the scene viewed by the performer. Professor Polli is currently collaborating with us to compose using EyeMusic.

5. HOW EYEMUSIC WORKS

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*, 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 from the video display. The coordinates (as well as the pupil radius) are made available in Max/MSP in the Macintosh computer. The coordinates are currently delivered via file transfer. Ultimately, each gaze sample will be transmitted in real time.

EyeMusic makes the eye movement data available within Cycling 74's Max/MSP, a graphical environment for creating music, audio, and multimedia that it is one of the most widely used software applications for electronic music composition. Max is typically used in conjunction with the Max Signal Processor (MSP) and thus referred to as Max/MSP.

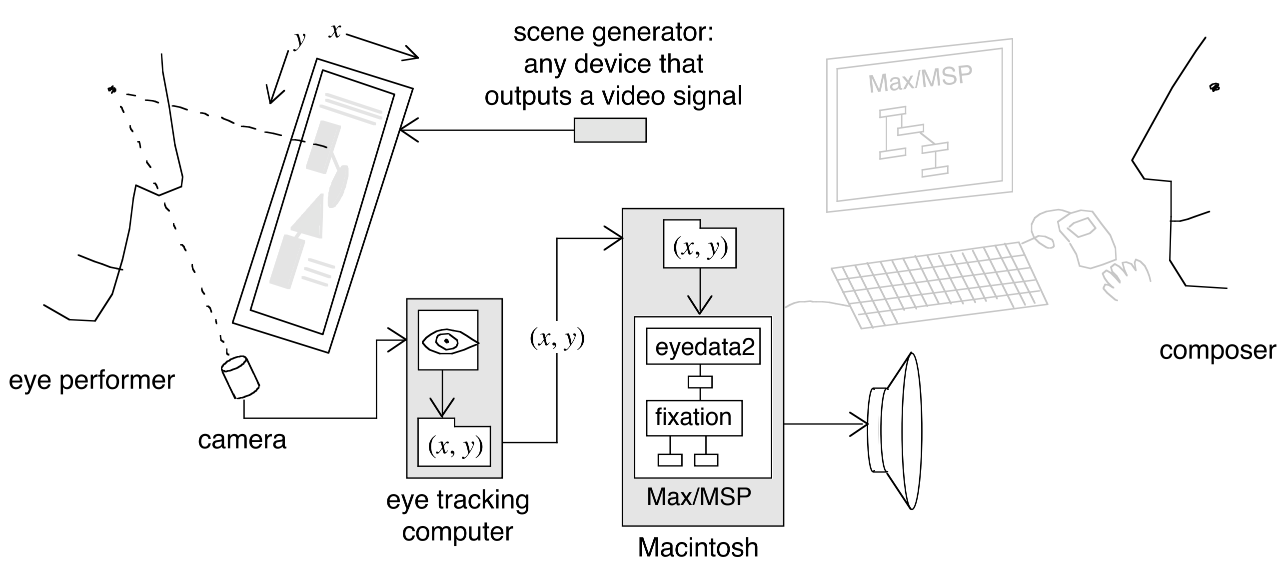

Figure 1. An overview of the major components in the EyeMusic system. Arrows indicate the flow of data.

The eye movement data are made available in Max/MSP by means of the *eyedata2* external and parsed into fixation data by means of *fixation* external, both of which were written by the authors and are discussed below in greater detail.

Alternative configurations to those shown in Figure 1 are possible. For example, the eye performer and the composer could be the same person, and the Macintosh could be used as the scene generator, with the scene changing as a function of where the eye performer looks.

EyeMusic is designed to work in two different modes: *playback* and *performance*. In playback mode, prerecorded eye movement data are read from disk into Max/MSP. In performance mode, the eye movement data are reported in real time, and the performer plays the music by just moving their eyes. Playback mode is fully implemented. Performance mode is currently being developed.

5.1. The *Eyedata2* External in Max/MSP

The basic building blocks in Max/MSP are data processing elements called *objects*. End-users can add customized objects called *externals* using the C programming language. In EyeMusic, eye movement data are read into the Max environment by means of an external called *eyedata2*. Figure 2 shows the eyedata2 external in use.

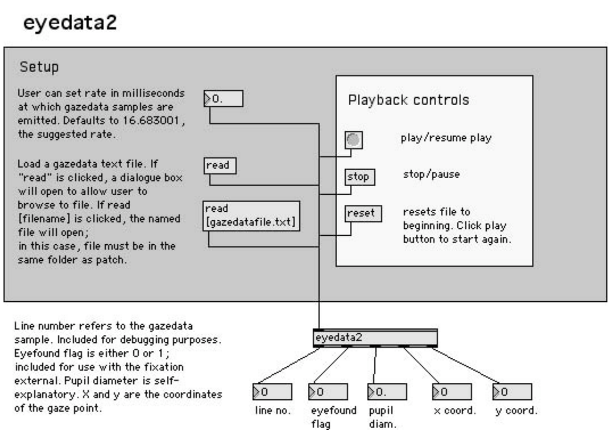

Figure 2. The eyedata2 external in Max/MSP

The eyedata2 external reads the eye movement data from disk, and sends it out of the external into any Max object that is connected to its outputs. Table 1 shows the form of the data that are output from the external: *Sample #* increments 60 times a second. *Eye found?* indicates whether the eye is successfully tracked for that sample ("1") or whether, for example, the eye was moved out of the camera range ("0"). The *x* and *y* are the screen coordinates of the gaze, with *(0, 0)* at the top left of the screen. *Pupil radius* is reported in mm. The samples shown in Table 1 were collected while a person was reading. The eyes finish a fixation at roughly *(553, 112)*, and then makes a horizontal saccade to the right, ending at roughly *(663, 113)*.

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
395	1	662	112	1.3
396	1	663	113	1.3
397	1	664	113	1.31
398	1	662	113	1.32
...	...	...	...	...

Table 1. Sample data from the eyedata2 external.

5.2. Video Playback

In playback mode, the *eyedata2* external outputs the recorded gaze position as (*x*, *y*) coordinates. This is adequate to play the musical composition but not for the composer or the audience to observe a correlation between where the person was looking during the eye tracking session and the sounds that are created by these eye movements. For example, the composition might be designed to play a bass note when the gaze lands on a red blob, but there will be no way for the audience to see this correlation. To address this problem, we record a video during the eye tracking session that shows the eye-performer's gazepoint as a set of crosshairs superimposed on the scene that he or she is viewing. (The video mixer and recorder are not shown in Figure 1.) The video is then played back within Max/MSP simultaneously with the data output from eyedata2. The composer and audience members could thus see and hear that the bass note is played when the gaze moves to the red blob.

5.3. The *Fixation* External in Max/MSP

The fact that the data flow out of the eyedata external at a rate of sixty samples per second is simply an artifact of the eye tracker used here. The primary relevant human physiological phenomenon is not this sample rate, but rather where the fixations occur. The *fixation* external parses the eye movement sample data, identifies where the fixations occur, and outputs each fixation and its location. This way, a composer can work more directly with the relevant human behavior. Figure 3 shows the fixation external in use.

The fixation external uses an established *dispersion-based* algorithm [3]. The algorithm is based on the fact that eye positions sampled during a fixation tend to cluster in a small region for a minimum amount of time. There are two parameters that must be set in the algorithm: (a) the *deviation threshold*, which is the size of the region in which the fixations must cluster to be identified as as fixation, and (b) the *duration*, which is the minimum amount of time that samples must be recorded in a cluster for a fixation to be reported. The fixation external defaults to commonly-used settings for both parameters: (a) a deviation threshold of 20 pixels, which corresponds to 0.5° of visual angle at normal viewing distances, and (b) a duration of 100 ms, which corresponds to 6 samples (60 per second). Both of these parameters can be modified in the composition permitting, for example, one fixation object to respond 100 ms fixations, and another to 500 ms fixations.

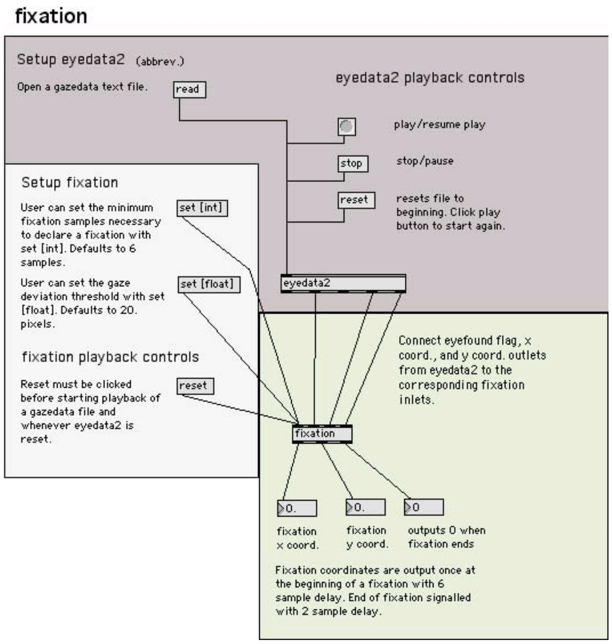

Figure 3. The fixation external in Max/MSP

6. RECORDINGS AND PERFORMANCES

Jeffrey Stolet, our collaborator in the School of Music, created an electronic composition using EyeMusic entitled "EyeMusic v. 0.9b" that was performed (in playback mode) at a Future Music Oregon concert on November 15, 2003. The composition uses the eyedata external to play back Stolet's eye movements recorded as he studied Kandinksy paintings during an eye movement recording session. A video of the fixation point superimposed on the stimuli was also played during the performance. An audio-video recording of "EyeMusic v. 0.9b" can be viewed on the EyeMusic web site, at http://www.cs.uoregon.edu/research/cm-hci/EyeMusic/.

In "EyeMusic v. 0.9b," the eye movement data output from the eyedata object (at a rate of sixty samples per second) were selectively sampled at slower rates in Max to produce a primary melody and (derived from the horizontal dimension), and a pointillistic counterpoint (derived from the vertical dimension). The resulting music, influenced in important ways by the Max processing, displayed clear parallels with the original data and produced a mysterious and lyrical ambiance.

The fixation external was developed after "EyeMusic v. 0.9b" was composed. To demonstrate how the external works, the authors composed a simple composition entitled "Reading/Typing." The composition is based on the fixations made while reading a text that describes the fixations that are typically made while reading a text. The text appears in a typewriter font. Every time a fixation occurs, a typewriter keypunch sound is played. Every time the eyes move back to the start of a new line, a old-fashioned typewriter bell and carriage return sound are played. The fixations do sound remarkably like typing because fixations occur at roughly the same rate that keys are pressed while tying, about four per second, and because the varied rhythm of the fixations also resembles that of typing. An audio-video recording of "Reading/Typing" is also available on the EyeMusic web site.

EyeMusic is available for download at http://www.cs.uoregon.edu/research/cm-hci/EyeMusic/.

7. FUTURE WORK

EyeMusic will continue to be developed on a number of different fronts. Live performance mode is currently being implemented. Once it is completed, we will user-test the system to see how a musician can control a musical passage with his or her eyes. We will continue to collaborate with Jeffrey Stolet and Andrea Polli to develop new and innovative musical compositions.

We will explore a range of possibilities to sonify eye movements for data analysis purposes. There are characteristics of human audition, such as a slower decay in working memory, that may make sonifications of eye movements more 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 communicate using an eye tracker.

ACKNOWLEDGMENTS

The authors thank Jeffrey Stolet for creative collaboration and technical support and Tim Halverson for assistance with the eye tracking. This work is supported by the National Science Foundation under Grant No. IIS-0308244. The eye tracking lab used for these projects is supported by the Office of Naval Research through Grant N00014-02-10440.

REFERENCES

1. Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises (Section commentary). In J. Hyona, R. Radach, & H. Deubel (Eds.), *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*. Oxford: Elsevier Science.
2. Polli, A. (1999). Active vision: Controlling sound with eye movements. *Leonardo*, 32(5), 405-411.
3. Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocol. *Proceedings of the Eye Tracking Research and Applications Symposium,* New York: ACM Press, 71-78.