Aktives Sehen: Kunstwerke unter Verwendung von Blickverfolgung und -analyse

Active Vision: Artworks Inspired by Vision using Video Tracking and Analysis

Andrea Polli
Hunter College, Film and Media
695 Park Ave.
New York, NY 10021
212.772.5589
apolli@hunter.cuny.edu
www.andreapolli.com
filmmedia.hunter.cuny.edu

Zusammenfassung:
Der vorliegende Beitrag setzt sich mit zwei künstlerischen Projekten auseinander, die durch das Sehen inspiriert wurden und Techniken von Blickverfolgung und –analyse nutzen. Das erste, ausführlich vorgestellte, Projekt „Intuitive Ocusonics" ist ein System, das für Liveperformances durch die Auswertung der Augenbewegungen konzipiert wurde. Über 50 Aufführungen und Präsentationen dieses Systems sind in den USA, Europa und Südamerika innerhalb der letzten 5 Jahre gezeigt worden. Das zweite Projekt, „Das Fliegenauge", Neuentwicklung des Jahres 2002, ist ein System, das für öffentliche Installation und Interaktion entwickelt wurde, inspiriert durch die Struktur eines Fliegenauges.

Abstract:
This paper discusses two artistic projects inspired by vision using techniques of video tracking and analysis. The first project, discussed in depth, is called Intuitive Ocusonics, is a system designed for live performance using eye movements. Over 50 performances and presentations with this system have been seen throughout the US, Europe, and in South America over the past five years. The second project, The Fly's Eye, newly developed in 2002, is a system designed for public installation and interaction inspired by popular notions of the structure of the eye of the fly.

In 1995, I traveled to Vilnius, Lithuania to work with a group of artists on an installation project in a 17th century Observatory on the campus of Vilnius University. One of the collaborators, a sculptor whose work involved transforming abandoned buildings, had fallen five stories from the roof of one of those buildings about one year before our meeting. Rasa had suffered a severe spinal cord injury and was unable to walk and control her movements fully. Meeting Rasa and talking with her was an eye-opening experience for me. Even though her movement and speech were extremely limited, she was able to communicate with everyone in a deeply personal way. Not only that, but she was able to communicate her ideas about her sculpture to the other artists and they were able to produce her artwork for her, even though she was unable to create detailed drawings. I was deeply affected by meeting Rasa and continue to be puzzled by one question: How was it possible that without detailed speech or movement one person could communicate so much?

The question was on my mind particularly that year because I was one of the first instructors at my college to teach a course using a videoconferencing system to students several hundred miles away. The course was structured to accommodate twenty-five students in a classroom with me in Chicago and ten students in a classroom in Springfield. The videoconferencing system in place was the state of the art. The system contained several remote controlled cameras including one that automatically followed the instructor's movements in the classroom. There were two large screens in which we could see the students in Springfield. From the main base in Chicago we could control the cameras in Springfield to focus on a specific student. I had several training sessions on the use of the equipment and techniques for effective videoconferencing. Everything was in place for the highest level of interaction, but I still found it incredibly difficult to communicate with the students in Springfield. Other instructors in this pilot program also expressed difficulties. How was it possible that a state of the art system that allowed the transfer of both visual and aural information in very nearly real time operated by experienced teachers could feel so unnatural and limit communication so severely? What were technologically mediated communication systems lacking?

I began to study, from an artist's perspective, aspects of person-to-person communication and how these might be explored as ways to enhance remote communication. One year later, *Gape*, my first performance using eye tracking, was created[1]. The system I developed used a simple software based eye tracking system that determined the position of light and dark pixels in a live video signal. As the performer, I would look at a grid of nine words. Through looking, I would trigger spoken words (digital audio samples of my own voice). The audience of *Gape* perceived a performer trying to 'speak' a coherent sentence. I used an eye tracking system and disabled my hands and body in this performance to comment upon the simultaneously enabling and disabling effect of technology in contemporary life [2].

In retrospect, I believe this performance related to metaphorical connections in my mind between the computer screen and the retinal image. The screen, made of a flat grid of pixels, is like the retinal image, a flat projection of color and pattern. The retinal image itself lacks the depth and meaning of the real world, and it is only through the process of interpretation that an understanding of the world is formed. In a similar way, information stored on the computer has no real meaning until it is interpreted. Information must not only be interpreted by the computer program, but also interpreted by a human interacting with the machine. The Platonic version of vision in which the eye emits a virtual fire, a material substance that formed the visible image called *simulacra*, that allows vision to take place is a very active one. [3] In *Gape*, the image of the eye, usually the receiver rather than the transmitter of an image, was received by the computer. The computer then took this very material information (the bits and bytes that make up the image) in real time and translated it into sound perceived by the ears of the viewer.

Eye Tracking Systems

Seeing is active. Vision itself cannot occur without finely tuned movements of the eye, taking in patterns of light and color on the retina which the mind must then translate into a coherent world. In all cultures, the eyes are used to convey a wide variety of messages. Currently, technological means of communication often lack the speed to communicate the subtleties of these movements or don't employ them at all (email and the telephone, for example).

Eye tracking systems have been developed to allow users without manual ability to do everything from controlling lights and appliances in their home to using the telephone to operating their computer. There are several kinds of devices currently used, from galvanometric sensors which measure voltages across the eye, to video image processors which examine images of the eye in real time. Systems based on electric skin potential around the eye, called electro-oculography, is the most invasive, requiring close contact of electrodes to the user or electro-magnetic contact lenses. Eye tracking using image processing is the most accurate and reliable. Image processing eye trackers can be head-mounted or remote. Remote image processing systems often use a pupil center/corneal reflection combination. A small, bright reflection is generated from a light near the camera on the surface of the eye's cornea. A computer calculates the gazepoint based on the relative positions of the pupil center and the corneal reflection. Other remote systems employ two cameras, a wide-angle camera for locating the position of the face, and a telephoto camera, robotically controlled by information from the wide angle camera, that locates and analyzes the user's eye.

Eye Movement and Attention

Visual perception is temporal, dependent on eye movements to examine a perceived object as a whole and in detail. Without eye movements, a stabilized image disappears from the perceiver's field of view. Eye movements have two major functions: fixation, to position target objects to the center of vision; and tracking, to keep fixated objects in the center of vision despite movements of the object or the observer. Eye movements can further be divided into three distinct types which can be under voluntary control: convergence, smooth pursuit, and saccades.

Convergence, also called vergence movement, is the process of following objects as they approach or move away. These movements are very slow, moving no faster than ten degrees per second in most cases. Smooth eye movements keep the line of sight on a selected object and compensate for motion on the retina that might be caused either by motion of the object or by motion of the head or body. In order to be initiated or sustained, smooth pursuit requires an external moving visual signal. Smooth pursuit movements are also relatively slow eye movements, they have a maximum speed of about 100 degrees of rotation per second.

The third type of movement, called saccadic movements, are rapid jumps of the eye used to shift gaze to a chosen object. Once a saccadic movement has been initiated, the movement cannot be interfered with until it reaches completion. [4] Saccadic movements are very fast, typically taking only 30 milliseconds to complete and reaching speeds of 900 degrees per second.

Smooth eye movements and fixation occur in the intervals between saccades. Intervals between saccades can be as long as several seconds during steady fixation; and in reading, about 3 times each second. Even when fixating, the eyes continue to move. They tend to drift and flick involuntarily and to oscillate back and forth continuously, although these movements are extremely small [5]. Sequences of fixation and saccades have been used to study cognitive processing in the coordination of eye and arm movements, during visual problem solving tasks, and during reading [6]. Hands-free eye tracking systems also utilize these intervals by employing a timed gaze as the manner of interaction. Topics are selected by the user fixing their gaze in a certain area for a specificed length of time (1/4 to 2/3 seconds is the most comfortable according to Nancy R. Cleveland of LC Technologies).

An increase in the speed of saccades can be learned or trained with daily practice, and many researchers indicate that saccades are planned, controllable activities. Some researchers suggest that shifting attention to another location while the eye remains stationary is the same as planning a saccadic eye movement. In other words, sequences of saccades and fixation are directly related to attention. Attention is allocated to a target shortly before the saccade is made to look at it [7].

Allocation of visual attention has been shown to be related to the content and meaning of the subject. Saccadic eye movements in particular are used to inspect a visual scene, requiring the integration of discrete time frames into a stable picture of the scene. In 1967, Yarbus conducted a series of experiments with adult subjects where recordings of eye movements made while looking at various paintings show systematic preferences in eye movements to repeatedly look at those elements that would seem to be most relevant to the content of the painting, or those elements that corresponded to questions asked of the subject prior to viewing. [8] There is not a clear demarcation between voluntary and involuntary eye movements, but it is known that the mechanism for eye movement is different than the mechanism for known voluntary movements of the body.

Interest and Emotion Sensitive Media

The exploration of eye movements as a way to understand and analyze human thought patterns has been a theme of some of the researchers and systems developers. Luciano da Fontoura Costa suggests that human cognition itself is structured along the lines of eye movements. In the following passage, he describes a metaphorical link between specific categories of eye movements and thought

"After pursuing some ideas for some time, our thoughts are directed somewhere else, into a new subject that is nonetheless associated in some way to what we were thinking before. If you close your eyes now, it is very likely that you will continue, for some time, thinking about the ideas presented in the last sentences. Then, you may start wondering about related ideas and concepts (`smooth pursuit'), sometimes focusing on specific themes, sometimes moving to completely new ideas (`saccades')." [9]

As if following Fontoura Costa's line of thinking with entreprenuership in mind, Arne Glenstrup and Theo Engell-Nielsen of the University of Copenhagen propose what they call 'Interest and Emotion Sensitive Media' (IES). Using the concept of selective attention and measuring patterns of fixation, IES proposes to allow for faster searching and more rewarding entertainment viewing. With IES, the user is able to seamlessly choose areas of interest. Glenstrup and Nielsen pose a mode of 'manipulatory looking,' in which the user is not only looking to gain more information, but looks with the intention of manipulating something in a scene. Manipulatory looking is already in practice, for example a parent looking at a child in a certain way to get them to do as they are told, or looking at a wristwatch during a meeting to encourage the speedy completion of the meeting. The eyes can be used as a pointing device, as has been shown for example in a mystery film scene in which the guilty party looks at where the stolen goods are hidden, telling the detectives where to look. Although the monitoring of interest and emotion may seem like an invasion of privacy, or a way to police 'thought crimes;' there is a missing element. Attention can be oriented toward information covertly in a complex scene. As was shown in the model of planning saccadic eye movements, it is possible to focus attention on various objects in a scene without directly looking at them.

Active Vision/Active Music

By 1999, free improvisational performance with the eye tracking musical system I called Intutive Ocusonics became very important to me. I believed that if I could control the instrument in such an uncontrolled situation as a musical improvisation, I would have determined that the eye could be used intuitively to control any instrument. At the same time I was performing and recording, I began looking at some of the research analyzing the process of improvisation in order to look for connections between intuition, improvisation, and human computer interaction.

David Rosenboom has written on the use of brain wave information in human-computer interactive musical systems. He posits that musical improvisation can be both a controlled and automatic process, the same way eye movements can be both voluntary and involuntary. My own work with eye movements also focused on connections between conscious and unconscious movements. For models, I have looked to activities that require complex movements to be performed extremely quickly, for example driving a car or participating in a musical improvisation.

Many musicians and artists have done research in creating intuitive systems for human computer interaction. Electronic artist and Jazz Musician George Lewis created one of the first computer driven jazz improvisation systems.

"It seems to me that focusing on the computer as an isolated entity is not the key question...You have to approach it on the level of emotion, on the level of creating dialogue, and on the level of passing information through sound...The difference between playing with people and playing with computers is in the realm of approaches to empathy." [10]

Improvisation and Automatic Processes

Similar to the eye movement 'repetoire' of two major functions and three types of movements, there are some commonly used improvisational musical strategies. In the short term, there is: simultaneity (performers playing the same thing at the same time), unpredictability (an attempt on the part of members in a group to surprise eachother), mimicry (one performer usually leads and the others attempt to copy), repetition (repeating a sequence of sounds), contrasting between loud and quiet, and trying to find the space or 'hole' and playing in that space or silence. On a macroscopic temporal scale, this contrast and repetition is similar to microscopic sequences of saccades and fixation that occur during the scanning of a visual scene. The sequences of contrast and repetition in a musical improvisational situation creates an auditory scene.

Conclusion

"Thus the concept of computer interactivity can be understood to not only include the interface to a user/perceiver and the redefinition of authorship which that implies but more fundamentally to include the potential for deep structural interaction between the different sensory modes of human perception." [11]

Performing with the Intuitive Ocusonics system over the past five years, I have formed more questions than answers about the nature of the visual and auditory scene. There were several insights that came with the performing of these works. First, and most important for me, I found that at several points during improvisational performance with the eye tracking device, I felt as if I was able to unconsciously control my eye in a very precise way to create specific sounds. In other words, the eye tracking instrument in improvisation could produce the same feeling for the performer as improvisation with a traditional instrument, i.e. the process became automatic.

In the design of the interface for eye tracking music, it seemed that the simpler the visual aspects of the interface, the more effective it was to play music. A complex visual can distract from the performer's ability to listen intensely while performing, and, in fact, there were some times during the playing of the instrument that I felt completely unaware of seeing anything at all, but purely focused on the sound.

In terms of similarities, a corellation between grouping and tracking was apparent and will be a further area of study with the instrument. The zone interface, organizing groups of amplitudes spatially, was a useful development in the design of the instrument. Do both the ear and the eye organize perceptual input into groups in similar ways?

The exploration of the process of musical improvisation led me to question how an auditory scene is constructed. Josephson and Carpenter suggest that the construction of the auditory scene is also an internal process. I discovered that this was not inconsistent with the construction of a visual scene. In the algorithmic model, there are four known stages to the visual perception of objects. The image-based stage, the surface-based stage, the object-based stage, and finally the category-based stage. The object-based stage is the stage in which the perceiver's internal knowledge informs how the subject is perceived, and the category-based stage combines the sequences of eye movements and their corresponding images on the retina combine to create the functional 'idea' of the visual scene.

My current work, The Fly's Eye, draws its inspiration from the structure, function and significance of the eye of the fly in relationship to the study of human sensation and perception. [12] Much of the human brain is devoted to processing visual information, and researchers believe more than half of the fly's brain is devoted to visual processing. A visual system such as the fly's, the simplest living eye, can help us to understand the properties of cells, the interpretation of visual information, and the representation and processing of information.

In the art work The Fly's Eye, multiple images are formed in positions projected in the gallery space based on the movement of viewers in the space. The Fly's Eye 'watches' the viewer in the space while the viewer simultaneously enjoys some control and direction of the location of the image and therefore the shape of the space. Each time the viewer changes position, the live video feed moves and a visible trail is left in the gallery space of all the events in the gallery that day. Time is built in layers of position and image. The resulting record of time and space presents the visitor with an unfamiliar level of complexity, but soon the viewer is able to 'read' the record with a surprising amount of comfort and accuracy .

I found so many connections between the temporal process of creating and listening to music and the process of seeing, that it is clear to me that an artist studying or working today has to consider visual media a temporal rather than static medium.

Endnotes

1. See http://www.khm.uni-koeln.de/~an/imagery/
created as part of the diploma project of An Reich at the Academy of Media Arts Cologne under the supervision of Dr. Seigfried Zielinski and Phillipp Heidkamp

2. This work was performed as a part of Meme Me: Identity and the Replication Age at Artemisia Gallery, in Cache at Columbia College Chicago in conjunction with ISEA97, and as part of a performative lecture at Imagina98. For a shockwave simulation of *Gape* see: http://homepage.interaccess.com/~apolli/gape.htm

3. Lindberg, David C. *Theories of Vision: From Al-Kindi to Kepler* (Chicago: The University of Chicago Press, 1976).

4. For more detail on saccades see the work of B. Fischer et. al:
Fischer, B. & Weber, H. "Express saccades and visual attention" *Behavioral and Brain Sciences* (16 3: 553-610 1993).

Fischer B, Ramsperger E "Human express saccades: extremely short reaction times of goal directed eye movements" *Experimental Brain Research* (57: 191-195 1984).

Fischer B, Ramsperger E "Human express saccades: effects of randomization and daily practice" *Experimental Brain Research* (64: 569-578 1986).

5. Rock, Irvin. *Indirect Perception* (Cambridge: MIT Press, 1997). p. 223

6. See Fischer, B. [2] and also Epelboim, J. and P. Suppes "Window on the mind? What eye movements reveal about geometrical reasoning" *Proceedings of the Cognitive Science Society* (18: 59 1996).

7. "The close link between attention and eye movements is supported by neurophysiology. Cortical centers containing neurons that are active before eye movements also contain neurons (sometimes the same ones) that are active before shifts of attention while the eye is stationary... Some investigators have gone as far as to suggest that shifting attention to an eccentric location while the eye remains stationary is equivalent to planning a saccadic eye movement." From *The MIT Encyclopedia of Cognitive Sciences* http://mitpress.mit.edu/MITECS

8. Yarbus, A.L. *Eye Movements and Vision* (New York: Plenum Press, 1967).

9. da Fontoura Costa, Luciano *Minding Vision* Cybernetic Vision Research Group IFSC-USP, 1996. See additional work: da F. Costa, L. "Dialoguing on Julesz's latest masterpiece: A personal review of Dialogues on Perception" *Real-Time Imaging* (1:159-162 1995).

da F. Costa, L. "Topographical Maps of Orientation Specificity" *Biological Cybernetics* (71:537-346 1994).

10. Lewis, George. Person to... person? Interviewed by Lawrence Casserley http://www.l-m-c.org.uk/texts/lewis.html

11. Dunn, David. Wilderness as Reentrant Form: Thoughts on the Future of Electronic Art and Nature. Leonardo: MIT Press Cambridge, Vol. 21, No. 4

12. The Fly's Eye http://www.andreapolli.com/studio/fly

General Bibliography

Bauby, Jean-Dominique *The Diving Bell and the Butterfly* (New York: Alfred A. Knopf, New York 1997).

Biscaldi M, Boch R, Fischer B. "Effects of directed attention on saccadic and manual reaction times" in: *Perception* (vol 18: pp. 521 1989).

Carpenter R. *Eye Movements* (New York: Macmillan Press, 1991).

Carpenter, R. H. S. *Movements of the Eyes* (New York: Pion, 1977).

Haddad GM, Steinman RM "The smallest voluntary saccade: Implications for fixation" *Vision Research* (13: 1075-1086 1973)

Hoffman, J.E. "Visual attention and eye movements" In H. Pashler (Ed.), *Attention* (London: University College London Press, 1997).

Kowler E "The role of visual and cognitive processes in the control of eye movement" *Reviews of oculomotor research* (4: p 1ff 1990)

Kowler, E., E. Anderson, B. Dosher and E. Blaser. "The role of attention in the programming of saccades" *Vision Research* (1897-1916 1995).

Rashbass C "The relationship between saccadic and smooth tracking eye movements" *Journal of Physiology* (159: 326-338 1961).

Rosenboom, David "Extended Musical Interface with the Human Nervous System Assessment and Prospectus" *Leonardo Monograph series* Published by the International Society for the Arts, Sciences, and Technology, 1990.