Jason Gotlieb
Computational Approaches to Human Vision
Psychology W3270
December 22, 2011

Gaze Tracking:
The Systems Behind it, Within it, and From it

1. Introduction

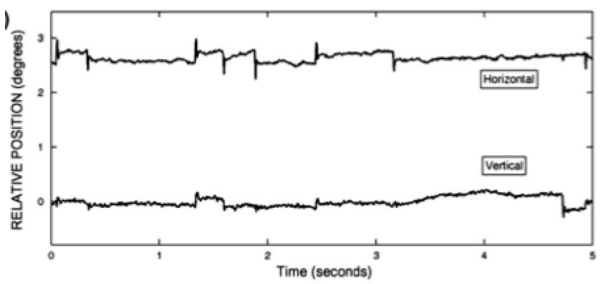
The rise of technology has brought researchers a new frontier to gaze upon. Scientific studies often rely on technology in their methods. This paper however, aims to discuss the way in which technology can utilize science. Knowledge of the human visual system—specifically patterns of eye movements—is instrumental in determining design practices for webpages and user interfaces. This information contributes to improving human computer interaction (HCI). In order to develop an understanding of this idea, one must first have command of the basics of eye movements. With this information, it is possible to understand the idea and implementation of eye tracking systems. This paper will explore one specific model that will form a generality of understanding that can be applied to other eye trackers. With the basics of the input components—the biology of eye movements and the technology of eye tracking—HCI can be examined. This paper will digress from eye tracking and eye movements to address a primitive HCI phenomenon called banner blindness. It will then turn to a better-aligned study that was conducted to determine the qualities of the web-based user interface of a particular website. HCI is an important reality of today's web-based world. Without quality interactions, machines cannot be used to their full potentials.

2. Eye Movements

An understanding of the science of eye movements is important in the development of improved HCI. This section aims to outline some basics of eye movement in order to inform the reader of eye movement essentials that will facilitate the understanding of further sections. One key term to know is arcminute or minute of arc that is defined as one sixtieth of one degree.

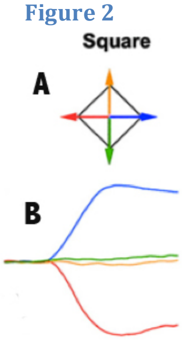
Gaze control or maintained fixation is the first relevant eye movement to HCI. This is the eye movement characterized by the smallest amount of actual displacement. The viewer is focused on a particular stationary object or target. In spite of perceived stillness during fixation, the eye is actually experiencing slow oscillations that are periodically interrupted by microsaccades. These microsaccades are defined as saccades with less than 15 arcminutes. They also happen simultaneously in both eyes (E. Kowler, 2011). Figure 1 shows eye movements during fixation of a stationary point target. The graph is recorded with the Dual Purkinje Eyetracker. Fixations are important to HCI because they can relate to attention. If a user fixates on a target on screen, it sometimes implies spatial attention (M. Carrasco 2011).

Figure 1



A second type of eye movement is smooth pursuit. This is the eye movement that occurs when the eye performs a smooth tracking of a moving object or target. Smooth pursuit cannot be initiated voluntarily over a stationary environment. In a hypothetical environment that consists of only moving objects, smooth pursuit cannot be fully suppressed. For these reasons, smooth pursuit is deemed involuntary. In Figure 2, A represents an object that moves horizontally

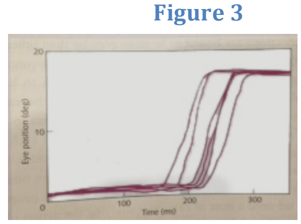
Figure 2



across a plane. B is the mapping of smooth pursuit. The colored lines of B correspond with the colored arrows in A. Smooth pursuit is less prevalent in HCI of static web pages but may prove important in interactive gaming or virtual reality simulations. In one study, a moving banner had little to no effect on the user's attention to it. This suggests that perhaps the involuntary quality of smooth pursuit has yet to be fully understood in its application to HCI (Benway & Lane, 1998). Again there is a connection here to selective attention. The person or smooth pursuit system can decide which target to pursue if given a choice. As mentioned, unless the entire environment is in motion, the path of smooth pursuit can be controlled.

A third and final type of eye movement—in the scope of this paper—is the saccade. Saccadic eye movements are rapid shifts in the line of sight made to bring the fovea from one selected location to another (E. Kowler, 2011). Saccades are fast. They have high speeds reaching upwards of 100°/second and once the movement's path is initiated, it cannot be altered. Saccades are characteristically similar. In Figure 3 saccadic paths can be seen. The subject's eye moves roughly 15 degrees and as shown in the figure, besides for latency, the saccades appear to be the same. The movements are also useful in sampling the visual environment. This is a particularly significant point in relation to HCI. As early as 1985 Koch and Ullman proposed a 'saliency map' that was defined as a 'global measure of conspicuity' (E. Kowler, 2011). Saliency maps are often overlaid onto websites or user interfaces in order to determine web or interface usability (C. Herendy, 2009). Saliency maps are said highlight distribution of attention by conjoining fixations with saccades across a

Figure 3



scene. A higher salience implies a higher likeliness that a given location will be fixated upon. Highly salient websites, therefore, presumably have high tendencies towards positive HCI.

3. Early Studies of HCI

The application of human computer interaction has undoubtedly been considered since the computer's inception. Until recently however, many of the design decisions were based on technological possibilities rather than tested and proven models. Early computers that used terminals with time shares were less concerned with usability because it would have meant a sacrifice in functionality. Studies of web usability are thus more recent; one relatively early study was conducted in 1998. The study, entitled, "Banner Blindness: Web Searchers Often Miss 'Obvious' Links" was conducted by Jan Benway and David Lane. The study aimed to respond to a common web-design hypothesis that states, "In general, the larger an item is, the greater its perceived visual importance and likelihood of attracting attention" (Detweiler & Omanson, 1996). Benway and Lane were able to dislodge this hypothesis with their research.

The researches built a special website upon which to conduct their experiment. The study instructed users to perform 24 searches for specific facts that could be extracted from that site. The study's remarkable findings concluded that users disregarded web information presented on banners. They were ostensibly blind to web-banners. Figure 4 outlines this phenomenon. The figure shows that 'banner tasks'—those searches that required the reading of a banner for success—

were only completed 58% of the time. 'Control tasks'—those searches that did not require banner based information for success—were completed 94% of the time. All users in the study were said to be familiar with computer and web use. Banners failed to attract attention in spite of their aesthetic 'pop.' That is to say that even though the banners fell in line with the hypothesis of Detweiler and Omanson—they were large, colorful, animated in some cases, and attention grabbing—they were determined to be more difficult to perceive than control tasks.

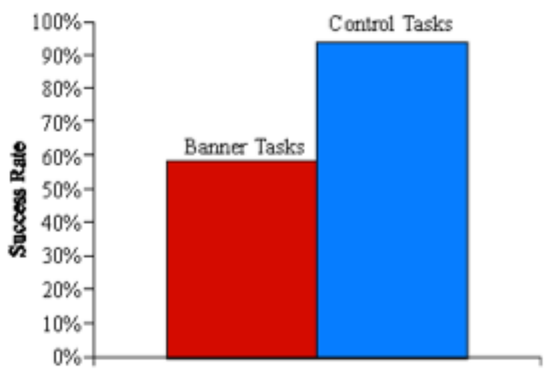


Figure 4

In a second experiment conducted by the same researchers, and published within the same paper, users were shown different types of banners and asked to perform similar searches. Some of these banners were animated yet the results were similarly grim. Of the 72 participants, only seventeen reportedly noticed banner-specific information. This finding is important to the development of HCI.

Benway and Lane did not use eye tracking in their experiment. Had they utilized this technology, perhaps we would have greater insight into their findings. Without eye tracking, the research they present seems faulty and incomplete. There is no way to actually know where the users were looking. Instead we are given results based on users' own memory and perceived self-perception. In other words, users had to report upon what they saw. There was no system to actually track their web experience. Luckily there are other experiments that are similar and have utilized eye tracking technologies. It is likely that the reason these early researchers decided to neglect the technology was because of the high costs that were once associated with it. We will see how gaze trackers work and their important application to developments in HCI.

4. Eye/Gaze Tracking Technology

Gaze tracking has come a long way since it's invention. Edmund Huey built one of the earliest intrusive eye tracking systems in the early twentieth century. This tracker used a contact like lens that was placed on the eye and had a small hole for the pupil to peep through. The lens was connected to a pointer that was able to respond to eye movements (Huey, 1908). Other early eye trackers involved, "Physically gluing something to a test subject's eyeballs" (Nielson & Pernice, 2009). Later, the precursor to the modern eye tracker strapped the user's head into the device so as to prevent head movements. Head position is an important factor in calculating eye position relative to the object being gazed upon; the fixed-head system took care of this computation by eliminating it. Today, eye tracking relies on much more technology and collected data. As we will see, systems often rely on artificial intelligence in tracking gaze. Tracking has also become significantly less intrusive, cheaper, and more efficient.

For the sake of research, it is important to understand modern eye tracking methods. Nielson and Pernice explain that all modern eye trackers work with the same basic principal, "focusing a light and a video camera on a person's eye." The light, often infrared, reflects off of the retina and helps determine the direction the person is looking. The camera records the interaction with the object being viewed. Computer screens are often the easiest to interface with because the camera is either built in or connected already and thus the screen being viewed is simpler to keep track of. (Nielson & Pernice, 2009)

Modern, low-cost systems have proven successful but are still not as precise as other high-cost, state-of-the-art, commercial trackers. One example of a low-cost eye tracking system was developed by ITU GazeGroup. They are a research group from the IT University of Copenhagen, Denmark. The group has developed an open source eye tracking system that can be downloaded for free and only requires a webcam and an infrared light. In a recent study conducted by ITU GazeGroup, their tracker had a mean error of 59 pixels compared to a commercial, higher cost, tracker that only had 31 mean pixels of error (Johansen et al., 2011). This shows that eye tracking technology is developing and becoming more accessible to the public.

In their paper, "Eye and gaze tracking for interactive graphic display," Zhiwei Zhu and Qiang Ji describe a specific tracker that they developed in 2004. Their paper acknowledges that gaze tracking is important for HCI. They also explain that usability itself is developed because intelligent graphics can be utilized. That is to say that not only does gaze tracking improve HCI, it redefines it.

The tracker proposed by Zhu and Ji is modern. It uses infrared beams and allows for users to make relatively major head adjustments. The system's definition of fixation is the point of intersection of the line of sight with the surface of the object being viewed; the screen. Gaze direction is determined by both face and eye orientation. Final gaze is the combination of face and eye orientation. These two orientations determine global and local direction respectively. (Zhu & Ji, 2004).

Figure 5 shows a visual algorithm of the major components of the proposed eye tracking system. The first step, 'eye tracking' involves the second step, which encompasses 'pupil and glint tracking.' The system accomplishes this by alternatively blinking rings of LED's to effectively retrieve dark and light pupil glints. The images that the camera captures with the blinking LED's are fed to a support vector machine (SVM) that recognizes eyes and specifically pupils. This is a form of artificial intelligence and thus the SVM must be ostensibly taught or loaded with what it needs to detect.

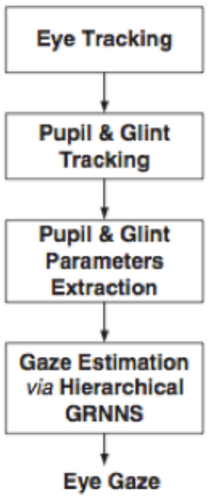


Figure 5

Figure 6 shows a flowchart algorithm of Zhu and Ji's pupil detection and tracking process (the first and second parts of the algorithm in Figure 5). If the SVM succeeds in detecting the pupil and glint, a Kalman filter is applied. The Kalman filter allows for real time tracking because it determines the current eye position based on the previous position. If the Kalman filter is successful the tracker has not lost track of the gaze. The filter itself is also augmented by way of mean shift tracking. This mean-shift accomplishes tracking via the image's brightness-intensity distribution.

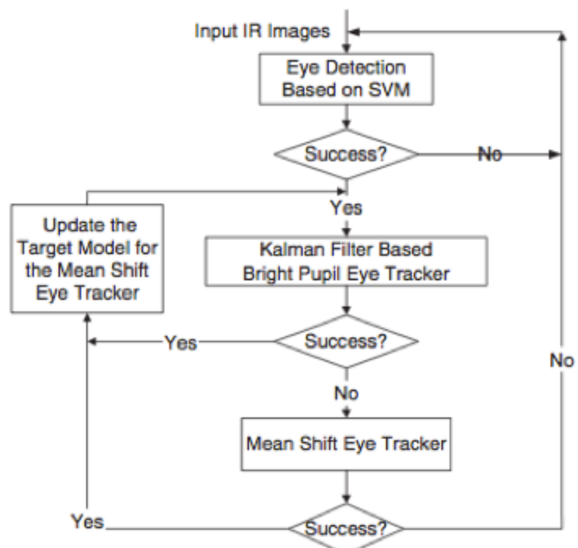


Figure 6

Depending where the eye is focused, the tracking system can use a mapping function to map the pupil-glint vector to screen coordinates. Head movement however must still be taken into account. Depending on the position of the face, the pupil-glint vector may be compromised. A face pose classification algorithm exploits the relationship between face orientation and pre-defined pupil parameters. Those parameters are: Δx, Δy, r, ø, gx, and gy. They are used to calibrate the gaze tracking with respect to head displacement. Delta x, and y represent pupil-glint displacement or local gaze. The r is the ratio of the major and minor axes of the ellipse that surrounds the pupil in real time. This number is equal to one when the face is looking forward and has no displacement. Theta is the pupil ellipse orientation. Theta changes when the user's face rotates within its plane. G of x and y represent the glint image coordinates. If the head undergoes translation within its plane, these values are affected (Zhu & Ji, 2004). Different facial poses are clustered together in pupil feature space (PFS). The PFS has five distinct facial orientations that are each determined based on differing pupil relationship parameters.

The last step in this gaze tracking system is the gaze calibration by way of generalized regression neural networks (GRNNs). Neural networks are used because it is difficult to, "analytically derive the mapping function that relates pupil and glint parameters to gaze under different face poses for different persons" (Zhu & Ji 2004). The GRNNs are feedforward networks that have fast training times and can model nonlinear functions. The GRNN uses the pre-defined pupil parameters as its input vector. The GRNNs in this system are trained using a one-pass learning algorithm that improves the overall gaze tracker's speed.

After training and testing, the gaze tracker proposed has 85% accuracy. Most important however is the knowledge of the tracker's functionality. Part of appreciating and improving HCI arguably involves an understanding of the tools that inform those interactions. By having a grasp on the interworking of an eye tracking system, a researcher is able to improve the interactions they have with that system and by association, with the interface in question. Now that the reader has a basic but valid understanding of the system, its application can be studied.

5. Improving HCI

Human computer interaction relies on systems that allow humans to interface with computers. Up to this point, the studies examined in this paper have not utilized eye-tracking systems in specific. The Banner Blindness study didn't track the user's gaze but rather relied only on the perceived interactions of its test subjects. Using gaze tracking, as we will see, greatly improves the researcher's ability to make informed decisions about webpages and user interfaces.

Csilla Herendy's study, "How to Research People's First Impressions of Websites? Eye-Tracking as a Usability Inspection Method and Online Focus Group Research" utilizes eye tracking and makes an informed critique of the Hungarian government's website portal. Without eye-tracking the study would be inconclusive unless it applied rigorous and ingenious testing. Eye tracking allows for quick usability tests that are effective and precise.

In her study, Herendy determined the Hungarian government website to be "too complicated, dull and difficult to apprehend at a glance" (Herendy, 2009). Two groups of users were instructed to interact with the website and to perform simple searches within the homepage presented (see Figure 7). Both groups answered online surveys about their interactions with the site and one of the groups also used gaze tracking to supplement the research. Like the Banner Blindness experiment, users were quizzed on their interactions with the interface. Herendy's experiment is more conclusive however because of the added eye-tracking information. The focus group that did not interface with eye-tracking presented interesting information but only with the supplement of the eye-tracking group is the information tangible and complete.

Figure 7

Figure 7 shows a print-screen photo of the Hungarian government website. Figures 8 and 9 show the scanpath of the user's eyes as they gaze at the site and try to find elements within it. Figure 8 is the scan path of a relatively simple search while Figure 9 represents a difficult search. The thin lines are representative of saccadic movement whereas the colored circles show fixations. Larger circles represent longer fixation times. The more fixations in a given area, especially longer

Figure 8

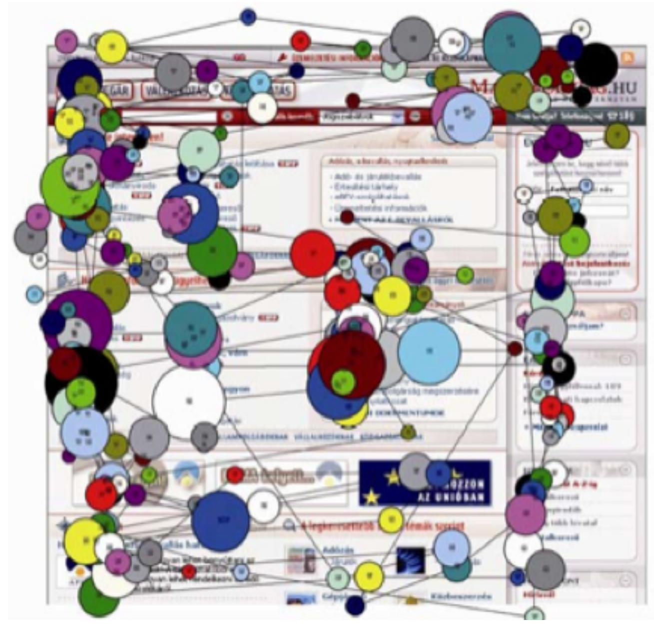
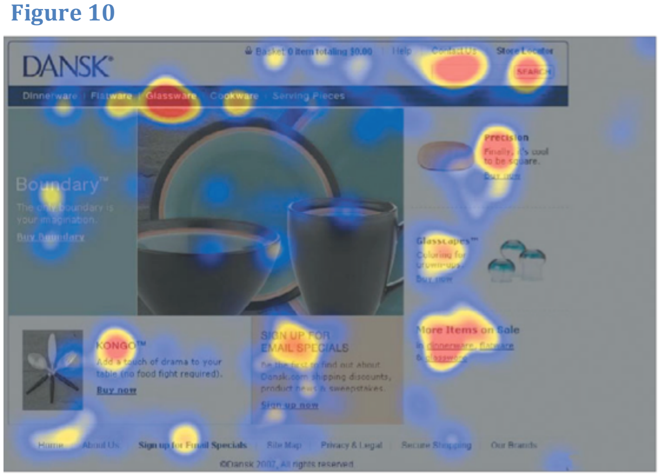


Figure 9

fixations, are representative of perceived attention. It is clear from these scan-paths alone that the Hungarian government website is convoluted and difficult to read. Such an example makes for difficult HCI. In Nielson and Pernice's research, they used gaze tracking to look at cluttered websites such as the ones shown above but also to look at relatively well-designed sites. Figure 10 shows an example of a simple and comprehendible site. The caption from the book reads, "the heat on the top and bottom utility navigation shows that users wanted to scan the entire homepage of this site—even the less likely areas—because it is uncluttered and easy to use" (Nielson & Pernice, 2009). Rather than following the hypothesis of Detweiler and Omanson that states that bigger is better, it seems like the real insight into HCI

Figure 10



involves simplicity with minimal options. A large banner might be ignored but probably not if it's the only thing on a page.

HCI is easily studied with the help of gaze tracking. Users have mice and keyboards in their hands but the majority of computer interaction involves visual searches of the screen. With gaze tracking, these searches are followed and can be studied. Both Figure 9 and Figure 10 show a lot of user interaction with the page in question. Figure 9's high level of interaction is saccadic in nature and thus representative of a user who was lost and confused. Figure 10 has a similarly large amount of interaction but it is uniform and clearly overlays features of interest rather than confusion. It is conceivable that had the user from the study behind Figure 9 been searching the site without instruction, she or he would have decided to leave the site entirely. That is to say that the information was probably only scanned with such vigor because it was within the scope of an experiment. Hypothetically, by the same logic, the user behind Figure 10 would have stayed on the site anyway because the information provided was well organized and interesting. The amount of visual information that eye tracking provides is astounding.

6. Conclusion

Human computer interaction is important to be aware of in today's modern society. As new sites and programs flood the market, it is vital for developers to understand the needs of the user. Technology no longer needs to wait for hardware to improve usability as it once did. Design is not a limiting factor in software development. In order to build a truly enjoyable user experience, it is important that developers have an understanding of the human visual system—specifically eye movements—and of the systems that exploit such knowledge (gaze tracking). Eye tracking is a unique technology in that it yields comprehensive and exact data about characteristics of the visual system. Eye trackers are now non-invasive and exploit a human feature without the need of physical contact.

It is possible that the future of computing will rid itself of the clunky mouse and instead utilize the expressions of human eyes to direct user interaction. The eye is fast and precise and thus capable of improving the efficiency of HCI. In this way eye tracking may not only inform HCI improvements but also define them. The tracker itself would be the system that improves human computer interaction. We saw this briefly with the implementation of Zhu and Ji's eye tracker. In their test cases, users' vision was not only tracked but used as a tool of interaction. This idea is already widely implemented with disabled peoples who are otherwise unable to communicate. Its commercial application however has not been fully realized.

For all these reasons, I am very interested in eye tracking and am glad to have studied the science behind it, the science within it, and the science gathered from it. This paper attempted to introduce the reader to the physical eye movements that eye tracking relies on. After that, it aimed to explain one eye tracking system whose structure is similar to most other modern systems. Once these concepts are made known, the reader is then able to appreciate the simple and straightforward research conducted with these powerful tools and to contemplate the future of HCI.

Sources cited

Detweiler, M., & Omanson, R. (1996). Ameritech web based user interface standards and design guidelines. (on-line) Available at http://www.Ameritech.com.

Eileen Kowler, Eye movements: The past 25 years, Vision Research, Volume 51, Issue 13, 1 July 2011, Pages 1457-1483, ISSN 0042-6989, 10.1016/j.visres.2010.12.014. (http://www.sciencedirect.com/science/article/pii/S0042698910005924)

Fethi Calisir, Demet Karaali, The impacts of banner location, banner content and navigation style on banner recognition, Computers in Human Behavior, Volume 24, Issue 2, March 2008, Pages 535-543, ISSN 0747-5632, 10.1016/j.chb.2007.02.019. (http://www.sciencedirect.com/science/article/pii/S0747563207000507)

Herendy, Csilla. 2009. "How to Research People's First Impressions of Websites? Eye-Tracking as a Usability Inspection Method and Online Focus Group Research." University of Pecs, Budapest, Hungary. IFIP International Federation for Information Processing.

Huey, Edmund Burke. (1968). *The Psychology and Pedagogy of Reading*. Cambridge, MA: M.I.T. Press. (Original work published 1908)

Marisa Carrasco, Visual attention: The past 25 years, Vision Research, Volume 51, Issue 13, 1 July 2011, Pages 1484-1525, ISSN 0042-6989, 10.1016/j.visres.2011.04.012. (http://www.sciencedirect.com/science/article/pii/S0042698911001544)

Nielson, Jakob and Pernice, Kara. (2009). *Eyetracking and Web Usability*. New Riders. Print.

Robert J. Snowden, Peter Thompson, Tom Troscianko, Basic vision: an introduction to visual perception, Oxford University Press, 2006, 382 Pages, ISBN 0199286701, 9780199286706 (http://books.google.com/books/about/Basic_vision.html?id=Kf1FAAAAYAAJ)

Sune Alstrup Johansen, Javier San Agustin, Henrik Skovsgaard, John Paulin Hansen, and Martin Tall. 2011. Low cost vs. high-end eye tracking for usability testing. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems* (CHI EA '11). ACM, New York, NY, USA, 1177-1182. DOI=10.1145/1979742.1979744 http://doi.acm.org/10.1145/1979742.1979744

Zhiwei Zhu, Qiang Ji. 2004. "Eye and gaze tracking for interactive graphic display." From *Machine Vision and Applications*. Rensselaer Polytechnic Institute.