Use of Eye Movement Gestures for Web Browsing

Kevin Juang, Frank Jasen,
Akshay Katrekar, Joe Ahn
Computer Science
Clemson University, Clemson, SC 29634
[kjuang | fjasen | akatrek | jahn]@clemson.edu

Andrew T. Duchowski
Computer Science
Clemson University, Clemson, SC 29634
duchowski@acm.org

ABSTRACT

Web browsing is one of the most common tasks performed by modern day computer users. Any improvement to typical web browser interaction could thus yield a wide-ranging benefit in terms of user efficiency or satisfaction. Studies have shown the advantages of traditional mouse-based gestures in web and hypertext browser applications, so a logical step would be to examine the potential for similar gestures using eye movement interaction. Here we implement such a system and study its effectiveness. Key issues when working with gestures are the quantity and complexity of the implemented gestures. Because of the established shortcomings in using eye gaze tracking for conventional motor control tasks, we severely limit the number of eye movement gestures in our system and constrain them to single-line motions. We believe that this allows our system to apply a more natural mapping between gesture and meaning. We test users in a simulated browser environment conducting a navigation task, comparing our eye movement gestures to traditional point-and-click mouse input. We contend that our eye gesture system compares favorably to mouse input in terms of user performance and satisfaction.

Author Keywords

Eye movements, eye tracking, gestures, web browsers, interaction techniques, alternative input.

ACM Classification Keywords

H.5.2; H.5.4 [Information Interfaces and Presentation]: User Interfaces — Input devices and strategies, interaction styles; Hypertext/Hypermedia — Navigation.

INTRODUCTION

One of the established uses for eye tracking technology is in the area of alternative input in a graphical user interface. Eye tracking as input has numerous uses for the general population, encompassing such diverse areas as Attentive User Interfaces (AUIs) [16], computer supported cooperative work [14], virtual reality [5], and general vision research [2]. Of course, eye gaze input is indispensable to users with severe motor disabilities, such as quadriplegics; it allows them to communicate with their eyes when manual typing is not an option [9].

However, there has been fair criticism of eye tracking as input by Zhai et al. [17] and others. Eye gaze is a visual perception channel and consequently is not well suited for motor control tasks. In addition, eye gaze input goes against users' mental model of hand-eye coordination where the eye searches for and receives information while the hand manipulates external objects. The chief advantage that eye input holds is that it is extremely fast, quicker than mouse-based, keyboard, haptic, or even speech input [2]. Along with this increase in speed, a decrease in accuracy often occurs: the classic speed-accuracy tradeoff.

The goal then, is to locate and examine a domain where the benefits of eye movement input are accentuated and the shortcomings are deemphasized. We hypothesize that one such area is in streamlining common tasks in a web or hypermedia browsing environment. Because tasks such as navigation in a web browser are so basic and ubiquitous, speed is the dominant issue [11]. Furthermore, these are extremely well learned tasks, helping to offset potential accuracy issues related to underdeveloped learning.

Web browsing is a particularly interesting area of study because it is perhaps the most prevalent aspect of modern computing [8]. It is true that eye tracking equipment is currently prohibitively expensive for the average user. Nevertheless, as prices inevitably drop in the coming years as the current PC and GUI paradigm moves toward AUIs and ubiquitous computing, advances made today in the field of web browsing could potentially affect billions of future users.

BACKGROUND

Ashmore et al. [1] identified four major issues with eye gaze input compared to manual pointing. They are:

- Eye tracker accuracy — The error associated with current eye tracking equipment limits the ability to measure eye gaze precisely.

- Sensor lag — The camera and equipment delay associated with motion tracking limits the speed of the system.
- Fixation jitter — Eye gaze, in contrast to a physically stationary mouse, is never perfectly motionless, which can hinder the effectiveness of dwell time input.
- Midas touch [7] — A classic eye tracking problem notes that it is difficult to differentiate between intentional and unintentional eye gaze selection from the user.

Sensor lag is not overly worrisome as long as the lag is imperceptible to the user in terms of feedback and eye movement input is still faster than other methods such as mouse input. Eye tracking accuracy, fixation jitter, and the Midas touch problem are all important concerns for traditional eye gaze input based on dwell time that do not have elegant solutions [17]. However, Isokoski asserted that using eye movement gestures in combination with off-screen targets greatly diminished the negative effects of low eye tracker accuracy and Midas touch [6]. The fixation jitter issue is avoided altogether as the system is using gestures for selection instead of dwell time.

The benefits and complications of traditional mouse gestures as they apply to web browsing have been documented. Their chief advantage is that the precise target acquisition of conventional Fitts' law tasks is circumvented [3]. Speed is greatly improved compared to point-and-click selection because gestures can be started without moving the cursor from its original position and because direction is leveraged instead of distance. Gestures also take advantage of muscle memory and consequently facilitate expert learning through rehearsal [11]. These are characteristics that we strive to duplicate in our eye gesture system.

There are some limitations to the universal use of mouse gestures, however. One is that complicated non-linear gestures arise in a domain with many possible selections, such as eye typing or word processing. Moyle and Cockburn identified web navigation as an area that was well suited to mouse gesture input in terms of speed [11]. They also noted that mouse gestures are not entirely context insensitive; the user has to make sure that the cursor is not over a link, image, button, or other object. This is not an issue with our eye gesture system, which is truly modeless. A further area of concern lies in sufficient feedback. Even with mouse gestures, many users have reported that they had accidentally operated a gesture and not discovered this until they were later confused. Eye gestures present an even greater concern since the gaze will unavoidably not be directed on the screen during operation. To overcome this, we utilize concise audio "click" response on selection for effective feedback.

While Isokoski's idea of off-screen targets for eye typing was an inspired approach to eye gaze input, there are some important concerns to using these gestures for typing. One major disadvantage is that typing sequentially, one character at a time, will necessarily be slower than the more parallel processing of manual typing [4]. Another disadvantage is the complexity of handling all of those characters, typically accomplished by either more targets or complicated hierarchical or non-linear gestures [6]. This negatively influences speed, accuracy, simplicity, and user satisfaction.

By keeping the operation of the targets simple and independent, our system avoids the issues associated with using eye movement for more difficult tasks such as eye typing. Web browsing is a domain with simple, repeated tasks that can also be naturally mapped to target locations: up and down for *page up* and *page down*, and left and right for *back* and *forward*. An additional benefit for the up and down mappings is that when it is time for the user to scroll the page in a certain direction, his or her gaze will typically already be close to the target in that direction. This also allows our gestures to be simple, linear movements.

A concern with physical targets is that users could possibly have too much of their attention drawn to these novel markers. Because of the natural mapping of web navigation to simple gestures, we are able to use off-screen "hot zones" in lieu of actual physical targets attached to the monitor. These hot zones are unique in that they do not depend on fixations at all. Even Isokoski's off-screen targets depended on a 100 ms dwell time [6].

We had two main hypotheses that we intended to examine with this study:

- Eye gestures should be both faster and less accurate compared to point-and-click mouse input due to the previously established characteristics of both eye gaze input and linear gestures.
- Eye fatigue should cause eye gesture activation to become slower as the number of commands in a particular task increases.

METHODOLOGY

Apparatus

The eye tracker that we used was the Tobii ET-1750 from Tobii Technology. The Tobii is a bright-pupil eye tracker that incorporates a 17-inch TFT flat panel monitor at a resolution of 1280 x 1024 and a pair of near infrared light emitting diodes for corneal reflection eye tracking. It is subsequently less intrusive to the user than a head mounted device or a set of visible cameras. It is capable of binocular tracking with 0.5° accuracy at a sampling rate of 50 Hz with a latency of 25-35 ms. The Tobii screen subtends a visual angle of 28° vertically and horizontally from the user's eyes at a distance of 60 cm. The Tobii ET-1750 is shown in Figure 1.

Figure 1. Tobii ET-1750 binocular eye tracker.

The TFT of the Tobii ET-1750 is driven by a 2.0 GHz Intel Xeon dual-processor PC with 2 GB RAM, running Red Hat Linux. An Intel Pentium 4 PC running Windows XP and Tobii software delivers real-time gaze coordinates via TCP/IP over a 100 Mb LAN to the Linux PC. The mouse used in the study was a standard Logitech optical mouse.

Stimulus

Our testing system (see Figure 2) was designed to model the look and feel of a standard web browser. The basic components of most graphical web browsers [10, 12, 13], which were each incorporated into our system, are:

- a drop-down menu bar containing such functions as *save*, *print*, *options*, and *exit*
- individual buttons for common functions such as *back*, *forward*, *reload*, and *stop*
- an address bar to display the current page and to allow for entering a new address
- a large main section to display the contents of the current page as interpreted by the rendering engine

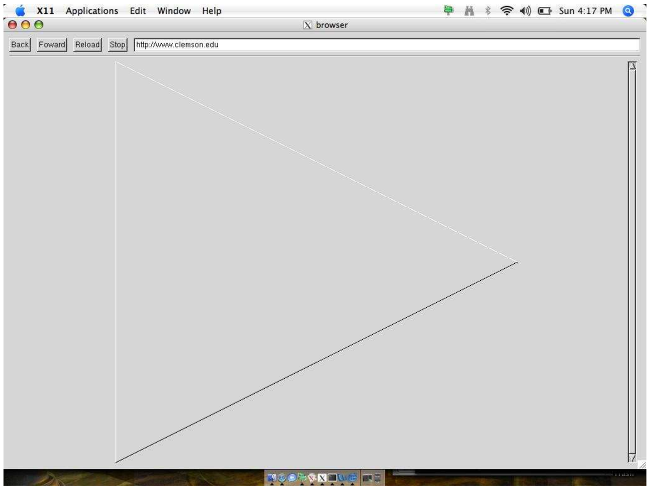

Figure 2. The system displaying a large forward arrow.

While the user interacts with our system, the only component that changes is the page content section, which updates accordingly as the user pages up or down. The content also changes completely whenever the user navigates back or forward to a different page. The different input methods we tested do not differ from one another with respect to how they affect the appearance of the system outside of the main section.

Our initial plan was to use high fidelity mock web pages and realistic search tasks, but in our pilot study, we noticed high variability within subjects with even mouse input. In the end, we decided to measure more low-level efficacy for this study and use simple arrows as seen in Figure 2.

Implementation

The current gaze point in our system is determined every 20 ms by averaging where the left and right eyes are looking. The eye gestures are set to activate as soon as gaze passes into one of the four hot zones. Through experimentation, we determined that errors were sufficiently minimized when the hot zones were placed at the middle 30% of each edge, extending from 15% of the screen width off the edge of the screen. This is illustrated in Figure 3.

Figure 3. Location of the eye gesture hot zones.

When a particular selection is made, a flag is set that prevents activation of any hot zone until gaze returns within the bounds of the screen. A timer is also set on a hot zone selection, and then that zone may not be selected again for another 100 ms. This delay prevents any unwanted selections that occur from the user's gaze rapidly drifting in and out of the hot zones.

For the mouse input, the *back* and *forward* buttons were used for selecting those commands, and clicking on the scroll bar anywhere above or below the position indicator was used to select *page up* and *page down* respectively.

Subjects

Sixteen participants (11 male, 5 female), ranging from ages 20 to 23 (mean 21.2), were recruited for the study from the computer science department of Clemson University. Each of the users possessed at least three years of web browsing experience and went online more or less daily. Five of the participants had previous experience with eye tracking technology but not with our eye gesture system. Of the sixteen participants, the data for one male user had to be discarded due to tracker error.

Experimental Design

A within-subjects, repeated measures two-by-two factorial design was applied in the study with point-and-click mouse input as the control condition and eye movement gestures using off-screen hot zones as a test condition. For each interaction technique, there was a short task consisting of ten commands and a long task consisting of thirty commands. These task lists were randomly generated on demand after it was determined in the pilot study that gesture direction had no significant impact on speed or accuracy. We measured our results in mean time and error rate per instruction.

Each participant was randomly assigned a specific order in which to perform the different trials according to a Latin Square. This was done in order to mitigate learning or fatigue effects that come with any within-subjects design.

Procedure

Participants were prescreened for normal or corrected vision and a minimum of three years of web browsing experience. On site, the subject consent forms were reviewed and completed. To help alleviate undue stress, it was emphasized to the participants that it was the system that was being tested, not them. The participants were instructed, however, to complete the navigation tasks as quickly as possible, while still attempting to maintain a high level of accuracy in their selected actions.

Participants were then seated roughly 60 cm away from the display and familiarized with the eye tracking equipment, the user interface of the system, and the interaction methods. The Tobii eye tracker was then calibrated for each participant using a nine-point calibration sequence, patterned after the calibration program provided in Tobii's own commercial ClearView software [15].

Following calibration, participants completed the four trials in their previously established order, with about fifteen to thirty seconds in between each trial in order to reduce eye fatigue. Each trial consisted of following a script of instructions displayed on the screen to navigate within and through a set of pages. Whenever an error occurred, the system alerted the user, who was then prompted to reverse the error before continuing with the current instruction. Quantitative data such as speed, errors, and eye gaze coordinates were logged by the system, and qualitative data such as observations of user frustration were noted by the evaluator.

After the trials were completed, the participants were given a brief, informal interview by the evaluator in order to get a general sense of their likes and dislikes. They were also asked to fill out a short questionnaire, with each question rated on a five-point Likert scale, which attempted to quantify the participants' opinions about the different interaction techniques. Finally, the participants were thanked for their time and cooperation and dismissed.

RESULTS

The mean times and error rates with standard deviations for mouse and eye gesture input for the short and long tasks are shown in Table 1. As could be expected, the variation between subjects for the eye gesture tasks was much greater than that for the mouse tasks with regard to both speed and accuracy.

	Mean Time per Instruction (SD)	**Error Rate per Instruction (SD)**
Short Mouse	1.90 (.21)	.67% (2.58%)
Long Mouse	1.63 (.20)	.44% (1.17%)
Short Gesture	1.13 (.80)	2.67% (5.94%)
Long Gesture	1.18 (.79)	2.22% (2.72%)

Table 1. Mean times (in sec) and error rates for each trial.

We did a multivariate analysis of variation on the data (see Table 2) and found that eye gesture input was very significantly faster than mouse input ($p < .01$). Our system was also less accurate compared to mouse input ($p < .05$). These findings supported our first hypothesis. We also hypothesized that perhaps eye fatigue would cause eye gestures to become somewhat slower as the number of instructions increased. The interaction between input method and trial length was very statistically significant ($p < .01$), supporting our second hypothesis.

	Mean Time per Instruction		**Error Rate per Instruction**	
	Wilks' λ	**p-value**	**Wilks' λ**	**p-value**
Input Method	.578	.006	.744	.045
Trial Length	.643	.015	.992	.734
Method * Length	.413	.001	.999	.889

Table 2. Significance measures for the independent variables.

We analyzed the gaze coordinate data that we collected to see if the participants were indeed operating the eye gestures as we expected. We averaged the coordinates for each command to form a composite scanpath. The four scanpaths were roughly rotationally symmetric, and they verified that the users did use eye gaze as a linear gesture. The users moved their gaze straight from the starting point to the appropriate hot zone and then returned their gaze back to the original point. This can be seen in Figure 4.

Figure 4. Composite scanpath for the forward gesture.

We also calculated mean values for responses to our survey questions, with 1 denoting "very unfavorable" and 5 denoting "very favorable." The prompts regarding our eye gesture system were:

- How would you rate how easy it is to *learn how to use* eye gestures? (4.20)
- How would you rate how easy it is to use eye gestures *given prior experience*? (4.53)
- How would you rate eye gestures compared to mouse input in terms of ease of use? (3.53)
- How would you rate eye gestures compared to mouse input in terms of speed and efficiency? (4.27)
- How would you rate eye gestures compared to mouse input in terms of overall satisfaction? (3.60)
- Overall, how would you rate the eye gesture system as an effective means of web browsing navigation? (3.93)

DISCUSSION

The objective of this study was to determine how eye gestures compared to mouse input in terms of speed, accuracy, and subjective satisfaction. Our results indicate that our eye gesture system was strongly significantly faster and slightly significantly less accurate compared to mouse input, a speed-accuracy tradeoff. Due to the participants' lack of fluency with eye gestures, the variation in results for the eye gesture tasks was much greater than that for the mouse tasks with regard to both speed and accuracy. Trial length and the interaction between trial length and input method did not seem to have any effect on the error rate.

We did not expect the significance of trial length or the interaction between trial length and input method to be nearly as high as it turned out. Mouse input was surprisingly faster across the board in the long task compared to the short task. We postulate that perhaps this is due to mouse users "getting into a groove" (as one participant put it), and that the effects of this are not linear but exhibit diminishing returns as the task length increases.

We expected the results that indicated that eye fatigue caused the long eye gesture task to take longer per instruction than the short task. Like with the mouse input, we also expect the curve of selection time over number of instructions to be logarithmic and not linear. Because the effect on speed of increasing instructions for eye gestures ran counter to the effect for mouse input, the resulting interaction between input method and trial length was remarkable. Although activation time by eye gestures was still much faster than mouse input in the long task, this implies that mouse input could potentially be a better choice for extremely long, sequential tasks or other situations where eye fatigue would be a major issue.

We examined the errors that were produced using our eye gesture system. Some of these errors occurred because the user occasionally selected the wrong command entirely, analogous to the errors that occurred with mouse input. Because of the speed of eye gestures, it is likely that a cognitive error with eye gesture input could have been caught before selection in a corresponding mouse input task. The second type of eye gesture error was reselection errors, where a user would accidentally select the same command twice instead of moving on to the next instruction. We established 100 ms as a good parameter for reselection delay, but increasing this value would reduce the number of reselection errors at the potential cost of speed.

In a domain such as HCI, user satisfaction is arguably every bit as important as objective measures of performance. Our system rated better than neutral for each subjective criterion that we measured. In particular, the participants found the eye gesture system both subjectively quick, and easy to learn and use. Comments that were given included:

- "Really cool, an amazing experience!"
- "great innovation"
- "so much faster and effortless"
- "I would like to see this expanded [beyond the four commands]"
- "I wish I had this at home right now!"

Most participants responded well to our system, operating roughly twice as fast with eye gestures while making no more than one error each. However, two participants struggled mightily, taking up to 50% longer than the mouse input and making multiple errors. If their data is discarded, the mean time per instruction becomes .89 for the short task and .95 for the long task, while the error rates become 1.34% for the short task and .88% for the long task. These are marked improvements, but we include the data for all fifteen measured participants in this study. This is because we do not know if maybe the eye tracker technology was not registering these participants well, or if simply these participants were not able to become accustomed to the unfamiliarity of using eye gaze as input. Furthermore, we do not know the extent to which this unfamiliarity would decrease with increased experience.

To reiterate, eye gestures were significantly faster even after considering that the correction time for any errors factored into the total time for completion. Thus, we contend that the immense speed benefit gained by use of our system far outweighs the resulting increase in errors. It is possible that the results could be even more favorable in the future given improved technology or increased user experience with eye tracking.

Now that we have established the benefits of eye gestures at a lower level, the next logical step is to test the system in a more real-world setting. The experimental tasks would be more representative of everyday web browsing tasks such as information search, message board reading, and form field entry. This could be done with high fidelity mockups or even better, by incorporating the eye gesture system into an existing web browser. Should this prove successful, we would then examine other domains outside web browsing that should be conducive to our system.

In addition, we would examine the effects of extending our linear gesture set to include one or more corners, as Isokoski did [6]. There are some common but non-navigational tasks that might be good candidates for diagonal gestures, including *print* and *save*. Another common task is following a hyperlink; we could test how a context sensitive *forward* gesture originating from a hyperlink would compare to hyperlink selection based on dwell time. Another idea is for the user to be able to control magnitude, such as how far the page scrolls up or down using one eye gesture. This could be done by dwelling or more interestingly, by the distance off the screen that the gesture travels.

We have reason to believe that the benefits of our system would be maintained or even improved with these modifications. Using the mouse for navigation in a web task is often disruptive; the user might have to move his or her hand to the mouse even to begin. When a command is selected, the user often must also return the point back to its original location, such as by clicking at the appropriate location within a text box. These considerations add to the time it would take to perform real web navigation compared to experimental sequential selection using the mouse.

A good reason why diagonal gestures for commands such as *print* and *save* might be useful is that the mouse-based equivalents involve clicking on a menu bar and selecting the appropriate action from a linear menu. Fitts' law and the keystroke level model would each predict a large speed increase over mouse input for these commands. Perhaps most beneficial of all, since navigation commands are usually spaced out more than in an experimental setting, eye fatigue should be much less of a problem in a real-world web browsing context.

CONCLUSION

We have presented a new system for web navigation based on eye gestures that leverages the speed benefits of both eye gaze input and linear gestures. Web navigation is a domain that suggests natural mapping of web space direction to input direction. This allowed our eye gesture system to utilize the unique concept of off-screen hot zones that do not depend on fixation time for selection and yet avoid the Midas Touch problem.

We evaluated our eye gesture system versus point-and-click mouse input in short and long navigation tasks. We found that our system was strongly significantly faster than mouse input, while suffering from a slightly significantly greater error rate. We also found that eye fatigue caused eye gestures to become slower per instruction as the number of instructions increased in a sequential, rapid-fire task. Overall, the participants seemed to prefer our system to traditional mouse input.

We conclude that the speed benefit is well worth the accuracy tradeoff since errors in web navigation can be corrected easily and harmlessly. Even with the increase in errors leading to more time spent correcting these errors, our system was still much faster than mouse input. If we exclude the data for the two outliers that experienced considerable difficulty, the time taken for eye gestures per instruction approaches half the time taken for mouse input.

REFERENCES

1. Ashmore, M. and Duchowski, A.T. Efficient Eye Pointing with a Fisheye Lens. In *Proc. of Graphics Interface*, CHCCS (2005).
2. Duchowski, A.T. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, Inc. (2003).
3. Dulberg, M., Amant, R., and Zettlemoyer, L. An Imprecise Mouse Gesture for the Fast Activation of Controls. In *Proc. INTERACT '99*, IOS Press (1999).
4. Fono, D. and Vertegaal, R. EyeWindows: Evaluation of Eye-Controlled Zooming Windows for Focus Selection. In *Proc. CHI '05*, ACM Press (2005), 151-160.
5. Hayhoe, M., Ballard, H., Triesch, J., Shinoda, H., Aivar, P., and Sullivan, B. Vision in Natural and Virtual Environments. In *Proc. ETRA '02*, ACM Press (2002), 7-13.
6. Isokoski, P. Text Input Methods for Eye Trackers Using Off-Screen Targets. In *Proc. ETRA '00*, ACM Press (1990), 15-21.
7. Jacob, R.J.K. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In *Proc. CHI '90*, ACM Press (1990), 11-18.
8. Lenhart, A. et al. The Ever-Shifting Internet Population. Pew Internet & American Life Project (2003).
9. Majaranta, P. and Raiha, K.J. Twenty Years of Eye Typing: Systems and Design Issues. In *Proc. ETRA '02*, ACM Press (2002), 15-22.

10. Microsoft Internet Explorer. http://www.microsoft.com/windows/ie/default.asp.
11. Moyle, M. and Cockburn, A. A Flick in the Right Direction: A Case Study of Gestural Input. In *Behaviour and Information Technology*, 24(4), Taylor and Francis Ltd. (2005), 275-288.
12. Mozilla Firefox. http://www.mozilla.org/products/firefox/.
13. Opera Web Browser. http://www.opera.com/.
14. Ou, J., Oh, L.M., Yang J., and Fussell, S.R. Effects of Task Properties, Partner Actions, and Message Content on Eye Gaze Patterns in a Collaborative Task. In *Proc. CHI '05*, ACM Press (2005), 231-240.
15. Tobii Technology AB. Tobii ET-17 Eye-tracker Product Description (v1.1), 2003.
16. Vertegaal, R. Attentive User Interfaces (Editorial, Special Issue on Attentive User Interfaces). *Comm. of the ACM*, 46(3), ACM Press (2003), 30-33.
17. Zhai, S., Morimoto, C., and Ihde, S. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proc. CHI '99*, ACM Press (1999), 246-253.