Empirical Evaluation of a Novel Gaze-Controlled Zooming Interface

Marc Pomplun, Nada Ivanovic, Eyal M. Reingold, and Jiye Shen

Department of Psychology, University of Toronto, 100 St. George St., Toronto, Canada M5S 3G3

ABSTRACT

In the present paper, we present a novel gaze-controlled interface. It allows the user to magnify and inspect any part of an image by just looking at the part in question and subsequently shifting gaze to another window. No manual input is required to control this process. The interface was empirically evaluated in a multi-session experiment employing a comparative visual search tasks that required several steps of zooming in and out of a search display. Each participant's performance was assessed separately for using gaze control and using a mouse as the input device and compared between conditions and across sessions. The results demonstrate that participants' performance with the gaze-control interface is quite comparable with a standard mouse input device and that using the gaze-control interface can be learned very quickly.

1. INTRODUCTION

When we interact with computers today, we typically feed input into the computer with our hands controlling a mouse or keyboard, while we employ our eyes to examine the computer's output on a screen. Intuitively, it seems beneficial to this process to also use the eyes as input devices - by measuring eye movements indicating shifts of attention – and thus eliminate the motor control loop for arm and hand. On the one hand, eye movements can indeed be used as a fast and direct means of controlling a computer, but on the other hand, they are not completely under conscious control, so when using gaze-control interfaces we may sometimes trigger actions unintentionally (the "Midas-Touch Problem"). It is a crucial point in the design of gaze-controlled interfaces to minimize such undesirable effects.

Several different kinds of gaze-controlled interfaces have been developed for handicapped persons (e.g., Frey, White & Hutchinson, 1990; Levine, 1981; Parker & Mercer, 1987; Spaepen & Wouters, 1989). The most widely employed paradigm in this field of research is "typing by eye", which enables the user to type text by fixating and thereby "pressing" keys on a virtual keyboard displayed on a computer screen (e.g. Stampe & Reingold, 1995). Although gaze-controlled interfaces are particularly useful for handicapped users, their applicability can be extended to facilitate human-computer interaction in normal population as well. A promising candidate for gaze control is one of the classic interface functions, namely zooming in and out to view an image (for instance, a map, diagram, or camera picture) at different resolutions.

Goldberg and Schryver (1995) proposed a gaze-controlled zooming interface analyzing the spatial and temporal distribution of preceding fixations to determine whether the user intends to zoom in, zoom out, or keep the current zoom level. Their methodology first collects samples of eye-gaze locations looking at the stimuli just prior to the user's intent to zoom, which are subsequently broken into temporal snapshots and connected into a minimum spanning tree, and then clustered according to user-defined parameters. A multiple discriminant analysis that uses cluster size, gaze position and pupil size statistics is then performed to formulate optimal rules for assigning observations into zoom-in, zoom-out or no-zoom conditions. Goldberg and Schryver did not report the results of an actual implementation of their proposal, but such an interface would inevitably have two drawbacks: First, the analysis of fixation patterns would require a sufficient number of fixations to determine the user intent. Since people make about two to four fixations per second, there would be a dragging delay of at least one or two seconds between a change in the user's intent and the system's response. Second, given the variability of eye-movement patterns across different situations and users, the system would perform a certain proportion of misinterpretations, causing actions that are unintended by the users who lose control over the system.

When implementing a different kind of interface, Jacob (1991) avoided such problems by dividing the computer screen into two windows presented side by side. Window *A* was a geographic display of ships, and window *B* showed some information about one of the ships, namely the last one the user had looked at in window *A*. This way, users could select one of the ships in window *A* by means of eye fixations and then read information about that ship in window *B*. Originally, Jacob (1991) predicted that computer commands triggered by eye movements would be difficult for subjects to get accustomed to, since eye movements are naturally used to screen the environment and are relatively hard to control. However, the results emphasized the advantage of using eye movements to activate system commands, in comparison to the more standard, button-press mouse device.

In the present paper, we used a similar two-window design to build a convenient and reliable gaze-controlled zooming interface. Our implementation employs two square windows *A* and *B* of the same size. Window *A* on the left shows the whole image. When users look at window *A*, a square, highlighted selection marker follows their eye movements through the picture. As soon as users switch their gaze to window *B* on the right ("zoom in"), it displays a magnified part of the image – the last one that was selected in window *A* by looking at it for at least 120 ms. As long as users inspect window *B*, the screen does not change, i.e., the selection marker in window *A* remains in the position corresponding to the image part shown in window *B* (see Figure 1). This enables users to switch back to window *A* ("zoom out") without losing their bearings. The interface only provides two magnification levels. More levels would require either more windows or gaze-triggered zoom selection buttons at the expense of resolution or efficiency respectively. However, our interface allows quick and reliable zooming operations that are suitable for a variety of tasks, given an appropriate choice of magnification factor between the two windows.

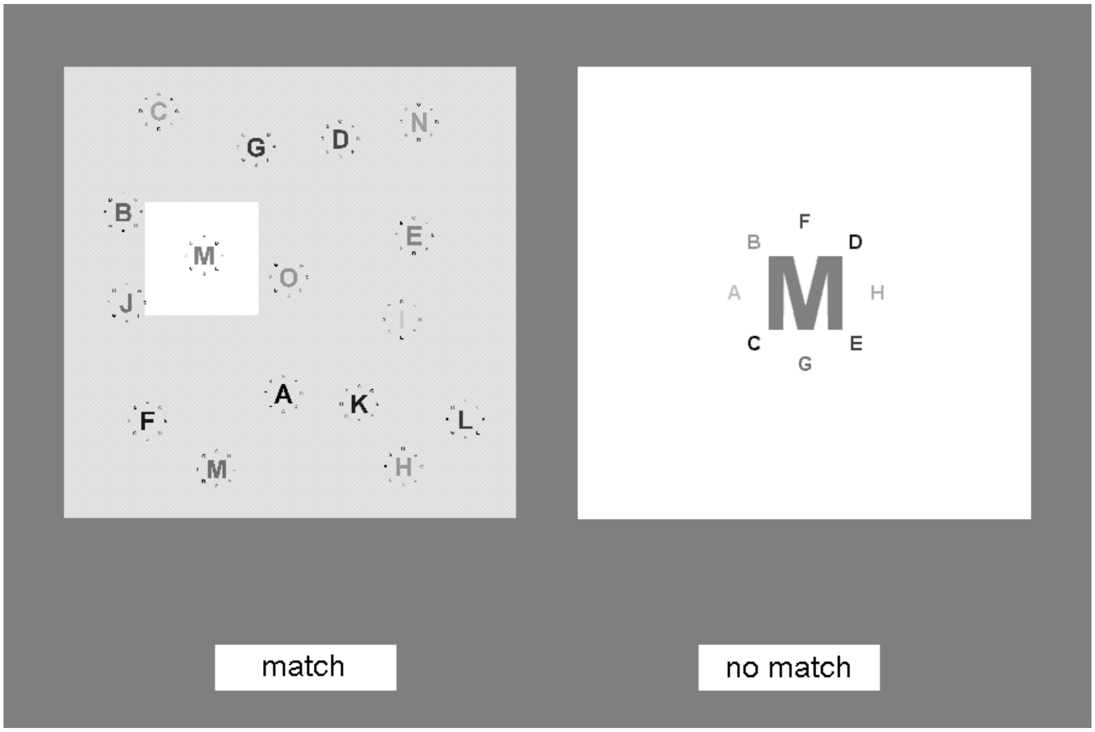

Figure 1: Layout of the zooming interface. While users inspect the left window showing an overview image, a highlighted square is continuously centered on their gaze position. As soon as the gaze switches to the right window, the area highlighted prior to the eye movement is shown magnified in the right image. The two fields labeled "match" and "no match" can be triggered by looking at them for a minimum duration and were used to register participants' response in the experimental task.

In order to evaluate the efficiency of gaze control with regard to our interface, we had participants perform in a dual-scale variant of comparative visual search (see Pomplun, Reingold & Shen, in press; Pomplun, Sichelschmidt, Wagner, Clermont, Rickheit & Ritter, 2001) requiring multiple zoom-in and zoom-out operations in each trial. The use of gaze control was compared to the use of a standard input device - a computer mouse – in the same task. To investigate practice effects, participants' performance was assessed in multiple sessions for each of the two control modes. Since all participants were used to controlling computers with a mouse, we expected only small practice effects for the mouse condition, but large effects for the gaze condition.

2. METHOD

2.1 Participants

Four undergraduate students from University of Toronto and York University participated in the study. All of them had normal or corrected-to-normal visual acuity and had no color vision defects. They were aware of the fact that the purpose of the study was to compare two different user interfaces.

2.2 Apparatus

Eye movements were recorded with the SR Research Ltd. EyeLink system, which is a video-based eye tracker operating at a sampling rate of 250 Hz (4 ms temporal resolution) and measures a participant's gaze position with an average error of less than 0.5 degrees of visual angle after a 9-point calibration at the beginning of the experiment. Before each trial, a 1-point re-calibration was performed to compensate for possible shifts of the system's headset. By default, only the participant's dominant eye was tracked in our study. The EyeLink system uses an Ethernet link between the eye tracker and display computers for real-time saccade and gaze-position data transfer. Stimuli were presented on a 19-inch Samsung SyncMaster 900P monitor with a refresh rate of 120 Hz and a screen resolution of 800 by 600 pixels. A gaze-contingent marker was implemented, which followed the participant's eye movements with an average delay of 14 ms.

2.3 Stimuli

On each trial, two square windows with a length of 10° appeared to the left and right of the center of the screen (see Figure 1). While the right window *B* was initially blank, the left window *A* contained a random distribution of 16 large letters with diameters of about 0.5°. All of these letters were different and had distinct colors, except for one pair of identical letters that also had the same color. Each of the large letters was surrounded by a set of eight circularly arranged small letters A to H (diameter 0.07°) in distinct colors. The positions of the small letters A to H and their colors were randomized for each large letter and across trials. Participants had to decide whether the two identical large letters shared an identical small letter in the same color - for example, a green B - which was the case in 50% of the trials. The task required zooming operations, because due to the screen resolution of 800 by 600 pixels, the small letters were illegible in window *A*. In order to magnify part of the search display, participants could move their gaze to the right window *B*. During such a saccade, a four-fold magnification of the area selected during the last fixation in window A was painted into window *B*. Thus, in window B, the large letters had diameters of 2° and the small ones diameters of 0.28 degrees, which made them clearly legible. After participants switched back to window *A*, window *B* was blanked again. Since it was virtually impossible to memorize all letter-color combinations at a time, many steps of zooming in and out were required to complete the task.

2.4 Procedure

Participants performed in two experimental conditions: In the "gaze" condition, participants controlled the computer with their gaze as described above, and in the "mouse" condition, they controlled the selection marker in window *A* with a computer mouse and pressed a mouse button to have the selected area magnified and displayed in window *B*. Once they determined the match or non-match, participants selected a corresponding key located in the bottom of the screen and made their response by clicking on it (mouse condition) or fixating on it for at least 500 ms (gaze condition). Each participant completed six sessions in intervals of approximately 48 hours. Each session consisted of a block of mouse trials and a block of gaze trials, each of which included 50 trials with short breaks after every tenth trial. The order of experimental conditions was counterbalanced across participants. Response times, error rates, and the number of magnifications per trial were taken as efficiency measures and compared across the two experimental conditions and across sessions.

3. RESULTS AND DISCUSSION

Each participant's response time, error rate, and number of magnifications (zoom-in operations) were recorded as efficiency measures across the two interface conditions and the six sessions (see Figure 2). To reduce noise, only data of the "no match" trials with correct response were considered, because in these trials, participants had to search through all eight small letters surrounding each of the matching large letters. Due to the small number of participants, no statistical analysis of the data was conducted.

Generally speaking, the results did not reflect strong practice effects on either of the two input modalities. Participant #1 exhibited faster response times in the gaze condition in the later sessions, but her error rate increased correspondingly, suggesting a potential speed-accuracy trade-off in performance. Interestingly, participant #2 presumably had short-term practice effects in the gaze condition. She produced longer response times and more magnifications in sessions 1, 3, and 5, in which the gaze condition trials preceded the mouse condition trials, than in sessions 2, 4, and 6, in which the mouse condition trials preceded the gaze condition trials, leading to a zigzag pattern in the corresponding diagrams. For participant #3, the only clear indication of a practice effect was a very slight negative slope in the number of magnifications along the six sessions in both conditions. Finally, participant

#4 demonstrated a considerable, almost linear practice effect across sessions resulting in approximately the same efficiency gain in both conditions. His error rate and number of magnifications, however, did not vary considerably over time. Taken together, very minimal practice effects were observed and were mostly similar for the gaze and the mouse interfaces. This indicates that practice effects were largely attributable to task demands rather than to differences between the two interfaces.

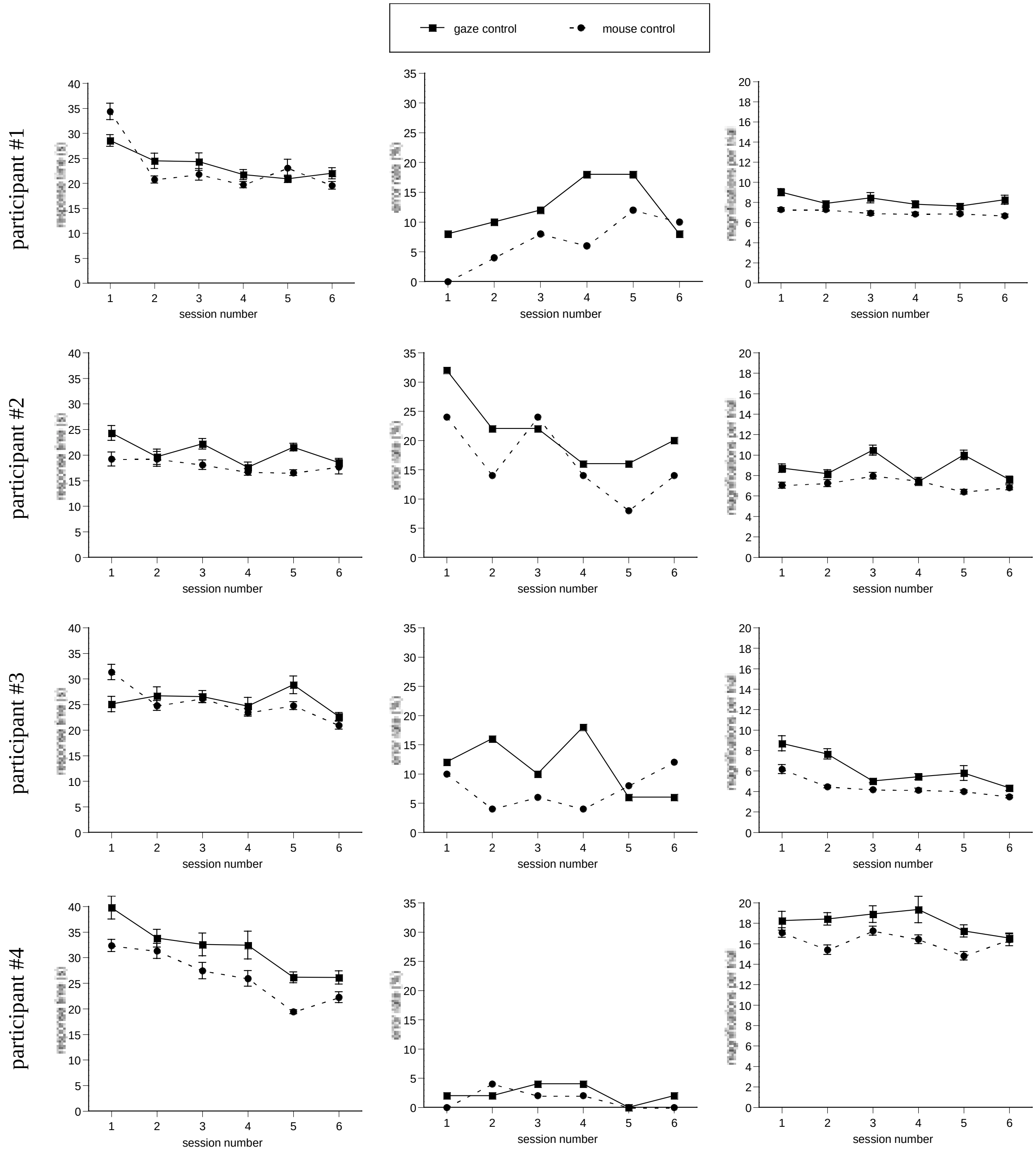

Figure 2: Response time, error rate, and number of magnifications per trial as functions of session number for each of the four participants and each interface condition (gaze vs. mouse). Error bars indicate standard errors within each session.

Comparing the absolute efficiency for gaze and mouse control, we find that, on average, more magnifications per trial were triggered in the gaze condition (10.28) than in the mouse condition (8.69). There are two potential factors contributing to this difference: First, in some trials, the gaze-position measurement may have been imprecise due to headset shifts, so the interface did not magnify the letter that the participant looked at, but instead magnified a neighboring item. Second, similar to the Midas-Touch Problem, when inspecting a particular letter in window *A* and intending to magnify it, participants may unintentionally have looked at another letter before switching to window *B*, so the wrong letter was magnified. When designing the gaze interface, we empirically tested a range of duration thresholds for triggering the magnification function in order to find the best compromise for precise control and quick responsiveness. The minimum dwell time of 120 ms was determined to provide both features appropriately, but obviously cannot completely eliminate unintentional selections.

The sporadic occurrence of unintentional selections may have contributed to the finding that using the gaze interface led to more incorrect responses (11.83%) than using the mouse interface (7.92%). Since all items in the comparative search task looked somewhat similar – large letters surrounded by small ones – in some cases participants probably did not notice when an unintended item was magnified and consequently produced an incorrect response. Conceivably, this might not happen so easily for different stimuli, for example, when applying the zooming interface to pictures or video streams showing real-world scenes, and therefore should not be considered a general shortcoming of the interface.

Finally, response times were slightly longer in the gaze condition (25.44 s) than in the mouse condition (23.19 s). This difference was smaller than 10%, demonstrating that, despite occasional unintentional selections for magnification, using the unfamiliar gaze interface and wearing the eye tracker headset did not considerably slow down participants' task performance as compared to the familiar mouse interface. Given that practice effects in the current experiment were small and similar across conditions, it seems that gaze control can be considered a valid alternative to standard control methods in suitably designed interfaces such as the one described here.

The findings of the current study suggest that the potential of gaze-controlled interfaces has not yet been fully realized. Contrary to Jacob (1991), who predicted that computer commands triggered by eye movements would be difficult for subjects to get accustomed to, and Goldberg and Schryver (1995), who suggested that a gaze-control interface is an inappropriate substitution for a mouse device, our interface shows that eye movements can in fact substitute manual input. Further empirical research is necessary to develop algorithms that can determine the users' intent from their spatiotemporal eye-movement patterns more reliably and hence minimize the Midas-Touch Problem. Appropriately designed gaze-controlled interfaces could then more and more replace conventional interfaces in areas of application where hands-free interaction with machines is desirable. Moreover, advanced gaze-controlled interfaces for handicapped people could considerably improve many aspects of their lives.

REFERENCES

Frey, L. A., White, P. K. & Hutchinson, T. E. (1990). Eye-gaze word processing. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(4), 944-950.

Goldberg, J.H. & Schryver, J.C. (1995). Eye-gaze determination of user intent at the computer interface. In J.M. Findlay, R. Walker & R.W. Kentridge (Eds.), *Eye Movement Research: Mechanisms, Processes, and Applications.* Amsterdam: Elsevier.

Jacob, R.J.K. (1991). The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems, 9* (3), 152-169.

Levine, J. L. (1981). *An Eye-Controlled Computer.* Yorktown Heights: IBM Thomas J. Watson Research Center, Res. Rep. RC-8857.

Parker, J. L. & Mercer, R. B. (1987). The ocular communicator: a device to enable persons with severe physical disabilities to communicate using eye movements. *Exceptional Child*, 34(3), 221-226.

Pomplun, M., Reingold, E.M. & Shen, J. (2001, in press). Investigating the visual span in comparative search: The effects of task difficulty and divided attention. *Cognition*.

Pomplun, M., Sichelschmidt, L., Wagner, K., Clermont, T., Rickheit, G. & Ritter, H. (2001). Comparative visual search: A difference that makes a difference. *Cognitive Science, 25* (1), 3-36.

Spaepen, A. J. & Wouters, M. (1989). Using an eye-mark recorder for alternative communication. In A. M. Tjoa, H. Reiterer and R. Wagner. (Eds.), *Computers for Handicapped Persons.* Vienna: R. Oldenbourg

Stampe, D.M. & Reingold, E.M. (1995). Selection by looking: A novel computer interface and its application to psychological research. In J.M. Findlay, R. Walker & R.W. Kentridge (Eds.), *Eye Movement Research: Mechanisms, Processes, and Applications.* Amsterdam: Elsevier.