Compensating for the Eye-Hand Span Improves Gaze Control in Human-Computer Interfaces

Mei Xiao[1], Judelande R. Hyppolite[1], Marc Pomplun[1], Sindhura Sunkara[1], Elena Carbone[2]

[1]Department of Computer Science, University of Massachusetts at Boston
100 Morrissey Boulevard, Boston, MA 02125-3393, U.S.A
[2]Department of Psychology, University of Bielefeld
Universitaetsstrasse 25, 33615 Bielefeld, Germany
Email: meixiao@cs.umb.edu, marc@cs.umb.edu, judhy@cs.umb.edu, ssunkara@cs.umb.edu, elena.carbone@uni-bielefeld.de

Abstract

Advances in the measurement of eye movements make it possible to construct gaze-controlled interfaces for physically challenged people and for enhanced human-computer interaction. The use of hybrid interfaces that combine manual and gaze control has the advantage of being effortless, quick, and intuitive: The user selects an item by looking at it and confirms this selection by pressing a button. Earlier studies have shown task completion times to be slightly longer for gaze-controlled interfaces compared to mouse-controlled interfaces. From these results we derived the hypothesis that the efficiency of conventional hybrid interfaces is limited by the fact that the users are already looking for the next object when pressing the button, causing an unintended selection. In the first experiment we investigated if such an eye-hand span really exists and which factors affect its size. In the second experiment we modified the gaze-controlled interface and showed that compensating for the eye-hand span substantially improves the subjects' performance.

1 Introduction

Although most of us are extremely experienced in using a computer mouse for the selection of objects in a display, the idea of selecting objects by simply looking at them is highly appealing. Eye movements are performed even faster and with less effort than mouse movements and such a selection by gaze would be even more beneficial for physically challenged people. The enhancement of eye movement measuring devices during the last decades has led to a growing interest in gaze-controlled interfaces (Frey, White & Hutchinson, 1990; Hyskykari, 1997; Jacob, 1991; Jacob, 1995; Levine, 1981; Parker & Mercer, 1987). The applications range from virtual typewriters to benefit persons with disabilities (Stampe & Reingold, 1995) to zooming interfaces for reading maps and diagrams to be used by an even much wider population (Goldberg & Schryver, 1995).

One problem that has to be solved by all of these gaze-controlled interfaces is that eye movements are not entirely under conscious control, and users do not just look at the to-be-selected objects but additionally scan the display to find them. Thus, a purely gaze-controlled human-computer interface would lead to inadvertent object selections. Usually, this problem is solved by setting a minimum dwell time for triggering gaze-controlled actions, but this procedure reduces the efficiency of the interface. Therefore, the research reported here focuses on hybrid interfaces that combine manual and gaze control. Such interfaces are similar to the mouse interface in that they allow the user to select an item from a graphical display – by looking at it – and confirm that selection by pressing a button. Hybrid interfaces, therefore, have the advantage of an effortless, quick and intuitive selection by moving one's eyes, and they avoid the problem of inadvertent triggering of actions.

Since eye movements require much less effort and time than mouse movements, we were surprised by the results of our previous studies that directly compared the efficiency of gaze- and mouse-controlled interfaces (e.g., Pomplun, Ivanovic, Reingold, and Shen, 2001; Sunkara & Pomplun, 2003). Across different tasks and interfaces, subjects were more or equally efficient using a mouse-controlled interface as compared to a gaze-controlled one. From observations in these studies we derived the hypothesis that the efficiency of conventional hybrid interfaces is limited by the fact that the users' eyes are already ahead of their manual input. In analogy to the eye-voice span in vocal reading, we termed this temporal discrepancy the eye-hand span in gaze-controlled human-computer interaction. We conducted Experiment 1 to obtain empirical evidence for the existence of the eye-hand span and to identify the factors that determine its size. In Experiment 2, we demonstrated that hybrid interfaces can be improved by compensating for the eye-hand span.

2 Experiment 1: Investigating the Characteristics of the Eye-Hand Span

In order to study the eye-hand span in gaze-controlled human-computer interaction we used stimuli showing letters of the English alphabet arranged on a 5×5 grid. The participants' task was to select the letters in alphabetical order by fixating on them and confirming each selection by pressing a button. We investigated two potential factors that might affect the scanning behavior and thus in turn the size of the eye-hand span, namely the order of the letters and the spacing between them (see also Hyppolite, Carbone & Pomplun, 2004).

2.1 Method

Apparatus and Stimuli. Eye movements were recorded with the SR Research Ltd. EyeLink-II system, which operates at a sampling rate of 500Hz and measures a participant's gaze position with an average error of 0.5 degrees of visual angle. Stimuli were presented on a 21-inch Dell Trinitron monitor with a refresh rate of 85Hz and a screen resolution of 1152 by 864 pixels. Each stimulus was composed of the letters A to Y arranged on a square 5×5 grid. The letters were black on a gray background, and each of them subtended a visual angle of about one degree. According to the variation of the factors alphabetical order and spacing of letters, the letters were either ordered alphabetically or scrambled, and the distance between neighboring letters was either large (4.5 degrees) or small (2.6 degrees).

Design and Procedure. We investigated the effect of the factors order of letters (ordered versus scrambled) and spacing of letters (large distance versus small distance). The resulting four types of stimuli were shown in blocks of five images each, preceded by a practice image, leading to a total of 24 trials. The images were presented one at a time and the participants were instructed to look at each letter in alphabetical order and confirm the fixation by pressing a designated button on a game pad. Each trial was terminated by the subjects' pressing another designated game pad button. The subjects were asked to complete each trial as quickly and accurately as possible; however, there was no time limit.

Participants. Ten students from the University of Massachusetts at Boston were tested individually. All participants had normal or corrected-to-normal vision. They were naïve with respect to the purpose of the study and were paid for their participation.

2.2 Results and Discussion

Figure 1 depicts the relationship between the dwell time on a letter (the viewing interval duration) and the time between starting to look at a letter and pressing the confirming button (the button delay) for each participant and each of the four conditions. The dotted line indicates the ideal confirmation time (i.e., at the center of the viewing interval), and the solid line indicates confirmation at the end of the viewing interval. Data points above the solid line correspond to unintended selections, because in those cases the eyes have already moved away from the intended selection target when the button is pressed. With a few exceptions, for all participants and all conditions, the confirmation of the selection occurs at the very end of the viewing interval, often even after the eyes have already moved on to the next target, which indicates a significant eye-hand span.

A two-way Analysis of Variance (ANOVA) of button delay with the factors ordering and spacing revealed that there was no significant main effect by either factor, and their interaction was not significant either, all $Fs(1; 9) < 5.07$, $p > 0.05$. Figure 1 suggests that across conditions the button delay was approximately proportional to the dwell time. This was confirmed by an analogous ANOVA of the quotient of button delay and dwell time, which showed no significant main effects or interaction of the factors ordering and spacing, all $Fs(1; 9) < 1$. Therefore, we used this quotient in Experiment 2 as the basis for estimating the size of the eye-hand span independently of individual subjects and tasks in our improved gaze-controlled interface.

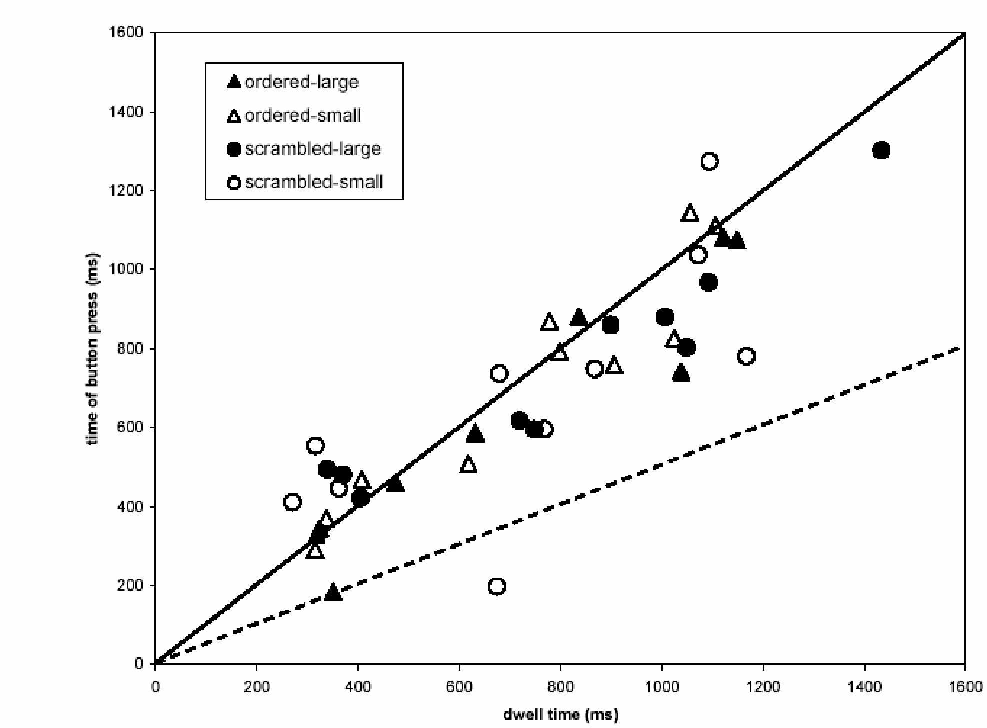

Figure 1: Relationship between dwell time (horizontal axis) and time of button press (vertical axis) across participants and conditions.

3 Experiment 2: Improving and Evaluating a Gaze-Controlled Interface

The results of Experiment 1 encouraged us to modify the standard hybrid interface to account for the users' eye-hand span. Our idea was to improve the interface by letting it estimate the range of the user's eye-hand span and consider it when the user manually confirms a gaze selection. The improved interface should not confirm the user's current gaze position, but the gaze position that was recorded a specific time before the button press. This time difference between confirmed gaze position and button press should depend on the estimated eye-hand span in order to compensate for it and thereby facilitate the interface control. Since the quotient of button delay and dwell time had been found to be relatively constant, it was used as the basis for this compensation. We conducted numerous pilot studies using a variety of approaches and parameters, and we determined the following algorithm to be most effective: If the button click press occurs during a fixation, and the time from the start of the fixation to the button press is shorter than one-ninth the duration of the previous fixation, then this button click is assumed to confirm the previous fixation position, otherwise the button click is assumed confirmed the current fixation position. If the button click occurs during a saccadic eye movement, then it is assumed to confirm the fixation immediately before that saccade.

We conducted Experiment 2 to test the efficiency of the modified hybrid interface against the initial (standard) hybrid interface and the mouse interface in a pair-detection task. We expected the performance to be better for the modified interface than for the standard interface – and possibly even better than for the mouse interface - because of the compensation for the eye-hand span.

3.1 Method

Apparatus and Stimuli. The apparatus was the same as in Experiment 1. We generated 12 different stimulus objects (four different geometrical shapes in three different colors). At the start of every trial, subjects were presented with a grid of 6×4 cells subtending the entire display. The grid was randomly filled with the 12 objects so that each distinct object appeared exactly twice, resulting in 12 randomly distributed pairs of identical objects.

Design and Procedure. We investigated the efficiency of the standard interface, the modified interface, and the standard mouse interface. In a training session prior to the experimental trials, participants were trained in using these different interfaces. In the subsequent experimental session, subjects performed 12 trials using the initial interface, 12 trials using the modified interface, and 24 trials using the mouse interface. The order of the three conditions was randomized and counterbalanced across subjects. The total time it took the participants to find all pairs in a stimulus image and indicate their positions with one of the interfaces was used as the measure of performance for that interface. In every gaze-interface trial, the participants' task was to select one object by looking at it, confirm the selection by pressing a game pad button, select its identical counterpart and complete the pair's selection by pressing the button again. The mouse interface worked analogously and intuitively: Objects were selected by positioning the mouse pointer on the desired object and pressing the left mouse button. If the selected objects were actually identical, they were deleted from the display; otherwise they remained, and the subject had to make new selections. A trial was finished when all pairs had been selected. As in Experiment 1, the subjects were asked to complete each trial as quickly and accurately as possible, and there was no time limit.

Participants. Ten students from the University of Massachusetts at Boston were tested individually. All of them had normal or corrected-to-normal vision, were naïve with respect to the purpose of the study, and were paid for their participation.

3.2 Results and Discussion

The average median of all the subjects' trial completion times for the old gaze interface, the new gaze interface, and mouse interface was 18.79s, 16.01s, and 16.42s, respectively (see Figure 2). To compare the efficiency of the three interfaces, we conducted three Bonferroni-adjusted paired t-tests for the median completion times across subjects in the three experimental conditions. The median trial time for the old gaze interface was significantly longer than for the new gaze interface, $t(9) = 4.01$, $p < 0.01$, and for the mouse interface, $t(9) = 2.87$, $p < 0.05$, while the difference between the new gaze interface and the mouse interface did not reach significance, $t(9) < 1$. These results indicate that modifying the gaze interface indeed improved the participants' performance. On average, the new gaze interface enabled the subjects to perform even faster than with the mouse interface, although this difference was not statistically significant.

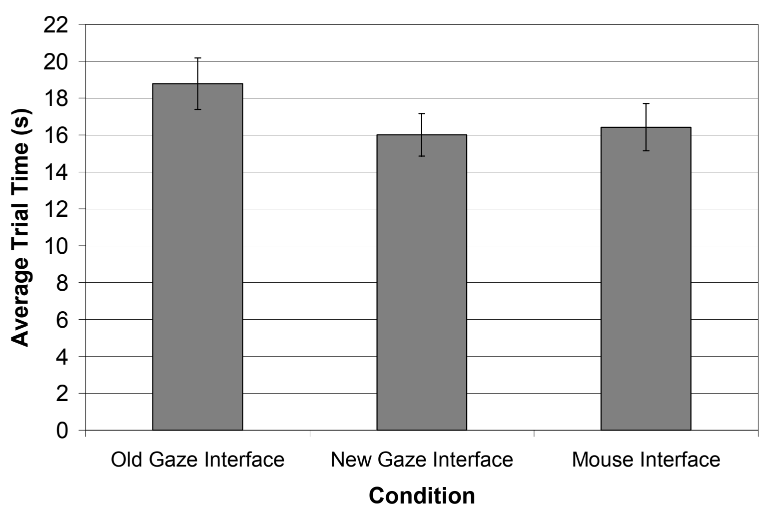

Figure 2: Mean task completion times and standard errors for the initial and the modified gaze interfaces as well as the mouse interface

4 General Discussion

In this paper we investigated the possibility that the efficiency of conventional hybrid interfaces is limited by the fact that the users are already looking for the next object when pressing the button, causing an unintended selection.

This eye-hand span was analyzed in two experiments. Experiment 1 established the existence of the span and showed that its size depended neither on the difficulty of the particular task nor on saccade amplitude. We found that the absolute value of the eye-hand span depended on the individual subject's average dwell time (i.e., duration of viewing intervals) on each letter. Based on these findings, we modified the original gaze-controlled interface to compensate for the eye-hand span and compared the participants' efficiency for the modified gaze-controlled interface, the original interface, and a standard mouse interface in Experiment 2. The results of Experiment 2 established the superiority of the modified interface: The median task completion time for the modified interface was shorter than for the other two interfaces, although its advantage over the mouse interface was statistically not meaningful.

Earlier research has already identified the benefits of gaze-controlled interfaces in human-computer interaction (e.g., Pomplun et al., 2001; Sunkara & Pomplun, 2003), but the results of the current experiments show that the efficiency of these devices can be further improved. The present study demonstrates the large extent to which the use of hybrid gaze-controlled interfaces is adversely affected by the eye-hand span, and it presents a simple mechanism for effectively attenuating the impact of this span. According to the results of Experiment 1, and partially confirmed through Experiment 2, this mechanism promises to be appropriate for the majority of computer users and for various types of tasks. The finding that the improved gaze interface can be used as efficiently as the mouse interface is impressive as subjects in the study had more than a decade of experience with using mouse interfaces, while it was their first time ever to use a gaze interface. Taken together, the results of our study might help to develop a new generation of highly efficient human-computer interfaces for a wide variety of applications.

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 (pp. 491-502). Amsterdam: Elsevier.

Hyppolite, J.R., Pomplun, M. & Carbone, E. (2004). The Eye-Hand Span in Human-Computer Interaction. Proceedings of the 18th National Conference On Undergraduate Research (NCUR) 2004, Indiana University/Purdue University (IUPUI), Indianapolis, Indiana, April 15-17, 2004.

Hyrskykari, A. (1997). Gaze Control as an Input Device, in Proceedings of ACHCI'97, University of Tampere, pp. 22-27.

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.

Jacob, R. J. K. (1995). *Eye tracking in advanced interface design.* In Barøeld, W., & Furness, T. (Eds.), Advanced Interface Design and Virtual Environments (pp. 258-288). Oxford: Oxford University Press.

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

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., Ivanovic, N., Reingold, E.M., and Shen, J. (2001). Empirical Evaluation of a Novel Gaze-Controlled Zooming Interface. In M.J. Smith, G. Salvendy, D. Harris & R.J. Koubek (Eds.), Usability Evaluation and Design: Cognitive Engineering, Intelligent Agents and Virtual Reality. Proceedings of the 9th International Conference on Human-Computer Interaction 2001, New Orleans, USA.

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 (pp. 467-478). Amsterdam: Elsevier Science Publishers.

Sunkara, S. & Pomplun, M. (2003). Evaluating Human-Computer Interface Usability based on Pupillary Response. 17th National Conference on Undergraduate Research, NCUR 2003, Salt Lake City, Utah.