Improving Hands-free Menu Selection Using Eyegaze Glances and Fixations

Geoffrey Tien[*]
Simon Fraser University

M. Stella Atkins[†]
Simon Fraser University

Abstract

A real-time eyegaze selection interface was implemented using a Tobii eyegaze tracking monitor. A hierarchical button menu was displayed on the screen and specified selections were made by eyegaze fixations and glances on the menu widgets. The initial version tested three different spatial layouts of the menu widgets and employed a dwell + glance method of selection. Results from the pilot interface led to usability improvements in the second version of the interface. Selections were activated using a glance + dwell method. The usability of the second study interface received a positive response from all 8 participants. Each selection gained more than a 100% speed increase using the revised interface.

A more intuitive selection interface in the second study allowed us to test users selection accuracy at faster dwell selection thresholds. Users quickly learned to achieve accurate selections in 180 ms, but made errors when selections occurred in 150 ms.

CR Categories: H.1.2 [Information Systems]: Models and Principles - User/Machine Systems—Human Factors; Human information processing; software psychology design; I.4.8 [Computing Methodologies]: Image Processing and Computer Vision- Scene Analysis; Tracking

Keywords: pointing and selection, eye pointing, eye tracking, hands free, surgery

1 Introduction

In many situations, it is useful to employ computer-human interaction without using the hands; this can be made possible by using gaze-contingent displays to interact using only eyegaze. For example, in laparoscopic (keyhole) surgery, a surgeon views the surgery site on a display monitor illuminated by an endoscopic camera entering the body through a small hole. The surgeon has both hands occupied and yet may wish to control movement of the endoscopic camera inside the body, or display a different medical image for reference. For sanitation and other reasons, some form of hands-free input mechanism is required for the surgeon to control the computer display system. Even for normal use, the ability to point and select using the eyes alone offers opportunities for a fast alternative input device, and for reducing repetitive stress injury in the wrist and fingers.

This paper details the iterative design process followed to develop an interaction technique for fast and accurate pointing and selection of menu items.

It is known that we are attentive when looking, and look before pointing at targets [Jacob 1991]. Gaze-contingent control relies on the fact that we usually gaze by focusing objects of interest onto the fovea, a highly detailed area of the eye with about 2 degrees diameter [Duchowski 2003]. Fixations are eye gaze movements which stabilize the retina over a stationary object of interest, typically lasting from 100-500 ms [Duchowski 2003; Wu and Remington 2004]. Glances occur when we look at something for a very short period, less than the duration of a fixation.

The user's eyes can be used as an input device, capable of replacing the mouse [Ohno 1998]. However, eyegaze control is limited by eye tracker accuracy, sensor lag, eye fixation jitter, unnatural staring [Zhai et al. 1999], and the important "Midas Touch" problem, where the user's wandering eyes cause unwanted interactions [Jacob 1991; Ashmore et al. 2005]. Various solutions to the usability problems have been proposed to reduce the Midas Touch errors, some of which require auxiliary input such as speech [Miniotas et al. 2004], or pressing a hot key such as used in the GUIDe (Gaze-enhanced User Interface Design) system [Kumar and Winograd 2007]. Augmented systems can perform selection times of around 1200 ms for a single selection task, although the fastest eyegaze only system evaluated had a selection time of around 2000 ms [Kumar et al. 2007].

Other solutions include designing suitable interface widgets for eye gaze pointing with larger buttons than are traditionally used for mouse GUI widgets [Ohno 1998], or using fish-eye expansion techniques [Kumar et al. 2007] by magnifying the area on the screen under the eyegaze. Eyegaze typing has been studied in applications for users with disabilities [cog 2005], and off-the-shelf cameras have been incorporated into such typing systems [Hansen and Hansen 2004], which with very large on screen buttons have achieved typing speeds of 3-5 words per minute. Another "typing by eye" project explores the effect of different feedback on modes with a short dwell time – as low as 300 ms [Majaranta et al. 2004].

Our work describes a novel 2-step interface using feedback and relatively large buttons based on [Ohno 1998; Majaranta et al. 2004] to explore the limits of speed for dwell based selection, and the tradeoffs between speed and accuracy where selection times of 150 ms were found possible.

2 General Method

A Tobii 17" integrated eyegaze tracking monitor [tob 2007] was used at 1280x1024 pixel resolution. Our application included a small timing test which measured a users eyegaze to be sampled at roughly 70 Hz; this sampling frequency was used for our timing data conversions.

All interfaces were implemented in Microsoft Visual Studio 2005 using C++ and the Microsoft Active Template Library (ATL) framework, and the gaze interaction was done using functions provided in the Tobii API [tob 2007]. The eyegaze coordinates were displayed on the screen as a cross-shaped icon with lines 16 pixels long and 2 pixels thick, so the cross was easily visible.

The task was to select a specified menu item from a 2-level menu list, using eyegaze only. Each interface shared the same 3x4 menu hierarchy. There were three items in the first level, each with four children in the second level for a total of twelve level 2 items.

---

[*] e-mail: gtien@cs.sfu.ca
[†] e-mail: stella@cs.sfu.ca

ETRA 2008, Savannah, Georgia, March 26–28, 2008.
© 2008 ACM 978-1-59593-982-1/08/0003 $5.00

We describe the development of the system interface in two experiments: the first, pilot experiment, gave us experience to measure the efficacy of gaze-contingent control in the second experiment, using an improved interface.

3 Experiment 1: Different menu layouts

3.1 Method

For this pilot study, we chose three different menu layouts differing only in widget placement, to compare usability for selection using a gaze-contingent display. Figures 1, 2, and 3 show stylized depictions of Type A, Type B, and Type C layouts respectively. The "M" represents the Menu button, light grey boxes are Level 1 menu items with "P" showing the currently selected Level 1 parent. Dark grey "C" boxes are children of "P". The "T" represents the target whose text must be selected from the menu, and "S" is the menu selection area. Type A layout resembles a typical drop-down mouse-controlled menu; type B layout was chosen from a tablet PC gesture-based menu, and type C layout was chosen to match requirements of eyegaze layouts such as large buttons and near distance, based on work by [Hornof and Halverson 2003].

Each button appears on screen as a 29 mm x 29 mm square, with centered caption text identifying the button's function. There is also a single visually different "menu selection area" shown as a 29 mm x 29 mm box with a grey outline and a top-aligned "Select" caption.

Selection in these interfaces was done in two stages. First, the user dwelled on the desired button until the button gained focus, identified by the button acquiring an extra black outline. A button gained focus if 30 of the previous 40 eyegaze samples lay within a button's pixel boundaries, followed by 20 cumulative samples within the boundaries, for a total number of 50 samples, about 700 ms. Selection of a top level button caused the button's children menu items to be displayed. Once a child menu button received focus, the user glanced at the menu's selection area to generate a mouse click at the focused button. Only one eyegaze sample was needed in the menu selection area to detect the glance. This extra step of confirmation before clicking was implemented to prevent Midas Touch errors.

12 graduate student volunteers (10 males, 2 females) participated in the study. 5 subjects used vision correcting lenses during the experiment, and none had prior experience using eyegaze-controlled selection. After briefing and calibration to the eyetracker, each subject performed the selection task on all six A-B-C permutations of the three layouts. For each of the three menu layouts, users were given one practice trial followed by four time-recorded trials where they were required to correctly select a sequence of 10 menu items.

We conducted a quantitative evaluation to compare the selection speed under the three menu layouts, and a qualitative evaluation to collect the users' feedback on using gaze-based pointing.

3.2 Quantitative Evaluation (experiment 1)

Table 1 shows the mean time per trial, for each menu layout. The times for each layout do not differ significantly. Because each trial involved three eyegaze "clicks", we infer that a single click took approximately 1.98 seconds. Further, the two stage "lock-and-confirm" mechanism yielded a 100% accuracy rate.

Layout	Mean (s)	Std. deviation
A	5.95	0.86
B	5.88	0.97
C	6.03	0.93

Table 1: *Trial times for each menu layout of pilot study.*

3.3 Qualititative Evaluation (experiment 1)

A number of participants reported difficulty in keeping the freely-moving cross icon within a button's pixel boundaries, especially near the boundaries of the monitor window. Several participants also reported difficulty focusing correctly in very densely packed layouts, as seen in Figure 2.

The subjects suggested ways to improve eyegaze cursor movement, placement of widgets, visual feedback, and selection activation. A subset of these suggestions were incorporated into the software for experiment 2, which focused on a variant of layout C, chosen for its relatively spacious placement of buttons.

4 Experiment 2: fast gaze control task

4.1 Method

The interface for the second study was improved in the following areas: cursor control, visual feedback, and task flow. Figure 4 shows a stylized screen shot of the task, using the same legend as Figures 1 to 3. The 110 x 110 pixel size of the buttons remained unchanged, though the caption text was enlarged to enhance readability.

To improve the speed from the pilot, fixations are now registered if at least 6 of the previous 10 eyegaze samples fall within a button's rectangular boundaries. A new "snap-on" function forces the eyegaze cursor to a button within a radius of 85 pixels from the centre of any button. With this feature, users require far less effort to keep the eyegaze cursor on the desired menu item.

The target buttons are now bitmap buttons with three different visual states, shown in Figure 5. The normal state shows the button name in black text against a grey background. On initially detecting a gaze, the button face becomes dark blue and the text becomes white. Halfway to the time needed to trigger a click, the text becomes yellow with an extra underline.

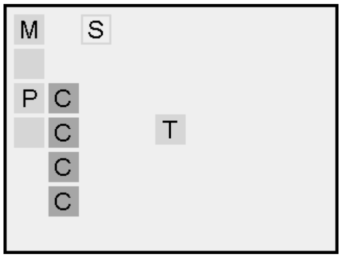

Figure 1: *Type A layout (experiment 1).*

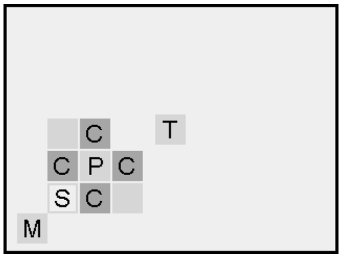

Figure 2: *Type B layout (experiment 1).*

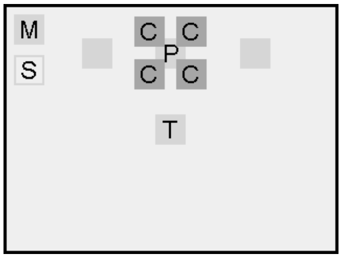

Figure 3: *Type C layout (experiment 1).*

The selection activation mechanism was changed to become purely dwell-time-based. When a top level item has been fixated longer than the dwell selection threshold, a click is generated and its four children appear in the configuration shown in Figure 4. A child can then be selected similarly by a sufficiently long dwell.

The common task of opening the menu is now done by simply glancing off-screen to the left. This allowed a great speed improvement and reduction of effort from the pilot.

Again we conducted a quantitative and qualitative evaluation. The quantitative task compared the speed and accuracy of the menu selection under 4 conditions: with the dwell time threshold at 370, 220, 177, and 147 ms, chosen as a set number of eyegaze samples divided by 70 Hz. The qualitative study included questionnaire data and our observations of users' eyegaze during the study.

The users' performance in each of the 4 conditions was tested, always in the same order of decreasing dwell time thresholds, because there was a learning effect.

8 graduate volunteers (7 male, 1 female) who had not participated in the pilot study and had not previously used eyegaze trackers were recruited for this study. The average age was 26 years. 4 subjects did not require any vision correction, 1 subject used contact lenses and 3 wore eyeglasses.

Again following a briefing and calibration, for each of the 4 conditions, subjects had a practice block of 10 trials followed by 5 blocks of 10 trials each. The sequence of 200 items to be selected throughout the entire study was a randomized sample, and the same sequence was used for every participant.

Since accuracy was being measured, there is a visual feedback showing the user the correctness of the last selection, lasting 110 frames after the selection and during which the user cannot activate the menu. The time for each block was measured as the amount of elapsed time between the start and end of the block, less the time that this feedback was displayed.

The post-study questionnaire asked participants to rate various aspects of the gaze interface on a 7-point Lickert scale and also gave them space to write their own comments and suggestions. These comments were appended by our own observations of the subjects' eyegaze patterns, since the eyegaze was always visible as an icon on the display.

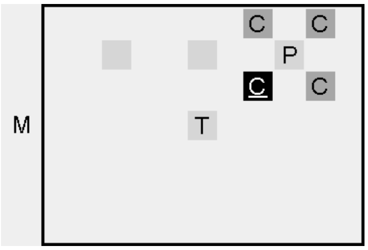

Figure 4: *A typical task screen (experiment 2). The menu opening widget has been moved off-screen to the left.*

Figure 5: *Different button states.*

4.2 Quantitative Evaluation - trial times and accuracy

Table 2 shows the mean time per trial, for each condition. The differences between each adjacent pair of dwell conditions are shown in Table 3. These differences are statistically significant, calculated from a paired-samples t-test. Trial times also do not decrease significantly as subjects performed additional blocks within a condition.

	Condition (ms)			
	370	220	177	147
Mean time per trial (s)	2.97	2.10	1.96	1.85
Std. deviation	0.93	0.56	0.53	0.58

Table 2: *Mean time per trial, per condition.*

	Condition		
	370 - 220	220 - 177	177 - 147
Difference (ms)	870	130	110
p	< 0.001	< 0.001	0.003

Table 3: *Trial time differences between conditions.*

Table 4 shows the mean accuracy per trial, for each condition. Using paired samples, there was a significant drop in accuracy from condition 1 to 2, and from condition 2 to 3.

4.3 Qualitative Evaluation

Table 5 shows results from two questions on the questionnaire for both studies. Both questions are rated on a 7-point Lickert scale, with 1 representing difficult/unclear, and 7 representing easy/clear. Question 1 asked the participants to rate the ease of using eyegaze as input, and Question 3 asked the subjects to rate the clarity of the feedback for marking when an item is ready to be selected.

5 Discussion

5.1 Speed & Accuracy

Table 2 shows that in experiment 2 where each trial involved 2 clicks, one click could be made in just under 1 second. With the pilot layout C one click took 2 seconds. Therefore the speed more than doubled with the same layout but new usability features.

As noted earlier, the task time for condition 1 is significantly longer than for the other conditions. This could be due to a learning effect, in addition to the longer dwell activation threshold.

	Condition (ms)			
	370	220	177	147
Mean accuracy per trial	0.98	0.87	0.81	0.79
Std. deviation	0.15	0.34	0.39	0.41

Table 4: *Mean accuracy per trial, per condition.*

Question	Mean (pilot)	Std.dev (pilot)	Mean (exp2)	Std.dev (exp2)
1	2.58	1.38	3.62	1.30
3	4.58	2.02	6.63	0.52

Table 5: *Comparison of questionnaire responses.*

There were still significant learning effects within the first 370 ms condition for each subject, indicating that the practice trials were not sufficient for the user to learn the task. Therefore we ascribe the extra time for the first condition to both the learning curve and to the extra system response time.

Recall that Midas touch errrors were not made in experiment 1 where a confirmation step was needed. However, using the purely dwell-based selection mechanism in experiment 2, errors were made, as shown in Table 4. Amazingly, a majority of the subjects (5/8) achieved accuracy of >80% even with the shortest dwell selection threshold of only 147 ms. This is likely only possible by employing a memorization strategy.

5.2 Questionnaire

The responses to question 1 in Table 5 show that subjects still had some trouble using the eyegaze interface. The feedback was very good (see question 3). One subject said the system was “fun to use”, and all the subjects enjoyed using the interface.

5.3 User Strategies

Participants of experiment 2 quickly developed their own ways to maintain a high selection accuracy. The most common strategy was to remember the positions of all the menu items, which was almost essential to maintain accuracy for conditions 3 and 4 where the selection thresholds are less than the time needed to cognitively process the button text seen during visual search.

After very quickly correctly memorizing each button’s location, Subject 1 would stare at the central box to visualize where the target would appear. Then in a single eye gesture, he saccaded off-screen with a glance to open the menu, followed by direct saccades to the targets parent and then the target itself after sufficiently long dwells.

Subject 8 also employed a memorization strategy, but noted that memorizing things incorrectly led to inaccurate selections, especially for the fast conditions.

Subject 3 was the fastest initially presumably due to a very fast reaction for visual search, but the slowest for the final condition. Without memorization of all the items positions, his strategy for reading and cognitively evaluating the text failed for the fast conditions. He reported using his peripheral vision to read the labels of the children items while staring at the parent button, to avoid triggering an incorrect response. This produced a consistently accurate result but cost additional time and concentration, which led to fatigue.

Subject 6 attempted to open the menu during the feedback, but glanced back and forth between the off-screen menu activation area and the central box, waiting for the new target to appear. He reported frustration and fatigue from moving his eyes between the off-screen menu activation area and the box for the next target. He was very slow by the end, and was the least accurate.

6 Conclusion

An iterative design process based on real user feedback enabled us to implement and improve an intuitive eyegaze-controlled interface that is empowering and “fun to use”. The rapid development of user strategies alludes to the ease with which the improved interface can be learned.

However, improved selection speed due to a simplified task flow, combined with decreasing dwell selection thresholds, introduced Midas Touch as a detriment to accuracy. Further suggestions from the study participants as well as further exploiting of user strategies can still be incorporated into future iterations, for a more efficient and reliable eyegaze-controlled menu selection interface.

References

ASHMORE, M., DUCHOWSKI, A. T., AND SHOEMAKER, G. 2005. Efficient eye pointing with a fisheye lens. In *Proc. Graphics Interface*, 203–210.

2005. Cogain. http://www.cogain.org/.

DUCHOWSKI, A. T. 2003. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag London Limited.

HANSEN, D., AND HANSEN, J. 2004. Eye typing with common cameras. In *Proc. ETRA 2004*, 55.

HORNOF, A. J., AND HALVERSON, T. 2003. Cognitive strategies and eye movements for searching hierarchical computer displays. In *Proc. CHI 2003*, 249–256.

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 (Apr.), 152–169.

KUMAR, M., AND WINOGRAD, T. 2007. Guide: Gaze-enhanced ui design. In *Proc. CHI 2007*, ACM Press, 1977–1983.

KUMAR, M., PAEPCKE, A., AND WINOGRAD, T. 2007. Eyepoint: Practical pointing and selection using gaze and keyboard. In *Proc. CHI 2007*, ACM Press, 421–430.

LANKFORD, C. 2000. Effective eye-gaze input into windows. In *Proc. ETRA 2000*, 23–27.

MAJARANTA, P., AULA, A., AND RÄIHÄ, K.-J. 2004. Effects of feedback on eye typing with a short dwell time. In *Proc. ETRA 2004*, 139–146.

MINIOTAS, D., ŠPAKOV, O., AND MACKENZIE, S. 2004. Eye gaze interaction with expanding targets. In *Proc. CHI 2004*, ACM Press, 1255–1258.

OHNO, T. 1998. Features of eye gaze interface for selection tasks. In *Proc. CHI 1998*, 176–181.

2007. Tobii technology. http://www.tobii.com/.

WU, S.-C., AND REMINGTON, R. W. 2004. Coordination of component mental operations in a multiple-response task. In *Proc. ETRA 2004*, 63–71.

ZHAI, S., MORIMOTO, C., AND IHDE, S. 1999. Manual and gaze input cascaded (magic) pointing. In *Proc. CHI 1999*, 246–253.