An Evaluation of an Eye Tracker as a Device for Computer Input

Colin Ware

School of Computer Science

Harutune H. Mikaelian

Department of Psychology

University of New Brunswick
Box 4400, Fredericton,
New Brunswick, E3B 5A3

Abstract

Since humans direct their visual attention by means of eye movements, a device which monitors eye movements should be a natural "pick" device for selecting objects visually present on a monitor. The results from an experimental investigation of an eye tracker as a computer input device are presented. Three different methods were used to select the object looked at; these were a button press, prolonged fixation or "dwell" and an on screen select button. The results show that an eye tracker can be used as a fast selection device providing that the target size is not too small. If the targets are small speed declines and errors increase rapidly.

Keywords: Input devices, eye movements.

Consider the following two observations.

Observation 1: One of the most frequent operations which occurs in human computer interaction is the selection of an object displayed on a terminal. The item selected may be a word of text in a typical editing application or a menu item in a menu based interface. This is the "pick" operation defined by Wallace [9].

Observation 2: One of the principal ways in which a human observer directs his or her visual attention to objects in the immediate environment is by foveating them [6]. The fovea of the eye is a small region of acute vision with which fine details can be discerned; outside of this region acuity falls off rapidly. In foveation, the eyeball of the observer is rotated in such a way that the object of interest becomes imaged on the fovea.

Since human observers presumably pay attention to objects they wish to select and since the observer pays attention to objects by foveating them, it seems that the most natural "pick" device from the user's point of view should be based on a device which detects his or her line of gaze. Using such a device to select a visual object, it is only necessary to look at the object and take some action denoting selection.

The above arguments provide a rationale for being interested in using eye movements for selection. However, there are a number of issues which need to be addressed in order to compare the functionality of eye trackers to that of other computer input devices.

An important issue concerns how the selection should be made. Should the observer stare at the item to be selected for some fixed interval or press a button on the keyboard? A second issue concerns the size of items that can be comfortably selected. In view of the inherent jitter in eye movements we can anticipate that very small items will not be selectable. In this paper we describe two experiments designed to address the above issues after first presenting some of the basic relevant psychophysical and physiological facts about eye movements.

Facts About Eye Movements

When we shift our gaze to fixate an object, we are causing that object to be imaged on the fovea of the retina. Outside of the fovea, acuity falls off rapidly. For example, at 10 deg. from the center of the fovea, acuity has dropped by 70 %. As a rule of thumb, the fovea is held to subtend 2 deg. of visual angle. Compare this with the approximately 0.6 deg of visual angle subtended by a character on the face of a monitor viewed at 14 inches.

The research literature on normal eye movements tells us that it is possible to fixate as accurately as 10 minutes of visual angle (0.16 deg) but that uncontrolled spontaneous movements cause the eye to periodically jerk off target [5]. It is likely, however, that when an observer is attempting to fixate numerous targets successively, accuracy is considerably reduced and misplacements as as large as a degree may become common [7].

Eye movements between points of fixation are ballistic; that is, no correction for errors in trajectory are made while the eye is in motion. Normal eye movements occur at a frequency of 3 to 5 per second, with the eye resting at each fixated location for approximately 200 ms. and requiring about 50 ms. to move to the next fixated location. Thus, it would seem that selection of visual objects based on eye movements could be very fast.

Researchers developing eye tracking devices for use by the disabled have obtained selection rates of approximately 60 targets per minute [3,4]. They have generally used a selection criterion based on how long the taget has been fixated [3]. Typically a user would fixate a target for approximately 0.5 seconds for it to be selected. This method is necessary for some members of the disabled population where eye movements may be the only body movement over which the person has reasonable control. For other disabled users and for non disabled users the selection can be made by means of a physical button such as part of a keyboard, or a micro-switch activated in any way desired. Another method of selection would be to use the fixation of a designated region of the screen as a select button.

Experiment 1: Target Distance and Selection Method

The first experiment was designed to measure selection time as a function of the distance to the target for three different selection methods.

1) Dwell time button. In this method the object fixated was selected if the observers gaze dwelled on the item for more than some fixed interval. After experimentation, we adopted a dwell time of 0.4 seconds.

2) Screen button. In this method we designated a large rectangular area of the screen as a button. A selection involved the observer looking at the item to be selected and subsequently fixating the button. A practical disadvantage of a screen button of this kind is that for a screen button to be effective there can be no menu items in the path between the item to be selected and the button. This prohibits a two dimensional matrix of menu items.

3) Hardware button. In this method the user pressed a physical button while fixating the item to be selected. The button was placed by the user in a convenient location.

Hardware

The research reported here was carried out using a Gulf and Western series 1900 eye view monitoring system. This system determines the direction of the observer's gaze by means of a television camera which digitizes an image of the observer's eye. An image processing sub-system detects the boundary of the observer's pupil and the boundary of the image of an infra-red source reflected from the cornea. Calculations based on the relative location of these two image components yield eye position accurate to about 1/2 deg. of visual angle. We enhanced the accuracy of the calibration somewhat using software techniques described in [8].

The version of this device which we used required that the observers maintain a fixed head position so that they did not move out of the narrow field of view of the camera. More expensive systems costing upwards of $100,000 can be obtained which do not necessitate a fixed head position. Freedom of head movement is obviusly highly desirable from the point of view of the user's comfort. However, there is no evidence to support the idea that freedom of head movement should improve performance on the target acquisition tasks we investigated. Thus we believe, although without hard evidence, that the studies we carried out can be generalized to other non-fixed head eye tracking devices having comparable accuracy.

Output from the eye tracker was fed into a microcomputer which was used to generate stimuli on a monitor, control the sequence of the experiment and store data.

Stimulus and Task Description

The screen layout is diagrammed in figure 1. It consisted of seven "menu items" arranged vertically in a column just to the right of the centre of the screen. Two forms of feedback were provided to the observer concerning his or her eye behaviour. The first was a screen cursor which tracked the current fixation position of the observer. The second was a square which appeared in the upper right hand corner of the last menu item fixated. When the screen button was used a large bar running the entire vertical extent of the screen was placed to the left of the menu items. If that bar was fixated, the last item fixated was taken to have been selected.

A single trial consisted of the following protocol. The observer fixated the central menu item and pressed a button to initiate the trial. After about 0.5 sec one of the

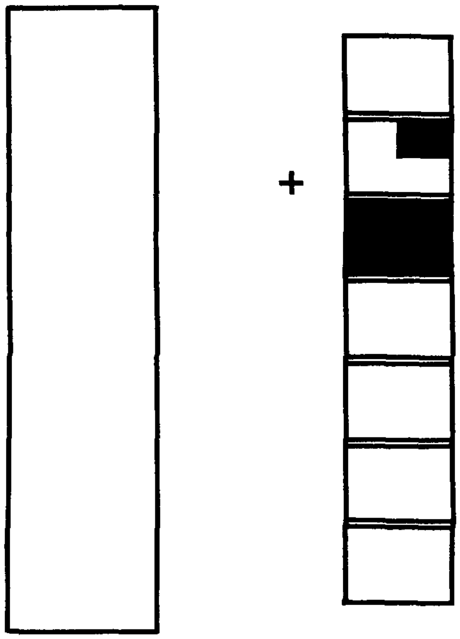

Figure 1. This diagram shows the screen layout used in experiment 1 with black and white reversed. Each menu item measured 3.2 cm by 2.6 cm which at the 90 cm viewing distance converts to a visual angle of 2.0 deg by 1.65 deg. The filled box denotes the target on the current trial. The box with the filled rectangle in the upper right corner denotes the last menu item fixated. The + symbol denotes the current location of fixation as measured by the eye tracker. The large rectangle to the left was the selection button used in the screen button condition. This rectangle was not present on the screen in the other conditions.

menu items changed from an outline square to a filled bright square. This was the target item which the observer immediately fixated causing the feedback square to appear at the upper right hand corner of the item. Finally the observer made one of the three selection actions using a dwell time button, the screen button or the hardware button depending on the experimental condition.

On all trials the interval between the appearance of the target menu item and the selection of that target was timed. An error was recorded if, on a given trial, the observer made a selection of any menu item other than the one highlighted.

In a given session an observer was given all three conditions in a different random order. In each condition 35 trials were given consisting of five trials with each of the seven menu items as target. The trials were randomly ordered. There were 12 experimental sessions carried out on different days for three of the observers. The fourth observer participated in 14 sessions because of problems with calibration.

Four observers were employed for the experiment. Three were paid assistants and one was the first author of this paper.

Results from Experiment 1

The following results are averages obtained from the last four well calibrated sessions obtained in each condition for each observer. Because all observers showed essentially the same pattern of responses we present data averaged across observers.

The selection time as a function of the distance of the target item from the initial point of fixation is plotted in figure 2. Following are the salient features of this data.

1) Irrespective of selection procedure, selection times were below one second. This makes the device faster than any of the selection devices reviewed in Card et. al. [2].

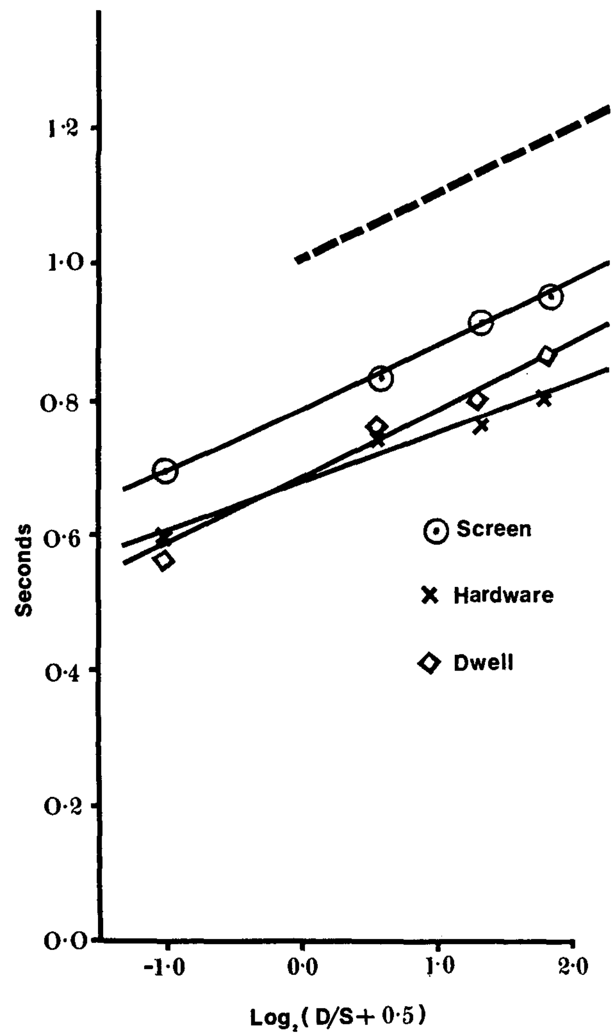

Figure 2. Data from experiment 1. The mean time to make selections using an eye tracker is plotted against the log transformed distance to the target for three button techniques. D is distance from the center menu item and S is the vertical size of the menu items. The dashed line represents data on the mouse taken from [2].

2) Selection using the hard button and dwell time were equally fast, and faster than selection using the screen button.

3) The data from eye tracker selection, like that from other selection devices, is well fitted by the modified Fitt's law [2]

$$T = C + I.\log_2 (D/S + 0.5).$$

Where C and I are constants, D is the distance to the target and S is the size of the target. This is a theoretical construct designed to account for eye-hand coordination. We use it here only as a convenient way of summarizing the results, not because we wish to make any theoretical claims. If we compare the data obtained from the eye tracker with that described for the mouse by Card et al., we find that the main difference appears to be in the additive constant term 'C'. For the mouse a large part of this constant would represent the time taken to grab the mouse. Since no grabbing is required with an eye tracker, the times are faster. For the mouse the fastest constants obtained were of the order of 1.0 sec. For the eye tracker this constant is of the order of 0.6 sec.

4) The errors were 22% in the screen condition, 8.5% in the hardware button and 12% in the dwell condition. None of these differences are statistically significant.

Although these error rates would be unacceptable in most practical applications they are comparable to the rates reported by Card et. al. for step keys, joysick and mouse. Presumably, there is a speed accuracy tradeoff in these experimental tasks which has different task contraints to those found in real applications. There is little doubt that more accurate performance can be achieved with the eye tracker and would be achieved if more emphasis were to be placed on accuracy.

Experiment 2: Target Size, Continuous Mode

In view of the well documented jitter in eye movements it is possible to state in advance that eye movements will not be suitable for picking small targets. The purpose of this experiment was to find out how target size effects response speed and error rates for two of the selection methods, namely the hardware button and dwell time button.

In addition, instead of having the observer separately initiate each trial, a new trial was begun immediately after the observer responded to the previous trial. Thus the observer ran through the entire sequence of trials in a burst of sustained continuous responding. The pattern of menu items used in this experiment was a 4 by 4 square matrix instead of the line of items used in the first experiment.

In a given session there was a sequence of 64 trials in which each of the 16 menu items came on exactly 4 times in a random order. The size of the menu was fixed for this entire sequence. After completing a sequence, the observer took a brief rest before responding to another sequence with a different size of menu. On a given day the observer was given all menu sizes in a different random order.

The sizes of the square menu items were 48 mm, 36 mm, 24 mm, 12 mm and 7.2 mm on a side. At a viewing distance of 90 cm this translates into approximately, 3.0 deg, 2.25 deg, 1.5 deg, .75 deg and 0.45 deg of visual angle. In the dwell time condition, dwell for selection was set at 0.4 sec. Three of the observers were the same as those used for experiment 1, the fourth was the second author.

Results From Experiment 2

Figures 3 and 4 summarize some of the results from this experiment. Once again the observers all showed the same pattern of responding and so we present only averages across observers. We summarize these results in point form.

1) Speed of selection increases dramatically from the 0.45 deg target to the 0.75 deg target. there is also an increase in speed from the 0.75 deg menu to the 1.5 deg menu. Above 1.5 deg there appear to be only minor changes.

2) For all sizes and for all observers the hardware button is faster than the dwell button. With the hardware button, continuous responding speeds for the three largest targets ranged from 0.6 sec per response for the fastest observer to 0.75 sec per response for the slowest. Dwell time response rates were even more consistent, and averaged close to 0.95 sec per response for all observers.

3) For all sizes and for all observers there are fewer errors for the dwell button than for the hardware button. This was in spite of instructions to respond as fast as possible while maintaining accuracy for both conditions. Errors were below 10% in all conditions where the targets subtended less than a degree of visual angle.

Superficially these results look like a speed/accuracy trade off, with observers adopting a strategy of greater speed at the expense of accuracy in the hardware button condition. But why the discrepancy between these results and those obtained in experiment 1? And why did observers ignore instructions to maintain accuracy in both conditions?

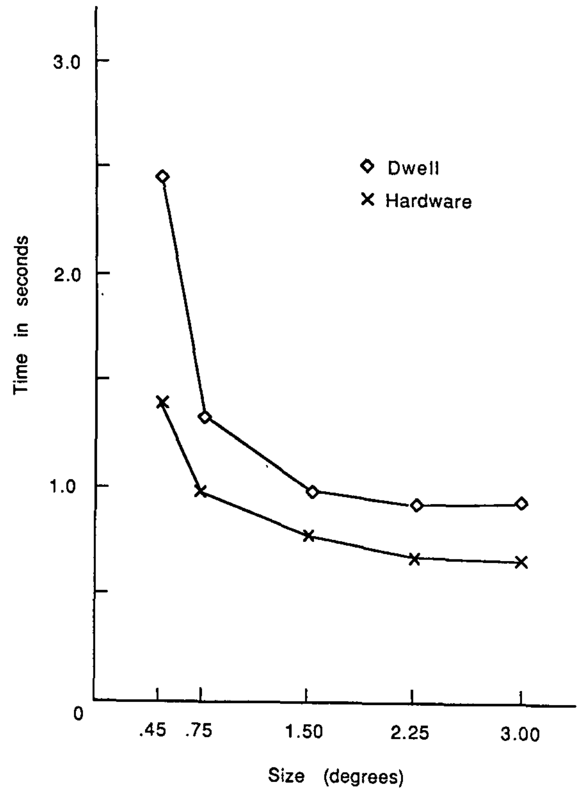

Figure 3. Mean response speed per selection is plotted against the size of the items for both dwell button selection and hardware button selection.

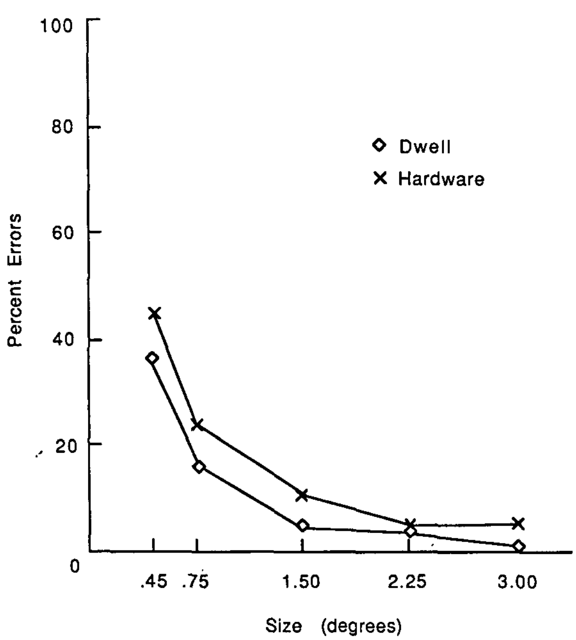

Figure 4. Mean error is plotted against the size of the items for both dwell button selection and hardware button selection.

Our account runs as follows: The most significant difference between experiments 1 and 2 was the fact that the second involved continuous responding. This seemed to produce a rhythmic sequence of button presses in the hardware button condition during which we suspect that the observers were in fact attempting to synchronize their button presses with the arrival of ballistic eye movements at their targets. The lower accuracy obtained in the hardware button condition can thus be explained by inaccuracies in the eye control system causing these eye movements to fall off target with a certain frequency, causing errors to be recorded.

If this account is correct, the subjects were essentially performing two tasks simultaneously in the hardware button condition, that is, pressing the button and making an eye movement. Conversely, in the dwell condition the subject was forced to perform the eye movement and the dwell selection sequentially and this accounts for the increased response times.

Discussion

What are the practical consequences of these results? The eye tracker is a fast selection device when compared to the mouse or other conventional pick devices. Experiment 1 suggests that the principal reason for this relative speed may be that it is unnecessary to remove a hand from the keyboard in order to use an eye tracker. The eye tracker is also limited in the size of targets which can be reliably selected. As a rule of thumb these should subtend at least a degree of visual angle. This is larger than the typical character, but smaller than most words of text on the screen.

As a device for the disabled, our results suggest that where the disabled user has the ability to make a button press this may be the selection technique of choice. Also, the button press may allow greater improvement over extended practice. In the dwell condition the half second or so needed to register a fixation provides a permanent barrier against a substantial speed-up. In the hardware button condition more substantial improvements may be obtained over extended practice as the user learns to synchronize button presses and eye movements.

In these two experiments we modelled our basic tasks on those designed by Card et. al. to invesigate other two-dimensional selection devices such as the joy stick and the mouse. These tasks employ target highlighting in order to eliminate noise generated by having the subject search for a target prior to selecting it. However, considering the issue of task validity, we note that this use of highlighting may in fact mask the real value of detecting line of gaze as a means of selection. In many applications the user would find an item to be selected by means of a visual search of the menu labels. Only after this search will the selection be made. With a typical mouse interface the user must remove look away from the item just located in order to locate the cursor and, using the mouse, drag the cursor to the item and select it. With an eye tracker it is much simpler, having located the item to be selected by visual search, the user can immediatly make the selection since the target is already being fixated.

To summarize all this in a sentence; where speed is of the essence, cost is no object, sizes are moderate, and it is important that the hands be reserved for other activities, the eye tracker may be the input device of choice.

Acknowledgements

The authors are grateful to Glenn Arsenault and Danny Jessome who produced the software for the experiments. This research was funded in part by grants A0289 and A8202 to Mikaelian and Ware respectively, from the National Sciences and Engineering Research Council of Canada.

References

1. Bolt, R.A. *The Human Interface.* Lifetime Learning Publications, 1984.

2. Card, S.K., Moran, T.P. and Newell, A. *The Psychology of Human-Computer Interaction.* Lawrence Erlbaum, 1983.

3. Downing, A.R. Eye for Independence. In *Proceedings of the Rehabilitation Engineering Society of North America* 8th Annual Conference, 1985.

4. Friedman, M.B., Killiany, G. and Dzmura, M. An Eye Gaze Controlled Keyboard. In *Proceedings of 2nd International Conference on Rehabilitation Engineering, Ottawa, 1985.*

5. Gaardner, K.R. *Eye Movements, Vision and Behaviour.* Hemisphere Publishing Corporation: Washington, 1975.

6. Monk T.H. Search. In *Sustained Attention in Human Performance,* J.S. Warm, Editor, Wiley, 1984.

7. Osaka, R. and Koga, K. 1980. Variations in Voluntary Fixations during Mental Tasks. In *Cognition and Eye Movements.* Edited by R. Groner and P. Fraisse. North Holland, 1980.

8. Sheena, D. and Borah, J. Compensation for Some Second Order Effects to Improve Eye Position Measurements. In *Eye Movements, Cognition and Visual Perception.* Edited by D.F. Fisher, R.A. Monty and J.W. Sendersa., 1981.

9. Wallace, V.L. The Semantics of Graphic Input Devices. In *Proceedings SIGGRAPH/SIGPLAN Conference.* Computer Graphics. 1976, 10, 61-65.