

EVALUATION AND ANALYSIS OF EYE GAZE INTERACTION

Linda E. Sibert

Human Computer Interaction Laboratory
Naval Research Laboratory Washington, DC 20375
sibert@itd.nrl.navy.mil

Robert J. K. Jacob

Department of Electrical Engineering and Computer Science
Tufts University Medford, MA 02155
jacob@cs.tufts.edu

James N. Templeman

Human Computer Interaction Laboratory
Naval Research Laboratory Washington, DC 20375
templema@itd.nrl.navy.mil

ABSTRACT

Eye gaze interaction can provide a convenient and natural addition to user-computer dialogues. We have previously reported on our interaction techniques using eye gaze [Jacob 1992]. In this paper, we assess the effectiveness of our approach in a controlled setting. We present two experiments that compare our eye gaze object selection technique with conventional selection by mouse. The results show that, for a simple task, it takes 60% less time to select an object with our eye gaze technique than with a mouse. We also use Fitts' Law to investigate the speed and quality differences between eye gaze selection and mouse selection. Our eye gaze selection technique is shown to produce a low slope, more like pure eye movement, which suggests that the technique preserves the inherent speed advantage of the eye over the hand. We find that eye gaze interaction is at least as fast as the mouse; and it is convenient in situations where it is important to use the hands for other tasks. It is particularly beneficial for the larger screens, workspaces, and virtual environments of the future; and will become increasingly practical as eye tracker technology matures.

Categories and Subject Descriptors: H.1.2 [Models and Principles]: User/Machine Systems—*human factors*; H.5.2 [Information Interfaces and Presentation]: User Interfaces—*input devices and strategies, interaction styles*; I.3.6 [Computer Graphics]: Methodology and Techniques—*interaction techniques*

General Terms: Design, Experimentation, Human Factors, Measurement, Theory

Additional Key Words and Phrases: Eye movements, eye tracking, input devices, interaction techniques

1. INTRODUCTION

We describe two experiments that compare our eye gaze interaction technique for object selection with the traditional method of selecting with a mouse. We find our eye gaze technique is measurably faster; we argue that eye gaze interaction has additional harder-to-quantify benefits as well. The experiments measure time to perform simple, representative direct manipulation computer tasks. The first requires the subject to select the highlighted circle from a grid of circles. The second asks the subject to select the letter named over an audio speaker from a grid of letters. We discuss physiological differences between eye and arm movement that are the basis of this speed difference and discuss how we made use of this information in our software architecture. We use Fitts' Law to show that our interaction technique design and implementation preserves the inherent speed advantage of the eye over the hand.

1.1 Eye vs. Hand

We know the eye can move faster than the hand; it is not our goal to verify this. We seek to compare two complete interaction techniques for selecting objects in a user interface, each with its various hardware, software, and interaction designs, in a simulated task setting. We have designed a set of interaction techniques for eye movement-based interfaces, incorporating real-time fixation recognition, local recalibration, nearest-neighbor selection, various dwell times and timeouts, and iteratively refined these interaction designs [Jacob, 1991]. The experiments we now report test our selection interaction technique. The challenge is to develop a robust interaction technique that preserves speed advantage of the eye over the hand. We want to test whether our selection technique does so. We believe the dwell times and other aspects of the interaction design enhance the operation of the interaction technique; but they cost some performance speed. We wish to evaluate whether the net result, after incurring these costs, still preserves the eye speed advantage, whether the combination of hardware, software, and interaction design in our technique still shows the speed advantage of the underlying raw physiology.

Our previous experience suggests that eye tracker technology is somewhat shaky, considerably less robust than the mouse. Eye tracking is still rarely used outside the laboratory, while the mouse is in wide use. We wish to see if we can overcome these technical difficulties (some are described below) sufficiently that we can make a head-to-head comparison of eye vs. mouse on the same tasks and under precisely the same conditions and rules of engagement.

We have also found in previous informal evaluation that, when all is performing well, eye gaze interaction can give a subjective feeling of a highly responsive system, almost as though the system is executing the user's intentions before he or she expresses them. We want to bring this benefit to the object selection interaction without incurring a speed penalty.

If the eye can "break even" with the mouse in a straightforward experimental comparison, we obtain the subjective benefits cited for free. If the eye interaction technique is faster, we consider it a bonus, but not the primary motivation for using eye tracking in most settings. Our results showed a distinct, measurable speed advantage for the eye movement-based selection technique over the mouse in a side-by-side comparison on the same tasks in the same experimental setting, and it was consistent in both experiments.

1.2 Fitts' Law

A byproduct of our experiment has been to gain some insight into how eye movements are modeled by Fitts' Law. Some previous investigation has suggested they follow it, rather like the hand but faster; while others have speculated that they do not, based on the nature of the muscles and their control mechanism (that is, that the Fitts' Law model for the eye would have a very small slope). Our data suggest the latter, that the time required to move the eye is only slightly related to the distance to be moved. This suggests eye gaze interaction will be particularly beneficial when the distances to be traversed are large, as with large- or multi-screen displays or virtual reality.

2. RELATED WORK

People continuously explore their environment by moving their eyes. They look around quickly and with little conscious effort. With tasks that are well-structured and speeded, research has shown that people look at what they are working on [Just and Carpenter 1976]; the eyes do not wander randomly. Both normal and abnormal eye movements have been recorded and studied to understand processes like reading [Just and Carpenter, 1980] and diagnosing medical conditions (for example, a link between vestibular dysfunction and schizophrenia shows up in smooth pursuit). People naturally gaze at the world in conjunction with other activities such as manipulating objects; eye movements require little conscious effort; and eye gaze contains information about the current task and the wellbeing of the individual. These facts suggest eye gaze is a good candidate computer input method.

A number of researchers have recognized the utility of using eye gaze for interacting with a graphical interface. Some have also made use of a person's natural ways of looking at the world, as we do. In particular, Bolt suggests that the computer should capture and understand a person's natural modes of expression [Bolt 1982]. His *World of Windows* presents a wall of windows selectable by eye gaze [Bolt 1981, 1992]. The object is to create a comfortable way for decision-makers to deal with large quantities of information. A screen containing many windows covers one wall of an office. The observer sits comfortably in a chair and examines the display. The system organizes the display by using eye gaze as an indication of the user's attention. Windows that receive little attention disappear; those that receive more grow in size and loudness. Gaze as an indication of attention is also used in the self-disclosing system that tells the story of *The Little Prince* [Starker and Bolt 1990]. A picture of a revolving world containing several features such as staircases is shown while the story is told. The order of the narration is determined by which features of the image capture the listener's attention as indicated by where they look.

Eye gaze combined with other modes helps disambiguate user input and enrich output. Questions of how to combine eye data with other input and output are important issues and require appropriate software strategies [Thorisson and Koons 1992]. Combining eye with speech using the OASIS system allows an operator's verbal commands to be directed to the appropriate receiver, simplifying complex system control [Glenn et al. 1986]. Goldberg and Schryver 1993] investigate whether there are consistent indications of a user's intent to zoom toward an object from characteristics of eye gaze, such as where the user is looking in the window. Their results are mixed but support the idea that information other than point-of-gaze is available from the behavior of the eyes. Ware and Mikaelian [1987] conducted two studies, one that investigated three types of selection methods and the other that looked at target size. Their results showed

that eye selection could be fast provided the target size is not too small. Zhai, Morimoto, and Ihde [1999] have recently developed an innovative approach that combines eye movements with manual pointing.

In general, systems that use eye gaze are attractive because they are easy to use, and they respond somewhat more like people, who commonly incorporate knowledge of where their conversational partner is looking into the dialogue. An information system at a science museum in Denmark using The Eyecatcher multimedia shell received positive response in early testing [Hansen et al. 1995]. They report problems, however, because the system caused people to become excited which often caused them to laugh or talk. Sometimes, in their enthusiasm, they moved out of the range of the eye tracker, a true conversation stopper!

3. BACKGROUND

3.1 Demonstration System and Software Architecture

Incorporating eye gaze into an interactive computer system requires technology to measure eye position, a finely tuned computer architecture that recognizes meaningful eye gazes in real time, and appropriate interaction techniques that are convenient to use. In previous research, we developed a basic testbed system configured with a commercial eye tracker to investigate interfaces operated by eye gaze. We developed a number of interaction techniques and tested them through informal trial and error testing. We learned that people prefer techniques that use natural not deliberate eye movements. Observers find our demonstration eye gaze interface fast, easy, and intuitive. In fact, when our system is working well, people even suggest that it is responding to their intentions rather than to their explicit commands. In the current work, we extended our testbed and tested our eye gaze selection technique through formal experimentation.

Previous work in our lab demonstrated the usefulness of using natural eye movements for computer input [Jacob 1991, 1992, 1993a, 1993b, 1994]. We have developed interaction techniques for object selection, data base retrieval, moving an object, eye-controlled scrolling, menu selection, and listener window selection. We used context to determine which gazes are meaningful within a task. We have built the demonstration system on top of our real-time architecture that processes eye events. The interface consists of a geographic display showing the location of several ships and a text area to the left (see Figure 1) and supports four basic tasks: selecting a ship, reading information about it, adding overlays, and repositioning objects.

The software structure underlying our demonstration system and adapted for the experiments is a real-time architecture that incorporates knowledge about how the eyes move. The algorithm processes a stream of eye position data (a datum every 1/60 of a sec.) and recognizes meaningful events. There are many categories of eye movements that can be tapped. Our current research uses events related to a saccadic eye movement, the general mechanism used to search and explore the visual scene. Other types of eye movements are more specialized and might prove useful for other applications, but we have not made use of them here. For example, pursuit motion partially stabilizes a slow moving target or background on the fovea and optokinetic nystagmus (i.e., train nystagmus) has a characteristic sawtooth pattern of eye motion in response to a moving visual field containing repeated patterns [Young and Sheena 1975]. These movements would not be expected to occur with a static display.

The eyes are rarely still because, in order to see clearly, we must position the image of an object of interest on our fovea, the high-acuity region of the retina that covers approximately one degree of visual arc (an area slightly less than the width of the thumb held at the end of the extended arm). For normal viewing, eyes dart from one fixation to another in a saccade. Saccades are the rapid ballistic movement of the eye from one point of interest to another whose trajectory cannot be altered once begun. During a saccadic eye movement, vision is suppressed. Saccades take between 30 and 120 msec. and cover a range between 1 to 40 degrees of visual angle (average 15 to 20 degrees). The latency period of the eye before it moves to the next object of interest is at least 100 to 200 msec., and after a saccade, the eyes will fixate (view) the object between 200 to 600 msec. Even when a person thinks they are looking steadily at an object, the eyes make small, jittery motions, generally less than one degree in size. One type is high frequency tremor. Another is drift or the slow random motion of the eye away from a fixation that is corrected with a microsaccade. Microsaccades may improve visibility since an image that is stationary on the retina soon fades [Boff and Lincoln 1988]. Likewise, it is difficult to maintain eye position without a visual stimulus or to direct a fixation at a position in empty space.

At the lowest level, our algorithm tries to identify fixation events in the data stream and records the start and approximate location in the event queue. Our algorithm is based on that used for analyzing previously-recorded files of raw eye movement data [Lambert et al. 1974; Flagg 1977] and on the known properties of fixations and saccades, and it is required to work in real time. The fixation recognition algorithm declares the start of a fixation after the eye position remains within approximately 0.5 degrees for 100 msec. (the spatial and temporal thresholds are set to take into account jitter and stationarity of the eye). Further eye positions within approximately one degree are assumed to represent continuations of the same fixation. To terminate a fixation requires 50 msec. of data lying outside one degree of the current fixation. Blinks and artifacts of up to 200 msec. may occur during a fixation without terminating it. The application does not need to respond during a blink because the user could not see such a response on the screen anyway.

Tokens for eye events - for start, continuation (every 50 msec. in case the dialogue is waiting to respond to a fixation of a certain duration), end of a fixation, raw eye position (not used currently), failure to locate eye position for 200 msec., resumption of tracking after failure, and entering monitored regions (a strategy typically used for mouse interaction) - are multiplexed into the same event queue stream as those generated by other input devices. These tokens carry information about the screen object being fixated. Eye position is associated with currently displayed objects and their screen extents using a nearest neighbor approach. The algorithm will select the object that is reasonably close to the fixation and reasonably far from all other objects. It does not choose when the position is half way between two objects. This technique not only improves performance of the eye tracker which has difficulty tracking at the edges of the screen (see discussion of the range of the eye tracker in the *Apparatus* section) but also mirrors the accuracy of the fovea. A fixation does not tell us precisely where the user is looking because the fovea (the sharp area of focus) covers approximately one degree of visual arc. The image of an object falling on any part of the fovea can be seen clearly. Choosing the nearest neighbor to a fixation recognizes that the resolution of eye gaze to be approximately one degree.

The interaction is handled by a User Interface Management System that consists of an executive and a collection of simple individual dialogues with retained state like coroutines [for details see Jacob 1993c]. Each object displayed on the screen is implemented as an interaction object and has a helper interaction object associated with it that translates fixations into the higher unit of gazes. This approach is more than an efficiency. It reflects that the eye does not remain still but changes the point of fixation around the area of interest.

4. STUDY OF EYE GAZE VERSUS MOUSE SELECTION

Our informal experience with eye gaze interaction has been positive; the present work attempts to make it more formal. In developing our demonstration system, we were struck by how fast and effortless selecting with the eye can be. We had developed the interaction techniques and software system after much studying, tinkering, and informal testing. To put our work on a firmer scientific footing, we conducted two experiments that compared the time to select with our eye gaze technique versus time to select with the most commonly used input device, the mouse. Our research hypothesis is that selecting with our eye gaze technique is faster than selecting with a mouse.

Our hypothesis that our eye gaze selection technique is faster than mouse selection might seem hardly surprising. After all, we must move our eyes to the target before we move the mouse. In addition, physiological evidence suggests that saccades should be faster than arm movements. Saccades are ballistic in nature and have nearly linear biomechanical characteristics [Bartz 1962; Abrams et al. 1989; Prablanc and Pelisson 1990]. The mass of the eye is primarily from fluids and, in general, the eyeball can be moved easily in any direction. In contrast, arm and hand movements require moving the combined mass of joints, muscles, tendons, and bones. Movement is restricted by the structure of the arm. A limb is maneuvered by a series of controlled movements carried out under visually guided feedback [Sheridan 1979].

However, we are not simply comparing the behavior of the eye with that of the arm in these experiments; we are comparing two complete interaction techniques with their associated hardware, algorithms, and time delays. For our research hypothesis to be true, our algorithm, built from an understanding of eye movements, plus the eye tracker we use which adds its own delay, must not cancel out the inherent speed advantage of the eye.

5. METHOD

We conducted two experiments that compared the two techniques. Each experiment tried to simulate a real user selecting a real object based on his or her interest, stimulated by the task being performed. In both experiments, the subject selected one circle from a grid of circles shown on the screen. The first was a quick selection task, which measured "raw" selection speed. The circle to be selected was highlighted. The second experiment added a cognitive load. Each circle contained a letter, and the spoken name of the letter to be selected was played over an audio speaker. The two experiments differed only in their task. The underlying software, equipment, dependent measures, protocol, and subjects were the same.

5.1 Interaction Techniques

Our eye gaze selection technique is based on dwell time. We compared that with the standard mouse button-click selection technique found in direct manipulation interfaces. We chose eye dwell time rather than a manual button press as the most effective selection method for

the eye based on previous work [Jacob 1991]. A user gazes at an object for a sufficiently long time to indicate attention and the object responds, in this case by highlighting. A quick glance has no effect because it implies that the user is surveying the scene rather than attending to the object. Requiring a long gaze is awkward and unnatural so we set our dwell time to 150 msec., based on previous informal testing, to respond quickly with only a few false positive detections.

6. EXPERIMENT 1: CIRCLE TASK

The task for the first experiment was to select a circle from a three by four grid of circles as quickly as possible (the arrangement is shown in Figure 2). The diameter of each circle was 1.12 inches. Its center was 2.3 inches away from its neighboring circles in the horizontal and vertical directions and about 3 inches from the edge of the 11 by 14 inch CRT screen.

Targets were presented in sequences of 11 trials. The first trial was used for homing to a known start position and was not scored. The target sets were randomly generated and scripted. One restriction was imposed that no target was repeated twice in a row. The same target scripts were presented to each subject. A target highlighted at the start of a trial; when it was selected, it de-highlighted and the next target in the sequence highlighted immediately. In this way, the end position of the eye or mouse for one trial became the start position for the next. No circle other than the target was selectable (although information about wrong tries was recorded in the data file). We presented the trials serially rather than as discrete trials to capture the essence of a real user selecting a real object based on his or her own interest. The goal was to test our interaction technique in as natural a setting as possible within a laboratory experiment.

6.1 Apparatus

The subject sat in a straight-backed stationary chair in front of a table (29.5 inches tall) that held a Sun 20-inch color monitor. The eye to screen distance was approximately three feet. The mouse rested on a 15-inch square table (28.5 inches tall) that the subject could position. The eye tracker hardware and experimenter were located to the subject's left, which dictated that only individuals that use the mouse right-handed could be subjects (otherwise we would have had to rearrange the equipment and recalibrate). The operator stood in front of the eye tracker console to adjust the eye image when needed and control the order of the experiment. The subject wore a thin, lightweight velcro band around the forehead with a Polhemus 3SPACE Tracker sensor attached above the left eye, which allowed a little larger range of head motion with the eye tracker.

The eye tracker was an Applied Science Laboratories (Bedford, MA) Model 3250R corneal reflection eye tracker that shines an on-axis beam of infrared light to illuminate the pupil and produce a glint on the cornea. These two features - the pupil and corneal reflection - are used to determine the x and y coordinates of the user's visual line of gaze every 1/60 second. Temporal resolution is limited to the video frame rate so that some dynamics of a saccade are lost. The measurable field of view is 20 degrees of visual angle to either side of the optics, about 25 degrees above and 10-degrees below. Tracking two features allows some head movement because it is possible to distinguish head movements (corneal reflection and center of pupil move together) from eye movements (the two features move in opposition to one another). We extended the allowable range that a subject could move from one square inch to 36 square inches by adding mirror tracking (a servo-controlled mirror allows +/-6 inches of lateral and vertical head motion). Mirror tracking allows automatic or joystick controlled head tracking. We

enabled magnetic head tracking (using head movement data from the Polhemus mounted over the subject's left eye) for autofocusing.

The position of gaze was transmitted to a stand-alone Sun SPARCserver 670 MP through a serial port. The Sun performed additional filtering, fixation, and gaze recognition, and some further calibration, as well as running the experiments. The mouse was a standard Sun optical mouse. Current eye tracking technology is relatively immature, and we did have some equipment problems, including the expected problem of the eye tracker not working with all subjects. Our eye tracker has difficulties with hard contact lenses, dry eyes, glasses that turn dark in bright light, and certain corneas that produce only a dim glint when a light is shown from below. Eye trackers are improving, and we expect newer models will someday solve many of these problems.

Our laboratory's standard procedure for collecting data is to write every timestamped event to disk as rapidly as possible for later analysis, rather than to perform any data reduction on the fly [Jacob et. al. 1994]. Trials on which the mouse was used for selection tracked the eye as well, for future analysis. We stored mouse motion, mouse button events, eye fixation (start, continuation, end), eye lost and found, eye gaze (start, continuation, end), start of experiment, eye and mouse wrong choices, eye and mouse correct choices, and timeout (when the subject could not complete a trial and the experiment moved on). All time was in milliseconds, either from the eye tracker clock (at 1/60 sec. resolution) or the Sun system clock (at 10 msec. resolution). We isolated the Sun from our network to eliminate outside influences on the system timing.

6.2 Subjects

Twenty-six technical personnel from the Information Technology Division of the Naval Research Laboratory volunteered to participate in the experiment without compensation. We tested them to find 16 for whom the eye tracker worked well. All had normal or corrected vision and used the mouse right-handed in their daily work (required because the eye tracker and experimenter occupied the space to the left). All participants were male, but this was not by design. The four women volunteers fell into the group whom the eye tracker failed to track, though women have successfully used our system in the past. The major problems were hard contact lenses and weak corneal reflections that did not work well with our system.

6.3 Procedure

Each subject first completed an eye tracker calibration program. The subject looked, in turn, at a grid of nine points numbered in order, left to right, top to bottom. This calibration was checked against a program on the Sun and further adjustments to the calibration were made, if needed, by recording the subject's eye position as they looked at 12 offset points, one at each target location. These two steps were repeated until the subject was able to select all the letters on the test grid without difficulty. The subject then practiced the task, first with the mouse and then the eye gaze selection technique. The idea was to teach the underlying task with the more familiar device. The subject completed six sets of 11 trials (each including the initial homing trial) with each interaction device. Practice was followed by a 1.5 minute break in which the subject was encouraged to look around; the eye was always tracked and the subject needed to move away from the infrared light of the eye tracker (the light dries the eye, but less than going to the beach) as well as to rest from concentrating on the task. In summary, the targets were

presented in blocks of 66 (six sequences of 11), mouse followed by eye. All subjects followed the same order of mouse block, eye block, 1.5 minute rest, mouse block, eye block. Because of difficulties with our set-up, we chose to run only one order. We felt this to be an acceptable, although not perfect solution, because the two techniques use different muscle groups, suggesting that the physical technique for manipulating the input should not transfer. Because of blocking in the design, we were able to test for learning and fatigue. Each experiment lasted approximately one hour.

6.4 Results

The results show that it was significantly faster to select a series of circle targets with eye gaze selection than with a mouse. The mean time for selection is shown in Figure 3. Figure 4 shows the median and spread of the distributions. Performance with eye gaze averaged 428 msec. faster. These observations were evaluated with a repeated-measures analysis of variance. Device effect was highly significant at $F(1,15) = 293.334$, $p < 0.0001$. The eye gaze and mouse selection techniques were presented in two blocks. While there was no significant learning or fatigue, the mouse did show a more typical learning pattern (performance on the second block averaged 43 msec. faster) while eye gaze selection remained about the same (about 4 msec. slower).

Only performance on correct trials was included in the analysis. We also observed that excessively long or short trials were generally caused by momentary equipment problems (primarily with the eye tracker; 11% of eye trials and 3% of mouse) and were therefore not good indications of performance. We removed these outliers using the common interquartile range criterion (any observation that is 1.5 times the interquartile range either above the third quartile or below the first was eliminated). An examination of the raw data suggested that this approach removed only questionable trials.

An issue is whether the stopping criteria, dwell time for the eye and click for the mouse, can be fairly compared. Does one take much more time than the other? When we first researched the question, we thought we would have to set our dwell time higher than 150 msec. because Olson and Olson [Olson and Olson 1990] reported that it takes 230 msec. to click a mouse. When we tested a click (mouse down - mouse up), we found it took less time in our setting. We confirmed our decision that using 150 msec. dwell time is reasonable by analyzing the time it actually took subjects to click the mouse in our circle experiment using the time-stamped data records we had collected. It took an average of 116 msec. Only four subjects averaged more than 150 msec., the highest being 165 msec. The fastest time was 83 msec. Olson and Olson's figure probably includes more than just the end condition we needed. We concluded that the 150 msec. dwell time compared with an average 116 msec. click for the mouse would, if anything, penalize performance in the eye condition rather than the mouse.

7. EXPERIMENT 2: LETTER TASK

The task for the second experiment was to select a letter from a grid of letters. Each letter was enclosed in a circle, and the circles were the same size and arrangement as in Experiment 1. The letters fit just inside the circles; each character was approximately 0.6 inches high, in a large Times font. The subject was told which letter to select by means of a prerecorded speech segment played through an audio speaker positioned to their right. When a letter was selected, it highlighted. If the choice was correct, the next letter was presented. If incorrect, a

"bong" tone was presented after 1250 msec. so that a subject who misheard the audio letter name could realize his or her mistake. We set the length of the delay through a series of informal tests. The delay we chose is fairly long, but we found if the signal came more quickly in the eye condition, it was annoying. (One pilot subject reported feeling like a human pin ball machine at a shorter duration!)

The apparatus used was the same as in the circle experiment with the addition of an audio speaker placed two feet to the right of the subject. The names of the letters were recorded on an EMU Emulator III Sampler and played via a MIDI command from the Sun. Playing the digitized audio, therefore, put no load on the main computer and did not affect the timing of the experiment. The internal software was the same and the same data were written to disk. The timing of the experiment was the same for the eye gaze selection condition and the mouse condition.

The subjects were the same 16 technical personnel. All completed the letter experiment within a few days after the circle experiment. The protocol for the letter experiment was identical to the first experiment: calibration, practice, alternating mouse and eye gaze blocks, all interspersed with breaks. The difference between the two experiments was the cognitive load added by having the subject first hear and understand a letter, and then find it. The purpose of the task was to approximate a real-world one of thinking of something and then acting on it.

7.1 Results

The results show that it was significantly faster to hear a letter and select it by eye gaze selection than with the mouse. The mean time for selection is shown in Figure 3. Figure 4 presents the median and spread of the distributions. Performance with eye gaze averaged 338 msec. faster. These observations were evaluated with a repeated-measures analysis of variance. Device effect was highly significant at $F(1,15) = 292.016$, $p < 0.0001$. The eye gaze and mouse selection techniques also were presented in two blocks. Again, there was no significant interaction. The mouse showed typical learning (performance in the second block averaged 17 msec. faster). Eye gaze selection showed some slowing (by 17 msec.). Again, only performance on correct trials was included in the analysis and outliers were removed as before (5% of eye trials and 3% of mouse).

8. DISCUSSION

Our experiments show that our eye gaze selection technique is faster than selecting with a mouse on two basic tasks. Despite some difficulties with the immature eye tracking technology, eye selection held up well. Our subjects were comfortable selecting with their eyes. There was some slight slowing of performance with eye gaze that might indicate fatigue, but there is not enough evidence to draw a conclusion.

We do not claim that the speed advantage we obtained is sufficient reason to use this technology. What the speed advantage shows is that our eye gaze interaction technique and the hardware we used works well. Our algorithm maintains the speed advantage of the eye. Our previous experience suggests benefits for eye gaze interaction in naturalness and ease. It is a good additional input channel, and we have now shown that its claimed benefits can be obtained without incurring any performance penalty.

In making our comparisons, we explicitly excluded the higher cost of the eye tracking equipment, which we view as a temporary obstacle as these costs continue to fall. We are concerned with the potential of eye movement-based interaction in general, rather than the performance of a particular eye tracker. For our results to be useful in practical settings, we postulate a better and cheaper eye tracker becoming available, but we simulate such with the hardware available today. Except for the most severely time-critical applications, we would not suggest deploying a duplicate of our laboratory configuration yet.

Because both experiments used the same design and subjects, we can say something about how the two different tasks responded to our techniques. The increment in time from the circle experiment to the letter experiment was similar for each device: 599 msec. for the eye and 509 msec. for the mouse. We suggest that this increment might account for a comprehension and search subtask in the letter experiment, which was not required in the circle one. That subtask is likely to be similar regardless of whether mouse or eye gaze is used. The speed advantage for eye gaze in the selection phase is about the same across tasks.

9. MODEL OF EYE GAZE

Fitts' Law has proven a useful predictor of target acquisition times on movement tasks where the goal is to reach a target region quickly and accurately [for example, Card et. al., 1978; Card et. al. 1983; Jagacinski et. al., 1980; Johnsgard 1994; Langolf et. al., 1976; MacKenzie et. al. 1991; MacKenzie 1992; Radwin 1990; Ware and Mikaelian 1987]. We applied Fitts' Law to results from the first experiment (selecting circles) to illustrate the difference between our eye gaze selection technique and mouse selection and compare our eye gaze technique with that of others.

9.1 Background: A Review of Fitts' Law

Fitts' Law relates movement time of a particular limb and set of muscles to the capacity of that motor system to process information [Fitts 1954; Fitts and Peterson 1964]. It unifies distance, movement time, and the variability of movements in one measure. Fitts' Law suggests that given a fairly constant rate of processing in the human motor system, a linear relationship holds between movement time (MT) and the index of difficulty (ID) of the movement task: .

$$MT = a + b \text{ ID} \quad (1)$$

ID is specified in terms of target width (W), and amplitude or movement distance (A), according to the formula: .

$$ID = \log_2 (2A/W) \text{ bits per response} \quad (2)$$

Fitts based his work on information theory and ID is measured in units of bits, the same as capacity of a communication channel. Fitts defined ID as the amount of information that the movement is required to generate [Fitts and Radford 1966]. Increasing the amplitude of a movement or decreasing target width increases the difficulty of the movement. Fitts also borrowed the information theoretic interpretation of the slope b as the inverse of the motor system information processing rate channel capacity. He called this the index of performance (IP). Langolf, Chaffin, and Foulke [1976] found that short distance finger and wrist motions showed much higher rates (38 and 23 bits per sec.) than longer-distance arm movements (10 bits per sec.), supporting Fitts' contention that various limb segments show different maximum information processing rates. MacKenzie [1992] provides an excellent table of results, including

the IP number, from a number of experiments studying user input devices. (IP or slope (b) of the ID-MT line is examined when comparing studies.) The original form of Fitts' Law is:

$$MT = a + b \log_2 (2A/W) \quad (3)$$

In equation (3), a and b are empirically fitted regression parameters. Scatter plots of ID against MT data reveal an upward curvature of MT away from the regression line for low values of ID. Welford [1968] proposed a modification to Fitts' formulation that improves the fit when ID is small and produces a slightly higher correlations with observed data [Card et al. 1978; Drury 1975; MacKenzie 1989]:

$$MT = a + b \log_2 (A/W + 0.5) \quad (4)$$

Welford also suggested that instead of the actual target width, a corrected estimate of W , adjusted for errors ($W \pm 2$ standard deviation), should be used in computing ID. This correction maintains a 4% error rate. Fitts [1964] supported both these arguments.

MacKenzie [1989] suggested further modifications that bring the formulation even closer to its information theoretic background, again improving correlation and the fit for small IDs:

$$MT = a + b \log_2 (A/W + 1) \quad (5)$$

Not all researchers favor Fitts' information-theoretic explanation of the logarithmic speed-accuracy tradeoff and propose alternative explanations [Crossman and Goodeve 1983; Sheridan 1979; Meyer et. al., 1988]. Regardless of the explanation however, there is strong agreement that average movement times conform well to Fitts' Law [Langolf 1976; Jagacinski 1985; Walker et. al., 1993]. Most would agree with Meyer, Smith, Kornblum, Abrams, and Wright [1990] who conclude that they expect performance to obey Fitts' Law approximately but not exactly.

Fitts' Law is a sound measure of aggregate performance and a valuable engineering model for understanding movement toward a target for human-computer interaction problems, an idea first suggested by Card, Moran, and Newell [1983]. Fitts' Law can be used to compare input devices performance on tasks that require absolute accuracy with unconstrained movement.

There are two cautions to using Fitts' Law. First, the nature of the task is important. There are many tasks that do not produce Fitts' results such as constrained arm movement [Kvalseth 1973] and movement to a target point within a pre-specified duration [Wright and Meyer 1983]. Second, there is a limitation to Fitts' Law for comparing studies. Different studies that test the same devices can produce different IP scores. For example, MacKenzie, Sellen, and Buxton reported an IP of 4.5 and our mouse version of the circle experiment produced 8.5 (the equation will be given in the *Results and Discussion* Section 9.4). The situation is even more complicated than this suggests. Devices can be called by the same name but constructed differently. Card, English, and Burr [1978] reported an IP of 10.4 for their mouse selection task but their mouse is not the standard optical or mechanical mouse in common use today. Also, the nature of the task used in comparing two devices influences the results as do the characteristics of the devices themselves. Fitts [1964] himself found that the slope of the function is less steep for discrete than serial responses. MacKenzie [1992] discussed the problem of comparing Fitts' Law results across studies and suggested a solution. He found that the ratio of IP values for two input devices within a study (using the same task) is in reasonable tolerance to the ratio of the same devices in another study.

9.2 Applying Fitts' Law

We will use Fitts' Law to compare eye gaze and mouse selection results from the first experiment and to compare our eye gaze selection results with those of Ware and Mikaelian [1987]. In developing this experiment, we were careful to craft a Fitts task. The subjects moved (either with eye or mouse) from one target area to another in an unconstrained manner, and the task was kept the same for both devices so that device characteristics rather than task influenced the slope of the equations.

Unfortunately we did not have a range of target sizes. We did have a range of distances, from adjacent targets to opposite ends of the screen. The distance for each trial is known because the starting point for a trial is the target from the preceding trial (providing the user hit the correct target on the preceding trial.) These precluded a complete Fitts' analysis, but they allowed us to investigate the time/distance relationship, which is our main interest here.

Our prediction is that, unlike the mouse, the eye selection data should not show a strong time/distance trade-off. If our approach preserves the physical characteristics of eye movement as we intended, the eye gaze selection data should have a low correlation and an almost flat relationship between time and ID. We expected the mouse data to be well correlated with the Fitts' model, showing a strong positive relationship consistent with past experiments.

9.3 Design

The set of distances available for the Fitts' analysis was determined by the layout of the target array. The targets were arranged in a three by four grid of circles, with the center of each circle 2.3 inches away from its neighbors in the horizontal and vertical directions. A sequence of targets could move in any direction: horizontal, vertical, or diagonal. Therefore, the set of all possible distances in inches ("A" in Fitts' formulation) was [2.3, 3.0, 3.7, 4.5, 5.4, 5.9, 6.4, 7.5, 8.9, 9.2, 10.0]. This set was crossed with the one target width (W) to produce the set of IDs: [4.8, 5.2, 5.6, 5.8, 6.1, 6.2, 6.3, 6.6, 6.8, 6.9, 7.0].

We chose circles for targets to simplify the calculation of target width from different approach angles. Fitts' original work examined horizontal movement only but other research has shown that Fitts' Law holds for target acquisition time in a two-dimensional array [Jagacinski et al. 1985]. Card, English, and Burr [1978] showed that approach angle makes no difference in time to select with a mouse. For the eye, some authors found that upward motions start somewhat sooner than downward motions and oblique movements start somewhat later than up or down motions [Boff and Lincoln 1988, p. 489]. The noise in the eye tracker recording system swamps such fine differences.

9.4 Results and Discussion

We have analyzed our data using both the Welford variation of Fitts' Law (equation 4) and that formulated by MacKenzie (equation 5) and obtained very similar results. We will report the Welford results because we have no low ID values and the use of Welford is consistent with many HCI researchers [Card et al. 1978; Ware and Mikaelian 1987]. We also do not need to correct for errors because only correct trials were included. Trials with wrong selection followed by correct selection were removed from this (but were included as part of total time in the first analysis). Our calculations take the mean of each subject's performance time and then the grand mean of all times to get average aggregate performance, consistent with most previous

researchers. We used a conventional regression procedure that fit a least squares multivariate regression to the data (rather than newer robust regression techniques, in order to be consistent with past research).

For the eye, the regression equation is $MT = 484.5 + 1.7 \log_2 (A / W + .05)$ ($r^2 = .02$). For the mouse, it is $MT = 155.3 + 117.7 \log_2 (A / W + .05)$ ($r^2 = .86$). These results support our predictions (see Figure 5). The mouse data are well modeled by Fitts' Law while the eye data are not. The eye results suggest a flat model, with approximately equal time to cover the set of distances.

The mouse results are similar to other Fitts' studies. Our eye results are more similar to those of Abrams, Meyer, and Kornblum [1989] who studied pure eye movement and showed some increase in time of saccadic eye movements with movement distance but a noticeable increase in velocity. We take this result as validating that our software reasonably preserves raw movement characteristics of the eye.

The slope of Ware's and Mikaelian's [1987] eye interaction data was a steeper slope, almost like a mouse's (they do not mention the Fitts equation but their graph includes a plot of mouse performance from Card, Moran, and Newell [1983]). One important difference from their work is that is that our task is a fairly pure movement one that does not involve long dwell times as does theirs. One of the cautionary aspects to using Fitts' Law is that the nature of the task influences the results as much as the characteristics of the input device.

This analysis helps explain why eye gaze selection could be a useful interaction tool. A technique that preserves the speed of saccadic eye movements means that movements would take about the same amount of time for a range of distances, unlike a device like the mouse which has a pronounced time/distance trade-off. As screens, workspaces, and virtual environments become larger, the speed advantage for the eye becomes more valuable.

10. CONCLUSIONS

Eye gaze interaction techniques are a useful source of additional input and should be considered when designing advanced interfaces. Moving the eyes is natural, requires little conscious effort, and frees the hands for other tasks. People easily gaze at the world while performing other tasks so eye combined with other input techniques requires little additional effort. An important side benefit is that eye position implicitly indicates the focus of the user's attention.

We presented two experiments to demonstrate the advantages of using a person's natural eye gaze as a source of computer input. The circle experiment attempted to measure *raw* performance, while the letter experiment simulated a real task in which the user first decides which object to select and then finds it. Our experimental results show that selecting with our eye gaze technique is faster than selecting with a mouse. The speed difference with the eye is most evident in the circle experiment. In it, selecting a sequence of targets is so quick and effortless that it almost feels like watching a moving target.

The Fitts analysis points out that, within the range we have tested, the farther you need to move, the greater the advantage of the eye because its *cost* is nearly constant. While the resolution of the eye makes it impractical for positioning tasks that require precision, it is excellent for jumping to distant regions of the screen quickly (where the hand might then be used for detailed work).

We argue for using natural eye movements and demonstrate constructing interaction techniques based on an understanding of the physiology of the eye. Our software extracts useful information about the user's higher-level intentions from noisy, jittery eye movement data. We also use Fitts' Law as a useful engineering tool to compare our results with previous work and show our approach preserves the natural quickness of the eye.

ACKNOWLEDGMENTS

We thank Debby Hix for all kinds of help with this research; Astrid Schmidt-Nielsen for help and advice on experimental design and data analysis; and our colleagues at the Naval Research Laboratory who took time from their work to serve as experimental subjects. Portions of this work were sponsored by the Office of Naval Research (at NRL) and ONR Grant N00014-95-1-1099, NSF Grant IRI-9625573, and NRL Grant N00014-95-1-G014 (at Tufts University).

REFERENCES

- ABRAMS, R. A., MEYER, D. E., AND KORNBLUM, S. 1989. Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 3, 529-543.
- BARTZ, A. E. 1962. Eye-movement latency, duration, and response time as a function of angular displacement. *Journal of Experimental Psychology*, 64, 3, 318-324.
- BOFF, K. R. and LINCOLN, J. E., eds. 1988. *Engineering Data Compendium: Human Perception and Performance*. AAMRL, Wright-Patterson AFB, OH.
- BOLT, R. A. 1981. Gaze-orchestrated dynamic windows. *Computer Graphics*, 15, 3 (Aug.), 109-119.
- BOLT, R. A. 1982. Eyes at the interface. *Proc. ACM CHI'82*, 360-362.
- BOLT, R. A. 1992. *Eyemovements in Human/Computer Dialogue*. AHIG Research Report 92-1, MIT Media Laboratory.
- CARD, S. K., ENGLISH, W. K., and BURR, B. J. 1978. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a crt. *Ergonomics*, 21, 8, 601-613.
- CARD, S. K., MORAN, T. P., and NEWELL, A. 1983. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, N. J.
- CROSSMAN, E. R. F. W. and GOODEVE, P. J. 1983. Feedback control of hand-movement and fitts' law. *Quarterly Journal of Experimental Psychology*, 35A, 251-278.
- DRURY, C. G. 1975. Application of fitts' law to foot-pedal design. *Human Factors*, 17, 4, 368-373.
- FITTS, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 6, 381-391.
- FITTS, P. M. and PETERSON, J. R. 1964. Information capacity of discrete motor responses. *Journal of Experimental Psychology*, 67, 2, 103-112.

- FITTS, P. M. and RADFORD, B. K. 1966. Information capacity of discrete motor responses under different cognitive sets. *Journal of Experimental psychology*, 71, 4, 475-482.
- FLAGG, B. N. 1977. *Children and Television: Effects of Stimulus Repetition on Eye Activity*. Thesis, Doctor of Education degree, Graduate School of Education, Harvard University.
- GLENN III, F. A., IAVECCHIA, H. P., ROSS, L. V., STOKES, J. M., WEILAND, W. J., WEISS, D., and ZAKLAND, A. L. 1986. Eye-voice-controlled interface, *Proc. of the Human Factors Society*, 322-326.
- GOLDBERG, J. H. and SCHRYVER, J. C. 1993. Eye-gaze control of the computer interface: discrimination of zoom intent. *Proc. Human Factors and Ergonomics Society*, 1370-1374.
- HANSEN, J. P., ANDERSEN, A. W., and ROED, P. 1995. *Eye-gaze control of multimedia systems*. Presented at the HCI International '95 Conference, Tokyo.
- 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 (April), 152-169.
- JACOB, R. J. K. 1992. What you look at is what you get: Eye movement-based interaction techniques. *Proc. ACM CHI'90*, 11-18.
- JACOB, R. J. K. 1993a. Eye-gaze computer interfaces: What you look at is what you get. *IEEE Computer*, 26, 7 (July), 65-67.
- JACOB, R. J. K. 1993b. What you look at is what you get: Using eye movements as computer input. *Proc. Virtual Reality Systems'93*, 164-166.
- JACOB, R. J. K. 1993c. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. In *Advances in Human-Computer Interaction*, Vol. 4, ed. H. R. Hartson and D. Hix, Ablex Publishing Co., Norwood, N.J., 151-190.
- JACOB, R. J. K. 1994. Eye tracking in advanced interface design. In *Advanced Interface Design and Virtual Environments*, ed. W. Barfield and T. Furness, Oxford University Press, Oxford.
- JACOB, R. J. K., SIBERT, L. E., MCFARLAND, D. C., MULLEN, M. P. 1994. Integrality and separability of input devices. *ACM Transactions on Computer-Human Interaction*, 1, 3-26.
- JAGACINSKI, R. J. and MONK, D. L. 1985. Fitts' law in two dimensions with hand and head movements. *Journal of Motor Behavior*, 17, 1, 77-95.
- JAGACINSKI, R. J., REPPERGER, D. W., MORAN, M. S., WARD, S. L., and GLASS, B. 1980. Fitts' law and the microstructure of rapid discrete movements. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 2, 309-320.
- JOHNSGARD, T. 1994. Fitts' law with a virtual reality glove and a mouse: effects of gain. *Graphics Interface '94*, 8-15.
- JUST, M. A. and CARPENTER, P. A. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 4, 329-354.

- JUST, M. A. and CARPENTER, P. A. 1976. Eye fixations and cognitive processes. *Cognitive Psychology*, 8, 441-480.
- KVALSETH, T. O. 1973. Fitts' law for manipulative temporal motor responses with and without path constraints. *Perceptual and Motor Skills*, 37, 427-431.
- LAMBERT, R.H., MONTY, R.A., and HALL, R.J. 1974. High-speed data processing and unobtrusive monitoring of eye movements. *Behavior Research Methods and Instrumentation*, 6,6, 525-530.
- LANGOLF, G. D., CHAFFIN, D. B., and FOULKE, J. A. 1976. An investigation of fitts' law using a wide range of movement amplitudes. *Journal of Motor Behavior*, 8, 2, 113-128.
- MACKENZIE, I. S. 1989. A note on the information-theoretic basis for fitts' law. *Journal of Motor Behavior*, 21, 3, 323-330.
- MACKENZIE, I. S. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7, 91-139.
- MACKENZIE, I. S. and BUXTON, W. 1992. Extending fitts' law to two-dimensional tasks. *Proc. ACM CHI'92*, 219-226.
- MACKENZIE, I. S., SELLEN, A., and BUXTON, W. 1991. A comparison of input devices in elemental pointing and dragging tasks. *Proc. ACM CHI'91*, 161-166.
- MEYER, D. E., ABRAMS, R. A., KORNBLUM S., WRIGHT, C E., and SMITH J. E. K. 1988. Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, 95, 3, 340-370.
- MEYER, D. E., SMITH J. E. K., KORNBLUM S., ABRAMS, R. A., and WRIGHT, C E. 1990. Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In *Attention and Performance XIII: Motor Representations and Control*, M. Jeannerof (ed.). Lawrence Erlbaum Associates, Hillsdale, N.J., 173-226.
- OLSON, J. R., and OLSON, G. M. 1990. The growth of cognitive modeling in human-computer interaction since goms. *Human-Computer Interaction*, 5, 221-265.
- PRABLANC, C. and PELISSON, D. 1990. Gaze saccade orienting and hand pointing are locked to their goal by quick internal loops. In *Attention and Performance XIII: Motor Representation and Control*, M. Jeannerof (ed.). Lawrence Erlbaum Associates, Hillsdale, N.J., 653-676.
- RADWIN, R. G., VANDERHEIDEN, G. C., and LIN, M-L. 1990. A method for evaluating head-controlled computer input devices using fitts' law. *Human Factors*, 32, 4, 423-438.
- SHERIDAN, M. R. 1979. A reappraisal of fitts' law. *Journal of Motor Behavior*, 11, 3, 179-188.
- STARKER, I., and BOLT, R. A. 1990. A gaze-responsive self-disclosing display. *Proc. ACM CHI'90*, 3-9.
- THORISSON, K. R., and KOONS, D. B. 1992. *Unconstrained Eye Tracking in Multi-Modal Natural Dialogue*. AHIG Research Report 92-4, MIT Media Laboratory.

- WALKER, N., MEYER, C. E., and SMELCER, J. B. 1998. Spatial and temporal characteristics of rapid cursor-positioning movements with electromechanical mice in human-computer interaction. *Human Factors*, 35, 3, 431-458.
- WARE, C. and MIKAELIAN, H. H. 1987. An evaluation of an eye tracker as a device for computer input. *Proc. ACM CHI'87*, pp. 183-188.
- WELFORD, A. T. 1968. Movement. In *Fundamentals of Skill*. Methuen & Co., London (1968), 137-160.
- YOUNG, L. R. and SHEENA, C.. 1975. Methods & designs: survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, 7,5 (1975), 397-429.
- ZHAI, S., MORIMOTO, C., and IHDE, S. 1999. Manual and gaze input cascaded (MAGIC) pointing. *Proc. ACM CHI'99*, 246-253.