A Fitts' Law Comparison of Eye Tracking and Manual Input in the Selection of Visual Targets

Roel Vertegaal
Human Media Lab, Queen's University
Kingston, ON K7L 3N6, Canada
roel@cs.queensu.ca

ABSTRACT

We present a Fitts' Law evaluation of a number of eye tracking and manual input devices in the selection of large visual targets. We compared performance of two eye tracking techniques, manual click and dwell time click, with that of mouse and stylus. Results show eye tracking with manual click outperformed the mouse by 16%, with dwell time click 46% faster. However, eye tracking conditions suffered a high error rate of 11.7% for manual click and 43% for dwell time click conditions. After Welford correction eye tracking still appears to outperform manual input, with IPs of 13.8 bits/s for dwell time click, and 10.9 bits/s for manual click. Eye tracking with manual click provides the best tradeoff between speed and accuracy, and was preferred by 50% of participants. Mouse and stylus had IPs of 4.7 and 4.2 respectively. However, their low error rate of 5% makes these techniques more suitable for refined target selection.

Categories and Subject Descriptors

H5.2. Information interfaces and presentation (e.g., HCI): User interfaces: Input devices and strategies.

General Terms

Human Factors

Keywords

Input Devices, Focus Selection, Fitts' Law, Eye Tracking, Attentive User Interfaces.

INTRODUCTION

The use of eye input as a means for operating graphical as well as ubiquitous computing interfaces has been a topic of interest in human-computer interaction for quite some time [2,10,22,33,35]. One quest has been, particularly in the area of interfaces for the disabled [13], to incorporate eye trackers as pointing devices in graphical user interfaces. There are a number of reasons why the use of eye gaze as a means for target selection might seem compelling [26]:

1) When the hands are occupied or unavailable, eye gaze provides an extra and independent channel of input.
2) The eyes have the fastest muscles in the human body. Moreover, during target acquisition, users tend to look at a target before initiating manual action [10]. This means that eye gaze theoretically provides one of the fastest possible input methods available.
3) Users can produce thousands of eye movements without any apparent fatigue. Use of eye gaze mitigates the need for repetitive manual actions, thus reducing the risk of repetitive strain injury.
4) Users are familiar with the use of their eyes as a means for selecting the context of their commands. E.g., eye gaze is used during communications to indicate who is being addressed [29].

However, it is important to note that humans are not very familiar with the use of their eyes as a pointing device. Firstly, the chief purpose of the eyes is to provide input to the human body, rather than provide output to the exterior environment. Secondly, the eyes move very rapidly between fixation points, to inhibit movement of the world on the retina [5]. Such movements between eye fixations are known as saccades, and typically complete within 60 ms [5]. This means eye input appears more suitable for the selection of discrete targets, than for pointing in a coordinate space. Perhaps as a consequence of this, some researchers have been disappointed with the use of eye trackers for interaction in Graphical User Interfaces.

There are indeed many arguments against the use of eye tracking as a pointing device. Firstly, eye trackers have been known to be sensitive to user head movements. With advances in computer vision, recent systems such as the Tobii 1750 [28] as well as the LC Technologies' EyeGaze System [13] now feature head movement tolerances of over 30x15x20 cm. Secondly, eye tracking can be noisy. That said, on-screen accuracies of 1 degree or better are now the norm [28]. It has been suggested by Jacob and others that the accuracy of eye trackers in pointing is fundamentally limited by the size of the human fovea, which is in the order of two degrees of visual angle [5].

ICMI'08, Oct 20–22, 2008, Chania, Krete, Greece.

However, accuracy limitations are mostly caused by limitations in computer vision of the pupil. As computer vision algorithms of the pupil improve, so should the accuracy of eye tracking input. Thirdly, eye trackers need to be calibrated. However, as work by Smith et al. demonstrates [26], discrete targets can now be tracked without any calibration whatsoever. Finally, eye trackers suffer from what is known as the Midas Touch Effect [10]. The Midas Touch Effect is caused by overloading the visual input function of the eye with a motor output task. Providing clicks with the eyes is useful particularly in cases where users do not have control over limbs other than their eyes. In such cases, the Midas Touch effect causes users to inadvertently select or activate any target they fixate upon. By issuing a click only when the user has fixated on a target for a certain amount of time (*dwell time click*), the Midas Touch effect can be mitigated. By issuing a click through another channel, such as with the hands or using the voice, it can be avoided entirely.

A New Perspective: Eye Tracking As A Device for Deixis

Given the above discussion, rather than thinking of eye tracking as a means for pointing in a coordinate space, we argue it is more useful to consider its use for disambiguating between targets that provide visual context to interaction. This is because eye input provides a very specific kind of information about the user: the objects that are subject to his or her visual attention. As Yarbus [35] showed, eye fixations tend to pertain to distinct objects, not arbitrary spatial coordinates. Fixations patterns in scenes are object-oriented, and different every time. They depend highly on context. This led Smith et al. to explore the notion of deixis as a metaphor for pointing [26]. Rather than providing a spatial coordinate, deixis specifies the referent, as in, "*that* object", within a spatial context [2]. The act of looking at an object may inform other kinds of input activity, such as a button click or speech command, as pertaining to the semantics provided by that object. This fits well with the role of the eyes in the kinematic chain model proposed by Guiard [9]. He saw the hands function as serially assembled links in a chain, with the left (or non-dominant) hand as a base link that provides context to the right (or dominant) hand, the terminal link. If we include the eyes in his model, we notice that their activity provides context to, and precedes, the activities of both hands. As such, we see eye input as best suited to provide context to the manual action that ends the kinematic chain.

Outline of the Paper

It is within the context of the above discussion that we performed a comparative study of the efficiency of a number of eye and traditional manual input techniques: not with the intent that eye input substitutes a manual device, but with the intent that eye input can be used to disambiguate between contexts of interaction, for example, as provided by two GUI windows. This fits a classic Fitt's Law paradigm, where the object is to iteratively select between two visual targets presented on screen. For fair comparison, and given the relatively large size of typical interaction contexts, we deployed targets with low Indices of Difficulty, as compared to prior studies. We will first discuss some examples of systems that use eye input for context selection, as well as prior empirical work. We then discuss our experimental evaluation, concluding that eye tracking input is indeed faster, but also more inaccurate, than manual techniques.

BACKGROUND

Recently, a number of desktop interfaces have been developed that use eye input primarily as a means of selecting context for other channels of input. Many of these interfaces were inspired by early work by Bolt et al. [2,27].

In [8], Fono et al. presented an elastic windowing interface that used an eye tracker for focus window selection. They argued eye trackers are most suitable for focus selection when windows do not overlap. Repositioning of a window, or scrolling, entails pointing in a coordinate space and is typically best left to the hands. They evaluated performance during window selection when hands were overloaded with a typing task. Results showed eye tracking with a manual click was twice as fast as the use of function keys or mouse for focus window selection. In [1], Ashmore et al. discussed the use of an eye tracker in fisheye views. They used eye input to determine the area of the screen that contained the region of interest. Fisheye magnification was subsequently activated with a manual click. Unfortunately, their evaluations did not compare performance between eye and manual input in this task. In [20], Qvarfordt and Zhai discussed iTourist, a system for city trip planning that senses users' interest in locations on a map based on eye-gaze patterns. Their system used a statistical visual interest score to determine which of a limited number of objects on a map the user might be referring to, e.g., when booking a hotel. However, manual pointing was deployed once the user committed to booking a place.

Similar efforts are underway in the area of Ubiquitous and Mobile Computing. In [26], Smith et al. presented a simple wearable calibration-free eye tracker. Their system allows any real-world object to be augmented with eye tracking capabilities by embedding a small infrared tag. They discussed application scenarios where tags broadcast URL information as the user looks at a real-world billboard. Users subsequently retrieve the URLs on a PDA via manual interaction techniques. Eye tracking sensors have also been used to set the context for speech and remote commands [11,16,19,23,21,30], as well as for communications [12,29]. In all these interfaces, the use of an eye tracker is *not* to obtain a spatial coordinate, but to determine which of a limited number of clearly distinguishable objects the user is interested in.

Empirical Studies

Unfortunately, there are few empirical studies on the efficiency of eye tracking input devices in object selection tasks. Part of the reason for this has been the limited availability of easy-to-use vision-based eye trackers. Perhaps because of technological problems, empirical results of target selection efficiency of eye tracking have traditionally been somewhat mixed. In one of the earliest available studies, Ware & Mikaelian [33] evaluated the use of an eye tracker in a pointing task. To assess performance, they used a standard Fitts' Law experimental paradigm [7]. In this paradigm, size (*width*) and distance (*amplitude*) of two targets is varied systematically during a pointing task. The task requires participants to reciprocally select targets using their pointing device. Fitts' Law relates their movement time (MT) in this task to the Index of Difficulty (ID) of the target pair, as given by the log term in the Fitts' Law equation [15]:

$$MT = a + b \log (A/W + 1) \qquad \text{(Equation 1)}$$

Here, a is a constant that describes the response time per selection, while b describes the relationship between movement time and ID of the task. A is the distance or amplitude between targets, and W the width of the targets. After measuring movement time for each target ID, regression is used to derive values for a and b. Performance of a device is given by the reciprocal of b, which is called the Index of Performance (IP), measured in bits/s. Ware and Mikaelian [33] compared three eye pointing styles for selecting targets on a CRT: (1) dwell time click, where the target was selected if the observers' gaze fixated on it for more than .4 s; (2) screen button, where the observer had to fixate on a button on the screen after looking at the target; and (3) hardware button, where the observers pushed a keyboard button while fixating on the target. Unfortunately, they did not investigate performance of manual pointing. However, results show click times compared favorably to those of the mouse, with an intercept approximately twice as small. They suggest the Index of Performance was similar to the one reported for the mouse in Card et al. [4]. In their experiment, movement time for eye input was significantly affected by target width, particularly with targets smaller than 1.5 degrees of visual angle. Error rates were high in all eye tracking conditions, with about 12% of trials discarded. Miniotas [17] did compare dwell time activated eye input with mouse pointing in a Fitts' Law task involving two widths (13 and 26 mm) and four amplitudes (26, 52, 104 and 208 mm). He found an Index of Performance of about 5.7 bits/s for dwell time activated selection, which was lower than that of the mouse (7.9 bits/s). This finding may be attributed to the fact that the experimenter used a relatively high dwell time threshold of 250 ms, which acted as a bottleneck during target selection. Unfortunately, the author did not report on error rates.

In one of the most favorable experiments to date, Sibert and Jacob [24] evaluated the use of a mouse and eye tracker with dwell-time activated click in a pointing task that involved selecting one of 16 circles on a screen. Circles were 1.12 inch in size, and placed 2.3 inches from each other. They found that trial completion time with the eye tracker was almost half that of the mouse. Again, a relatively high error rate of 11% was reported for the eye tracking condition, compared to only 3% for the mouse.

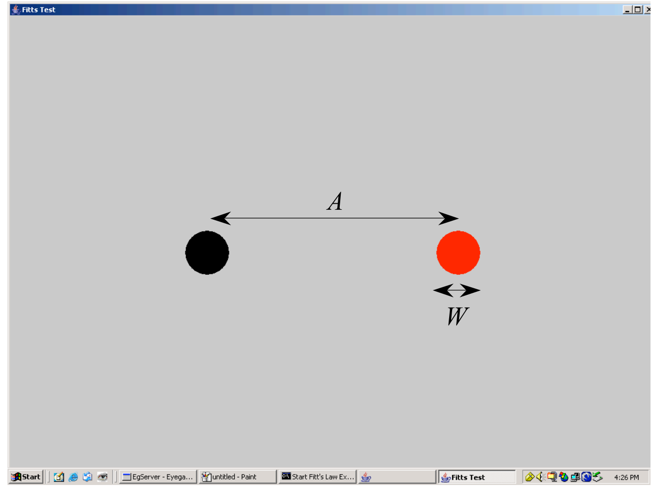

Figure 1. Screen shot of the experimental task.

Wang et al. [32] discussed an evaluation of eye-based selection of Chinese characters for text entry. In their task, users chose one of 8 on-screen Chinese characters by looking at the character while pressing the space bar. Results showed eye-based selection was not significantly faster than traditional key-based selection. They attributed this to lag present in their eye tracking equipment, and the fact that the overall time required to complete their task was dominated by decision time, rather than movement time.

Zhai et al. [35] evaluated the use of eye input in gaze assisted manual pointing. In their MAGIC pointing technique, an isometric joystick was used to select targets on a screen. However, to speed up isometric pointing, they positioned the cursor to a location close to the current eye fixation point whenever the user initiated movement with the joystick. Results were somewhat disappointing. MAGIC pointing only marginally improved movement time in a Fitt's Law task, with a mean completion time 8% faster than manual pointing. The Index of Performance of MAGIC was reported to be higher than that of manual isometric pointing, with 4.6 bits/sec vs. 3.2 bits/sec.

Fitts' Law Fit

Although manual pointing typically provides an excellent fit to the Fitts' Law model, there has been some debate as to whether the same would be true for eye input. Although they did not report a correlation value, Ware & Mikaelian [33] found that eye input does indeed follow Fitts' Law. Zhai [35] found correlations that were fairly low, in the order of r^2=.75. The most likely cause for this was the presence of eye tracker noise in their experiment. Perhaps most interestingly, Sibert and Jacob reported little variance in movement time over distance [24]. Miniotas [17] reported the highest fit with an r^2 = .98.

EXPERIMENTAL EVALUATION

For our experiment, we selected a state-of-the-art Rev.II LC Technologies desk-mounted eye tracker capable of performing with an on-screen accuracy better than 1 degree of visual arc. We used this tracker to compare performance of a number of input techniques in a Fitts' Law task.

Input Conditions and Training

We selected four pointing techniques for comparison, including two manual, and two eye tracker conditions:

1) *Mouse.* In this condition, we used a Wacom mouse mapped to the screen in absolute coordinates [31]. Users used the center button of the mouse to click a target.
2) *Stylus:* Here, users used a Wacom stylus to point at a target. Targets were selected by touching the tablet with the tip of the stylus.
3) *Eye tracking with manual click.* In this condition, the eyes were used for pointing, and a mouse button was used to click.
4) *Eye tracking with dwell time click.* Here, clicks were issued automatically after a fixation of 100 ms. While we realize this is too low for practical use, we purposefully chose this theoretical lower limit to investigate the Fitts' Law performance of the eyes themselves, when not throttled or encumbered by large dwell times.

In order to minimize lag in both eye tracking conditions, we performed as little processing as possible on the raw gaze coordinate obtained from the LC Technologies system. While this meant the on-screen cursor, in the form of a small crosshair, was moved to the location of the eyes without any form of temporal averaging, it reduced overall eye tracking latency by a factor two. All participants were trained with the use of every device prior to the experiment. Measurement started when their mean improvement in movement time was less than 10% between trials.

Apparatus

In all conditions, subjects were seated behind a 17" Dell LCD screen at about 60 cm. distance. This screen, with a resolution of 1024x768, was connected to a Rev. II LC Technologies Eyegaze system [13], which was used to constitute the eye tracking conditions. For the manual conditions, we used the standard stylus and mouse of the Wacom Intuos 2 [31] (model XD-0912-U, 30.4 *cm* x 45.7 *cm* active area, 2540 lpi resolution). The tablet area was mapped to the screen area using a linear mapping and a control-display gain of 1 for both devices. The software used to constitute the Fitts' Law task was a standard java package designed by Smith [25] (see Figure 1), and ran on the LC computer. In the eye tracking conditions, cursor movements and clicks were generated by a custom package that used the LC API to obtain eye tracking coordinates. Participants were calibrated for the eye tracker at the beginning of the experiment using a standard 15 point calibration that lasted 15 seconds. Only participants that achieved a calibration bias error of .15 inch were allowed to continue. We did not include any participants with glasses or contact lenses. Participants' head movements were not restricted in any way during trials.

Experiment Design and Task

Ten participants, 4 female and 6 male, carried out the experimental task. Participants consisted mostly of university students and faculty, with an average age of 27. 90% of participants were right-handed, with all participants using their dominant hand. All participants had prior experience with a mouse, 50% had prior experience with eye tracking, and 60% had used a Wacom stylus before. Users were asked to use their input device to point and click alternately between the two presented targets, as rapidly and accurately as they could. There were three constraints on the selection of target stimuli. Firstly, the on-screen accuracy of the eye tracker with a calibration error of .38 cm is about 1 cm. Target sizes should be double this in order to allow a good aim at the center of the target. This resulted in an absolute minimum target size of 70 pixels or 2 cm on screen. Our second constraint was size and resolution of the screen given the 40 degree angular range of the eye tracker, which limited the distance between the largest targets to about 800 pixels on-screen. Our third constraint was to obtain the broadest and most evenly distributed range of IDs possible.

Index of Difficulty

This resulted in a stimulus set which varied target width and amplitude fully crossed as follows:

> **Target width:** 70 pixels (2 cm or 2 degrees of visual angle at 60 cm distance), 100 pixels (3 cm or 3 degrees of visual angle at 60 cm distance) and 140 pixels (4 cm or 4 degrees of visual angle at 60 cm distance);
>
> **Target amplitude:** 200 pixels (6 cm or 6 degrees of visual angle at 60 cm distance), 400 pixels (12 cm or 12 degrees of visual angle at 60 cm distance) and 800 pixels (24 cm or 24 degrees of visual angle at 60 cm distance).

This led to an even distribution of ID, with targets ranging between 1.28 and 3.6 bits in difficulty. To not overly disadvantage the eye tracker, we removed duplicate ID targets involving a width of 70 pixels, the absolute minimum accuracy of the eye tracker. We used a within-subjects design in which all participants used all input techniques on all presented stimuli. Each target pair was presented 5 times per device. To counter order effects presentation of target stimuli was randomized. The order of input device presentation was counterbalanced using a Latin square design.

Handling Errors

An error was defined as a click outside the current target area, which was indicated using a red circle (see Figure 1). If the subject clicked off-target, a miss was logged but the

Input Technique	Mouse	Stylus	Eye with Manual Click	Eye with Dwell Time Click
Movement time ***(s.e.)***	*.66 (.03)*	*.63 (.03)*	*.57 (.04)*	*.45 (.02)*

Table 1. Mean movement times (s) and standard error per input technique.

Input Technique	Mouse	Stylus	Eye with Manual Click	Eye with Dwell Time Click
Error rate ***(s.e.)***	*4.6% (1.3%)*	*6.2% (1.5%)*	*11.7% (3.5%)*	*42.9% (3.7%)*

Table 2. Mean error rates and standard error per input technique.

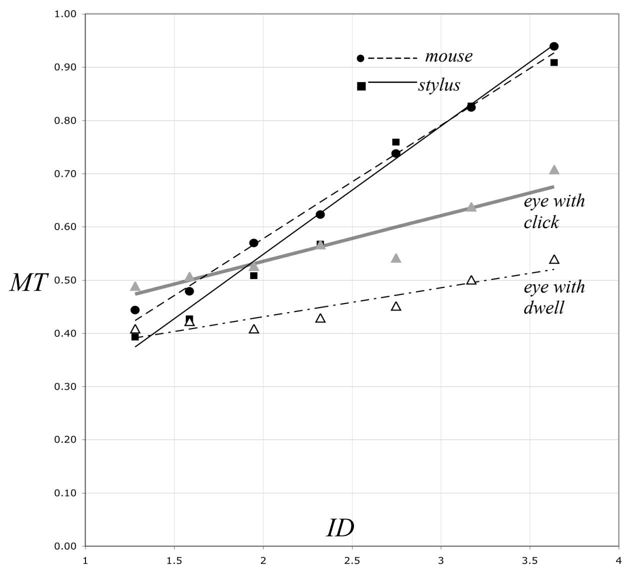

Figure 2. Regression lines per input device comparing Movement Time (s) with Index of Difficulty (in bits).

trial was continued until a target was hit. An extra trial was added to make up for the missed trial. In the eye tracking with dwell time condition, after an error, subjects had to move the cursor at least 32 pixels before dwell time clicking was reactivated. Only trials with no errors were collected for performance analyses.

Hypotheses

We expected both eye tracking techniques to outperform manual techniques in terms of IP and movement time across trials. We also expected eye tracking with dwell time click to outperform eye tracking with manual click. In terms of error rates, we expected eye tracking with dwell time click to be the outlier, with a higher percentage of errors than the other techniques. We expected both manual techniques to display superior accuracy.

RESULTS

Table 1 provides an overview of the mean movement times and standard errors during the experimental trials, per input technique. Movement time varied significantly with input technique: $F(3,252)=52.22, p<.001$. There was a significant interaction for ID *x* input technique: $F(18,252)=5.68, p < .001$. Post-hoc Bonferroni corrected results show highly significant differences in movement time between all input techniques, except between mouse and stylus $(p=.344)$. Eye tracking with dwell time click was faster than eye tracking with manual click $(p<.001)$, stylus $(p<.001)$ and mouse $(p<.001)$. Eye tracking with manual click was faster than stylus ($p=.007$) and mouse ($p<.001$). We also found significant effects of target width ($F(2,252)=19.62, p<.001$), target amplitude ($F(2,252)=80.96, p<.001$), as well as Index of Difficulty ($F(6,252)=56.15, p<.001$) on movement time. We found a significant interaction effect of target amplitude *x* input technique $(F(6,252)=11.79, p<.001)$, but not of target width *x* input technique $(F(6,252), p=.81)$, most likely because targets were large. Eye tracking with dwell time click featured significant effects of both target width $(F(2,63)=3.46, p=.038)$ and target amplitude $(F(2,63)=4.16, p=.02)$ on movement time.

Regression Models

Regressions showed completion time closely followed Fitts' Law for both manual devices. For both eye tracking conditions the fit was somewhat lower but still good. Figure 2 shows the regression models and mean data points for each device:

Mouse: $MT= .15 + .22 \log_2(A/W+1)$ $(r^2=.99)$

Stylus: $MT= .07 + .24 \log_2(A/W+1)$ $(r^2=.98)$

Eye with click: $MT= .36 + .086 \log_2(A/W+1)$ $(r^2=.87)$

Eye with dwell: $MT= .32 + .055 \log_2(A/W+1)$ $(r^2=.86)$

All regression models were highly significant. The Index of Performance (IP) was much lower for manual devices in comparison with eye tracking conditions, with 4.7 bits per second for the mouse and 4.2 bits per second for the stylus. The Index of Performance for eye tracking with manual click was about 2.5 times higher at 11.7 bits per second. Eye tracking with dwell time click performed at 18.3 bits per second. Performance of the stylus was somewhat lower than typically reported in literature. If we look closely at the data points of the stylus (■) in Figure 2, we observe that this is probably due to the higher ID data points, which appear to fit on a different regression line than the cluster of low ID points. This is most likely because subjects used wrist action for targets with lower amplitudes, and elbow movement for targets with high amplitudes. For the lower four data points, the index of performance of the stylus is indeed closer to 5.8 bits per second (r^2=.99).

Error Rates

Table 2 shows the error rates as a percentage of trials per input technique. As expected, we found a significant effect

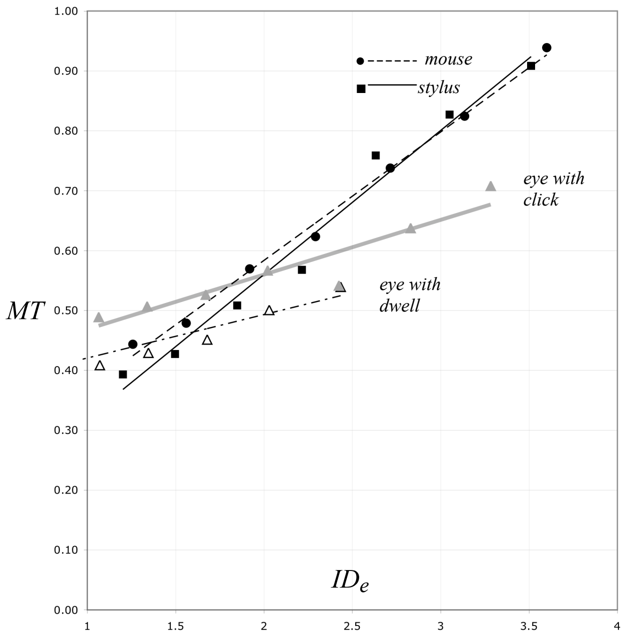

Figure 3. Regression lines per input device comparing Movement Time (s) with Effective Index of Difficulty.

for input technique on error rates: $F(3,36)=42.62$, $p<.001$. Bonferroni corrected pair-wise comparisons showed that eye tracking with dwell time click was indeed the outlier ($p<.001$), with all other comparisons insignificant.

Welford Corrected Models

According to Welford [34], error rates above 4% may lead to an inappropriate increase in IP (see also [5]). For fair comparison between eye tracking and manual conditions, our results required normalization to account for the larger effective width (W_e) aimed for by participants in the eye tracking conditions. We used a method proposed in [34] and popularized by MacKenzie [14]. For each subject and condition, we obtained the effective Index of Difficulty (ID_e) by normalizing all target widths to a standard 4% error. ID_e was calculated by multiplying the target width with the *two-tailed* z-score of 4% (i.e., 2% or 2.054) divided by the two-tailed z-score of the obtained error percentage per trial. Figure 3 shows the regression models and mean data points for each device:

Mouse: $MT= .15 + .21 \log_2(A/W_e+1)$ $(r^2=.99)$

Stylus: $MT= .08 + .24 \log_2(A/W_e+1)$ $(r^2=.98)$

Eye with click: $MT= .38 + .091 \log_2(A/W_e+1)$ $(r^2=.88)$

Eye with dwell: $MT= .35 + .073 \log_2(A/W_e+1)$ $(r^2=.89)$

All regression models were again highly significant, with a slightly better fit for the eye tracking conditions. The corrected Index of Performance (IP) was again much lower for manual device conditions, with 4.7 bits per second for the mouse and 4.2 bits per second for the stylus. By comparison, the Index of Performance for eye tracking with manual click was about 2.6 times higher than the stylus at 10.9 bits per second. IP for eye tracking with dwell time click was 3.3 times higher than that of the stylus at 13.8 bits per second. Figure 3 also shows how Welford correction caused a shift in the evaluated ID, particularly for eye tracking with dwell time click. Eye tracking with manual click was less affected by the operation. Differences in the resulting sets of IDs per device made it difficult to make a fair comparison of absolute movement time between manual input and eye tracking conditions after correction. We do note that even after correction, mean movement time for both eye tracking conditions appeared lower than for manual input when comparing a similar set of IDs. This can easily be observed in Figure 3: eye tracking regression lines are largely below those of manual conditions.

Input Technique	Mouse	Stylus	Eye with Manual Click	Eye with Dwell Time Click
Fast	*3.7*	*3.1*	*3.9*	*4.3*
Accurate	*4.2*	*4.1*	*3.4*	*2.7*
Easy to Learn	*4.5*	*4.3*	*4.1*	*3.8*

Table 3. Mean score per questionnaire item for each input technique.

Questionnaire

Table 3 shows the mean scores per input technique for three Likert-style questions on the questionnaire, with scales varying from 1 (strongly disagree) to 5 (strongly agree). Non-parametric within-subjects analysis of variance (Friedman Test) showed differences in perceived speed between input techniques were significant. Eye tracking with dwell time click ranked as the fastest technique: $\chi^2(3)=11.9$, $p<.008$. Input techniques also differed significantly in perceived accuracy, closely following results for error rates, with eye tracker with dwell time click obtaining the lowest score: $\chi^2(3)=12.6$, $p<.006$. Input techniques did not appear to differ significantly in learnability: $\chi^2(3)=6.9$, $p=.07$. 50% of participants ranked eye tracking with manual click as their preferred technique.

DISCUSSION

Our hypotheses were confirmed. However, the performance comparisons in this study underline both the strengths as well as the shortcomings of eye tracking as a technique for selecting even large sized targets. Eye tracking with dwell time click provides the fastest technique, with a very high IP of over 13.8 bits per second after Welford correction. This agrees with the general consensus that the eyes provide the fastest muscle set in the human body, and shows that considerable bandwidth gains can be made by incorporating eye tracking input in computer interfaces. Results generally agree with observations by Sibert and Jacob, with eye input with dwell time click on average 46%

faster than the mouse in the uncorrected case. The speed advantage of eye tracking is almost entirely explained by performance with higher indices of difficulty. We believe there are three possible explanations for this. Firstly, in our study, the intercept for both eye tracking conditions was about twice that of the manual conditions. This constant takes up a greater proportion of movement time in low ID cases. Secondly, speed gains of the eye muscles are greater over larger distances. This is because in those cases, manual input relies on the use of relatively slow muscles in the elbow, rather than the muscles of the wrist. Thirdly, when targets are further apart, participants are more likely to fixate on each target before initiating hand movements. While the fit to a Fitts' Law model of our eye tracking conditions is not as good as that of manual conditions, it is sufficient. Contrary to Sibert and Jacob [24], we did find an effect of target amplitude as well as width on movement time for eye tracking with dwell time click. The intercept of both eye input techniques was considerably lower than that reported by Ware and Mikaelian [33], but higher than the intercept of the mouse. The intercept in the eye tracking conditions can, in large part, be attributed to eye tracker latency. If we subtract the dwell time of 100 ms from the intercept of eye tracking with dwell time click, we obtain a remaining system latency of approximately 200 ms. We believe future eye tracking systems may reduce this latency to more acceptable levels. The higher regression line of eye tracking with manual click, reflects, in part, a speed penalty of 100 ms with this technique, as compared to dwell time clicking.

The Trouble With Eye Tracking

Our positive results are, however, still affected by the inaccuracies observed in the eye tracking conditions. A 5% error rate is generally regarded as acceptable in manual input studies. How useful is an extremely fast input device if it cannot be used to reliably select targets several centimeters in diameter? Welford-corrected results may show that eye tracking is superior in terms of speed when we disregard error, but this will not help the user selecting a small target using an interactive system.

Error rates were consistent with those reported in prior experiments. Our hypothesis that the eye tracker with dwell time click would be the outlier in terms of error rate was confirmed. At 43% overall, the error rate in this condition was unacceptably high. The high error rate of the dwell time condition can be attributed to the short threshold of only 100 ms, which we purposefully chose to observe maximum speed benefits. This does not necessarily mean that accuracy was simply traded for speed in this condition. High error rates in dwell time activated eye tracking should be considered a function of the interaction between noise present in the eye tracking data and the Midas Touch effect: users happened to click wherever their gaze appeared to rest. While the bulk of this noise is generated by current limitations in computer vision technology, and not by the muscles of the eyes themselves, the Midas Touch effect and associated high error rates make dwell time activated eye tracking less suitable for performing the kind of selection tasks common in graphical user interfaces.

At 10.9 bits per second after Welford correction, eye tracking with manual click provides the most reasonable tradeoff between accuracy and speed for use in selection of large targets (i.e., larger than 2 cm on screen). This is reflected in the preference of the majority of participants for this technique. While limitations in the accuracy of this technique may disqualify it for use in the kind of high resolution pointing tasks common in GUIs, eye tracking with manual click provides a fast alternative for the selection of large visual contexts several centimeters in size, such as windows, without a Midas Touch effect. We believe that for eye tracking to become a mainstream input device, continued refinements of computer vision algorithms, in terms of lag as well as spatial resolution, range, bias error and noise, remain necessary. This is worth the effort as the eye does appear to provide one of the most efficient forms of spatial input the human body has to offer.

CONCLUSIONS

In this paper, we presented a Fitts' Law comparison of the efficiency of eye tracking and manual input as a means for selecting large visual targets on a display. We compared performance of two eye tracking techniques, one with a manual click and one with a dwell time click, with that of mouse and Wacom stylus. Our findings indicate eye tracking with manual click outperforms the mouse by about 16%, with eye tracking with dwell time click about 46% faster. Due to the large intercept in both eye tracking conditions, speed gains pertained mostly to higher-ID targets. Eye tracking conditions also suffered from a high error rate of 11.7% for manual click and 43% for dwell time click. After Welford correction, eye input still appears to outperform manual devices, with dwell time click the fastest technique at an IP of 13.8 bits/s. Eye tracking with manual click performed at an IP of 10.9 bits/s, and appears to provide the best tradeoff between speed and accuracy. This input technique was also preferred by 50% of participants. The mouse and stylus performed relatively poorly by comparison, with IPs of 4.7 and 4.2 respectively. However, these techniques provided superior accuracy, with error rates in the order of 5%. We conclude that accuracy of eye input remains an issue even with large targets, thus limiting its suitability for refined target selection tasks.

REFERENCES

1. Ashmore, M., Duchowski, A., Shoemaker, G. Efficient Eye Pointing With a Fisheye Lens. In Proceedings of Graphics Interface 2005, Victoria, BC, Canada, 2005.
2. Bolt, R. A. "Put-That-There": Voice and Gesture at the Graphics Interface. In Proceedings of SIGGRAPH'80 Conference on Computer Graphics and Interactive Techniques. ACM Press: New York, NY, 1980, pp. 262-270.
3. Bolt, R. A. Gaze-Orchestrated Dynamic Windows. In Proceedings of the 8th Annual Conference on Com-puter

Graphics and Interactive Techniques. Dallas: ACM Press, 1981, pp. 109-119.
4. Card, S.K., English, W.K., and Burr, B. J. Evaluation of Mouse, Rate-controlled Isometric Joystick, Step Keys and Text Keys for Text Selection on a CRT. Ergonomics Vol. 21 No. 8, 1978, pp. 601-613.
5. Crossman, E. R. F. W. and Goodeve, P. J. Feedback Control of Hand Movement and Fitts' Law. In Quarterly Journal of Experimental Psychology 35A, 1983, pp. 251-278, (original unpublished paper 1963).
6. Duchowski, A. *Eye Tracking Methodology: Theory & Practice*. Berlin: Springer-Verlag, 2003.
7. Fitts, P.M. The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Journal of Experimental Psychology Vol. 47, 1954, pp. 381–391.
8. Fono, D., and Vertegaal, R. EyeWindows: Evaluation of Eye-Controlled Zooming Windows for Focus Selection. In Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems. Portland, OR: ACM Press, 2005.
9. Guiard, Y. Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. Journal of Motor Behavior 19, 4, 1987, pp. 486-517
10. Jacob, R.J.K. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At Is What You Get. ACM Transactions on Information Systems, vol. 9 (no. 3), 1991, pp. 152-169.
11. Kembel, A. Reciprocal Eye Contact As an Interaction Technique. In Extended Abstracts of ACM CHI'03 Conference on Human Factors in Computing Systems. Fort Lauderdale: ACM Press, 2003.
12. Kuno, Y., Miyauchi, D., Nakamura, A. Robotic Method of Taking the Initiative in Eye Contact. In Extended Abstract of ACM CHI 2005 Conference on Human Factors in Computing Systems. Portland, OR: ACM Press, 2005.
13. LC Technologies, Inc. *Eyegaze System Manual*. www.eyegaze.com, 2004.
14. MacKenzie, I.S. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. Human-Computer Interaction Vol. 7, 1992, pp. 91-139.
15. MacKenzie, I.S., Sellen, A., and Buxton, W. A Comparison of Input Devices in Elemental Pointing and Dragging Tasks. In Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems. New Orleans: ACM Press, 1991, pp. 161-166.
16. Maglio, P., Matlock, T., Campbell, C., Zhai, Z., Smith, B. Gaze and Speech in Attentive User Interfaces. In Proceedings of the International Conference on Multimodal Interfaces. Berlin: Springer-Verlag, 2000.
17. Miniotas, D. Application of Fitts' Law to Eye Gaze Interaction. In Extended Abstracts of ACM CHI 2000 Conference on Human Factors in Computing Systems. The Hague, The Netherlands: ACM Press, 2000, pp. 339-340.
18. Morimoto, C. et al. Pupil Detection and Tracking Using Multiple Light Sources. Image and Vision Computing 18, 2000.
19. Oh, A., Fox, H., Van Kleek, M., Adler, A,. Gajos, K., Morency, L-P, and T. Darrell. Evaluating Look-to-Talk: A Gaze-aware Interface in a Collaborative Environment. In Extended Abstracts of CHI 2002 Conference on Human Factors in Computing Systems. Seattle: ACM, 2002, pp. 650-651.
20. Qvarfordt, P., and Zhai, S. Conversing with the User Based on Eye-Gaze Patterns. In Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems. Portland, OR: ACM Press, 2005.
21. Selker, T., Lockerd, A., Martinez J., Eye-R, a Glasses-Mounted Eye Motion Detection Interface, In Extended Abstracts of CHI '01 Conference on Human Factors in Computing Systems. Seattle WA: ACM Press, 2001.
22. Shell, J.S., Selker, T., and Vertegaal, R. Interacting with Groups of Computers. Communications of the ACM Vol. 46 No. 3 (March 2003), ACM Press, New York; pp. 40-46.
23. Shell, J.S., et al. ECSGlasses and EyePliances: Using Attention to Open Sociable Windows of Interaction. In Proceedings of ACM Eye Tracking Research and Applications Symposium 04, San Antonio, TX, 2004.
24. Sibert, L. and Jacob, R.J.K. Evaluation of Eye Gaze Interaction. In Proceedings of ACM CHI 2000. The Hague, The Netherlands: ACM Press, 2000.
25. Smith, B. *Cursor Motion Test Application.* http://www.almaden.ibm.com/cs/people/basmith/testapp.html , 1999.
26. Smith, D., Vertegaal R., Sohn, C. and Cheng, D. ViewPointer: Lightweight Calibration-Free Eye Tracking for Ubiquitous Handsfree Deixis. In Proceedings of ACM Symposium on User Interface Software Technology (UIST) 2005, ACM Press, 2005.
27. Starker, I., and Bolt, R. A Gaze-Responsive Self-Disclosing Display, Proceedings of the ACM CHI'90 Conference on Human Factors in Computing Systems, Seattle, WA: ACM Press, 1990, pp. 3-10.
28. Tobii Technology. *Tobii 1750 Manual.* www.tobii.se, 2004.
29. Vertegaal, R., Slagter, R., Van der Veer, G., and Nijholt, A. Eye Gaze Patterns in Conversations: There is More to Conversational Agents than Meets the Eyes. In Proceedings of CHI'01 Conference on Human Factors in Computing Systems. Seattle, WA: ACM Press, 2001, pp. 301-308.
30. Vertegaal, R., Cheng, D., Sohn, C. and Mamuji, A. Media EyePliances: Using Eye Tracking For Remote Control Focus Selection of Appliances. In Extended Abstract of ACM CHI 2005 Conference on Human Factors in Computing Systems. Portland, OR: ACM Press, 2005.
31. Wacom, Inc. *Intuos 2*. www.wacom.com, 2004.
32. Wang, J., Zhai, S. and Su, H. Chinese Input with Key-board and Eye-Tracking - An Anatomical Study. In Proceedings of ACM CHI'01 Conference on Human Factors in Computing Systems. Seattle, WA: ACM Press, 2001, pp. 349-356.
33. Ware, C., and Mikaelian, H.T. An Evaluation of an Eye Tracker as a Device for Computer Input. In Pro-ceedings of the ACM CHI + GI'87 Human Factors in Computing Systems Conference. Toronto, Canada: ACM Press, 1987, pp. 183-188.
34. Welford, A.T. *Fundamentals of Skill.* London: Methuen, 1968.
35. Yarbus, A.L. Eye Movements During Perception of Complex Objects. In Riggs, L.A. (Ed.), Eye Movements and Vision, Ch. VII. New York, NY USA: Plenum Press, 1967, pp. 171-196.
36. Zhai, S., Morimoto, C., and Ihde, S. Manual and Gaze Input Cascaded (MAGIC) Pointing. In Proceedings of CHI'99 Conference on Human Factors in Computing Systems. Pittsburgh, PA: ACM Press, 1999, pp. 246-253.