Designing for the Eye – Design Parameters for Dwell in Gaze Interaction

Abdul Moiz Penkar, Christof Lutteroth, Gerald Weber
Department of Computer Science
University of Auckland
38 Princes Street, Auckland 1010, New Zealand
{moiz, christof, gerald}@cs.auckland.ac.nz

ABSTRACT

Eye gaze tracking provides a natural and fast method of interacting with computers. Many click alternatives have been proposed so far, each with their own merits and drawbacks. We focus on the most natural selection method, i.e. the dwell, with which a user can select an on-screen object by just gazing at it for a pre-defined dwell time.

We have looked at three design parameters of the dwell click alternative, namely dwell time, button size and placement of content. Two experiments, with similar user interfaces, were designed and conducted with 21 and 15 participants, respectively. Different combinations of dwell times and button sizes were tested in each experiment for each participant. One experiment had content placed on the buttons to be gazed at, while the other had content placed above the buttons.

One important finding is that moving the content outside the clickable areas avoids accidental clicking, i.e. the Midas Touch problem. In such a design, a combination of big buttons and short dwell times are most suited for maximizing accuracy and ease of use, due to a phenomenon identified as the 'gaze-hold' problem.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces—*Evaluation/Methodology*

General Terms

Design, Experimentation, Human Factors

Keywords

Eye gaze tracking, dwell time, Midas Touch

1. INTRODUCTION

Eye gaze tracking is a technology that tracks the point of gaze, i.e. where the user is looking. It is considered to be a promising component of future 'natural user interfaces' mainly because of its inherent ease of use [21] and speed [24, 1] advantages, "potentially revolutionizing the way we use computers" [18]. Gaze also reflects a user's attention and intention [15]. Research has shown that people look at what they are working on [11] and that the eye movements are highly task-dependent as the gaze is mostly directed towards task-relevant objects [4, 13]. The fact that eye movements can be used to infer a user's attention and intentions is extremely useful for HCI practitioners and interface designers. Gaze tracking can be used as a pointing input device as well as to perform actions or selections. Pointing with gaze tracking is relatively straight-forward, but using it for performing actions or selections is challenging.

One of the many ways actions (or clicks) can be encoded with a gaze tracker is by looking at a point of interest (e.g. a button) for a specified amount of time – known as fixating or dwelling. It is a very obvious click alternative but suffers from the problem of excessive inadvertent clicking, i.e. the Midas Touch problem [7]. We are interested in the following questions that arise when using dwell as a click alternative.

Q1. How do dwell times and button sizes affect efficiency and accuracy?

Q2. What is the effect of having content (the label) inside or outside of the clickable areas on efficiency and accuracy?

The motivation for these questions is to better understand the design parameters of dwell and their relationships with each other, and to help designers in making better user interfaces for gaze interaction. A dwell time that is too short results in incorrect selections (Midas Touch problem), while a dwell time that is too long slows down the interaction and undermines the main advantage of using an eye gaze tracker. Different researchers have investigated the minimum size as well as minimum dwell times required for selection [8, 21].

As part of a wider study on how the eye gaze can be utilized to enhance the user experience, this paper reports the initial findings from two similar (albeit independent) experiments. Both experiments use a combination of three dwell times and three button sizes and measure the accuracy and difficulty in specific ways, addressing Q1. The two experiments differ in the way the content that is to be selected is placed: in the first experiment the selectable content is placed on the clickable buttons, whereas in the second experiment it is placed outside (above) the buttons. This allows us to investigate Q2 by comparing the results of both experiments. We make the following contributions:

- Placing the content to be selected (text, in this case) outside the clickable area (the buttons) circumvents the Midas Touch problem.
- A problem, which we call the 'gaze-hold' problem, was found in the case of long dwell times. They resulted in less accurate and more difficult selections.
- If the selectable content is outside the clickable area, large button sizes and short dwell times are best. If the content is inside the clickable area, the best choice of design parameters depends on whether accuracy or ease of use is more important.

Section 2 summarizes other research in this area, starting from generic and ending with some more specific work. Section 3 reports on the experiment where content is placed on the buttons, and Section 4 reports on the experiment with content placed outside the buttons. The paper ends with a conclusion (Section 5) which summarizes the findings and identifies further research questions.

2. RELATED WORK

Eye gaze has been studied for a long time. The first reported study of how the gaze behaves in a reading task dates back to 1878, and was probably the first research to analyse the gaze as a combination of saccades and fixations [9]. In 1967, it was shown that the gaze behaviour or pattern depended on the context and task being performed [25]. The same artwork was shown to different participants. The participants were asked several questions and the gaze pattern was clearly dependent on the question asked.

In order to use eye gaze for computer interaction, a lot of effort has been put into finding efficient and accurate click alternatives [26, 14]. The most straightforward and commonly-used click alternative is considered to be fixating or dwelling on a clickable area. Blinks, winks, glancing and different muscle sensors have also been experimented with, especially for limited-accessibility applications [16, 6]. Another alternative is to use a physical button which the user presses to indicate 'action' at the point being gazed at. This was found to be faster than using a dwell time of 0.4 seconds, but less accurate due to the way users start pressing the button before even looking at the area of interest [24].

As an improvement over the dwell click alternative, another technique, making use of histograms to figure out which object a user is more interested in, has been experimented with in 3D virtual environments. This study concluded that "the eye movement based interaction was faster than pointing, especially for distant objects" [22] in a 3D virtual environment. MAGIC is another technique, demonstrated as an alternative solution, which makes the user rely on explicit commands using other input devices (e.g. mouse) while also benefiting from the speed of pointing with gaze trackers [26]. This approach moves the mouse pointer quickly to the area a user looks at, but relies on the mouse for finer adjustments and clicking. Research has also been conducted on small-target selection with eye gaze using a discrete zoom tool, to cater for the problem that gaze cannot be used reliably on small target areas [20]. The "aging" technique, as another improvement over fixation, has been demonstrated in an AR Gallery application [19].

Some research has also used the gaze data implicitly to indicate the order of interest when multiple objects are visible to a user and how further information about an object of interest can be shown in an augmented display [12]. Such techniques do not use gaze to perform any explicit command or selection. Other interesting alternatives use eye gestures [17] or anti-saccades [5] to indicate a click. These require the users to look in a particular direction (e.g. away from a pop-up button) to click or select it.

But despite the availability and research on many different ways of performing 'actions' using the eye gaze as a pointing mechanism, according to an earlier independent study, people prefer techniques that are natural, requiring minimum deliberate eye movements [8].

As far as using dwell as a click alternative is concerned, research has been done separately either to find the minimum size of such buttons or the ideal dwell time needed, in a gaze-tracked user interface. The different dwell times being used in such experiments, ranging from 0.01 to 3.0 seconds [3] indicate that the optimal dwell time might be dependent on other variables including subject, task, context, cost of errors and user interface layout. Different dwell times (0.01 to 0.1 seconds) have also been proposed based on different types of selections in the same user interface (e.g. by [8]) whereas some researchers have found longer dwell times to work better (e.g. [21]).

As far as the button sizes are concerned, smaller sizes are difficult to point at with the eye [24] while bigger ones take up more screen real estate. An area covering 1 degree of the visual angle is considered to be the minimum required for gaze-based interaction due to the size of the fovea and the gaze jitter, which corresponds to about 1.1 cm at a distance of 65 cm from the display screen. The error of the gaze tracking equipment also affects this adversely. There seems to be a gap in research when it comes to finding a relationship between the button size and dwell time or measuring the effect of placing content outside the clickable areas.

3. EXPERIMENT WITH CONTENT ON THE BUTTONS

3.1 Design

In this section we report on the experiment where content is placed on the buttons. The experiment has two independent variables, "button size", "dwell time". The dependent variables are "accuracy" and "difficulty."

The three button sizes being used for this experiment are: 170x150 pixels (5x4 cm), 120x100 pixels (3.5x2.5 cm), and 70x50 pixels (2x1.3 cm) whereas the three dwell times are 0.2, 0.4 and 1.0 seconds. These two independent variables (size and time) were investigated by a 'within-group' design: each participant performed the tasks using different combinations of these factors.

3.2 Setup

The software applications used in both experiments were developed in Java. The stereo infrared cameras of the gaze tracker were mounted below the 15 inch LCD screen running at a resolution of 1280 by 1024 pixels. Gaze direction was determined by the gaze tracker software using the pupil-centre-corneal-reflection (PCCR) method. The raw gaze coordinates were smoothed by the Java application to cater for gaze jitter and gaze tracker error. A fully adjustable chair with headrest was used to help maintain the participant's

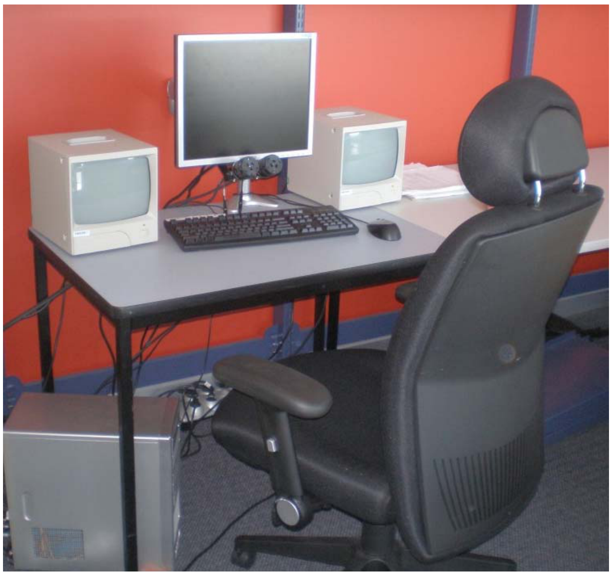

Figure 1: Gaze tracking setup

head position as the maximum freedom of head-movement allowed by this gaze tracker was 2.7 cubic inches. The LCD screen and the gaze tracker cameras were mounted on a movable arm (complete setup shown in Figure 1).

Before the experiment, the participant was comfortably seated and the chair's height and distance from the gaze tracker were adjusted, if needed. The LCD screen and cameras were then positioned more precisely, according to the participants' position. The gaze tracker was calibrated for each user, requiring about 15 seconds, followed by training tasks and the main tasks.

3.3 Task

A simple task of answering multiple-choice questions was chosen. For the experiment, it was not important what the question was about. The questions had the sole purpose of enabling the experimenter to observe the participants in a realistic selection process. All questions were kept very easy to minimize the cognitive processing involved in answering. The questions were trivial ones involving simple maths and choosing rhyming words. For our experiment, it did not matter whether the user actually recognized the correct answer or not. The only important point was that the participant was certain about the answer, right or wrong. The experiment, as will be detailed later, is designed to measure the accuracy between the participant's intended answer and the answer selected by the new click alternative.

The user was presented with one question at a time. Each question was displayed across the top of the window, in a large clear font. The four answers appeared below this, in a horizontal row of rectangular buttons (sample screenshot in Figure 2). We chose a horizontal row of answers instead of a vertical list because we assume that the latter would imply too strongly an order that needs to be followed while evaluating the answer choices. All the buttons were the same size, chosen randomly from the three possible options (shown in Tables 1 and 2). The distance between the buttons was kept the same for all button sizes. Each of the buttons was a rect-

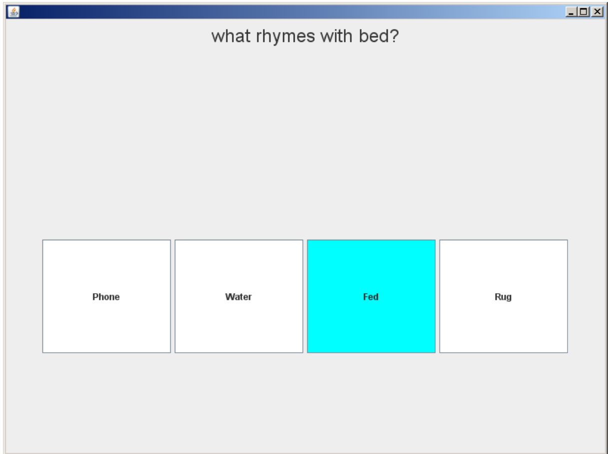

Figure 2: Large button sizes with answer highlighted

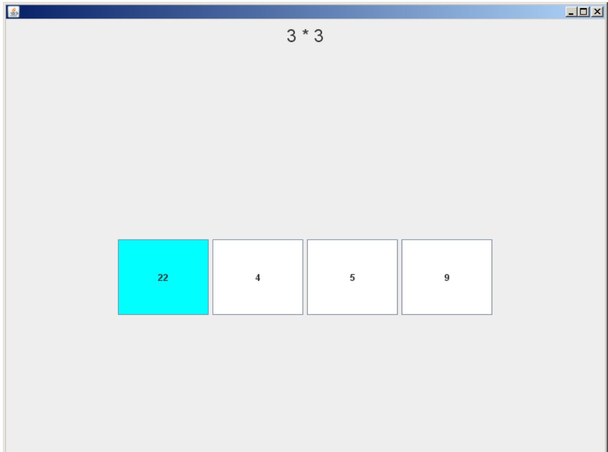

Figure 3: Medium button sizes with incorrect button highlighted

angle, with the width being slightly longer than the height. This is because the jittery eye movements that occur during a fixation happen more in the horizontal direction than the vertical direction [10].

After the calibration, first the training tasks and then the main tasks were performed. Each participant was shown 3 questions during training and 45 questions during the main tasks. The software went through a sequence of states for each question. In the first state only the question was shown, but the answers were hidden until the user hit the space bar. Pressing the space bar showed the possible answers and the user would then look (dwell) at the correct answer to select it. Once the correct answer was selected, the user would press the space bar again to submit the response. The software would then show the next question and so on. The wait for a space bar before showing the answers gave the user time to read and understand the question. It also made the user move their eyes off the answer area between questions. This was important to avoid users selecting the answer to the next question as soon as it had been shown, only because their eyes were still in that position from the previous question.

The answer texts were visible in the centre of the clickable buttons.

An answer button would be highlighted after it was fixated for the specified dwell time, after which the user would press space bar to 'submit' and move on to the next question. The 'submit' step provides us the opportunity to measure the unintentional selections and is therefore used to measure the first dependent variable, accuracy. Due to the ease of the questions it is reasonable to expect that the user would be able to decide on the correct answer right away without much thinking, sometimes even before they would see the answer choices. The answers were very short (a character or a commonly used word) and did not require much reading on the buttons.

3.4 Data Collection

The first dependent variable, accuracy, was measured by the number of selections made while answering a question. In the ideal case, the user would select just one button, the correct answer. However, more buttons could get selected inadvertently before the right one was selected. The accuracy of each combination of dwell and size was measured by averaging the number of selections made while answering a question. An average selection of 1 would imply an accuracy of 100%.

The second dependent variable, difficulty, was measured as the time it took to make a selection. As a measure of difficulty, we recorded the time between when the user pressed space bar to show the answers to the question, and when the user had selected their final answer by looking at it, before submitting it by pressing space bar again. That is, the end of the measured interval is not the hitting of the space bar, but the last selection. It was observed in the pilot as well as the actual experiments that when the users were trying to highlight or select the correct answer, there was a possibility of the gaze exiting the buttons and coming back in, which would reset the dwell timer. Because the dwell time is given and not a consequence of the users' efforts, we subtract the given dwell times from each task completion time to find the 'time to select'. One has to be aware that the time to select contains the 'scan time' that the user needs to scan the answer choices. But since this time should not be dependent on button size or dwell time, and since we make only relative comparisons, the fact that the scan time is included in our measured time amounts to noise and not to a systematic error.

A post-test questionnaire was used to ask some quantitative as well as open questions to collect the participants' views and preferences about this input method.

3.5 Results

Out of the 21 participants who performed this experiment, all were students between the ages of 19 and 27, with one third having commerce and two thirds having computer science as their majors. All of them reported being very confident with using computers, using them several times a day and were also competent with traditional input devices such as mouse and keyboard.

In the questionnaire after the tasks were performed, four participants reported the eye gaze tracker to be more difficult to use compared to the mouse while three reported it to be of similar difficulty. The rest (14) reported the gaze tracker to be either easier or much easier to use than the mouse (as shown in Figure 4). Almost all participants wanted to use it more often and some of them mentioned the benefit of avoiding RSI (Repetitive Strain Injury) while others reported it was futuristic and exciting. When asked if they would like to use the gaze tracker more often, 13 participants either agreed or strongly agreed (as shown in Figure 5). Three neither agreed nor disagreed while five disagreed.

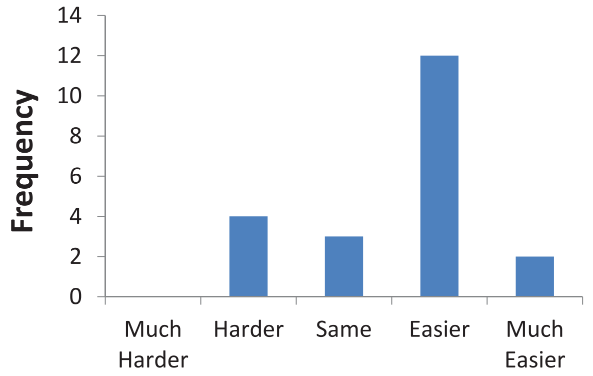

Figure 4: User responses to "How easy is using the eye gaze tracker compared to the mouse?"

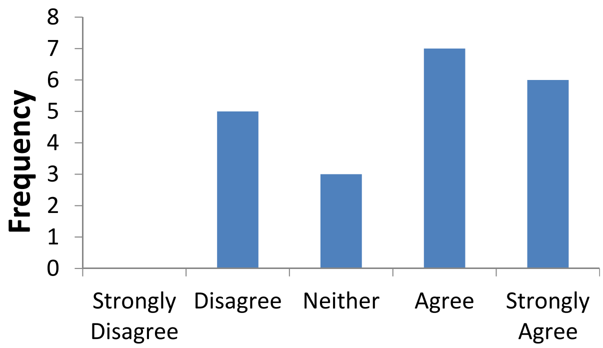

Figure 5: User responses to "I would like to use the gaze tracking technology more often"

Although the gaze tracker was supposed to work with participants wearing glasses or contact lenses, there were a few disappointing experiences. For example, one participant stated, "Not very accurate... easy to make errors, and makes my eyes really dry and tired (I'm wearing contact lenses)."

Table 1 shows the accuracy for each combination of dwell time and button size. As mentioned earlier, each value in this table represents the average number of selections or highlights by the user for each combination of button size and dwell time. The number of observations for each combination of dwell time and button size is 106 in this experiment. It is clear from this table that the combination of the smallest button and longest dwell time is the most accurate

Table 1: **Accuracy: Average number of selections (with standard deviation)**

	Button Size			
Dwell Time	70x50	120x100	170x150	Avg. by Dwell
0.2s	1.59 (0.8)	1.64 (0.7)	1.66 (0.8)	1.63
0.4s	1.30 (0.6)	1.22 (0.5)	1.13 (0.4)	1.21
1.0s	1.01 (0.1)	1.03 (0.2)	1.03 (0.1)	1.02
Avg. by Size	1.30	1.29	1.27	

Table 2: **Difficulty: Average time to select, excluding dwell time, in seconds (with standard deviation)**

	Button Size			
Dwell Time	70x50	120x100	170x150	Avg. by Dwell
0.2s	2.28 (2.5)	1.88 (1.8)	1.99 (2.3)	2.04
0.4s	3.26 (3.5)	1.81 (2.3)	1.72 (1.9)	2.26
1.0s	3.60 (3.5)	2.09 (2.5)	1.68 (1.4)	2.45
Avg. by Size	3.04	1.92	1.79	

one. The number of selections increases if the dwell times are decreased, for all button sizes.

The effect of size appears to be dependent on the dwell times. For the short dwell time of 0.2 seconds, although the average number of selections does increase slightly with size, all sizes appear to be equally inaccurate and the size of the button does not affect the accuracy much. But for a dwell time of 0.4 seconds, the smallest button is most inaccurate and the accuracy increases with the button size. Lastly, in case of the long dwell time of 1.0 second, the accuracy varies only a little, unlike the case of the short dwell time, and remains almost equally and fully accurate for all sizes. In the case of the smallest button and longest dwell time there was only a single instance out of 106 data points, where the number of selections was two.

Overall, the analysis of variance (Table 3) shows that the change in number of selections due to button sizes does not seem to be significant whereas the change due to dwell times is very significant. It also shows that the change in averages due to both the factors (dwell and size) is not statistically significant.

The average time to select (as a measure of difficulty) is shown in Table 2 for each combination as well as aggregated by sizes and dwell times. Note that these values were calculated by subtracting dwell time from the complete answer selection time for each task for comparison purposes. These results show that the worst or most difficult combination is that of a long dwell time and a small button size and the easiest to select is that of a long dwell time and a large button size. In contrast to the 'number of selections' results, the difference in average 'time to select' for each button size is statistically very significant (Table 3). The smaller button is much harder to select, while the larger button is just slightly better than the medium one. Shorter dwell times seem to be better than longer ones but do not seem to differ significantly. The change in averages of all combinations does seem to be statistically significant for the time to select.

According to Table 2, small-sized buttons are easier to select with shorter dwell times. Conversely, large button sizes are easier to select with longer dwell times. For the medium-sized buttons (120x100 pixels), the medium dwell time (0.4 seconds) seems to be the easiest though the short and long dwell times are not much worse.

3.6 Discussion

3.6.1 Accuracy

Longer dwell times seem to be significantly more accurate as they minimize inadvertent clicks. However, the button size does not significantly affect the accuracy. Even the smallest button was bigger than 1 cm, the minimum recommended to accommodate for gaze jitter and gaze tracker error [24]. Additionally, the text inside the button probably provided an anchor for the gaze to hold on to, reducing the chances of the gaze drifting away.

3.6.2 Difficulty

We chose to subtract the dwell time from the overall task completion time. If one would be primarily interested in usability of a certain setting over the course of many interactions, then one could argue that the dwell time should be included since it slows down repeated interactions. But our interest is rather on the inherent difficulty for the user to communicate to the machine the intended command. We are therefore not interested in a time-like property, we are just using a time measure to capture this property.

When we analyse the difficulty, represented by the 'time to select', we see that the large buttons are overall easier to select as compared to smaller ones, and difficulty increases with decreasing sizes. Furthermore, overall selection seems easier with shorter dwell times than with longer ones, though the difference is not statistically significant. This is according to our expectation since the gaze has more chances of going out of the clickable area with smaller button sizes, resetting the dwell timer frequently.

What is very interesting is that although selection is easiest with large button size and long dwell time (an intuitive finding), the worst combination is that of a small button size and a long dwell time. This shows that the long dwell times are not good by themselves: it seems to be difficult to hold the gaze in the clickable area for even just one second, unless the clickable area is big enough. This difference, although

Table 3: Two-factor analysis of variance (p values)

Source of Variation	Number of Selections	Time to Select
Dwell Time	4.65×10^{-35}**	0.128
Button Size	0.822	2.16×10^{-10}**
Interaction	0.329	0.012*

not statistically significant, could be due to jitter, drift or gaze tracker error.

We will relate our results to Fitts' law in the discussion of the next experiment.

3.6.3 Limitations

Our findings are based on a simple task and it is not clear if they can be generalized to more complex tasks and contexts. This is becuase gaze behaviour can be very task and context dependent [25, 4, 13]. This is especially applicable to this experiment with content placed on the buttons. Difficult tasks might require longer inspection of the button content, which may interfere with dwell as a click method. The findings are also based on a particular demographic profile of university students who are proficient with computers.

We incorporated smoothing of the gaze coordinates as a necessary component to take care of jitter and equipment error. The particular smoothing method used might have an effect on the results.

The gaze tracker used in our experiments had a limited freedom of head movement. This resulted in some participants requiring longer times for the seat and screen positioning and the gaze tracker training steps. This disappointed some of the participants before the experiments could even start. For a few participants, the experiment had to be restarted as they mistakenly moved their heads too much and were not able to select any buttons on screen.

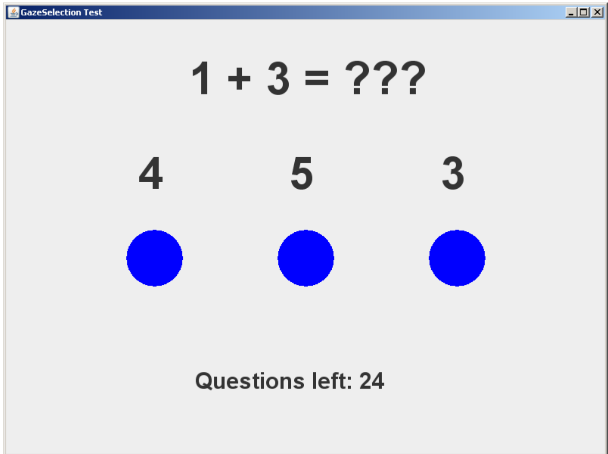

Figure 6: Screenshot with small button sizes

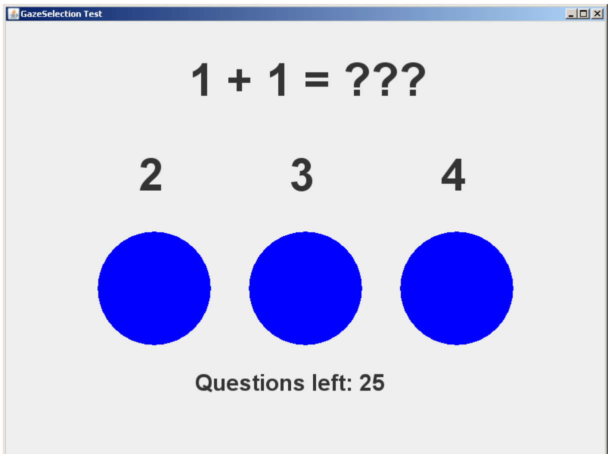

Figure 7: Screenshot with large button sizes

4. EXPERIMENT WITH CONTENT ABOVE THE BUTTONS

4.1 Design

In this experiment, the placement of content was outside (above) the buttons. As in the experiment described in the previous section, the 'button size' and 'dwell time' for selecting a button were the two independent variables. The possible diameter of each clickable button was 75 pixels (4 cm), 100 pixels (5 cm) or 150 pixels (8 cm) in this experiment and the three possible dwell times were 0.2, 0.6 and 1.0 seconds. As in the experiment above, the two independent variables were investigated using a within-group design.

The dependent variables being measured were 'accuracy', 'difficulty' and the 'number of attempts' made while selecting an answer. Accuracy and difficulty are as defined in the first experiment. The number of attempts is defined as the number of times the user has to start the dwell timer in order to select the answer. If the gaze entered the clickable area but went outside before the dwell time had passed, then came back and stayed inside the clickable area for the respective dwell time, it would result in a 'number of attempts' value of 2. As explained earlier, such repeated attempts are necessary if the gaze inadvertently leaves the clickable area.

4.2 Setup and Task

The setup of this experiment was exactly the same as in the other experiment. The task was also similar but had some minor, though important, differences in the way the user interface was designed. In addition to placing content above the buttons, circular buttons were used for this experiment.

Participants had to answer 27 randomly generated simple addition questions, one by one. To keep the questions simple, the two numbers to be added ranged from 1 to 3 only (sample screenshot in Figure 6). Three answers appeared below each question. Once the user had looked at the options and decided on the correct answer, they had to look at the corresponding button, which got selected according to the dwell time chosen randomly for that question. On selection of an answer, a blank screen was shown for 2 seconds to allow participants time to reset their focus and get ready for the next question. This was necessary due to a problem observed in a pilot study where participants unintentionally selected the answer of the following question before having a chance to read it. This problem happened particularly with short dwell times.

4.3 Data Collection

We measured the number of times a participant answered incorrectly (either by coming up with or inadvertently selecting an incorrect answer). The number of ‘incorrect attempts’ was the primary measure of ‘accuracy.’

The task completion time was measured for each question. For comparison between different dwell times, the respective dwell time for each task was subtracted from it to calculate the ‘response time’. This is used as a measure of difficulty in this experiment but cannot be compared to the ‘time to select’ measured in the ‘on-button’ experiment as that did not include the time spent on reading a question.

Lastly, we counted the number of times each button was started to be gazed upon as the ‘number of attempts’. This measure was included only in this experiment because the accuracy turned out to be 100% during the whole experiment, as will be detailed later.

After all tasks had been completed, each participant filled out a questionnaire.

4.4 Results

15 participants performed this experiment, out of which two thirds were male. All of them were students ranging in age from 18 to 23 years, and all of them reported using computers more than three hours every day. For this experiment, there were 45 samples for every combination of dwell time and button size (each cell in the Tables 4 and 5).

As far as ‘incorrect selections’ are concerned, there were none in this experiment, i.e. we encountered 100% accuracy. Not only were no incorrect buttons highlighted, but an incorrect button was not even gazed upon once by any participant.

The average number of attempts and response times are shown in Tables 4 and 5, respectively. There seems to be a gradual increase in both these measures when grouped with the dwell times, indicating increasing difficulty and number of attempts. This trend also seems to be true for all button sizes.

The button sizes also show a slight trend towards decreasing number of attempts with increasing button sizes. The

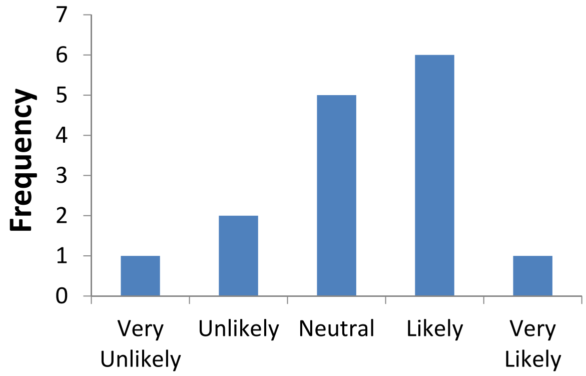

Figure 8: User responses to ”I would use the gaze tracker again”

size does not seem to affect the number of attempts for the short dwell time. For the long dwell time the opposite is true, i.e. number of attempts is bad generally for all sizes. For the medium dwell time, the number of attempts decreases with the button size. The best combination is the large button size and short dwell time, while the worst is the small button size and long dwell time. The analysis of variance (Table 6) does not show any significant effects on the number of attempts, although the effect of dwell time is almost statistically significant at the 5% level.

Selection was generally more difficult with longer dwell times (Table 5). Also, selection became more difficult with decreasing button sizes. The least difficult combinations are the ones with a large button and a short or medium dwell time, whereas the most difficult combinations seem to be the small button with medium or long dwell times. The analysis of variance (Table 6) shows that the effects of dwell time and button size on response time are both statistically significant, but their interaction is not.

In the questionnaire, most participants answered that they would like to use the gaze tracker again for computer interaction (Figure 8). Interestingly, the responses did not show a clear preference for any button size (Figure 9). The comments were also mixed, e.g. ”Not very practical until technology gets better. Need to keep head in a fixed place for it to be effective” and ”Great, when it works.”

4.5 Discussion

4.5.1 *Accuracy*

In this experiment, none of the participants made a wrong selection, meaning we had 100% accuracy. By analyzing the log files, which recorded when the gaze entered a button, we could confirm that the participants never even looked at the buttons for the wrong answers. Since incorrect selections were not even attempted in this experiment, it essentially means that the Midas Touch problem was completely circumvented by having the content separated from the clickable areas. The fact that none of the participants looked at an incorrect button once is a significant finding, which was unexpected. Since users had already performed the training task before starting the actual tasks, they already knew that

Table 4: **Average number of attempts (with standard deviation)**

	Button Size			
Dwell Time	75 px	100 px	150 px	Avg by Dwell
0.2s	1.23 (0.6)	1.23 (0.6)	1.09 (0.3)	1.17
0.6s	1.44 (0.8)	1.33 (0.6)	1.11 (0.3)	1.29
1.0s	1.46 (0.8)	1.29 (0.7)	1.42 (1.1)	1.39
Avg by Size	1.37	1.28	1.20	

Table 5: Difficulty: Average response time in seconds (with standard deviation)

	Button Size			
Dwell Time	75 px	100 px	150 px	Avg by Dwell
0.2s	1.64 (1.3)	1.42 (0.9)	1.34 (0.5)	1.46
0.6s	2.30 (1.7)	1.94 (2.0)	1.32 (0.4)	1.85
1.0s	2.23 (1.6)	1.87 (1.3)	1.81 (1.7)	1.97
Avg by Size	2.05	1.74	1.49	

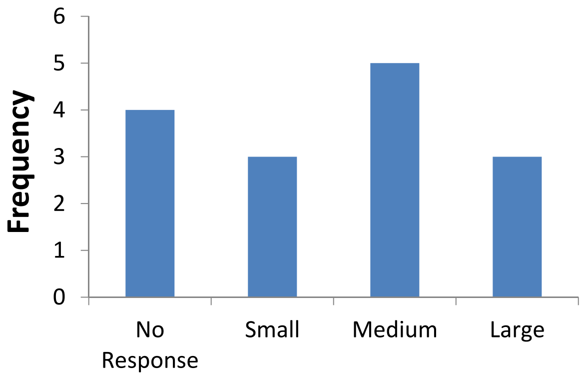

Figure 9: User responses to "Which button size did you like the most?"

the buttons would be present below the questions, and that they would change their size during the experiment.

4.5.2 Number of Attempts

It is evident that the average number of attempts increases with dwell times for all button sizes. We call this the 'gaze-hold' problem. A plausible explanation for this is that users find it difficult to look at a button for a long dwell time, even if they intend to. This phenomenon makes longer dwell times less suitable for a gaze-based interaction, regardless of button sizes, if content is placed outside the buttons.

4.5.3 Difficulty

The effect of dwell time on difficulty is almost statistically significant at the 5% level, probably due to the gaze-hold problem. For the long dwell time, selection takes the longest, even after subtracting the second taken additionally by the dwell. The users clearly seem to have difficulties gazing at a blank button for a long dwell time. One reason might be that since there is nothing inside the buttons to look at, the gaze might be moving to the edges of the buttons more often, making it more difficult to select using a long dwell time. It might be easier for participants to hold their gaze if a visual anchor feature was provided. A short dwell time is the easiest to use, for all button sizes. The effect of the button size is statistically significant, showing that the large buttons are the easiest and the small buttons the most difficult to select. In summary, if the content is outside the buttons, it seems better to have bigger buttons and the easiest combination seems to be a large button and a short or medium dwell time. This is consistent with earlier research which reported longer dwell times as being unnatural [8].

Finally, we relate the findings to Fitts' law. In the experiment with content outside the buttons, the time to select a button grows with decreasing button sizes. This indicates a possible connection with Fitts' law. There is currently a lack of consensus over the validity of Fitts' law for gaze-based interfaces [2]. What seems more relevant is the presence of distinct anchor points in the buttons (e.g. text labels) to look at, and the absence of other visual distractions (e.g. colorful borders). The effect of button size on response time in this experiment is larger than predicted by research based on Fitts' law for gaze tracking [23]. This may be because in our experiments the gaze-hold problem dominates the time to select. Further investigation of the relation between Fitts' law and the gaze-hold problem would be useful.

4.6 Limitations

The limitations due to equipment are the same as the ones mentioned earlier, i.e. the limited freedom of head movement and the use of a smoothing filter. Similarly, due to the sample demographics, the findings might only be applicable to users who use computers on a daily basis. However, the findings of this experiment are not limited to simple tasks since the content (text, in this case) was placed outside the buttons. Our results indicate that more cognitive processing or analysis of the content would not affect the click method in this case (e.g. it would not cause the Midas Touch problem).

The way this experiment was conducted, the total task completion time includes the reading time as well as the

Table 6: Two-factor analysis of variance (p values)

Source of Variation	Number of Attempts	Response Time
Dwell Time	0.056	0.007**
Button Size	0.163	0.003**
Interaction	0.469	0.434

thinking time. Since all questions were similar and very simple, the reading and thinking times are expected to be very small and similar for all questions. However, variations in the reading and thinking times, possibly due to boredom and/or fatigue, may have introduced some noise into our results. A task design that separates the reading and thinking times from the time required for clicking, as used in the first experiment, may yield clearer results (e.g. show that the effect of dwell time on the number of attempts is significant).

The lack of responses to the question asking for button size preference can have several reasons. One would be that participants did not like any of the sizes and would have preferred some size other than the three used. Another would be that participants did not have a clear preference, e.g. because the effect of dwell time was much more obvious to them.

Another limitation is that the effects caused by the different button shapes were not analyzed. They did not seem to affect the accuracy or difficulty, but may still have been a significant confounding factor. Lastly, we have only performed a two-factor analysis of variance and some of the results are close to being statistically significant. Pairwise testing of the differences between the experimental conditions may reveal more structure in the data.

5. CONCLUSION

One major finding is that in a gaze tracked interface the content (the labels) should not be on the buttons if the dwell click method is used. Placing the content outside the buttons seems to reduce the Midas Touch problem significantly. It is plausible that this will become even more important with more complicated content, since detailed evaluation of the content will not cause an accidental click. The combination of a large button size and short dwell time (150 pixels and 0.2 seconds in our experiment) appears to be the easiest as well as the most accurate in this case. Short dwell times seem not as bad as one might expect, and in fact long dwell times seem worse due to the 'gaze-hold' problem.

If, for some reason, content has to be placed on the buttons, then a combination of large buttons and long dwell time (around 1.0 second) should be used. Buttons that are too small suffer from the gaze-hold problem, and short dwell times exacerbate the Midas Touch problem.

Overall, we conclude that dwell is a feasible click method as long as the buttons do not contain the content that should be selected ('off-button' content). There is certainly more analysis possible from the data collected in these experiments, e.g. for the change in accuracy for individual participants over time. Based on the results, using 'off-button' content should be explored further, possibly with a dot in the center of the buttons as a visual anchor point. We expect that such a gaze anchor would increase accuracy and ease of use. Since only three distinct dwell times and sizes were considered in our experiments, there is also room for further refinement of these design parameters.

6. ACKNOWLEDGMENTS

We would like to acknowledge the assistance of Remy Lim, Simon Liu, Christian Sheehan and Hitarth Sharma in designing and performing the user evaluations, and the participation of all the users without whom this study would not have been possible.

References

[1] J. S. Agustin, J. C. Mateo, J. P. Hansen, and A. Villanueva. Evaluation of the potential of gaze input for game interaction. *PsychNology Journal*, 7(2):213–235, 2009.

[2] H. Drewes. Only one fitts' law formula please! In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, CHI EA '10, pages 2813–2822. ACM, 2010.

[3] J. P. Hansen, A. S. Johansen, D. W. Hansen, K. Ito, and S. Mashino. Command without a click: Dwell time typing by mouse and gaze selections. In *INTERACT*, 2003.

[4] M. Hayhoe and D. Ballard. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4):188 – 194, 2005.

[5] A. Huckauf, T. Goettel, M. Heinbockel, and M. Urbina. What you don't look at is what you get: anti-saccades can reduce the midas touch-problem. In *Proceedings of the 2nd symposium on Applied perception in graphics and visualization*, APGV '05, pages 170–170. ACM, 2005.

[6] A. Hyrskykari, H. Istance, and S. Vickers. Gaze gestures or dwell-based interaction? In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 229–232. ACM, 2012.

[7] R. J. K. Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*, CHI '90, pages 11–18. ACM, 1990.

[8] R. J. K. Jacob. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans. Inf. Syst.*, 9:152–169, April 1991.

[9] E. Javal. Essai sur la physiologie de la lecture. *Annales d'ocullistique*, 80:61–73, 1878.

[10] Q. Ji and Z. Zhu. Eye and gaze tracking for interactive graphic display. In *Machine Vision and Applications*, pages 139–148, 2002.

[11] M. A. Just and P. A. Carpenter. Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4):441 – 480, 1976.

[12] M. Kandemir, V.-M. Saarinen, and S. Kaski. Inferring object relevance from gaze in dynamic scenes. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 105–108. ACM, 2010.

[13] M. F. Land. Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, 25(3):296 – 324, 2006.

[14] C. Lankford. Effective eye-gaze input into windows. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, ETRA '00, pages 23–27. ACM, 2000.

[15] S. P. Liversedge and J. M. Findlay. Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1):6 – 14, 2000.

[16] P. Majaranta and K.-J. Räihä. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, ETRA '02, pages 15–22. ACM, 2002.

[17] E. Møllenbach, M. Lillholm, A. Gail, and J. P. Hansen. Single gaze gestures. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 177–180. ACM, 2010.

[18] C. H. Morimoto and M. R. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4 – 24, 2005.

[19] H. M. Park, S. H. Lee, and J. S. Choi. Wearable augmented reality system using gaze interaction. In *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*, pages 175 – 176, sept. 2008.

[20] H. Skovsgaard, J. C. Mateo, J. M. Flach, and J. P. Hansen. Small-target selection with gaze alone. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 145–148. ACM, 2010.

[21] D. M. Stampe and E. M. Reingold. Selection by looking: A novel computer interface and its application to psychological research. *Studies in Visual Information Processing*, 6:467–478, 1995.

[22] V. Tanriverdi and R. J. K. Jacob. Interacting with eye movements in virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '00, pages 265–272. ACM, 2000.

[23] R. Vertegaal. A fitts law comparison of eye tracking and manual input in the selection of visual targets. In *Proceedings of the 10th international conference on Multimodal interfaces*, ICMI '08, pages 241–248. ACM, 2008.

[24] C. Ware and H. H. Mikaelian. An evaluation of an eye tracker as a device for computer input. In *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface*, CHI '87, pages 183–188. ACM, 1987.

[25] A. L. Yarbus. *Eye Movements and Vision*. Plenum Press, 1967.

[26] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, CHI '99, pages 246–253. ACM, 1999.