Completion Time Predictions of Mobile Touch-Screen Interactions in Dual-Task Situations

Martin Schedlbauer
Computer Science
University of Massachusetts
Lowell, MA 01854 (USA)
mschedlb@cs.uml.edu

Abstract. *As the use of touch input in mobile and kiosk devices is becoming more ubiquitous, the validity of human performance models such as Fitts' Law must be investigated to determine whether they still hold in situations where users are performing other tasks while interacting with the device. This paper describes an experiment that explores the effect of split attention on the completion time and accuracy of an interactive touch selection task. It was found that the task completion time, the error rate, and the spatial variability of target selections increased significantly when attention was not solely focused on the interactive task. Consequently, the paper suggests specific guidelines for the design of ubiquitous computing devices used in split attention environments.*

Keywords. *Fitts' law, human performance modelling, completion time predictions, engineering models.*

1. Introduction

The mobile computing work context often differs substantially from that of the traditional office environment. For one, the interaction with the mobile computing device, such as a handheld PC, a personal digital assistant (PDA), or a kiosk system, is not the only task. Instead, the user must split his attention between the interaction and another, perhaps more important task [3]. For example, while checking into a flight using an airline self-service kiosk, a passenger is required to enter a flight confirmation number that he perhaps needs to read from a sheet of paper, requiring a constant back-and-forth between the cognitive task of reading and the motor behavioral task of data input. Another example is input of a GPS coordinate into a navigation computer. While the operator enters the coordinate, he has to continually look up to visually determine if there is danger in the road ahead. Once again, the user of the mobile device has to divide his attention between two tasks leading to a *dual-task situation*.

Most mobile devices rely on touch screens as their primary user interface control mechanism, because indirect input devices, such as a mouse, are not practical. Interactions are principally in the form of selecting, swiping, and dragging. The screens on mobile devices are generally small and user interface design requires controls to be laid out carefully to maximize the limited screen real estate. Consequently, the design of mobile devices must rely on the results of empirical usability studies.

2. Related Work

Cursor positioning models allow interface designers to understand and predict human aiming performance, such as the activation of controls in graphical user interfaces. The most frequently applied model for human aiming is Fitts' Law [8]. It expresses a logarithmic relationship between task completion time (TT) and the ratio of the distance to (A) and the size of (W) the target. More precisely,

$$TT = a + bID \tag{1}$$

and

$$ID = \log_2\left(\frac{A}{W} + 1\right) \tag{2}$$

The constants a and b of Equation (1) are experimentally derived regression coefficients. The logarithmic term of Equation (2) represents the *Index of Difficulty* (ID). It characterizes the complexity of the target selection. The target width, *W*, is the width or extent of the target, as illustrated in Figure 1.

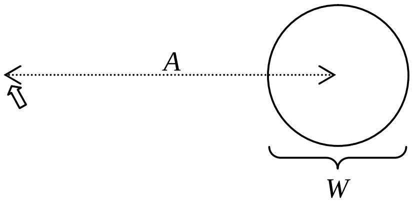

Figure 1. A one dimensional Fitts movement from the cursor along a direct path of length *A* to a circular target of width *W*.

There is research by Hoffmann and Lim [1] that suggests interference between motor and cognitive performance in dual-task situations. Specifically, Hoffmann and Lim point out that when manual and cognitive tasks are done sequentially, the combined task time is the sum of the individual task completion times. However, when both tasks are done concurrently, the total task completion time increases. In particular, Hoffmann and Lim theorize that corrective submovements during target selection are impeded when cognitive tasks such as choosing are made while the movement occurs. Their experiments show that the total time it takes to complete manual and decision tasks concurrently is dependent on the complexity of the decision. That is, movement time increases as the number of choices gets larger. In addition, the interference is amplified by the difficulty of the task. Hoffmann and Lim postulate that cerebral interhemisphere effects are the cause for the interference.

Studies by Shin and Rosenbaum [7] find similar interactions between cognitive and perceptual-motor processes. In particular, their experiments reveal that when cognitive tasks are done in parallel with aiming tasks, the aiming tasks take longer. In their experiments, subjects were asked to hit an on-screen target with a mouse while carrying out arithmetic calculations. Their research concludes that cognitive tasks are typically scheduled to start before perceptual-motor processes, so problems occur when both tasks are interleaved and scheduled to occur concurrently. Interestingly, their evaluation of the empirical data concludes that the overall cognitive task time does not increase when a perceptual-motor task is executed at the same time. This implies that the time to complete the motor task increases.

The main goal of this study is to determine if Fitts' Law holds when attention is split between two tasks, *i.e.*, in a dual-task situation. Presently, Fitts' law has only been tested in single-task situations where the user is entirely focused on the aiming task.

The experiment described in this paper asked participants to glance at a monitor prior to starting a target selection trial, read a number and make a decision based on the value of the number. The cognitive work is done prior to the motor task. Therefore, given the results by Shin and Rosenbaum as well as Hoffmann and Lim, the following *null* hypothesis can be stated: There is no difference in completion time or accuracy of the selection in a dual-task situation, if the two tasks are carried out serially.

3. Methods

3.1 Participants

Eight right-handed participants (7 men and 1 woman) were recruited to complete the experiment. The participants had a mean age of 37 years and a mean height of 172*cm*. All participants had extensive experience using computers and had normal or corrected-to-normal vision and reported no other physical impairments. The participants received compensation in the form of a gift certificate. Only right-handed users were selected because the test apparatus was not reconfigurable for left-handed use.

3.2 Apparatus

The experiment was conducted on a Gateway M275 PC (Windows XP, 1.4GHz CPU, 512MB RAM) with an attached 15” LCD touch screen manufactured by Elo Touch Systems (Model ETL150P-8PWA-1) using surface-acoustic technology. A second 17” CRT displayed a navigation simulator running on a Dell Optimax PC. Figure 2 shows a photograph of the experimental apparatus.

The trials were presented using a research workbench developed by the author [5][1].

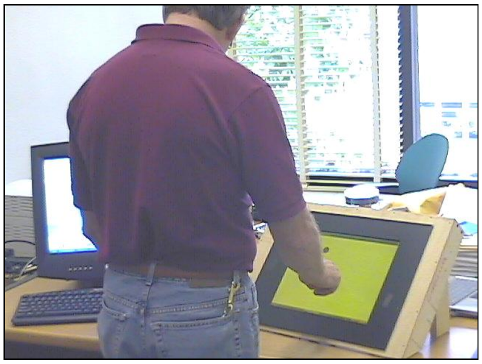

Figure 2. Experimental apparatus for testing the effect of secondary tasks on aiming performance. To the left of the Elo touch screen is a 17” monitor displaying a mock-up navigation screen on which participants have to monitor a depth indicator.

3.3 Experiment Design

Since the experiment measured the effect that the presence of another task might have, the participants were asked to select the targets under two different scenarios: performing the selection task only and performing the selection task while carrying out a secondary motor-cognitive task. Subjects were asked to select, in a standing posture, four differently sized circular targets (6, 12, 20, and 25*mm*) using finger touch. Each subject was presented with 20 repetitions for each target size. During each repetition, the position of the target was varied. While the aiming tasks were carried out, subjects had to monitor a changing depth indicator located on a second screen to the left of the touch screen (see Figure 3). Whenever the depth indicator reached a value below 10.0, subjects had to move a mouse device with their left hand, which reset the depth indicator to a value above 10.0.

The subjects were instructed not to look during the aiming task. Verification of the instructions was done by observation. Since the subjects did not glance at the second screen while they were aiming, the trial completion time does not include glance time. Rather, the glance occurred principally at the end of each movement trial and before the start of the next one.

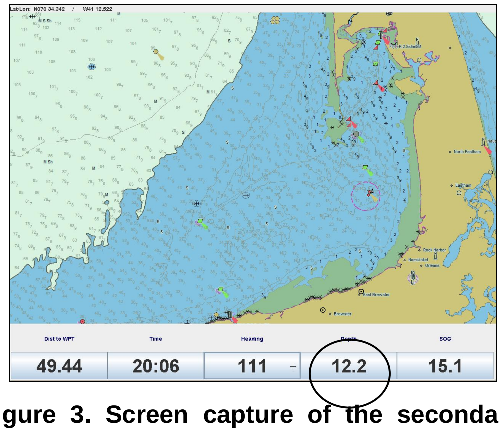

Figure 3. Screen capture of the secondary task simulator. Subjects are instructed to monitor the “Depth” figure (circled above) while selecting targets.

The goal of the experiment was to determine whether carrying out a cognitive task in conjunction with a selection task has an effect on the completion time of the selection task. Both experiments were conducted in a standing posture since that posture is most common with mobile and kiosk devices. It has been shown previously that accuracy of selection decreases as the posture is changed from sitting to standing, but that the overall task completion time remains unaffected [6].

Because the targets were positioned at various angles, a circular target shape was used in all experiments. This approach avoids a potential confounding effect with a varying target width depending on the approach angle. A circular target presents an equal width from all approach angles.

In keeping with the recommendations of Soukoreff and MacKenzie [8], the experiment varied the selection difficulty by testing a broad range of *ID* values (*min* = 0.28, *max* = 4.64, *mean* = 2.70), *i.e.*, targets placed at different distances and having different widths.

3.4 Procedure

Before testing, participants were instructed to hit the targets as quickly as possible while minimizing errors. Any tap outside the target area was recorded as an error. The participants were allowed to rest before each block of 20 trials. An additional rest period was provided between task type changes. A target acquisition trial consisted of clicking a home region at the center of the screen which started the timing and caused the home region to be hidden. This was followed by

---

[1] The experiment software is available at `http://www.cs.uml.edu/~mschedlb/mte`.

clicking on the target. Auditory feedback confirmed successful acquisition of the target.

All experiments were designed as repeated-measures designs on the same population. This design makes the significance of statistical tests more accurate, but carries with it the possibility of practice and fatigue effects. The practice effect was ameliorated through the use of warm-up trials and practice runs, while the possibility of fatigue was reduced through frequent breaks. The order of the trials was randomly varied.

3.5 Data Analysis

Time measurements were taken at a resolution of 10*ms*. Amplitudes were calculated using the Pythagorean distance between the starting point and the end point of the movement. The recorded movement time was not adjusted to remove reaction time so that the measured time more accurately reflects the total interaction time [9].

The collected data contained a few outliers some of which were removed from the data set. For each trial, the distance between the starting and end point was measured and, along with the target width, was inserted into Equation (2) to calculate an Index of Difficulty (*ID*) for the task. The correlation coefficient (R^2) was calculated by linearly correlating the average task completion time values (*TT*) over 20 *ID* ranges.

There is considerable debate over whether to use the raw data values in the correlation calculations or averaged *TT* values over fixed *ID* ranges [5]. From a statistical perspective, the use of the raw data shows the shape of the data more clearly and the correlation results are more meaningful. However, a few far outliers can affect correlation results markedly. Using averaged values attenuates the effect of outliers by bringing them closer to the mean, but this may have the unintended consequence of hiding the effect of certain factors. For instance, for finger touch, the smallest target size had a much higher selection time. When using averaged *TT* values, this effect may be hidden. Therefore, threshold values that markedly affect performance may not be detectable. On the other hand, most published studies on Fitts' law report correlations based average *TT* over a fixed range of *ID* values, so the publication of the correlations obtained from the averaged data allows more meaningful comparisons with prior results.

Soukoreff and MacKenzie [8] state that obvious outliers should be removed from the calculation of *ID*, which they define as being farther than three standard deviations from the mean. They attribute the presence of outliers to 'misfires' where a subject accidentally double-clicks on a target or pauses during the movement. The outliers observed in this experiment do not fall into one of these two categories. Rather, they appear to be caused by the imprecision of touch input for small targets or poor touch screen calibration. Touch screens report a single (x, y) position to the testing software even though the probe (finger or stylus) covers much more than a single pixel on the screen. The reported position is most commonly an average of the covered pixels. Therefore, targets that are smaller than the probe often require repeated attempts before a successful selection occurs. Consequently, the high trial completion time measured by *TT* captures the actual difficulty of the task and outliers generally represent selections of small targets.

Due to uneven calibration of the touch screen, 18 targets were difficult to select and had to be excluded, leaving 622 out of 640 scores (97.2%) for analysis.

4. Results

The collected task completion time and error rate data were not found to be normally distributed (Shapiro-Wilk normality test, $p < 0.05$). As a log_{10} transform of the scores failed to normalize the distribution, the nonparametric Wilcoxon test was used to determine statistical significance at an alpha of 0.05.

4.1 Performance

As shown in Figure 4 and confirmed by a Wilcoxon rank sum test ($W = 113916$, $p < 0.0001$), the task completion time increased significantly from 703*ms* ($sd = 670$) to 1058*ms* ($sd = 893$) in the presence of a secondary task. Pairwise Wilcoxon tests with Bonferroni corrections reveal that the differences in task completion times were significant across all target sizes. Table 1 shows the mean task completion times by task type for each target size.

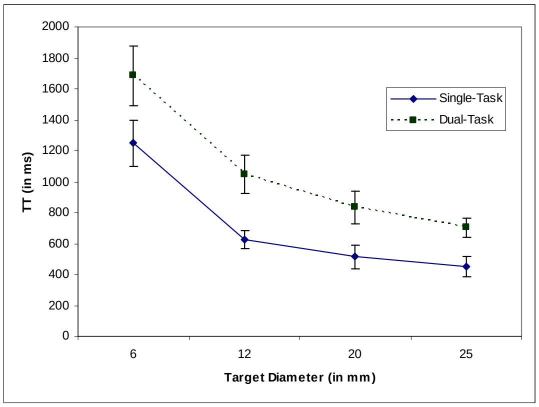

Figure 4. Mean task completion times (*TT*) for each task type by target diameter. Shown with 95% confidence interval.

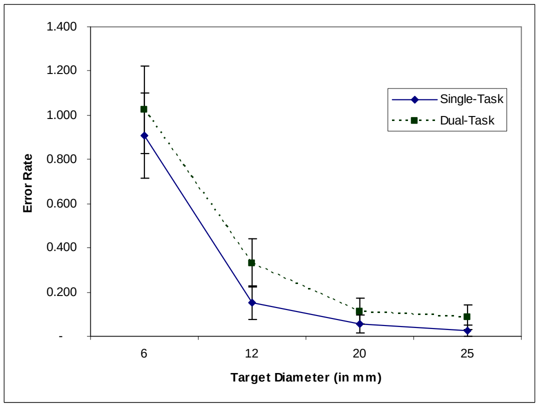

Figure 5. Mean error rates for each task type by target diameter. Shown with 95% confidence interval.

Table 1. Task completion time by task type and target size plus the statistical significance of the differences.

	Single-Task		Dual-Task		
Target Size	*mean*	*sd*	*mean*	*sd*	*p*
6	*1251*	*929*	*1684*	*1213*	***
12	*624*	*377*	*1047*	*777*	***
20	*514*	*490*	*833*	*676*	***
25	*451*	*418*	*702*	*382*	***
all	*703*	*670*	*1058*	*893*	***

*** $p < 0.0001$

Table 2. Error rate by task type and target size plus the statistical significance of the differences.

	Single-Task		Dual-Task		
Target Size	*mean*	*sd*	*mean*	*sd*	*p*
6	*0.907*	*1.196*	*1.026*	*1.233*	‡
12	*0.152*	*0.486*	*0.331*	*0.671*	**
20	*0.056*	*0.257*	*0.113*	*0.372*	**
25	*0.025*	*0.157*	*0.088*	*0.361*	**
all	*0.278*	*0.745*	*0.381*	*0.829*	**

** $p < 0.001$, ‡ $p > 0.05$

4.2 Accuracy

The accuracy of an input method can be captured by two measures: error rate and spatial dispersion. The error rate is the average number of selections outside the target area for some number of trials, while spatial dispersion is the standard deviation of the distances of the selection end points from the mean. The greater the distances between the selection end points, the less accurate the selection.

The error rate, as illustrated in Figure 5, changes significantly from 27.8% in a single-task situation to 38.1% in a dual-task situation [$W = 179908$, $p = 0.00469$], an increase of approximately 37%. As shown in Table 2, the differences are significant for all but the smallest target size. For the 6*mm* target diameter, the error rate is approximately 100% for both task types, which means that it takes on average two attempts to select a target correctly.

On the other hand, a one-way ANOVA with task-type as the sole factor does not reveal a significant increase in the spatial dispersion (deviation from the mean) of the valid selection end points in the dual task situation ($F_{276} = 0.147$, $p = 0.706$, $power = 0.75$). Figure 6 provides a visual illustration of the selection end points for a small subset of the target selections.

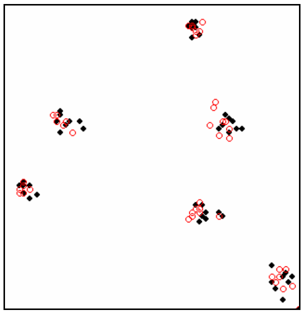

Figure 6. Spatial dispersion (variability) of the selection end points by task type (open circles = single task, filled circles = dual task).

4.3 Performance Models

Linear correlation of the averaged values for task completion time against the 20 tested *ID* values results in $R^2 = .83$ ($p < 0.001$) for the single-task data and $R^2 = .61$ ($p < 0.001$) for the dual-task data. Regression analysis reveals the following Fitts performance models for task completion time (*TT*):

$$TT_{Single-Task} = -38 + 238 \log_2 \left(A/W + 1 \right) \qquad (3)$$

$$TT_{Dual-Task} = 315 + 237 \log_2 \left(A/W + 1 \right) \qquad (4)$$

where *A* is the distance to the target and *W* is the width along the approach vector to the target. Figures 7 and 8 show scatter plots of the averaged data values with the regression lines superimposed.

The negative intercept for finger touch is unusual because the most common explanation for the value intercept is that it captures the non-informational factors in the pointing process, such as decision time or button click time [9]. On the other hand, Hoffmann and Sheikh [2] also report negative intercept values for finger touch. The most likely explanation is that the high *TT* values for the smallest target widths force the intercept to be negative during regression.

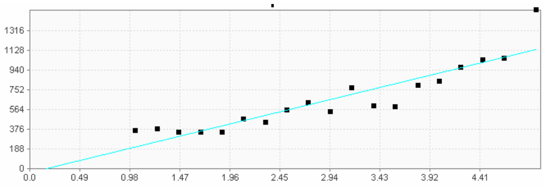

Figure 7. Scatter plot and regression line for the averaged *TT* values (y-axis) over 20 *ID* values (x-axis) for single-task selection.

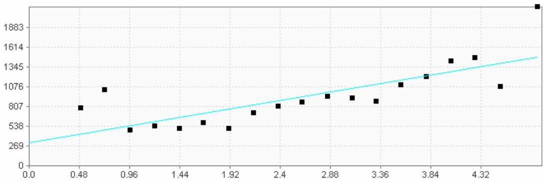

Figure 8. Scatter plot and regression line for the averaged *TT* values (y-axis) over 20 *ID* values (x-axis) for dual-task selection.

5. Discussion

The selection of the smallest target (6*mm*) required on average two attempts. Therefore, the mean task completion time is substantially higher for that target size. The average error rate across all target sizes is 38.1%, which is substantially higher than the average error rate of about 4% reported for mouse input [4]. In the single-task scenario, an error rate of 5.6% is reached once the target size increases to 20*mm*, while for dual-task scenario the error rate for 20*mm* is still at 11.3% and only decreases to a more tolerable rate of 8.8% for a target size of 25*mm*.

It is interesting to note that the intercepts of the two regression equations (3) and (4) differ substantially. An increase in that element suggests that the glance to the second monitor increases the planning of the movement. Even though subjects did not start the task (and the timing of the trial) until they looked back to the experiment screen, they still required additional time to refocus on the screen. The average increase in task completion time between single and dual-task types was 355*ms*. The intercepts of the two regression equations differ by 353*ms*, suggesting that the additional time taken to complete the task can be attributed primarily to movement planning time.

More importantly, the slopes of the two regression equations are essentially the same. This implies that as the complexity of the task increases, *i.e.*, the *ID* gets larger, the increase in the task completion time is the same for both task types. In other words, the presence of a secondary task has a linearly additive effect on Fitts' law.

Based on the decreased accuracy and increased completion time, the initial hypothesis must be rejected. Furthermore, the results suggest that performance is negatively affected in dual-task situations even when the tasks are not carried out simultaneously. The observations of this study point to the importance of testing human performance models, such as Fitts' law, in the actual work context.

6. Conclusions

While Fitts' law holds reasonably well in situations where attention is split between two tasks, an overall increase in the non-informational element – the intercept – of the regression equation is observed. Besides the increased task completion time, the accuracy of the selections lessens as well, particularly for small targets. This implies that designers of kiosk, in-vehicle navigation, and mobile systems must be prepared for an increased error rate and must make destructive actions not easily selectable. Extrapolating from the analysis of error rate versus target size, a minimum user interface control size of 25*mm* in diameter is recommended for devices that are used in dual-task situations. In contrast, if the user can focus on the interaction task alone, a smaller minimum target size of 20*mm* is acceptable.

7. Acknowledgements

The author wishes to thank Dr. Jesse Heines for his helpful comments on early drafts of the paper. Also, the contributions of the many volunteers participating in the experiment are gratefully acknowledged.

References

[1] Hoffmann, E., & Lim, J. (1997). Concurrent manual-decision tasks. *Ergonomics*, 40(3), 293-318.

[2] Hoffmann, E. R., & Sheikh, I. (1991). Finger width corrections in Fitts' Law: Implications for speed-accuracy research. *Journal of Motor Behavior*, 23(4), 259-262.

[3] Kristoffersen, S., & Ljungberg, F. (1999). Making place to make IT work: Empirical explorations of HCI for mobile CSCW. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work*, Phoenix, AZ, 276-285.

[4] MacKenzie, S. (1995). Movement time predictions in human-computer interfaces. In *Readings in Human-Computer Interaction*, 2nd Edition, 483-493, Los Altos, CA: Morgan Kaufman.

[5] Schedlbauer, M., Pastel, R., & Heines, J. (2005). An extensible and interactive research platform for exploring Fitts' law. Technical Report 2005-014, Department of Computer Science, University of Massachusetts Lowell.

[6] Schedlbauer, M., Pastel, R., & Heines, J. (2006). Effects of posture on target acquisition with a trackball and touch screen. In *Proceedings of Information Technology Interfaces ITI 2006*, June 2006, Dubrovnik, Croatia, 257-262.

[7] Shin, J., & Rosenbaum, D. (2002). Reaching while calculating: Scheduling of cognitive and perceptual-motor processes. *Journal of Experimental Psychology*, 131(2), 206-219.

[8] Soukoreff, W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. International *Journal of Human-Computer Studies*, 61(2004), 751-789.

[9] Zhai, S. (2004). Characterizing computer input with Fitts' law parameters – the information and non-information aspects of pointing. *International Journal of Human-Computer Studies*, 61(6), 791-809.