Investigating Selection and Reading Performance on a Mobile Phone while Walking

Bastian Schildbach[1], Enrico Rukzio[2]
[1]Media Informatics Group, University of Munich, Germany, schildbach@cip.ifi.lmu.de
[2]Computing Department, Lancaster University, UK, rukzio@comp.lancs.ac.uk

ABSTRACT
More and more people interact with their mobile phone while walking. The presented research analyzes; firstly, the negative effect of walking when considering reading and target selection tasks, such as weaker performance and higher workload. Here, we focused on one-handed interaction with a touch screen whereby the thumb is used as the input device. Secondly, we analyze how these negative effects can be compensated by increasing the text size and the size of the targets to select on the mobile phone. A comparative user study was conducted with 16 participants who performed target acquisition and reading tasks while standing and walking. The results show that whilst performance decreases, cognitive load increases significantly when reading and selecting targets when walking. Furthermore, the results show that the negative effect regarding target selection can be compensated by increasing the target size, but the text reading task did not yield better performance results for a larger text size due to the increased demand for scrolling. These results can be used to inform future designs of mobile user interfaces which might provide a dedicated walking mode.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Input devices and strategies; Prototyping.*

General Terms
Human Factors, Performance, Design, Experimentation.

Keywords
Mobile interaction, walking, target selection, reading.

1. INTRODUCTION
More and more people use their mobile phone while walking in order to call somebody, to browse the web, to read texts or for social networking. This leads to the problem that pedestrians are distracted from the main walking and navigation task which increases the likelihood of injury from stumbling more often and run into others or objects [18]. As this is an increasing problem, it should be analyzed how mobile user interfaces could be improved or adapted so that they reduce the time the user is engaged with them and the cognitive load while walking. A mobile user interface that adapts itself when the user is walking could also be considered as an important feature in the future, leading to a unique selling point for the first handsets supporting such a function effectively.

MobileHCI'10, September 7–10, 2010, Lisbon, Portugal.
ACM 978-1-60558-835-3/10/09.

Mobile user interfaces are mainly designed, developed and tested with having a user in mind who is sitting or standing, but not walking, running or travelling as a passenger. When on the move, the whole body moves, in particular, the user's hand as well as their head is constantly moving. This makes it more difficult to read a given text or to interact with the mobile phone via a touch screen or keypad.

We conducted a comparative user study in order to analyze the effects of walking on text reading and target selection. 16 participants took part and walked on predefined routes while reading text and selecting targets on a mobile phone with their finger. We varied the size of the text and targets from standard sizes to 120% and 140% as we assume that such an increase would compensate the negative effects of walking. Furthermore, we were interested how extensively the size of text and targets should be increased when considering the tradeoff between their size and available screen space.

The target selection task, in which a user had to select rectangles of different sizes with their finger on a touch screen, showed an 31% increase in the time needed to select a target, an 23% increase in error rate and a 24% increase in task load between a standing and walking condition. The results also show that increasing the target size by 20% leads to significant improvements and with an increase of 40% can the negative effects of walking be almost completely compensated.

Similar effects were observed when the users read text while walking. Here, the reading speed in words per minute decreased by 19% and the task load increased by 16% compared to a standing condition. Interestingly, an increase of the text size by 20% and 40% did not lead to any actual improvements as the advantages of the large text size were completely compensated by the increased demand for scrolling.

The results presented in this paper can be exploited by user interface designers and developers in order to develop a walking mode in which the button size should be increased by a range between 20% and 40%. In addition, the text size should not be increased as this leads to a corresponding increase in the user's demand for scrolling.

2. RELATED WORK
This section discusses related work towards the evaluation of mobile interfaces while walking, research on target acquisition on mobile devices, reading using small displays and cognitive psychology of reading.

2.1.1 Use-in-motion Evaluation of Mobile Devices
Barnard et al. [2] conducted an empirical comparison of use-in-motion scenarios for mobile computing devices. They pointed out

the importance of choosing the right evaluation method which suits the chosen mobile computing scenario. They show that a controlled walking scenario as well as a treadmill task may be appropriate depending on the study objectives. They argue that the usage of a treadmill is suitable to analyze performance aspects, whereas it seems to be not sensitive enough for finding contextual factors. This is why, in order to achieve more realistic measures in terms of performance, accuracy and workload, a walking scenario should be used.

In [9], Kjeldsov and Stage mention that it is difficult to perform controlled studies (capture key situations, applying established evaluation techniques, data collection) in a real world-setting. Their work examines laboratory evaluation techniques for mobile devices in comparison to a real world-setting (walking in a pedestrian street). Results show that laboratory studies provide a good approximation for the user performance, but more realistic settings should be used to analyze the workload.

2.1.2 Target Acquisition

The following papers address questions regarding the optimal target size on a touch screen in general and in particular when walking. All of them have the basic assumptions that the larger the targets are the shorter the target selection time is due to the general principles formulated in Fitts's Law.

The size of soft buttons of e.g. 6.74 x 6.74 mm (as on the iPhone) is a tradeoff between the goal of achieving relatively accurate selections and the desire to have sufficient space for other user interface elements. For almost perfect accuracy, those soft buttons would require a size of more than 20 x 20 mm which would not allow the design of optimal user interfaces on small devices [6].

Lin et al. [11] examined stylus-based tapping behavior while walking in order to find out about the effects of walking on performance. Furthermore, they compared two different ways of simulating realistic walking and let their participants walk on a treadmill and on an obstacle course. Their results show that the usage of the obstacle course leads to more realistic results as the user has to perform the navigation and walking task in parallel to the target acquisition task. When analyzing the data, they saw that walking had no effect on the time needed to select a target, but they saw an increase towards overall task completion time, error rates and workload. Furthermore, they showed that participants reduced their walking speed in order to select the targets accurately. In addition, they analyzed the effect of different target size and their results show that error rate and selection time decrease the larger the target is. A key difference to our work, lays in the fact that in their study a stylus and therefore two handed interaction was analyzed. In our study, we considered one-handed interaction in which the user uses their thumb to select a target as this is currently the most commonly used technique with touch screen phones.

Mizobuchi et al. [13] analyzed the effects of walking and keyboard sizes when performing text input. A stylus was used to interact with a virtual keyboard displayed by a PDA. Their results show that the walking condition had no significant effect on text input speed, but lead to a higher error rate. Furthermore, they showed that the text input speed increases and the error rate decrease the larger the keys are. The difference to our research was again that they used two-handed input using a mobile device and a stylus. Moreover, no effect on text input speed was shown which stays in contrast to our findings.

Kane et al. [8] examined the effect of walking and the adaption of the user interface on the performance when using two hands and the thumb to interact with soft-buttons on a mobile. The first reported experiment showed a decrease in error rate when increasing the target size while the other results are not conclusive (as discussed by the authors) due to the low number of participants and the high variance in the study results. Their second experiment focused on a user interface that automatically increases the target size once the user is walking. Their study did not provide any evidence for the advantages of this approach when compared to a non-adaptive user interface with large targets. The authors argue that this was due to some issues with the prototype design (e.g. some buttons in the adaptive interface were just too small) and due to the tradeoff between screen size and button size.

Parhi et al. [16] performed a user study comparing different target sizes for thumb-based one-handed interaction with a touch screen while standing. Their results show that task times and error rates decreased with larger key sizes. The authors argue that a target size of 9.2 to 9.6 mm is an optimal tradeoff between target selection time and error rate. Unfortunately, those rather large buttons are relatively big when compared with current screen sizes of mobile devices. Consequently, the iPhone, for example, has a minimal target size of 6.74 mm leading according to Parhi et al. [16] to an error rate of circa 8-12% and according to our research even to 23% (see Figure 7).

2.1.3 Reading

The standard text height of 2.20 mm found in many mobile phones is the result of looking for a minimal text size considering the distance between mobile phone and the eyes and the resolution of the eyes and the mobile phone screen.

Darroch et al. [3] analyzed the effect of different font sizes on the reading performance while standing when using the small screen a mobile device provides. Their results show that reading performance did not increase above 6 point (2.12 mm), but the participants actually preferred fonts in the range of 8-12 point (2.82 – 4.23 mm). Barnard et al. [2] showed that a larger font size of 12 point (4.23 mm) when compared with 10 point (3.53 mm) leads to a greater subjective readability, lower levels of perceived difficulty in reading and was also more preferred by the users.

Mustonen et al. [14] analyzed different methods for studying legibility of text displayed on a mobile phone while walking. The main results are that an increase in walking speed leads to a deterioration of visual performance, and therefore, to a lower reading performance. In addition, they compared two different settings in order to study reading performance while walking. In the first one, the participants searched for a number of target characters in a piece of pseudo text and in the second one a realistic passages of text was used which was read by the participants. Their results show that reading a realistic passages of text is a more useful measure of legibility due to the increased external validity as the participants used rather unrealistic approaches to search for and count the target characters when using the pseudo text.

Vadas et al. [20] compared the performance of reading passages of text on a mobile device and listening to passages of text via a synthesized speech audio display while on the move. Their task contained the reading of passages of text while standing and while walking. Participants were asked to read through different passages and to answer two multiple choice questions about each text. Results suggest that the mobility condition had a significant negative effect on the questions being answered correctly. The overall workload also shows that the user feels more stressed in the walking condition when compared to the standing condition. There were also significant differences between the walking speeds with the natural walking speed being significantly faster than the walking speed when reading passages of texts.

Another proposition to compensate the negative effects of walking was introduced by Rahmati et al. [17]. The idea was to shift the user interface contents according to the movement of the device by using a physically inspired model of springs and dampers providing through this an anti-shake feature. Their prototype did not perform very well when held by the test persons themselves because users compensated the expectable bumps subconsciously by themselves. On the other hand, the prototype performed well in conditions where the mobile was mounted in a car or held by another person.

2.1.4 Cognitive Psychology of Reading

Cognitive psychologists have conducted a great deal of research in the area of visual processing in order to understand the process of reading. The visual system is able to compensate "for the movement of the image on the retina" caused by the movement of the head or the object [4].

When considering legibility in a static posture from paper, there is no change in performance between font sizes of 9pt (3.18 mm) and 12pt (4.23 mm) as well as for line lengths between 2.3 inch and 5.2 inch. This important guideline shows that the length of one text line also has an effect on the legibility of texts as well as on the number of words, which can be displayed within one line [19].

2.1.5 Discussion of Related Work

The related work suggests that evaluations of the influence of walking on the usage of mobile services should be conducted in rather realistic settings and that treadmill approaches should be avoided. Previous target acquisition research focused rather on two-handed input scenarios and the usage of a stylus. Furthermore, previous research reports that walking has no influence on target selection time and text input speed; this is in contrast to our findings. Although research towards the benefits of larger text size for reading speed exists, no research is reported on the effects of increasing text size in order to compensate the negative effects caused by walking. This paper is the first one showing the negative effects of walking on target acquisition and reading in thumb-based usage scenarios, and how this can or cannot be compensated by larger targets or text size.

3. FIELD EXPERIMENT

The goal of the user study was to understand the effects of walking on reading and target selection performance and secondly to analyze the effects of increasing text and target size while walking. The experiment was conducted on an outdoor test track built-up in front of – anonymized for blind review –.

3.1 Participants

16 participants, 8 females and 8 males, took part in our user study and received an incentive for their participation. They were recruited from campus of – anonymized for blind review – University. Their age ranged from 19 to 38 years with a mean of 24 years. None of the participants had visual impairments and all of the participants owned a mobile phone whereas 4 of them used touch screen technology. There were 15 right-handed participants and 1 left-handed participant.

3.2 Experimental Design

The experiment was divided into two phases consisting of four different settings each. A within-subjects design with repeated measures was used in both phases of the user study. The order of the test cases was counterbalanced over the participants using Latin squares, which also avoided first order carryover effects [22]. We choose to use a control condition for each of the two different tasks (reading and selecting) in order to compare our results of the walking conditions to a standing condition.

3.2.1 Target Acquisition Task

The independent variable, **target size**, contained three levels. The standard target size of 6.74 x 6.74 mm was based on the Apple iPhone Human Interface Guidelines [1]. The three target sizes being compared were 6.74 x 6.74 mm, 8.18 x 8.18mm and 9.50 x 9.50 mm (see Figure 1). The two larger target sizes are 20% respective 40% larger than the standard one as it was one aim of the study to show whether the user would benefit from such increases. The three target sizes will be referred to as "small" (6.74 x 6.74 mm), "medium" (8.18 x 8.18mm) and "large" (9.50 x 9.50 mm).

Figure 1. Relationship between the three different target sizes.

3.2.2 Reading Task

The independent variable, **text size**, contained three levels. The standard text size being used was based on an analysis of text sizes being used for displaying SMS messages, web pages and menus in the three different mobile phones: Apple iPhone, HTC G1 and Nokia 5800. The default standard text size was in average 2.20 mm based on the height of the upper case letter "H". Two additional text sizes with a height of 2.64mm and 3.08mm were selected in order to test text sizes which are 20% and 40% larger than the standard (see Figure 2). The three text sizes will be referred to as "small" (2.20 mm), "medium" (2.64mm) and "large" (3.08mm).

Figure 2. Relationship between the three different text sizes.

3.3 Materials

A Nokia 5800 with a resistive touch-screen was used during the user study. For the two different tasks, two different prototypes were implemented using Java ME and the Nokia Series 60 SDK for Java in order to get information about the location the user touched on the screen (see Figure 3). The latter SDK includes libraries for the standard Java ME class *GameCanvas* and provides a hardware specific implementation of the methods *pointerPressed()*, *pointerReleased()* and *pointerDragged()*.

Figure 3. Relationship between target sizes (left, showing the three different target sizes), text sizes (right, showing the three different text sizes) and the mobile phone used in the study.

3.3.1 Target Acquisition Task

For the target acquisition task, the prototype was implemented in accordance to the multi-directional pointing task described in ISO 9241-9 [7]. If the user successfully selected a target, visual (target color changed from black to orange) and tactile feedback (the mobile vibrated) was provided, as literature suggests that a combination of different feedbacks can improve the response times [4].

3.3.2 Reading Task

For the prototype for the reading task we used the Lightweight User Interface Toolkit (LWUIT) for Java ME because it allows an easy way to switch between different text sizes and offers built-in text scrolling support.

For the passages of text, eight standardized reading comprehension texts of about 200 words each were used. These passages of text and reading comprehension questions were taken out of [10], which is a reading skills training book. Each participant was asked to answer two multiple-choice questions on a sheet of paper about the passage they just read. The passages of text were displayed to the participants in a randomized order.

3.4 Test Track

In order to realistically test reading and target selection while walking, we designed two different track layouts (see Figure 4) of 35m and 29.5m length for the two phases of the user study. The idea of conducting the test in a real world setting (e.g. inner city centre) was rejected as our pre-studies had shown the likeness of participants walking against lamp posts, columns and pedestrians. The area covered by the two track layouts measured about 9 m x 7.2 m.

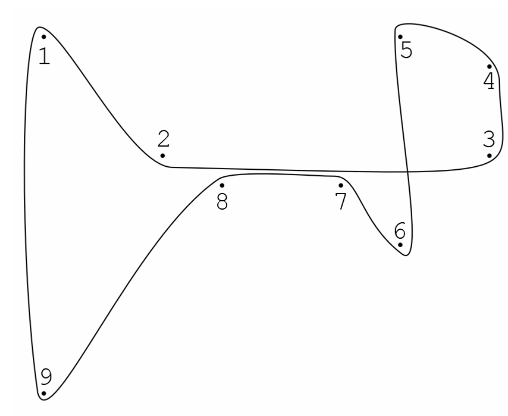

Figure 4. Example of a test track layout.

The track was built up by poles having numbers and arrows attached on the top part indicating the order and direction the participants had to walk around the track (see Figure 5). This setup was designed to simulate the navigation of mobile phone users in an urban environment in order to produce controlled yet realistic results to best effect. Users had to orientate themselves within the track in order to not walk into the wrong direction. Participants were asked to perform the reading and target selection tasks while walking on one of the two tracks and were motivated to walk with their normal walking speed. The participants performed a test walk though the course before participating in the study in order to learn how to navigate using the numbers and arrows on the poles.

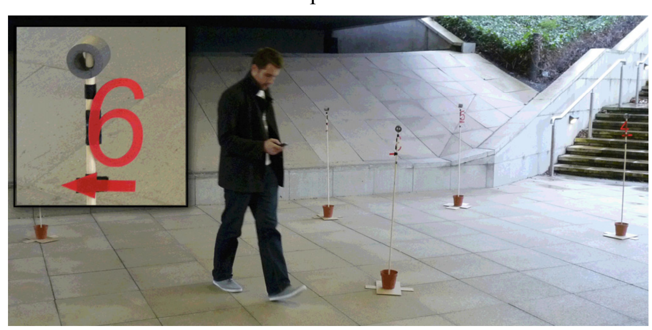

Figure 5. Participant walking around the test track (right) and orientation help on top of the poles (top left).

3.5 Dependent Measures

During the user study, quantitative as well as qualitative information was measured. Video footage of the participants was recorded during both phases of the study in order to collect data about the walking speed, head movements and walking errors (e.g. going the wrong way, slowing down, standing still, hitting a pole) participants made. After each phase, the participants were asked to fill in a questionnaire in order to find out about how they estimate their own performance during the different walking conditions regarding *accuracy* and *speed*. Additionally, a workload assessment – inspired by the NASA Task Load Index (TLX) [15] – was conducted. Therefore, a five-point scale between *very low* and *very high* was used and categories measured were *mental demand*, *physical demand* and *frustration level.* The categories were weighted equally.

3.5.1 Target Acquisition Task Measures

The performance during the target acquisition task was measured by different quantitative variables: target selection time, error rate, and overall task completion time. This data was collected automatically by the prototype application running on the mobile phone. Each participant was asked to perform one run consisting of 90 targets that they were supposed to select by touching them with their thumb. Distance between two consecutive targets varied and was either 31.5 mm and 42.5 mm.

In summary, the study design was:

16 Participants x
90 Targets x
4 Settings (1 standing + 3 walking conditions)
= 5760 measures overall

3.5.2 Reading Task Measures

The additional automatic measures that were performed during the reading task were task completion time and the scrolling behavior. The prototype used a smooth scrolling technique that fades out the scrolling speed before the text stays in its fixed position. Thus, it was possible to flick the text. The scrolling was measured in pixels and shows the overall distance of the thumb movement on the screen. The reading comprehension results between the different settings were also collected. Every single participant had to read two passages of text in each setting.

In summary, the study design was:

16 Users x
2 Passages of text x
4 Settings (1 standing + 3 walking conditions)
= 128 measures overall

3.6 Procedure

At the beginning of the study, the participants filled in a pre-questionnaire about their demographic data and their mobile phone usage history. Following this, an introduction to the task (target acquisition or reading) was given. Participants were requested to find the best trade-off between speed and accuracy. The track was explained to the users so that they got an idea of how to navigate the track. Participants received the same training of the track and were allowed to walk around the track once without the mobile. The walking speed observed during this lap was used as a control for the walking speed measures. Each participant started with the target acquisition task. After this task, the track was changed and again, the participant was allowed to walk around the track once in order to give the same training to every single participant.

3.6.1 Target Acquisition Task

The prototype was explained to the participants and they got a short training. One run contained the selection of 90 targets (e.g. 45 times with small distance and 45 times large distance) on the touch screen using the thumb in each setting. Participants were also asked to remember the size of the targets currently being tested in order to be able to compare the four different target sizes afterwards in the questionnaire. In accordance to the counterbalanced order, every participant performed four settings and was told which setting they are going to perform next in each run. These settings contained three walking conditions with varying target sizes and one standing condition as a control, which used the smallest of the three target sizes. After the four runs, the participants were asked to fill in a questionnaire examining the participants' accuracy and speed estimation as well as the subjective workload.

3.6.2 Reading Task

After the target acquisition task, the second phase of the user study was started examining the reading of passages of text. Therefore, the track layout was changed. After walking around the new track layout once without the mobile, the participants received a training of the scrolling behavior of the text display. As with the target acquisition task, the users performed all the four different settings, but now had to answer reading comprehension questions after reading each passage of text. For each setting, two texts were shown to the users so that every participant had to read eight passages of text in total. Having finished the reading tasks, the participants had to fill in another questionnaire comparing the subjective workload and the self-estimations regarding *speed* and *accuracy* when reading the passages of text in the three different sizes.

Each participant was asked to fill in a post-questionnaire about their ideas concerning the two different tasks and the different target/text sizes.

4. RESULTS

This section will show the results of both tasks previously discussed. Error bars (where present) display the 95% confidence intervals.

4.1 Target Acquisition Task

In the following section the results of the target acquisition task will be presented. First, the effect of walking will be analyzed before analyzing the differences regarding the three different targets sizes when walking.

4.1.1 Quantitative Measures

Effect of walking on target selection time. Walking had a large effect on target selection time, which was increased by 31.25% between the standing and the walking condition (see Figure 6). The control condition *stand/small* was compared to the condition *walk/small* using a dependent t-test. On average, the target selection time while walking ($M = 459$, $SE = 17.25$) was significantly higher than while standing and is also showing a large effect size, ($M = 603$, $SE = 47.59$), $t(15) = -3.66$, $p < .01$, $r = .62$. This stands in contrast to the findings of Lin et al. [11] and Mizobuchi et al. [13] who did not find a significant effect of movement on target selection time. This might be due to the more realistic test track used in this study, which forced participants to concentrate on the navigational task. Another reason might also be that there is a difference between stylus-based input and thumb-based input.

Target selection time while walking. The average time needed to hit a target decreased with larger target sizes (see Figure 6). When comparing the values for the different target sizes, then there was a decrease in target selection time of 11.5% between the small and the medium size, and of 7.8% between the medium and the large target size. A one-way repeated measures ANOVA was used to compare the results between the three different walking conditions. Mauchly's test violated the assumption of sphericity ($\chi^2(2) = 7.06$, $p < .05$). This violation means that a Greenhouse-Geisser correction had to be used ($\varepsilon = .72$) in order to correct the degrees of freedom. The results show that there was no significant main effect of target size on target selection time ($F_{1.43, 21.5} = 2.80$, $p = .10$).

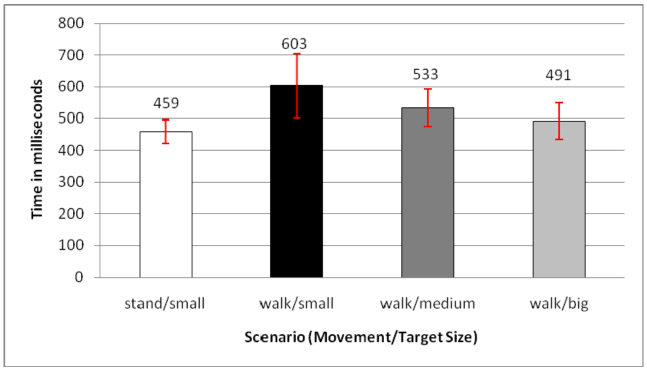

Figure 6. Average time needed to select a target.

An error was counted if the participant was not able to hit the target. In this case they had to select the same target again as long as it was successfully selected.

Effect of walking on error rate. Users made 6.77% less errors when standing (see Figure 7). A dependent t-test, analyzing the effects of movement on error rate between the standing and the walking condition, shows a significant difference between the standing ($M = .23$, $SE = .03$) and the walking condition ($M = .30$, $SE = .04$), $t(15) = -3.04$, $p < .01$, $r = .53$).

Error rate while walking. In Figure 7, it can be observed that the error rate decreased with larger target sizes. The decrease was higher between the small and the medium targets (29.58%) as between the medium and the large targets (23.19%).

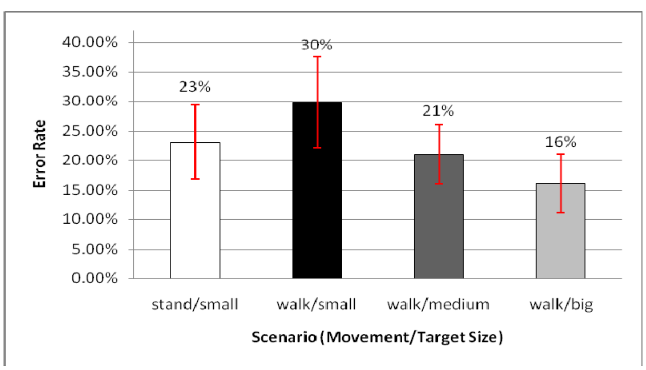

Figure 7. Percentage of how often users were not able to hit the target.

Mauchly's test did not meet the assumption of sphericity ($\chi^2(2) = 10.51$, $p < .05$) so that the degrees of freedom were corrected using a Greenhouse-Geisser correction ($\varepsilon = .65$). Results show a significant effect of target size with a large effect size ($F_{1.31, 19.63} = 21.54$, $p << .01$, $r = .76$). Bonferroni post-hoc tests, which were performed in order to see where the differences between the different target sizes can be found, revealed significant differences between all the target sizes (all $p \leq .01$).

Effect of walking on task completion time. Users needed 40% more time when walking in order to complete the target selection task (see Figure 8). A dependent t-test revealed a significant difference between the standing ($M = 56701$, $SE = 3499$) and the walking condition ($M = 79746$ $SE = 6308$), $t(15) = -5.18$, $p << 0.01$, $r = .73$.

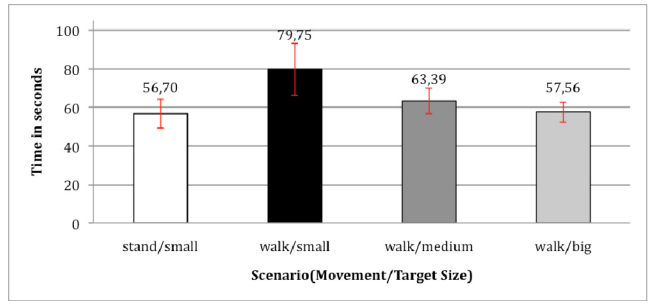

Figure 8. Overall Task Completion Time for one test run consisting of 90 Targets.

Overall task completion time while walking. The overall task completion time decreased with larger targets. On average, the increase in target size of 20% between the small and the medium target size decreased the completion time by 20.51%. This effect was more than half the size smaller between the medium and the large target size (9.20%) A one-way repeated measures ANOVA between the different target sizes was performed to analyze the effects on task completion time. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 7.37$, $p < .05$) and so a Greenhouse-Geisser correction was conducted ($\varepsilon = .71$) and revealed a significant effect of target size ($F_{1.42, 21.29} = 9.381$, $p < .01$, $r = .59$. Bonferroni post-hoc tests revealed a significant difference between the small and the large target size, $CI_{.95} = 5706$ ms (lower) 38672 ms (upper), $p < .05$.

Effect of target acquisition task on walking speed. Users slow down their walking speed when using a mobile phone (see Figure 9). A control condition was used in order to investigate the walking speed of participants around the test track layout when not using a mobile phone.

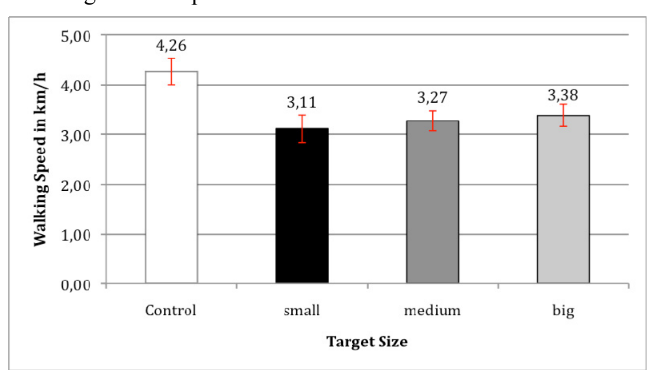

Figure 9. Walking speed measures for different target sizes and the control condition without mobile.

On average, participants decreased their walking speed by approximately 25% when using a mobile phone. Pairwise comparisons between the control condition and the different target sizes indicate that there is a main effect of the mobile usage condition for all target sizes (all $p \ll .01$). However, there was no effect of target size on walking speed ($p > .05$).

Video analysis. Overall, the performance of walking improved for larger target sizes. Participants made more walking errors when the target size was small (see Figure 10). This effect can be seen in all the different categories. One-way repeated measures ANOVA shows that the number of times participants stopped walking ($F_{2,30} = 5.68$, $p < .01$, $r = .48$) as well as the number of times they slowed down ($F_{2,30} = 8.74$, $p \ll .01$, $r = .57$) was significantly affected by the target size. The number of times users walked on the wrong path and looked up did not reach significance (*p > .05)*. Bonferroni post hoc tests revealed a significant difference between the small and the medium target size for the times users stopped walking ($CI_{.95} = .164$ (lower) 2.336 (upper), $p < .05$). For the times users slowed down, there were significant differences between the small and the medium ($CI_{.95} = .29$ (lower) 4.46 (upper), $p < .05$) as well as between the small and the large size ($CI_{.95} = .71$ (lower) 4.42 (upper), $p < .05$).

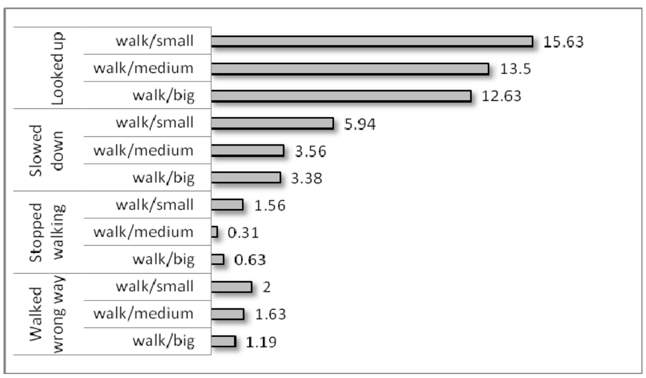

Figure 10. Average number of different walking errors during target acquisition task.

4.1.2 *Qualitative Measures*

Accuracy Estimation. 12 of 16 participants rated their accuracy best for the large target size. All the 16 participants thought to have performed worst for the small target size. A non-parametric Friedmann test shows that the participants' rating of their own accuracy was significantly affected by target size ($\chi^2(2) = 7.59$, $p < .03$). Wilcoxon tests were used in order to follow-up this result. A Bonferroni correction was used and all the results are reported at a significance level of .025. It appeared that the small target size compared to the medium target size significantly affected the accuracy estimation ($Z = 3.76$, $r = .66$) as well as the large target size compared to the medium target size did ($Z = 2.32$, $r = .41$).

Speed Estimation. For the large target size, 9 participants thought to be fastest, whereas 5 estimated to be fastest with the medium and 2 with the small size. 13 participants thought to be slowest with the small size and 10 participants rated their own speed to be second fastest with the medium size. A non-parametric Friedmann test shows that the participants' rating of their own speed was significantly affected by target size ($\chi^2(2) = 11.63$, $p \ll .01$). Wilcoxon tests were used in order to follow-up the results. It appeared that the small target size compared to the medium target size significantly affected the speed estimation ($Z = 2.70$, $p < .025$, $r = .48$), whereas the comparison between the medium and the large target size did not meet significance.

Subjective Workload. Results of the workload assessment show that the task was not rated as a highly demanding task (see Figure 11). On a five-point scale between very low (-2) and very high (+2) the overall task load was in average rated -1.19 for the *standing/small*, 0.02 for the *walking/small*, -0.67 for the *walking/medium* and -1.06 for the *walking/large* condition. A non-parametric Friedmann test shows that the subjective workload while walking was significantly affected by target size ($\chi^2(2) = 20.11$, $p \ll .01$). Follow-up Wilcoxon signed-rank tests show that there was a significant difference between the small and the medium ($Z = 3.02$, $p < .025$, $r = .53$), as well as between the medium and the large size ($Z = 2.64$, $p < .025$, $r = -.47$). A non-parametric Wilcoxon signed-ranks test shows that the workload during the control condition (*stand/small*) ($Mdn = -1.19$) was significantly lower than during the condition *walk/small* ($Mdn = 0.02$), $Z = 3.42$, $p \ll .01$, $r = .60$.

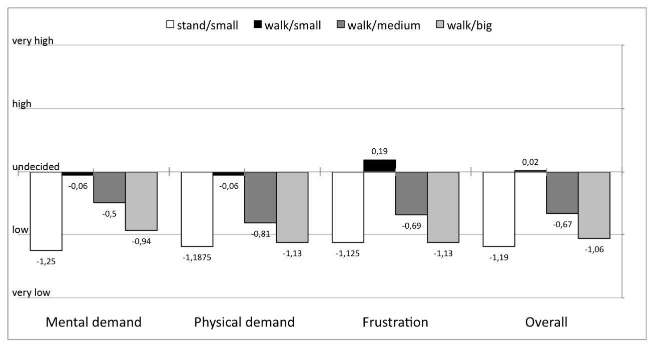

Figure 11. Subjective workload measures for target acquisition task.

4.2 Reading Task

4.2.1 *Quantitative Measures*

Effect of walking on reading speed. The reading speed measured in words per minute decreased by 18.6% when walking compared to the standing condition (see Figure 12). A dependent t-test between the control condition *stand/small* ($M = 190.12$, $SE = 12.40$) and the condition *walk/small* ($M = 154.69$, $SE = 11.77$) shows a significant main effect of movement on reading speed, $t(15) = 2.95$, $p = .01$, $r = .61$.

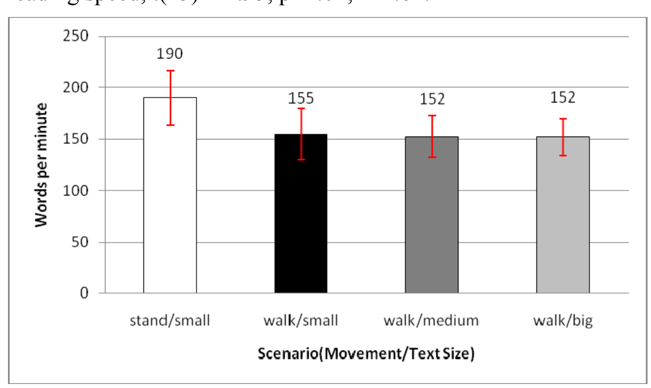

Figure 12. Reading speed during reading task.

Reading speed while walking. It is notable that there was no difference in reading speed between the three different text sizes ($p > .05$) while walking (see Figure 12). Reading speed ranged from 155 words/min for the small text size to 152 words/min for the large text size.

Effect of movement on scrolling behavior. The scrolling behavior is a measure of the thumb movement on the screen. With larger text sizes, it is necessary to move the text more than with a small text size, according to the increase in text wrapping. However, the smooth scrolling technique used in the prototype, made flicking possible so that it was not necessary to use the thumb all the time. Participants scrolled 19% less during the walking condition compared to the control condition, which demonstrates that there is a trade-off between the navigational and the reading task (see Figure 13). However, the results of a dependent t-test between the standing ($M = 313.56$, $SE = 39.53$) and the walking condition ($M = 252.47$, $SE = 42.90$) revealed that there were no statistically significant differences ($p > .05$).

Scrolling behavior. The scrolling behavior increased in average by 119% (300 pixels) between the small and the medium text size and by 42% (231 pixels) between the medium and the large size. This effect of text size on scrolling is much higher than the increase in text size of 20% between the different settings (see Figure 13). A one-way repeated measures ANOVA was conducted in order to compare the scrolling behavior between the different text sizes. The results show, that the text size has a very large effect on scrolling behavior ($F_{2, 30} = 33.24$, $p << .01$, $r = .82$).

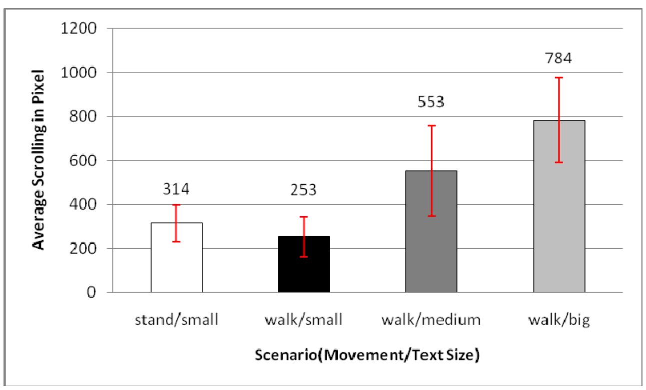

Figure 13. Amount of pixels scrolled during reading task.

Walking speed. People slowed down their walking speed by more than 1 km/h in average when reading passages of text on the mobile compared to a situation where no mobile was used. The difference between the control condition and the three text sizes were tested by dependent t-tests. The results show a large main effect of the reading task on walking speed ($p << 0.01$). The walking speed was quite similar for the different text sizes. It was between 2.83 km/h for the large text and 2.86 km/h for the small and medium text. Statistical analysis showed that there were no significant differences in walking speed between the different text sizes ($p > .05$) (see Figure 14).

Video Analysis. Results of the video analysis regarding the walking errors did not meet expectations and were quite random for the times users walked the wrong way, slowed down or looked up (see Figure 15). For the times users stopped walking there seems to be a tendency, indicating that for a larger text size, users need to stop walking less often. However, this finding did not reach significance ($p > .05$).

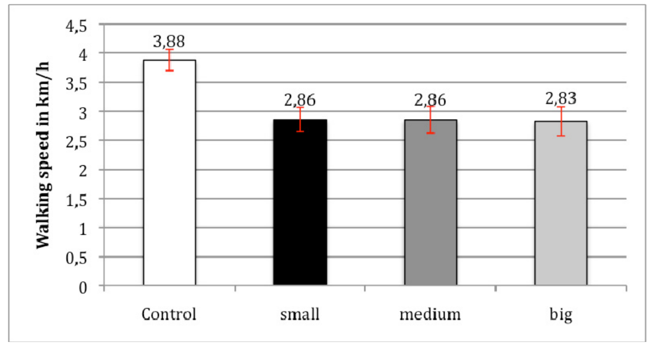

Figure 14. Walking speed during reading task.

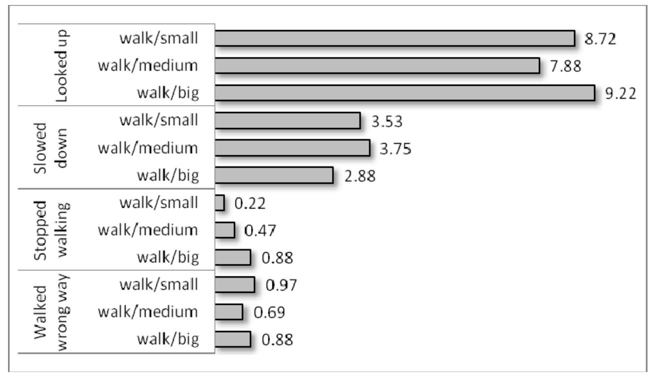

Figure 15. Walking errors and behavior during reading task.

Reading comprehension. Regarding reading comprehension, users were able to answer most of the questions right while standing (84.37%). While walking, the performance was best for the medium text size (79.69%). A one-way repeated measures ANOVA comparing the walking conditions showed that there was no main effect of text size on reading comprehension ($p > .05$). In parallel, a dependent t-test between the control and the condition *walk/small* did not reach significance.

4.2.2 Qualitative Measures

Accuracy estimation. The results of the accuracy estimation show that there is no consensus between the participants on which text size is best if someone is to read texts accurately. 6 of the 16 participants rated the small text size best, while 6 participants rated it worst. The same effect can be seen for the large text size, where 7 rated it as best and 8 as worst. Only for the medium size there is a more obvious result with 11 participants rating their accuracy in second place. A non-parametric Friedmann test shows that the participants' rating of their own accuracy was not affected significantly by text size ($p > .05$).

Speed estimation. The results of the speed estimation show similar results compared to the accuracy estimation with no obvious tendency. Seven participants rated the small text size best, whilst eight rated it third. The large size was rated five times first, five times second and six times third. For the medium size, 10 participants thought to be second best with this text size. A non-parametric Friedmann test shows that the participants' rating of their own speed also was not affected significantly by text size ($p > .05$).

Subjective Workload. The biggest differences in workload measures can be found in the category *mental demand* where the large text size was rated even better than the small text size in the standing condition (see Figure 16).

The physical demand was rated least in the standing condition. The frustration level is highest in the walking condition with a small text size and gets less with a larger text size. Overall, the workload also decreases with larger text sizes and a comparison between the control condition and the *walk/small* condition shows that the effect of movement on workload is larger than the change in text size. A non-parametric Friedmann test did not show a significant effect of text size on overall workload ($p > .05$). However, regarding just the mental workload, the results reach significance ($\chi^2(2) = 7.24$, $p < .05$). A Wilcoxon test, used to follow-up this result, shows that between the small and the medium text size as well as between the medium and the large text size no significance was reached ($p > .05$). A non-parametric Wilcoxon signed-ranks test shows that the workload during the standing condition (Mdn = -1.14) was significantly lower than during the walking condition (Mdn = -.33), $Z = 3.42$, $p << .01$, $r = .50$.

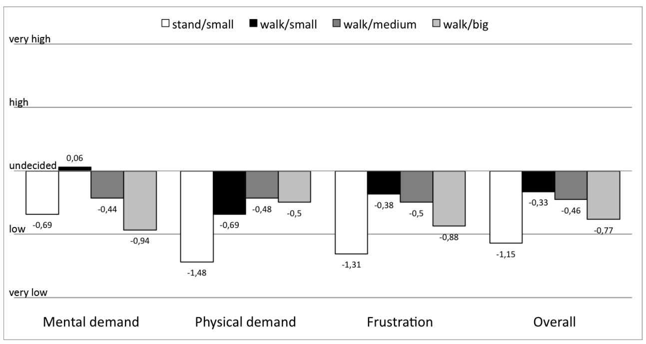

Figure 16. Subjective workload measures for reading task.

5. DISCUSSION

The results of the two tasks showed that there are different effects between the target acquisition task and the reading task. A change in target size has a much greater influence on performance than a change in text size.

5.1 Target Acquisition Task

The comparison between the control condition and the walking condition yielded significant results for all the different measures. This shows that walking conditions lead to serious problems concerning the usability of mobile applications.

The biggest effect of changing the target size was found for the error rate measures. The target selection time became smaller for larger targets; however, this finding did not reach significance. Anyhow, the error rate has a much larger impact on the usability of mobile applications while walking as for every wrong selection, in most cases, the user has to undo their wrong input and so will again have the chance to make another error when trying to undo their action.

Another main effect was found for the overall task time needed to complete one run consisting of 90 targets, which had to be hit correctly during the tasks. There was a significant difference between the small and the large target size showing that less errors also means that tasks can be performed faster if fewer errors are made by users. Target size also has to be increased by a large amount to really avoid errors. Especially in a mobile situation this is an important result because users should not be distracted for too long from the much more important navigation task.

The participants estimated large target sizes as best for accuracy and speed and small target sizes as the worst. These estimates match the quantitative data regarding accuracy and speed. The same pattern can also be found for the subjective workload assessments. Users felt less stressed for the big target sizes and the ratings were best for the standing condition. Even if the overall workload was not seen as high, the results reached significance between the different settings and the users prefer bigger target sizes.

The change in target size also had an effect on the navigation task that users had to solve while walking around the test track. Users tended to stop walking more often when the buttons were the smallest size. This effect was especially large between the small and the medium size so that it could be concluded, that the small target size forces the user to change the priority of tasks between the navigation and the targeting task. The targeting task becomes the primary task and the navigation task, which always should be the primary task, becomes the secondary task. Users in a real-world situation could get into seriously dangerous situations, if they do not take care of where they are walking. The high error rate and the longer overall task completion time for small targets both yield to an increase of the possibility to have an accident. This is reinforced by the number of times users slowed down their walking speed. This measure reached significance for all comparisons performed and makes the result even more alarming.

5.2 Reading Task

The effect of walking was significant in all measures, except the scrolling behavior and the reading comprehension. Users scrolled slightly less when using the mobile while walking compared to the standing condition. The reading comprehension was best for the standing condition, followed by the medium text size.

It was an expectable result that the scrolling behavior also increased significantly with a larger text size. But in addition, this had no effect on the reading speed needed to go through the passages of text. This leads to the question if the additional work needed to scroll through the texts removes a performance gain of a larger text size. Another possible reason for not improving the performance of reading by adjusting text size might be, that with a larger text size only a very small number of words could be put within one line of text. This makes the passages of text less readable, which was pointed out by some participants and corresponds to [19].

The text size did not have an effect on walking speed. Regarding previous research results, our outcome conforms to these results. There seems to be a specific amount of cognitive load having to be shared between the navigational and the reading task. This amount can be named by the reduction in walking speed, which was at about 1 km/h. As well as for the walking speed, there was no main effect of text size on the walking errors users made. The results were quite random and no tendency was observable except for the times users stopped walking. Regarding workload, the sub-category ‘mental demand’ was the only measure that reached significance. Users found it easier to read passages of text with larger letters.

Overall, participants were not able to say, which text size allowed them to be more accurate and faster. This fits well again to our quantitative measures. A lot of participants annotated that for them, the change in text size did not have a large effect.

6. CONCLUSION AND FUTURE WORK

The results presented in this paper show the negative effect of walking on target selection and reading when interacting with mobile devices. It was shown that this can be compensated by increasing the target size when it comes to selection tasks. Increasing the text size does not provide any benefits as the positive effects of this are again compensated by the increased demand for scrolling. Those research results can influence future work towards the provision of a walking mode in mobile devices.

Especially the interaction with small targets on mobile devices has gained quite some interest in the last years. This lead to new interaction techniques such as Shift [21] of Escape [23] being based on fine granular zooming techniques or gestures. We assume that, based on the results presented in this paper, those techniques can't effectively be used when walking as very accurate finger movements or gestures have to be performed. Our future work will focus therefore on the development of adaptive user interfaces allowing the user to accurately select small targets on mobile phone screens. At this we will consider especially the provision of semi-transparent information on top of the previously selected information, which helps the user to perform the correct selections without losing or changing the context.

7. ACKNOWLEDGEMENT

This work is supported by the NoE INTERMEDIA (funded by the European Commission, NoE 038419) and Nokia.

8. REFERENCES

[1] Apple Inc. iPhone Human Interface Guidelines, 2009.

[2] Leon Barnard, Ji Soo Yi, Julie A. Jacko, and Andrew Sears. An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. International Journal of Human-Computer Studies, 62(4): 487-520, 2005.

[3] Iain Darroch, Joy Goodman, Stephen Brewster, and Phil Gray. The effect of age and font size on reading text on handheld computers. In INTERACT 2005, 253–266, 2005.

[4] Alan Dix, Janet E. Finlay, Gregory D. Abowd, and Russell Beale. Human-Computer Interaction. Prentice Hall, third edition, December 2003.

[5] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47(6): 381–391, June 1954.

[6] James R. Lewis. Literature Review of Touch-Screen Research from 1980 to 1992. IBM Technical Report 54.694, August 1993.

[7] International Standard for Organization. ISO 9241-9: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices, 2000.

[8] Shaun K. Kane, Jacob O. Wobbrock, and Ian E. Smith. Getting off the treadmill: Evaluating walking user getting off the treadmill: Evaluating walking user interfaces for mobile devices in public spaces. In MobileHCI '08, pp. 109–118.

[9] A. K. Kjeldskov and J. Stage. New techniques for usability evaluation of mobile systems. International Journal of Human-Computer Studies, 60(5-6):599–620, 2004.

[10] Learning Express, editor. 501 Reading Comprehension Questions. Learning Express, Third edition, 2006.

[11] Min Lin, Rich Goldman, Kathleen J. Price, Andrew Sears, and Julie Jacko. How do people tap when walking? An empirical investigation of nomadic data entry. International Journal of Human-Computer Studies, 65(9):759–769, 2007.

[12] I. Scott MacKenzie and William Buxton. Extending fitts' law to two-dimensional tasks. In Proc. CHI '92, pp. 219-226.

[13] Sachi Mizobuchi, Mark Chignell, and David Newton. Mobile text entry: relationship between walking speed and text input task difficulty. In Proc. MobileHCI '05, pp 122–128.

[14] Terhi Mustonen, Maria Olkkonen, and Jukka Häkkinen. Examining mobile phone text legibility while walking. In CHI '04 extended abstracts, pp. 1243–1246.

[15] NASA Ames Research Center, 1987. Manual of NASA Task Load Index (TLX), v. 1.0 Computerized Version.

[16] Pekka Parhi, Amy K. Karlson, and Benjamin B. Bederson. Target size study for one-handed thumb use on small touchscreen devices. In Proc. MobileHCI '06, pp. 203–210.

[17] Ahmad Rahmati, Clayton Shepard, and Lin Zhong. NoShake: Content stabilization for shaking screens of mobile devices. In Proc. Percom 2009, pp. 1-6.

[18] Matt Richtel. Forget gum. Walking and using phone is risky. The New York Times. 17.01.2010

[19] Miles A. Tinker. Bases for Effective Reading. University of Minnesota Press, November 1965.

[20] Kristin Vadas, Nirmal Patel, Kent Lyons, Thad Starner, and Julie Jacko. Reading on-the-go: a comparison of audio and hand-held displays. In Proc. MobileHCI '06, pp219– 226.

[21] Daniel Vogel and Patrick Baudisch. Shift: a technique for operating pen-based interfaces using touch. In Proc. CHI '07, pp. 657–666.

[22] Williams, E. J. Experimental designs balanced for the estimation of residual effects of treatments. Australian Journal of Scientific Research, Ser. A 2, 1949, pp. 149-168.

[23] Koji Yatani, Kurt Partridge, Marshall Bern, and Mark W.Newman. Escape: a target selection technique using visually-cued gestures. In Proc. CHI '08, pp 285–294.