Exploring Methods to Improve Pen-based Menu Selection

First Author Name (Blank if Blind Review)
Affiliation (Blank if Blind Review)
Address (Blank if Blind Review)
e-mail address (Blank if Blind Review)

Second Author Name (Blank if Blind Review)
Affiliation (Blank if Blind Review)
Address (Blank if Blind Review)
e-mail address (Blank if Blind Review)

ABSTRACT

Though Tablet PCs and stylus-based PDAs are gaining popularity, many individuals (and in particular older individuals) still struggle with pen-based devices. One type of error, missing just below, occurs when a user's tap distribution is downwardly shifted, such that, he/she selects the region just below the target relatively often, while rarely selecting the corresponding region of the target itself. This paper attempts to address this problem and presents the results of laboratory experiment to evaluate two interfaces designed to address missing just below, relative to each to a control interface. Our results found that one of the proposed interfaces was effective, but that some participants disliked it. We discuss possible reasons for this disconnect between performance and subjective response, and ways of addressing the negative feedback.

Author Keywords

Pen-based Target Acquisition, Tablet PC, Older Users, Interaction Techniques

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Direct pen-based input takes full advantage of hand-eye coordination, and offers a familiar form of interaction [Greenstein, 1997]. With current-day Tablet PCs and stylus-based PDAs gaining popularity, pen input now seems more than ever in a position to succeed as a mainstream form of input.

The main motivation for this research comes from a previous study [Moffatt, 2007], in which we examined the types of difficulties older and younger users encounter while tapping to acquire targets on a Tablet PC. In this paper, we investigate methods for addressing, missing just below errors, one of the difficulties identified in that study. Missing just below refers to erroneous selections of the top edge[1] of the item below the targeted item. In [Moffatt 2007], the data across all age groups suggested that a selection in this edge region is 11 times more likely to be intended for the item above the selected item than the selected item itself. As such, it seems possible to use this information to reduce target acquisition errors.

We have identified two possible approaches for addressing missing just below errors: reassigned edge, and deactivated edge. In the *reassigned edge (RE)* approach, the top edge of each menu item is reassigned such that taps in this region result in selection of the item above. This approach effectively shifts the target region (in motor space) of each menu item down by two pixels, while leaving the visual appearance unchanged. In the *deactivated edge (DE)* approach, the top edge of each item is deactivated such that taps in this region are ignored. This approach effectively shrinks the height (in motor space) of each item to 90% of the original height, without changing the visual appearance, and adds an invisible menu separator between each item.

In this paper, we present the results of a controlled laboratory experiment designed to evaluate the RE and DE approaches relative to each other and to a *Traditional Edge (TE)* control condition. Specifically, the experiment was designed to test the following hypotheses:

H1. Both DE and RE will be superior to TE in terms of reducing top edge errors.

H2. DE will be superior to RE in terms of reducing top edge errors, but will require an overall increase in the number of taps required to make a selection.

H3. Both groups will benefit from DE and RE, but the older age group will benefit more so.

H4. RE will be preferred to DE, especially by the older participants.

Our rationale for these hypotheses comes from our previous observations. Given the distribution of taps observed in [Moffatt, 2007] (shown in Figure 3), we would predict that both conditions will achieve an overall reduction in top edge errors. However, because the RE condition does

[1] Throughout this paper, "top edge" refers to the top 10% (i.e., the top 2 pixels of a 20 pixel high item). This definition comes from the vertical tap distributions observed in [Moffatt, 2007].

introduce some errors (whenever the user taps on the top edge of the target item), we would further predict that in terms of top edge error reduction, it would be inferior to DE. However, there is a tradeoff. Although DE does not introduce any new errors, it requires the user to re-tap for *every* top edge selection. Thus, we predict that there will be an overall increase in the number of taps required to select items in DE.

Furthermore, from our observations in [Moffatt, 2007], we noticed that users typically do not wait to see if their taps are successful. On the few occasions where taps did not register (for example, because the user hit a real menu separator), we noticed subjects try to move on to the next trial, subsequently realize they had not finished the current trial, and then go back to try again. Thus, we predict that although DE will result in the greatest reduction in errors, its incremental advantage over RE will be sufficiently small compared to the disruption caused by having to re-tap that users will prefer RE. We predict that this will be especially true for older users as older users have previously been shown to be less able to adapt to adapting to changing task requirements [Heath, 1999].

EXPERIMENTAL METHODOLOGY

Participants

For the study, we recruited 24 participants from two age groups: *younger* (aged 19–30, mean 24; 5 male, 7 female), and *older* (aged 66–81, mean 73; 6 male, 6 female). Participants received $10 per hour of participation, and to additionally motivate individuals to perform the task quickly and accurately, a $10 incentive was awarded to the top 1/3 of the participants in each age group.

All of the participants were free of diagnosed impairment to their hands, and had normal or corrected-to-normal eyesight. To control for biases between age and Tablet PC experience, we limited participation to individuals with no Tablet PC experience and no or limited PDA experience. However, differences in general computer experience did exist, with the younger participants being more experienced (in terms of frequency of use, breadth of applications used, and self-ranking) than the older participants.

Task

Three menu types were used for the evaluation: *reassigned edge (RE)*, *deactivated edge (DE)*, and *traditional edge (TE)*. For each type of menu, participants completed a shorted practice block followed by six blocks of trials with an enforced 45 second break between blocks. Each block consisted of a 36-item randomly ordered selection sequence from a single 12 item menu (each item was selected three times). For each trial, the item to be selected was displayed across the top of the screen, above the menu.

Menu contents remained consistent within each menu condition, but changed between conditions. Each menu contained three groups of 4 semantically related items. These schemes were randomly generated using the approach presented by Cockburn, Gutwin, and Greenberg [2007]. Each item was 20 pixels (4.8 mm) high.

Procedure

The experiment was designed to fit into a single 120 minute session. For one participant in the older age group, keeping the study length within 120 minutes required modification to the design. This participant only completed four (out of six) blocks for each of the three conditions. All other participants finished in between 75 and 120 minutes.

Participants started the study by completing a series of standardized tests of their sensory-perceptual and motor-skills, and a brief questionnaire about their background and computer experience. They were briefly introduced to the Tablet PC and the tablet was calibrated to them. We note that beyond the instructions given in this introduction, participants were not instructed to use the pen in any particular manner. We explicitly wanted to observe how individuals would naturally approach the task.

Participants then completed the menu conditions (in a counter-balanced presentation order). Participants were instructed that they were going to be using three different menu programs, but not told how the programs differed. After each menu condition, participants were asked to complete the ISO9241-9 independent ratings questionnaire[2] [ISO, 2000], and between conditions, participants completed short verbal distracter tasks. Finally, a feedback questionnaire was used to rank the conditions on qualitative dependent variables and to record additional comments.

Measures

Our main variable of interest was error rate, although for completeness we also recorded trial time. In particular, we were interested in the effect of our three designs on errors involving taps on the top edge of a menu item. Thus, for each interface we calculated the Net Benefit of top edge selections (TES) as follows:

$$NET_{TES} = Correct_{TES} - Incorrect_{TES}$$

To clarify, for the TE condition, a correct top edge selection occurs when the top edge of the target item is selected; for the RE condition, it occurs when the top edge of the item below the target is selected; and for the DE condition, it occurs if the subsequent item selection (after a tap on any top edge) is within the active region of the target item.

To quantify the cost of deactivating pixels, we recorded the number of taps per trial. Furthermore, we recorded the vertical tap distribution of all trials to allow us to make informal comparisons to the distributions observed in [Moffatt, 2007]. From the poll-style questionnaire, we recorded overall preference, efficiency, error rate, and frustration, in addition to free-form comments.

[2] The purpose of this questionnaire was to emphasize the switch between conditions and to encourage participants to reflect. For analysis purposes, we focused on the final poll-style questionnaire.

Design
The experiment used a 2x3 factorial design with age (younger, older) as a between-subjects factor, and interface (RE, DE, TE) as a within-subjects factor. A within-subjects design was chosen for its increased power and because it allows for comparative comments on the interfaces.

Apparatus
We used a Fujitsu LifeBook T3010D Tablet PC with a 1.4 GHz Pentium M processor and 768 MB RAM, running the Windows XP Tablet Edition operating system. The display was 12.1 inches large, with a resolution of 1024 x 768. The experimental software was written in Java, using the Standard Widget Toolkit (SWT). For the experimental tasks, the Tablet PC was placed on a stand, which positioned the screen at a comfortable viewing angle (based on previous pilot studies) of approximately 35 degrees from horizontal. Participants were encouraged to adjust the position of their chair and the placement of the stand.

RESULTS
In this section, we present the experimental results. As expected, there were no significant main or interaction effects of the presentation order of our interfaces, so we examine only the effects of age and interface in this section. To determine if there were differences in performance between the menu conditions, two-way ANOVA's (Age x Interface) and post-hoc comparisons using Bonferroni adjustments were calculated for the Net $_{TES}$ (defined in the previous section) and for the average number of taps needed to make a selection. In all of our repeated-measures analyses, sphericity was an issue; thus, Greenhouse-Geisser adjustments were used. Along with statistical significance, we report partial eta-squared (η^2), a measure of effect size. Effect size is a measure of practical significance, and is often more appropriate than statistical significance in applied research in Human-Computer Interaction [Landauer 1997]. To interpret partial eta-squared, .01 is a small effect size, .06 is medium, and .14 is large.

Net Benefit for Top Edge Selections
There was a significant main effect of interface ($F(1.22,26.90) = 5.88$, $p = .017$, $\eta^2 = .211$) and a marginally significant interaction between interface and age ($F(1.22,26.90) = 3.19$, $p = .078$, $\eta^2 = .127$) on Net $_{TES}$. Post-hoc pair comparisons revealed that for the older age group, the DE condition had a significantly higher Net $_{TES}$ than both the TE condition (p = 0.010) and the RE condition (p = 0.001) for the older age group. This interaction is shown in Figure 1. There was no main effect of age on Net $_{TES}$.

Taps per Trial
During the study, we observed a number of the older participants struggling to open the menu. Thus, for this analysis, we exclude the taps required to open the menu, and count only the number of taps required to select the item once the menu was successfully opened. There was a significant main effect of interface ($F(1.18,25.92) = 5.43$, $p = .028$, $\eta^2 = .198$), a significant main effect of age ($F(1,11.20) = 5.55$, $p = .038$, $\eta^2 = .202$), and a marginally significant interaction between interface and age ($F(1.18,25.92) = 3.29$, $p = .075$, $\eta^2 = .130$) on taps to select. Post-hoc pair comparisons revealed that for the older age group, the DE condition required significantly more taps per trial to make a selection than both the TE condition (p = 0.010) and the RE condition (p = 0.030). Note that this is across all trials not just trials involving a top edge tap. This interaction is shown in Figure 2.

Vertical Distribution of Taps
Figure 3a shows a histogram of the vertical distribution of taps relative to the center of the target item. We note that this distribution is not consistent with that reported in [Moffatt, 2007] (see Figure 3b). Most notably, the data in [Moffatt, 2007] showed a downward shift in the distribution (the mean was 25% below the center of the target), and correspondingly, there were a sizeable number of selections on the top edge of the item below the target (44), with relative few selections on the top edge of the target item itself (4). In contrast, in this study, the mean was close to the target center, and there were in fact more selections on the top edge of the target item than on the top edge of the item below. This data helps explain the unsatisfactory performance results seen for the RE condition. We further discuss this distribution and potential reasons for the difference in the following section.

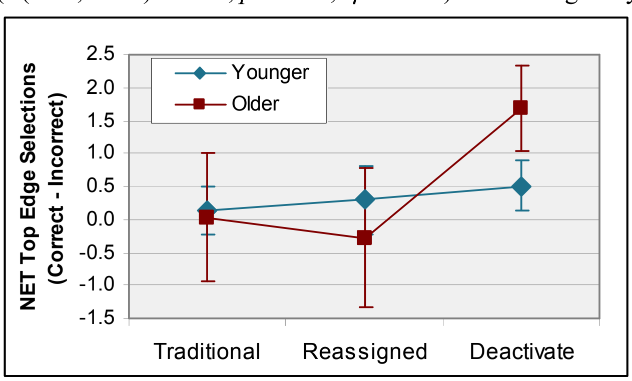

Figure 1: Average Net Benefit per block of each interface, by age group (*N*=24). Error bars represent 95% Confidence Intervals.

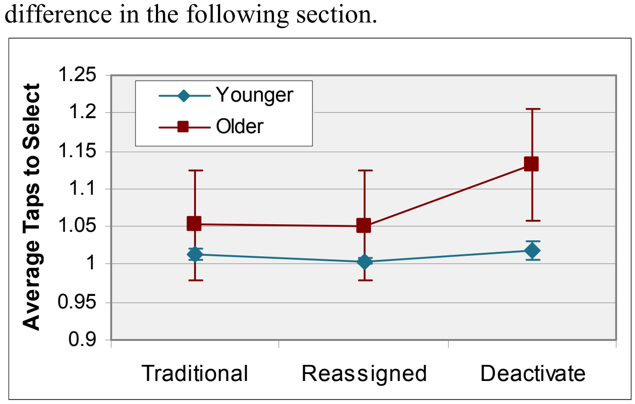

Figure 2: Average number of taps needed to select an item, by age group (*N*=24). Error bars represent 95% Confidence Intervals.

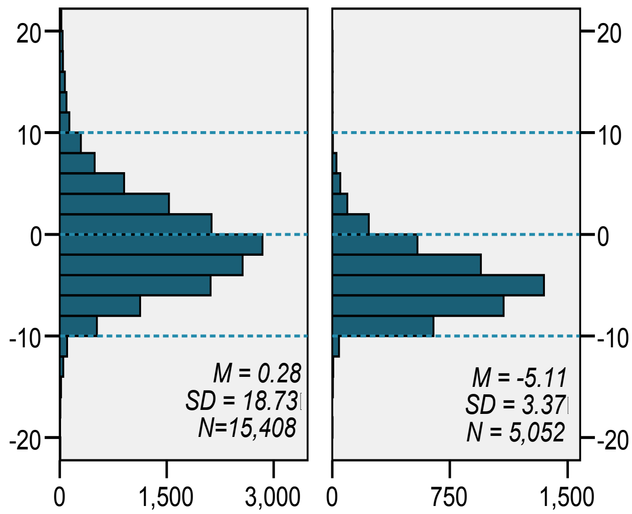

Figure 3: Histograms of the vertical position of taps (occurring on the target item, and the lower/upper half of the item above/below) for this study (left), and the study reported in [**] (right).

Self-reported Measures

A Chi-square analysis on the frequency with which each menu condition ranked first on our subjective measures (preference, efficiency, accuracy, and frustration) revealed no differences, with most participants reporting no difference on two or more measures. However, spontaneous comments made by 6 participants (4 older) told a different story, and reflected a strong dislike for the DE condition. In contrast, there was only one negative comment made for TE and none made for RE.

These participants reported confusion and distraction. As one individual put it, "[It] really throws you off when you have to click more than once." Another described it with, "I kept thinking I had tapped the right thing and then had to go back." Others made less specific comments such as "[DE] seems to be a little more awkward," and "[With DE, it] was harder to make selections." Other comments reflected a misconception that more or more sustained force was required in the DE condition. For example one individual reported, "This one seems to need you to press harder," while another speculated, "I think you need to hold it [the pen] for quite a while [with DE]."

DISCUSSION AND FUTURE WORK

Although the DE condition did reduce errors on the top edge of the menu items, it is concerning that there was a negative response to it. One possible reason for this reaction is confusion over what exactly was happening when taps were ignored. For the purposes of evaluation, we did not explain the assistance to participants. It is possible that a better understanding of why taps are being ignored, coupled with feedback to let the user know their taps are registering (i.e., feedback indicating that they are using enough force) may rectify the negative assessment.

The poor performance of the RE condition, are not surprising when we consider the differences in the tap distributions between this study and [Moffatt, 2007]. One possible reason for the divergence in the data, is differences between the tasks used in the study. In {Moffatt, 2007], we used a discrete task, which required the user to return to the center of the screen after making a selection. In this study, we used a continuous task (to reduce trial time, and increase the number of trials per condition). However, as a result, some participants may have been starting their upward motion to the menu head (for the next trial) before fully completing the item selection (of the current trial). Although the continuous task is more realistic, this finding does suggest that in real life the tap distributions are likely to more varied, and less clear cut than the data in [Moffatt, 2007] would suggest.

Another factor may be the smaller number of menu items used in this study (12 versus 36). One explanation for the distribution observed in [Moffatt, 2007] is that hand occlusion caused participants to move past target items before selection. With fewer items to learn, participants may have been relying less on visual search, and thus did not need to target below the text label.

RELATED WORK

[Will need to find some room to put in some related work]

CONCLUSION

This paper presented the findings of an experiment comparing the effectiveness of two techniques designed to address missing just below errors. In contrast to [Moffatt, 2007], we did not see a clearly defined downward shift in the tap distributions as we expected. This is reflected in the poor performance results for the RE condition. On the other hand the DE condition was able to provide assistance. However, it was unpopular with users.

REFERENCES

1. Anderson, R.E. Social impacts of computing: Codes of professional ethics. *Social Science Computing Review 10*, 2 (1992), 453-469.
2. Zellweger, P.T., Bouvin, N.O., Jehøj, H., and Mackinlay, J.D. Fluid Annotations in an Open World. *Proc. Hypertext 2001*, ACM Press (2001), 9-18.

The columns on the last page should be of approximately equal length.