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, Older Users, Interaction Techniques

ACM Classification Keywords
H5.2. Information interfaces and presentation: User Interfaces — *Input Devices and Strategies*.

INTRODUCTION

Direct pen-based input takes full advantage of hand-eye coordination, and offers a familiar form of interaction [2]. With current-day Tablet PCs and stylus-based PDAs gaining popularity, understanding and addressing pen-based interaction difficulties is more important than ever before.

Despite many advantages, some individuals (and in particular older individuals) still experience difficulties with pen input [5]. This research seeks to address one such difficulty: *missing just below*. In a menu selection task, missing just below occurs when a user's tap distribution is downwardly shifted, resulting in frequent erroneous selection of the top edge[1] of the item below the target item, and infrequent selection of the corresponding region of the target item itself. One possible explanation for this downward shift is that with a pen, hand occlusions forces users to move beyond the target center before being able to read the item label. Previous data [5] suggests that a selection along the top edge of a menu item is 11 times more likely to be intended for the item above.

We have identified two possible approaches for addressing missing just below errors: reassigning selections along the top edge, and deactivating the 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, 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, and adds an invisible menu separator between each item. It also leaves the visual appearance unchanged.

We suggest that the main advantage of the RE approach is that it least changes the current interaction. We predict that most users would not notice the small shift, but would simply benefit from fewer errors. Its disadvantage is that it turns a small number of correct selections into errors. On the other hand, the DE approach does not introduce any new errors, but changes the interaction, requiring the user to re-tap every time they hit the top edge of an item. Moffatt and McGrenere noted [5] that users typically do not wait to see if their taps register, but rather move on, subsequently realize they have not, and then have to go back to try again.

Thus, we predict that although DE will result in the greatest reduction in errors, it will have other costs. We further predict that these costs will particularly affect older users as they have previously been shown to be less able to adapt to changing task requirements [3].

EXPERIMENTAL METHODOLOGY

To examine whether it is better to introduce a small number of new errors (as with RE) or potentially disrupt the user with unregistered taps, we ran a controlled laboratory experiment to compare the RE and DE approaches to each other and to a *Traditional Edge (TE)* control condition.

[1] We define the top edge based on the results in [Moffatt, 2007] to be the top 10% (i.e., the top 2 pixels of a 20 pixel high item).

Participants

For the study, we recruited 24 participants from two age groups: *younger* (aged 19–30, mean 24; 7 female), and *older* (aged 66–81, mean 73; 6 female). All participants received $10 per hour of participation, and an additional $10 incentive was awarded to the top 1/3 performers to motivate quick and accurate performance. 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, all were novices to pen-based interaction. However, younger participants did have greater general computer experience (in terms of frequency of use, breadth of applications used, and self-ranking) than older participants.

Design, Procedure, Task, and Apparatus

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. It 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. 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 ISO 9241-9 independent ratings questionnaire[2] [4], 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.

The menu task was as follows. For each interface, 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 constant within each menu condition, but changed between conditions. Each menu contained three groups of 4 semantically related items, and was randomly generated using the approach presented in [1]. Each item was 20 pixels (4.8 mm) high.

[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.

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. Bonferroni adjustments were used in all post-hoc comparisons. Furthermore, in all of our repeated-measures analyses (except trial time), sphericity was an issue; thus, Greenhouse-Geisser adjustments were used.

Net Benefit for Top Edge Selections

To measure the effect of our three designs on errors involving taps on the top edge of a menu item, we need to consider both the number of errors prevented and the number of errors introduced. Thus, we calculated a Net score 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.

A two-way ANOVA (Age x Interface) for NET_{TES} revealed 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 ($F(1.22,26.90) = 3.19$, $p = .078$, $\eta^2 = .127$). Post-hoc pair-wise comparisons further revealed that for the older age group, DE had a significantly higher NET_{TES} than both TE ($p = .01$) and RE ($p = .001$), as shown in Figure 1. There was no main affect of age ($p = .398$)

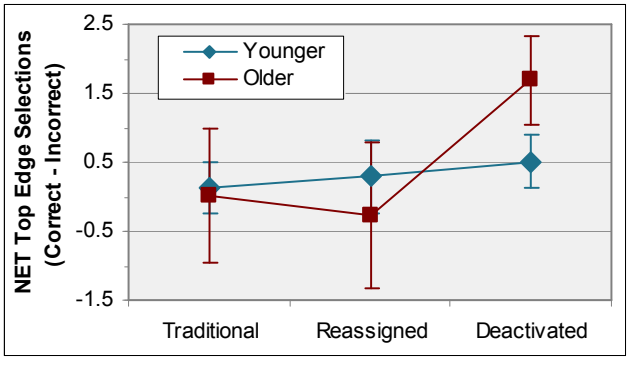

Figure 1: Average Net Benefit per block of each interface, by age group (N=24). Error bars represent 95% Confidence Intervals.

Cost of Re-tapping

To measure the potential negative impact of DE, we performed two-way ANOVA's (Age x Interface) on both the average number of taps required to select an item (excluding those required to open it), and the average trial time.

In terms of the average number of taps to select, 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 ($F(1.18,25.92) = 3.29$, $p = .075$, $\eta^2 = .130$). Post-hoc pair comparisons further revealed that for the older age group, DE required significantly more taps per trial than both TE (p = .01) and RE (p = .03), as shown in Figure 2. In contrast, there was no significant main effect of interface ($p = .537$) or interaction between interface and age (p = .705) for trial time (As we would expect there was a significant main effect of age, $F(1,11.252) = 45.88$, $p < .001$, $\eta^2 = .676$).

Thus, although the cost of re-tapping in DE did result in an overall increase in the number of taps required (for the older group), the cost was not measurable in terms of an increase to the time required.

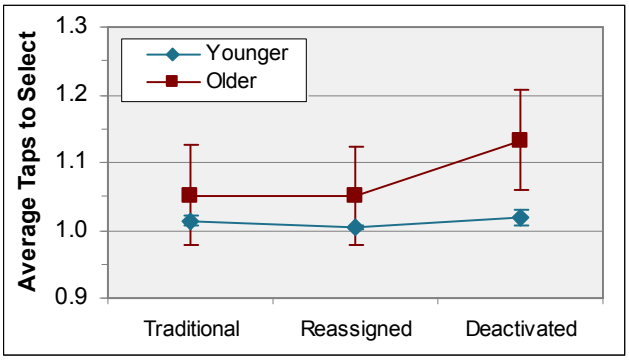

Figure 2: Average number of taps needed to select an item, by age group (*N*=24). Error bars represent 95% Confidence Intervals.

Vertical Distribution of Taps

Figure 3 (left) 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 [5] (see Figure 3 (right)). Most notably, the data in [5] 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.

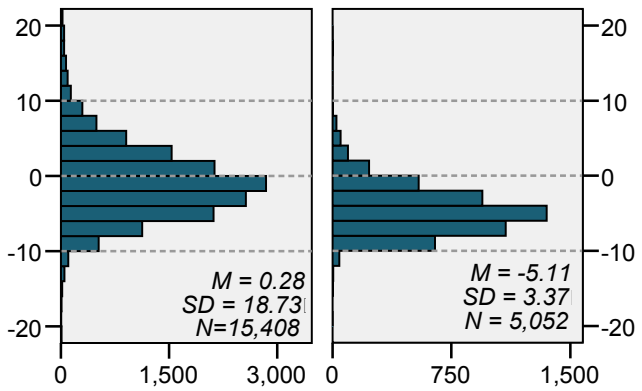

Figure 3: Histograms of the vertical position of taps (occurring on the target item, and the lower/upper half of the item above/below) for this study (left), and the study reported in [5] (right).

Self-reported Measures

Most participants reported that they were unable to detect any differences between the conditions. Thus, not surprisingly, a Chi-square analysis on the frequency with which each menu condition ranked first on our subjective measures revealed no significant differences. This result is not entirely unexpected. Many individuals have compact tap distributions, and make infrequent use of the top edge (correctly or erroneously). For these individuals, the three conditions are almost identical. However, we were surprised that even participants who did use the top edge were unable to distinguish between conditions; rather they tended to attributed performance differences to practice, fatigue, or particular word combinations.

However, analysis of the spontaneous comments made throughout the study provided more insight. Six participants (4 older) commented negatively on DE, while only one commented negatively on TE, and no one commented on RE. The comments on DE reflected confusion and disruption. As one individual put it, "[It] really throws you off when you have to click more than once." Another described it as, "I kept thinking I had tapped the right thing and then had to go back." Others were less specific, making comments such as "[DE] seems to be a little more awkward," and "[With DE, it] was harder to make selections." Some comments reflected a misconception that more or more sustained force was required in DE. 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, is not surprising when we consider the differences in the tap distributions between this study and [5]. One possible reason for the divergence, is differences between the tasks used in the study. In [5], 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 [5] 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 [5] 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.

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 [5], 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. Cockburn, A., Gutwin, C., and Greenberg, S. (2007). A predictive model of menu performance. *Proc. CHI '07,* 627–636.
2. Greenstein, J.L. (1997). Pointing devices. In M.V. Helander, T.K. Landauer, & P.V. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd Ed.). Amsterdam: Elsevier, 1317–1348.
3. Heath, M., Roy, E.A., & Weir, P.L. (1999). Visual-motor integration of unexpected sensory events in young and older participants: A kinematic analysis. In *Developmental Neuropsychology*, 16(2): 197–211
4. ISO (2000), Ergonomic requirements for office work with visual display terminals (VDTs)—Part 9: Requirements for non-keyboard input devices, (ISO 9241-9:2000). ISO.
5. Moffatt, K., and McGrenere, J. (2007). Slipping and Drifting: Using Older Users to Uncover Pen-based Target Acquisition Difficulties. *Proc. ASSETS '07*, (To Appear).