Slipping and Drifting: Using Older Users to Uncover Pen-based Target Acquisition Difficulties

Karyn Moffatt and Joanna McGrenere
Department of Computer Science, University of British Columbia
ICICS/CS Building, 201-2366 Main Mall, Vancouver, BC, CANADA V6T 1Z4
{kmoffatt, joanna}@cs.ubc.ca

ABSTRACT

This paper presents the results of a study to gather information on the underlying causes of pen-based target acquisition difficulty. In order to observe both simple and complex interaction, two tasks (menu and Fitts' tapping) were used. Thirty-six participants across three age groups (18–54, 54–69, and 70–85) were included to draw out both general shortcomings of targeting, and those difficulties unique to older users. Three primary sources of target acquisition difficulty were identified: slipping off the target, drifting unexpectedly from one menu to the next, and missing a menu selection by selecting the top edge of the item below. Based on these difficulties, we then evolved several designs for improving pen-based target acquisition. An additional finding was that including older users as participants allowed us to uncover pen-interaction deficiencies that we would likely have missed otherwise.

Categories and Subject Descriptors: H.5.2 [Information interfaces and presentation]: User Interfaces — *Input devices and strategies.*

General Terms: Design, Experimentation, Human Factors.

Keywords: Target Acquisition, Pen-based Interaction, Tablet PC, Older Users, Inclusive Design, Universal Usability.

1. INTRODUCTION

Direct pen-based input takes full advantage of hand-eye coordination, and offers a familiar form of interaction [6]. These benefits have been shown to be particularly beneficial for older adults [2, 3]; however, until relatively recently the only available direct input pen device was the light pen. Despite its advantages, the pen's high cost relative to a standard mouse, coupled with the fatigue associated with using a pen on a vertical display surface (i.e., a standard monitor), left it largely unadopted. With current-day Tablet PCs and stylus-based PDAs gaining popularity, it now seems that pen input is finally in a position to succeed. However, in our own work designing mobile technology for older and motor-impaired individuals [17], we have informally observed many struggling with target acquisition (e.g., selecting an icon or a menu item) using a stylus. These observations have motivated us to gain a better understanding of the challenges inherent to pen interaction, and to ascertain the extent to which age is a factor.

Although there has been a great deal of research aimed at developing improved target acquisition techniques, including a sizeable amount directed specifically to the pen [1, 10, 16, 18], room for improvement remains: many users still experience difficulties, and standard point and tap (i.e., selection by (i) tapping down, (ii) possibly moving the pen, and (iii) tapping up, with selection determined based on the location of the tap up) remains the dominant technique. We note three limitations that span the majority of pen-based target acquisition research: (1) the narrow focus on young-healthy adults, who can more easily adapt to different techniques, (2) the focus on evaluation with a single, typically highly-constrained, task, and (3) the focus on designing and evaluating novel techniques over developing a deeper understanding of how users manage basic tapping.

In terms of the first limitation, there are many parameters, including a user's sensory and motor ability, that are likely to affect target acquisition and manipulation skill. Thus, a broader perspective can be gained by examining a range of users and abilities. Since aging leads not only to reduced capability, but also to greater variability, older adults may provide especially rich information. Such information could lead to improvements that benefit not only the older demographic but perhaps younger ones as well.

The second limitation relates to restricting the evaluation of techniques to one task. Although not exclusively used, the standard for comparing interaction techniques is a Fitts' tapping task [7, 20]. Its main advantage is that it provides well-understood measures of speed and accuracy. However, it only reflects very simple interaction with a single isolated target. Real world applications require much more complicated forms of interaction. But more complex interactions are less well understood, harder to analyze, and often lead to less clear conclusions. Thus, we believe it is important to include multiple tasks to capture both concrete comparative measures, and complex interaction.

With respect to the third limitation, focusing on developing new techniques and evaluating them against the status quo (point and tap) has led research towards gross measures of overall speed and accuracy. These measures provide comparative data about which technique is superior, but when the results are inconclusive, they do not give us the richness of information required to know why. For example, it can be unclear whether the problem was with the initial homing in on the target, or with staying on the target while completing the selection. Or, it can be unclear whether the technique was unintuitive or too cognitively complex, or just required more practice or training. As a result, we do not know which limitations to address, or where innovation is still needed.

The work described in this paper attempts to fill this niche by gathering information on the underlying causes of target acquisition difficulty. We used two tasks (a Fitts' tapping task and a menu task) to provide a range of interactions to examine. We involved users from three different age groups to help us understand both the

ASSETS'07, October 15–17, 2007, Tempe, Arizona, USA.

general shortcomings, and those unique to older users. Specifically, the goal of this work is threefold:

1) to perform a detailed analysis of the types of difficulties users encounter while tapping to acquire targets,
2) to determine if these difficulties vary over task situations, and
3) to determine if these difficulties vary in terms of their nature and severity with age.

The results revealed three primary sources of target acquisition difficulty: slipping off the target, drifting unexpectedly from one menu to the next, and missing a menu selection by selecting the top edge of the item below. Slipping mostly affected older users, while drifting and missing just below the target item impacted both younger and older users, alike. An additional finding was that including older users as participants allowed us to uncover pen-interaction deficiencies that we would likely have missed otherwise. Drifting and missing just below were not behaviors we predicted; rather our observations of the older users during the experimental sessions prompted us to investigate them, thus revealing their general impact.

From our results, we evolved several designs for improving pen-based target acquisition to address the shortcomings we identified. We discuss these designs in detail, taking into account changes across the lifespan.

2. RELATED WORK

We begin our coverage of the literature with an overview of the general effects of aging on motor skill to highlight the reasons for age-related differences in targeting ability. We then describe previous research investigating novel pen techniques. Because the mouse has been the subject of much more attention historically, we briefly review work in that area, focusing specifically on techniques that may have applicability to the pen.

2.1 Effects of Aging on Targeting Ability

There is a considerable body of literature that has examined the negative effects of aging on the aspects of motor control that pertain to general targeting ability, both with respect to mouse use and interaction in the physical world. Research has found that older adults use different strategies concerning the speed-accuracy tradeoffs involved in movement control. They tend to be more conservative, and make more corrective sub-movements [24]. Older adults have also been found to cover less distance with their primary movement [13], to make many more sub-movements en route [12], to make less smooth movements [27], and to have difficulty staying on the target while clicking [19]. In addition, slower selection speeds have been attributed to lower peak velocities [12, 13], longer deceleration phases [13], and more pauses while homing in on the target [12].

2.2 Pen-based Acquisition Techniques

There has been a small body of work devoted to developing improved pen-based target acquisition techniques, with modest results. Novel techniques often only slightly out-performed standard tapping, or only in specific constrained situations.

Ren and Moriya [18] compared six selection strategies and found that for targets smaller than 1.8 mm that Slide Touch (selection at the moment the pen-tip first touches a target after landing) was best in terms of speed, accuracy, and participant preference. However, they cautioned that this technique would not be suitable for dense displays, for which they recommended either Direct On (which relies on the pen landing on the target), or Direct Off (standard tapping). They further noted that both Direct On and Direct Off require good hand/eye coordination. Thus, it is unclear whether they are suitable for older users.

Mizobuchi and Yasumura [16] compared tapping to circling for a multi-target selection task. They hypothesized that circling would be faster and more accurate than tapping, but found that it was only better in the specific situation where targets formed a cohesive group with low shape complexity.

Accot and Zhai [1] compared tapping to crossing and found crossing was at least as fast and had similar accuracy. Although not outright better than the status quo (tapping), they concluded crossing is a viable interaction technique and suggested there may be special situations in which it has specific advantages, including support for elderly or motor impaired users. However, they did not follow up on these ideas.

As already mentioned, a notable limitation of the above body of work is that it is all based on the evaluation of young healthy adults. One exception in this domain is the work by Hourcade and Berkel [10], which compared the accuracy performance of 18–22, 50–64, and 65–84 year olds for tapping and touching (selection if the pen touches the target at any time before tap up). They found that for the smallest target size examined (3.8 mm), the oldest group was somewhat more accurate using touch, but noted that some users reported touching to be more tiring.

2.3 Mouse-based Acquisition Techniques

In this section, we discuss mouse techniques that may also be applicable to the pen. Specifically, we exclude techniques that manipulate the ratio between mouse and cursor movement (for an example, see Sticky Icons [26]), as the direct mapping between the pen and the cursor makes these techniques inappropriate.

Most work on improving mouse interaction has focused on easing cursor positioning. One technique that has shown some success is to dynamically expand targets as the cursor approaches [4, 15]. However, this requires surrounding targets to either move or be occluded, and some research suggests that older adults may be incapable of adapting their initial motor response to take advantage of the increased target size [9]. Moreover, the ability to remain outside the detectable range of the tablet until late in the interaction, may delay expansion and hinder the support provided.

A related approach, area cursors [8, 11, 26], has been shown to have some promise specifically for older adults [26]. Area cursors replace the standard single hot spot cursor with a cursor that covers a larger area. In order to support selection from multiple proximate targets, several ideas have been proposed [8, 26] Most notably, Bubble Cursor [8] dynamically resizes the cursor such that only one target is selectable.

More recently, an approach has been proposed that does not deal with easing the initial positioning of the cursor, but rather with keeping it steady once it is in place. Steady Clicks [23] is intended to help individuals who find it difficult to hold the mouse still while clicking. It prevents slipping by freezing the cursor at the mouse down position until either the button is released (resulting in a steadied click) or the mouse is moved beyond a freeze threshold (returning the mouse to normal operation). An evaluation of Steady Clicks found that it enabled participants to select targets using significantly fewer attempts, and for those with the highest slip rates, to select them significantly faster.

3. EXPERIMENTAL METHODOLOGY

To address the goals outlined in the introduction, we performed a multi-task evaluation of pen-based target acquisition across multiple age groups. Specifically, we had the following hypotheses for this study:

H1. Speed and accuracy will decrease as age increases.
H2. Age will impact the types of errors made.
H3. Task will impact the types of errors revealed.

3.1 Participants

Thirty-six participants from three age groups (12 each) were included in the study:

- *Young:* 18–54 (5 male, 7 female; mean age 31.7)
- *Pre-old:* 55–69 (4 male, 8 female; mean age 62.1)
- *Old:* 70–85 (3 male, 9 female; mean age 76.3)

The justification for these groupings rests on the age related changes that occur in cognition [5], notably that higher cognitive function remains relatively stable up to about age 55, after which there is a small decline, followed by a much steeper one after 70.

All participants were right-handed and free of diagnosed impairment to their right hand, and had normal or corrected-to-normal eyesight. To control for any biases between age and previous Tablet PC experience, we limited participation to individuals with no Tablet PC experience and no or limited PDA experience. None of our participants had previously owned a PDA, but some reported having tried a friend's, or participating in other studies involving PDAs. Furthermore, within and across each age group, participants had a wide range of computer experience, from novice to expert.

Additionally, we screened participants using the North American Adult Reading Test [21] to ensure sufficient English fluency to follow our instructions. Three participants from the 18–54 age group did not meet our minimum criterion (not included in the 36 above). They were allowed to finish the study, but their data was not included in our analysis.

3.2 Motor Skill

Because motor skill is known to be one of the main factors accounting for age-related differences in targeting ability [19], we administered three standardized tests to gather data about our participants' motor abilities. As a measure of perceptual speed we used the Digit Symbol Substitution Test [25], as a measure of motor-coordination we used the Purdue Pegboard test [22], and as a measure of steadiness we used a 9-hole steadiness tester [14].

3.3 Task

To gain a better understanding of how task might affect targeting ability, we used two tasks in this study: a multi-dimensional Fitts' tapping task [20] and a menu selection task. The tapping task was selected because it is the gold standard for evaluating input techniques, and provides well understood measures of speed and accuracy. The menu task was selected because it provides a greater degree of realism. Additionally, we believe it may require slightly more cognitive effort. We hypothesized that these factors might affect performance, especially accuracy.

Multi-dimensional Tapping Task: For the tapping task (Figure 1a), each trial started with a single blue 'start' circle in the middle of the screen. Once tapped on, the 'start' circle faded to a light grey, and a red 'target' circle appeared. The next tap ended the trial, regardless of whether the tap successfully acquired the target or not. An audible beep provided feedback when the trial was unsuccessful. Participants were instructed to tap on the target circle as quickly as possible while remaining accurate.

Target width, amplitude (i.e., distance to the target), and angle (of motion) were varied. Targets were presented at three diameters: 14, 28, and 42 pixels (3.3, 6.7, 13.4 mm); three amplitudes: 120, 240, and 360 pixels (28.8, 57.6, 86.4 mm); and eight angles: 0, 45, 90, 115, 180, 225, 270, and 315 degrees. The task was broken into four consecutive blocks with an enforced one minute break between blocks. Each block consisted of 72 randomly ordered trials representing one of each possible combination of width, amplitude, and angle.

Menu Task: Each trial in the menu task (Figure 1b) also began with a single blue 'start' circle. When the participant tapped on it, it faded to a light grey (as in the tapping task), and a prompt appeared above the menu bar indicating which menu-item pair was to be selected. The trial ended when the participant successfully selected a menu item, regardless of correctness. An audible beep provided feedback when the wrong item was selected. Again, participants were instructed to make selections as quickly as possible while remaining accurate.

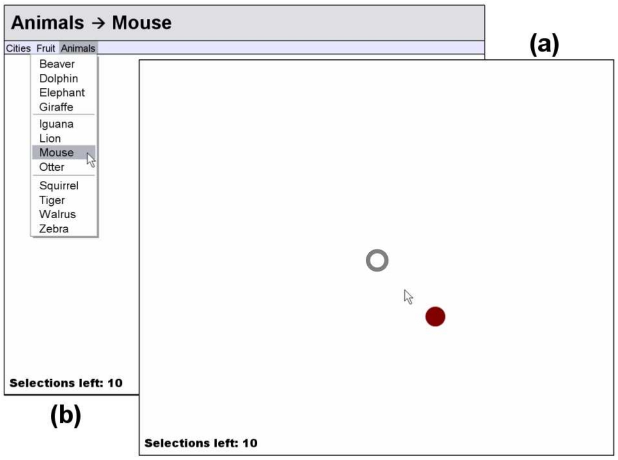

Figure 1: Screen shots mid-trial: (a) the multi-dimensional tapping task, (b) the menu task (note, the 'start' target is occluded by the screen shot of the tapping task).

The study used three menus grouped by category (Animals, Fruit, and Cities). Each menu contained 12 alphabetically-ordered items, separated into three even groups of four menu items. A length of 12 items was chosen by taking the average menu length of three common applications: FireFox 1.5, Microsoft Word 2003, and Adobe Reader 7.0. Each item was 20 pixels (4.8 mm) high. As with the tapping task, there were four blocks of trials with an enforced one minute break between blocks. Each block consisted of 36 trials representing one selection of each menu-item pair, ordered randomly.

3.4 Measures

We included measures of speed and accuracy. For speed, we measured trial time as the time from the pen up action off the 'start' circle to the pen up action that ended the trial. We included several measures of accuracy, as we were interested in not only the numbers of errors but also the types.

For the tapping task, we modified for pen interaction the classification described by Trewin, Keates, and Moffatt [23], based on a study of older and motor-impaired mouse users [12]:

- *Slips:* the pen lands on target, but slips off before it is lifted.
- *Near misses:* the pen lands off target, and lifts at a distance less than 50% of the target radius away (from the target boundary).
- *Not-so-near misses:* the pen lands off target, and lifts between 50% and 100% of the target radius away.
- *Other (or unclear):* the pen lands off target, and lifts between 100% and 200% of the target radius away.
- *Accidental taps:* the pen lands off target, and lifts at a distance greater than 200% of the target radius away.

Note the key difference between slips and misses is whether the pen initially lands within the target. Near and not-so-near misses are interpreted as being intentional taps directed at the target. Accidental taps are interpreted as unintentional taps made en route to the target. Other taps are those where the intent is unclear.

For the menu task, we also considered slips and misses as distinct error types, but the above sub-categorization of misses does not apply to this task. Instead, we specify two categories of misses in addition to slips:

- *Slips:* the pen lands on the target item, but slips off before lifting, or the pen lands on the menu head, and slips off (resulting in selection of the top menu item).
- *Correct-menu misses:* the pen lands on (and lifts from) an incorrect item of the correct menu
- *Incorrect-menu misses:* the pen lands on (and lifts from) an item of an incorrect menu

For correct-menu misses, we further recorded the proximity to the correct item. In both tasks, for slips we recorded the distance traveled between the pen down and up (i.e., the distance slipped).

3.5 Design

This experiment used a mixed design with two counterbalanced tasks (menu, tapping). Because the structure of each task was different, we present them here as two separate sub-designs.

The tapping task used the following design: 3 (age groups) x 4 (blocks) x 3 (target widths) x 3 (target amplitudes) x 8 (angles). The presentation order of each combination of target width, amplitude, and angle was randomized.

For the menu task, the design was: 3 (age groups) x 4 (blocks) x 3 (menus) x 12 (items). Each participant was assigned one of the six possible menu order permutations at random, and the presentation order of the menu-item pairs was randomized.

Age was the only between-subjects factor. Thus, each participant completed 288 trials in the tapping task, and 144 in the menu task.

3.6 Procedure

The experiment was designed to fit into a single 120 minute session. All participants finished in between 75 and 120 minutes.

We began with the motor tests, which were given in the order: Digit Symbol Substitution, Purdue Pegboard, and Steadiness. Next was the North American Adult Reading Test. We then introduced the Tablet PC. Participants were asked to complete the first 8 steps of "Get Going with the Tablet PC", the native tutorial that introduces new users to the Tablet PC and using the pen. (Steps 9–17 concern text input and were not relevant to the study.) Once the participant finished the tutorial, we presented the first task (either menu or tapping). After completing all four blocks of the first task, participants were given a brief questionnaire about their background and computer familiarity. They then completed their second task.

We note that beyond the instructions given in the tutorial, participants were not instructed to use the pen in any particular manner. We explicitly wanted to observe how participants would naturally approach the task.

3.7 Apparatus

All experimental conditions were run on 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 standard inductive pen that came pre-packaged with the machine was used for all computer tasks; however, the button on the side of the pen was removed to ensure participants did not accidentally use it as it was not required for the study tasks. 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 (approximately 35 degrees from horizontal). We chose this setup, because pilot studies indicated difficulty viewing the screen when it was horizontal on the table, and we felt asking participants to hold the tablet would unfairly disadvantage the older groups. Participants were encouraged to adjust the position of their chair and the placement of the stand for comfort.

4. RESULTS

In this section we present our results. Unless otherwise noted, Bonferroni corrections were used for all post-hoc pair-wise analyses. Where Levene's test revealed significant heterogeneity of error variance, we used a Welch's ANOVA for testing the main effect and Games-Howell corrections for post-hoc pair-wise comparisons. Both are robust against unequal error variances. Finally, in all our repeated measures analyses, sphericity was an issue; thus, Greenhouse-Geisser corrections were used.

Not surprisingly, analysis of the motor tests confirmed overall motor decline with age. Unexpectedly however, we did not detect differences between the pre-old and old groups, which foreshadows a lack of significant differences between these age groups in our analyses of the target acquisition tasks.

As a final note, in some of our analyses we encountered outliers, which we define as scores more than two standard deviations from the mean. Analyses where outliers have been removed are noted.

4.1 Tapping Task

Speed decreased with age. As H1 predicted, older users were slower. (A significant effect of age was revealed by a one-way ANOVA on median trial time, $F(2,32) = 4.255$, $p = .023$, $\eta^2 = .210$, which excluded 1 outlier from the young group). However, post-hoc pair-wise comparisons only detected a difference between the young and old groups ($p = 0.019$), though the trends (as shown in Figure 2a) did indicate a general slowing with age.

Everyone misses, but older users also slip. Although, we initially intended to examine accuracy using five categories of error, slips and near-misses accounted for 90% of the errors observed (with no other category accounting for more than 5% percent). Thus, we focused our analyses on them.

Previous research [10, 19] has found interactions between age and target size for tapping accuracy measures. Thus, we performed repeated-measures analyses (for target width and age) on each of slips and near-misses. We found that while slipping clearly increases with age, near-missing remains relatively constant, as shown in Figure 2b. (There was a main effect of age for slipping, $F(2,15.92) = 3.860$, $p = .043$, $\eta^2 = .185$, but post-hoc pair-wise comparisons did not produce any significant results. In contrast, there was no effect of age for near-misses). As we would expect, we also found main effects of target width for both slips ($F(1.18,39.03) = 18.341$, $p < 0.001$, $\eta^2 = .357$) and near-misses ($F(1.17,38.52) = 40.337$, $p < 0.001$, $\eta^2 = .550$), indicating that both these errors increased as targets got smaller.

In addition, we also found that older users did have greater difficulty with slipping from smaller targets than larger ones, whereas slipping was infrequent for young users across all widths. However, there was no such effect for near-misses. (There was an interaction between age and width for slipping, $F(2.37,39.03) = 5.874$, $p = 0.004$, $\eta^2 = .263$, but none for near-misses.)

It is also interesting to note that slips were overall relatively short. On average they were 12 pixels (2.9 mm) long (median: 9 pixels, standard deviation: 7), and while the largest slip was 49 pixels (11.8 mm), over 90% were less than 25 pixels (6.0 mm).

To summarize the tapping task results, older users were indeed slower, and combining misses and slips, they made many more errors, supporting our hypothesis that speed and accuracy would decrease with age (H1). In addition, we saw that while missing itself remained relatively constant across age, slipping clearly increased, supporting our hypothesis that older adults do not just make more errors, they make different errors (H2).

4.2 Menu Task

Although not one of our planned measures, a dominant pattern observed during the sessions was that of accidentally drifting to the adjacent menu. As with a mouse, moving the cursor over a menu while one is open causes the open menu to switch. However, on the Tablet PC, this occurs regardless of whether the pen is touching the screen or hovering above it. Moreover, when using a pen, the hand often occludes menu items, requiring users to lift their hand up and away to see. Depending on the distance lifted and the angle of this action, the pen may accidentally drift to the next menu. In consideration of this dominant behavior, we chose to consider drifting in our analysis.

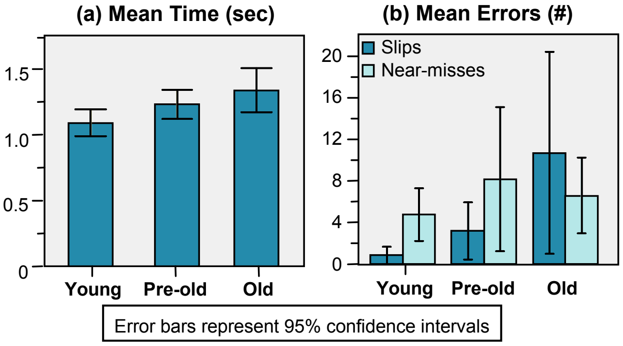

Figure 2: Tapping task results, by age group: (a) mean trial time ($N = 35$), (b) mean errors for slips & near misses ($N = 36$).

Drifting impeded task performance. Participants were often confused when drifting occurred: they reported not knowing why the wrong menu was open and not being sure how to proceed. Many participants would attempt to re-open the desired menu; however, when the pen neared that menu, the hovering would trigger it to re-open. But, the users would not notice, and they would tap on it anyway, which actually resulted in it closing. Needless to say, this led to considerable confusion.

Thirty-five out of 36 participants drifted at least once, and 31 responded to a drift by re-tapping (and thus closing) the target menu at least once. Moreover, performance was significantly impeded by drifting in terms of slower trial times, as can be seen in Figure 3a. However, overall accuracy was not affected. (Paired t-tests on the 35 participants who drifted for both speed and overall accuracy revealed a significant effect on speed, $t = 5.115$, $df = 34$, $p < 0.001$, but no effect on accuracy, $p = 0.164$).

Older users drifted more. Although drifting affected all age groups, pre-old and old participants drifted more than participants from the young group. Figure 3b shows mean drifts by age group. (A one-way Welch's ANOVA excluding three outliers, 2 young and 1 old, revealed a significant effect of age, $F(2,16.51) = 9.351$, $p = .002$, $\eta^2 = .221$. Post-hoc pair-wise comparisons further showed significant differences between the young and both the pre-old, $p = 0.038$, and the old, $p = 0.008$ groups.)

We note that this further supports our hypothesis that accuracy would decrease with age (H1). Although drifting did not have an explicit affect on overall task accuracy, it does represent a difficultly in accurate interaction.

Drifting did not decrease with learning. It is also interesting to note that drifting behavior did not improve over the course of the menu-task; i.e., participants did not get used to the designed interaction. (A repeated-measures ANOVA on block, with age as a between-subjects factor was not significant for either the main effect of block or the interaction between block and age).

Drifting aside, older users were still slower. In consideration of our findings for drifting, we performed our analysis of age on trial time based solely on drift-free trials to determine if there was an effect independent of that caused by age-related differences in drifting behavior. Comparing only drift free trials, both the pre-old and the old were significantly slower than the young group, which is also supported by Figure 3a. (A one-way ANOVA on age for drift-free trials revealed a significant main effect of age, $F(2,32) = 4.678$, $p = .017$, $\eta^2 = .226$, while post-hoc pair-wise comparisons revealed that the youngest group was significantly faster than both the pre-old, $p = 0.04$, and the old, $p = 0.03$).

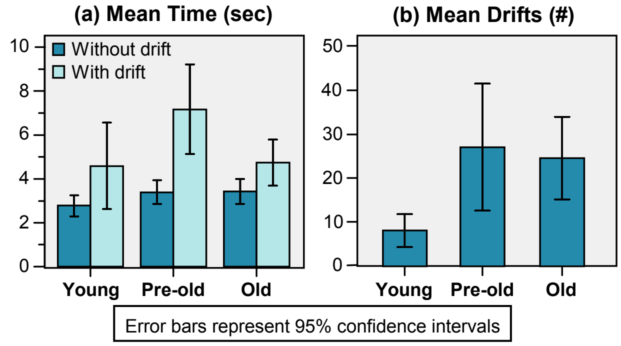

Figure 3: Drifting results, by age group: (a) mean trial time, without & with drift ($N = 35$), (b) mean drifts ($N = 33$).

Few errors overall, but misses were the main source. Overall errors were lower in the menu condition than we expected. Of 5184 trials (36 participants x 144 trials) there were only 135 errors, and 60 (44%) of these were committed by 3 individuals, one in each of the three age categories. If we exclude these individuals, the overall error rate was only 1.4%, which is less than the 4% generally expected in any Fitts'-like experiment. As we were including a much broader age range, we expected the error rates to be even higher. Thus, we did not attempt to make any age comparisons for errors. Instead we examined our data for general trends in the types of errors observed.

Excluding the three outliers mentioned above, correct-menu misses and slips accounted for 70% and 29% of the errors, respectively. Incorrect-menu misses were exceptionally rare, accounting for only 1% of the errors. Slip length was comparable to what we observed for the tapping task. Slips were on average 10 pixels (2.4 mm) long (median: 8 pixels, standard deviation: 5). However, the maximum slip length (22 pixels) was much shorter.

Missing occurred just below. We further analyzed the correct-menu misses based on their proximity to the target item. Across all 36 participants, 58 of 71 (82%) correct-menu misses were on the item below the target. If we look at the vertical distribution of tap ups, we see that 44 of these misses (62% overall) were on the top two pixels of the item below (i.e., 10% of the item height, or 0.5 mm). In contrast, only 4 trials involved a tap up on the top two pixels of the target item itself. In other words, a tap up occurring on this top region was 11 times more likely to be intended for the item above the selected item, than the selected item itself. Figure 4 shows a histogram highlighting how the distribution of tap ups was offset such that it was much more likely for a tap to occur on the top of the item below than the top of the targeted item.

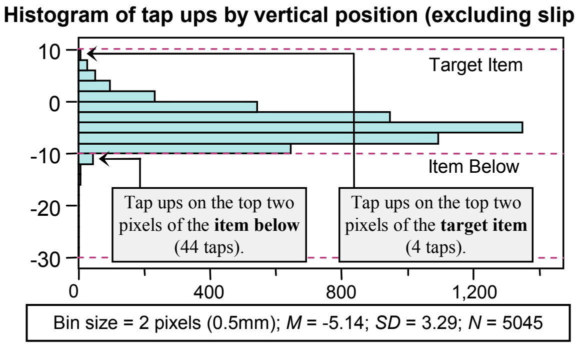

Figure 4: Histogram of the vertical position of tap ups occurring on the target item and the item below ($N = 5045$).

To recapitulate the menu task results, an unexpected dominant pattern observed in the menu task was drifting to the next menu. Although, it did not affect overall task accuracy, it did have a significant negative impact on speed. Drifting was not unique to older users, affecting everyone. However, older users did drift disproportionately, and were slower, even when the effects of drifting were factored out. Furthermore, although overall errors were low in the menu condition, missing the target item by selecting the topmost region of the item below was a major source of the errors observed.

4.3 Summary of Results

In this section, we bring together our results from each task and discuss how they contributed to confirming our hypotheses.

H1: Speed and accuracy will decrease as age increases. This hypothesis was supported. In both tasks, we saw overall main effects of age on trial time. In the tapping task, we also saw an overall decline in accuracy with age. Although we did not see differences in accuracy in terms of overall error rates for the menu task, we did see that older users drifted more. Drifting, though not formally a task error, is indicative of greater difficulty accurately performing the interaction.

H2: Types of errors made will be impacted by age. In the menu task, there were too few errors to examine the effects of age for this hypothesis. However, the tapping task clearly provided support: we saw that while there was no effect of age on missing, slipping clearly increased with age.

H3: Task will impact the types of errors revealed. Each task informed us of different types of targeting difficulties, confirming this hypothesis. Because the tapping task was the simplest task, it was best for uncovering low-level interaction difficulties (e.g., slips and misses). In contrast, the menu task was more realistic and revealed difficulties pertaining to combinations of widgets and interactions (e.g., drifts and menu closes). Thus, including both tasks led to richer findings.

5. IMPLICATIONS AND FUTURE WORK

Our results revealed three primary sources of target acquisition difficulty: one, specific to older users, and two, which apply generally to all ages. In this section we discuss those difficulties and suggest directions for providing better interaction support. In some cases, the solution is relatively straightforward, while others warrant deeper investigation. In the final part of this section, we discuss how including older users helped identify target acquisition difficulties for both older and younger users, and its implications for other research.

5.1 Support for slipping

Slipping was a problem for our older users, a result that is consistent with research on the mouse [12, 19]. It is interesting to note that with a mouse, however, slipping has generally been attributed to an inability to hold the mouse steady while clicking. As tap selection does not have an analogous button clicking action, it is surprising it was also a problem here.

One approach to preventing pen-based slip errors would be to adapt Steady Clicks, which we described in Section 2.3, to work with a pen. But, this is not without challenges. Steady Clicks assists the user by freezing the cursor at the mouse down position. However, the direct mapping between the cursor and the tip of the pen makes this technique less ideal for pen interaction. On the other hand, slips were generally short (on average 10–12 pixels, 2.4–2.9 mm), and our participants appeared mostly unaware of them (none reported noticing slips, and many reported confusion over errors they thought were accurate). One possibility would be to not manipulate the cursor, and handle the freezing internally.

Another problem with freezing is that it may impede other pen targeting strategies. In previous work [17], we informally observed individuals using inactive space around targets as a "landing zone" from which they could drag the stylus to the desired target. This compensation strategy would be in direct conflict with freezing, which relies on a correct tap down. Although we did not observe any use of this landing strategy in our current study, it is likely due to the nature of our tasks: participants did not have to keep trying until they succeeded, as they did in our previous study, and so they may have been less motivated to explore different approaches.

One way of overcoming some of the limitations described above would be to combine freezing with area cursors. With an area cursor [8, 11, 26], it is not the tip of the cursor that defines the object selected, but rather a larger selection area centered on the tip. This small degree of separation may provide the flexibility needed to allow a natural form of freezing. On pen down the area cursor would freeze, but the pen would remain free to move within the area cursor. Freezing would break, if the pen crossed the edge of the cursor, returning the interaction to normal.

Area cursors have already been shown to be helpful for older adults with the mouse [26], and, although this approach would not directly support the land and drag strategy we observed previously, it may circumvent the need for it, as the main advantage of an area cursor technique is that it reduces the precision needed to position the cursor.

We note that this interpretation of freezing is slightly different from the implementation used for Steady Clicks. In Steady Clicks, the freeze threshold was based on empirical data of slipping behavior; here, we propose using the radius of the area cursor. As previously mentioned (Section 2.3), the size of the area cursor must be chosen to ensure proximal targets are selectable. We propose using Bubble Cursor [8], which dynamically resizes the cursor such that only one target is selectable, while maximizing the area cursor size (and thus, the slip threshold).

5.2 Support for Drifting

In contrast to slipping, drifting was a problem for young and old users alike. The simplest way to prevent drifting would be to turn off the ability to switch menus by hovering. Although this would clearly fix the problem, it may have implications for other aspects of menu interaction. In our study, participants always knew exactly which menu contained the target item (both the menu and the item were specified in the task prompt). It may be the case that when the user is browsing through menus for an item, the ability to trigger menus while hovering above the screen is useful (because the hand otherwise occludes the menu).

Another approach would be to introduce some form of delay to the switch. This could either be done with a time delay such that if the pen is only briefly hovering over another menu, it does not switch, or by implementing a distance threshold such that the menu does not switch until the pen has covered some percentage of the menu head. The rationale for using a distance threshold is that when browsing, right-handed users often bring the pen across towards the rightmost edge to read the menu, whereas with drifting, they stay more towards the leftmost edge.

Whether the best approach is to turn hovering off or to use one of the suggested delay mechanisms requires further investigation. It could be that the benefits of preventing drift outweigh any cost that would result, or that it is not possible to distinguish between accidental drifts and intentional hovers by either time or distance.

As for the ability to close menus by re-tapping on the menu head, it seems likely that turning this behavior off would benefit most users, with minimal negative impact. Menus can also be closed by tapping on inactive screen space, which is familiar to most users.

5.3 Support for Missing Just Below

Across all age groups, the majority of correct-menu misses occurred at the very top of the item below the target, while very few correct selections involved the corresponding region of the target item itself. We suggest two possible ways of modifying the interface to prevent these errors.

The first is to shift the target region of each item (the motor space) by two pixels (i.e., 10% of the menu item height, or 0.5 mm), while leaving the visual appearance unchanged, such that selections occurring on this top region of an item are interpreted as selections of the item above. In our data, this would remove 44 errors, while introducing only 4 new ones.

The second approach would be to deactivate the top two pixels of all menu items, such that taps in this region would be ignored, much like taps on menu separators. In this case, all 44 errors would be removed. However, from our observations, we noticed that users typically do not wait to see if their taps are successful. On the few occasions where taps did not register (e.g., because the user hit an actual menu separator), we noticed participants try to move on to the next trial, subsequently realize they had not finished, and then go back to try again. Thus, although this approach would reduce the greatest number of errors, there are potential negative implications for speed.

Further investigation is required to determine whether it is better to introduce a small number of new errors (as with the first approach), or to risk delaying the user with unregistered taps (as with the second approach). It is also worth noting that both of these approaches not only potentially help prevent missing just below errors, but could also eliminate short slipping errors: 12 of the 60 slip errors in the menu task involved a tap up on the top two pixels of the item below.

5.4 Learning from Older Users

One of our key results was that the behavior of older participants enabled us to uncover difficulties common across the lifespan. The most prominent example of this was drifting, although it also applies to missing just below. Drifting and missing just below were not behaviors we predicted; rather our observations of the older users during the experimental sessions prompted us to investigate them in detail. It was only upon closer examination of the data that we discovered that they impacted all participants.

For drifting, the reason for our initial bias was that the effect was more pronounced in the older population. Because they moved more slowly overall, it was easier to follow their actions and catch inefficiencies. Also, they were more overt about their interactions, making comments such as, “Now what happened here?” (Upon realizing the wrong menu was open.) Or, “No. I want that one!” (Before meticulously re-tapping on the desired menu, causing it to close.) Younger users, on the other hand, recovered more quickly and were considerably less vocal about their experience.

For missing just below, the effect was more subtle. Articulated confusion by the older participants over incorrect selections similarly prompted us to more closely investigate the vertical tap distributions leading to the discovery that the majority of miss errors occurred on the very topmost region of the item below.

As a final point about age, we note that our use of three distinct age groups did not impact the results as we had anticipated. Significant differences were not often found between the young and the pre-old groups, and no differences were found between the pre-old and the old. Further investigation is required to explore alternate groupings.

6. CONCLUSION

This paper presented the findings of an experiment designed to gather information on the underlying causes of target acquisition difficulty, with a particular focus on how age affects targeting ability. Specifically, we performed a detailed analysis of the types of errors incurred on two tasks across three age groups. From this analysis, we identified three sources of pen-based target acquisition difficulty: slipping, drifting, and missing just below. Slipping was unique to our older users, while drifting and missing just below affected all age groups. To address these difficulties, we evolved several detailed design possibilities for improving pen-based target acquisition that take into account changes across the lifespan. Additional research is needed to implement and evaluate these designs. An additional finding was that including older users as participants did allow us to uncover pen-interaction deficiencies that we would likely have missed otherwise.

7. ACKNOWLEDGEMENTS

This research was made possible by support from the Canadian Institutes of Health Research (CIHR) and the Natural Sciences and Engineering Research Council of Canada (NSERC).

8. REFERENCES

[1] Accot, J. & Zhai, S. (2002). More than dotting the i's—foundations for crossing-based interfaces. In *Proc. CHI '02*, 73–80.

[2] Charness, N., Bosman, E.A., & Elliott, R.G. (1995). Senior-Friendly Input Devices: Is the Pen Mightier than the Mouse? Presented at the *103rd Annual Convention of the APA*.

[3] Charness, N., Holley, P., Feddon, J., & Jastrzembski, T. (2004). Light pen use and practice minimize age and hand performance differences in pointing tasks. In *Human Factors*, 46(3):373–384.

[4] Cockburn, A. & Firth, A. (2003). Improving the Acquisition of Small Targets. In *Proc. HCI '03*, 181–196.

[5] Craik, F.I.M., & Salthouse, T.A. (Eds.) (1992). The Handbook of Aging and Cognition. (2nd Ed). Hillsdale, NJ: Erlbaum.

[6] Greenstein, J.L. (1997). Pointing devices. In M.V. Helander, T.K. Landauer, & P.V. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd edition.). Amsterdam: Elsevier, 1317–1348.

[7] Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of human movement. In *J. of Experimental Psychology*, 47: 381–391.

[8] Grossman, T. & Balakrishnan, R. (2005). The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proc. CHI '05*, 281–290.

[9] 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

[10] Hourcade, J. P. & Berkel, T. R. (2006). Tap or touch? Pen-based selection accuracy for the young and old. In *Ext. Abs. CHI '06*, 881–886.

[11] Kabbash, P. & Buxton, W. A. (1995). The "prince" technique: Fitts' law and selection using area cursors. In *Proc. CHI '95*, 273–279.

[12] Keates, S. & Trewin, S. (2005). Effect of age and Parkinson's disease on cursor positioning using a mouse. In *Proc. ASSETS '05*, 68–75.

[13] Ketcham, C., & Stelmach, G. (2004). Movement control in the older adult. In *Technology for Adaptive Aging*, Washington DC: National Academies Press, 64–92.

[14] Lafayette Instrument. (2007). Model 32011. Steadiness Tester—Hole Type. (http://www.lafayetteinstrument.com)

[15] McGuffin, M. J. & Balakrishnan, R. (2005). Fitts' law and expanding targets: Experimental studies and designs for user interfaces. In *ACM TOCHI*, 12(4):388–422.

[16] Mizobuchi, S. & Yasumura, M. (2004) Tapping vs. circling selections on pen-based devices: evidence for different performance-shaping factors. In *Proc. CHI '04*, 607–614.

[17] Moffatt, K., McGrenere, J., Purves, B., & Klawe, M. (2004). The participatory design of a sound and image enhanced daily planner for people with aphasia. In *Proc. CHI '04*, 407–414.

[18] Ren, X. & Moriya, S. (2000). Improving selection performance on pen-based systems: A study of pen-based interaction for selection tasks. In *ACM TOCHI*, 7(3):384–416.

[19] Smith, M.W., Sharit, J. & Czaja, S.J. (1999). Aging, Motor Control and the Performance of Computer Mouse Tasks. In *Human Factors*, 41(3): 589–596.

[20] Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. In *International J. of Human-Computer Studies*, 61, 751–789.

[21] Spreen, O. & Strauss, E. (1998). A compendium of neuropsychological tests. NY: Oxford University Press.

[22] Tiffin, J. & Asher, E.J. (1948). The Purdue Pegboard: Norms and studies of reliability and validity. In *J. of Appl. Psychol.*, 32: 234–247.

[23] Trewin, S., Keates, S., & Moffatt, K. (2006). Developing Steady Clicks: A method of cursor assistance for people with motor impairments. In *Proc. ASSETS '06*, 26-33.

[24] Walker, N., Philbin, D.A., & Fisk, A.D. (1997) Age-related differences in movement control: adjusting submovement structure to optimize performance. In *J. of Gerontology*, 52B(1): P40–52.

[25] Wechsler, D. (1981). Wechsler Memory Scale-Revised Manual. New York: The Psychological Corporation.

[26] Worden, A., Walker, N., Bharat, K., & Hudson, S. (1997). Making computers easier for older adults to use: area cursors and sticky icons. In *Proc. CHI '97*, 266–271.

[27] Yan, J.H. (2000). Effects of Aging on Linear and Curvilinear Aiming Arm Movements. In *J. of Experimental Aging Research*, 26(4):393–408.