Exploring Methods to Improve Pen-Based Menu Selection for Younger and Older Adults

KARYN MOFFATT and JOANNA MCGRENERE
University of British Columbia

Tablet PCs are gaining popularity, but many individuals still struggle with pen-based interaction. In a previous baseline study, we examined the types of difficulties younger and older adults encounter when using pen-based input. The research reported in this article seeks to address one of these errors, namely, missing just below. This error occurs in a menu selection task when a user's selection pattern is downwardly shifted, such that the top edge of the menu item below the target is selected relatively often, while the corresponding top edge of the target itself is seldom selected. We developed two approaches for addressing missing just below errors: reassigning selections along the top edge and deactivating them. In a laboratory evaluation, only the deactivated edge approach showed promise overall. Further analysis of our data revealed that individual differences played a large role in our results and identified a new source of selection difficulty. Specifically, we observed two error-prone groups of users: the low hitters, who, like participants in the baseline study, made missing just below errors, and the high hitters, who, in contrast, had difficulty with errors on the item above. All but one of the older participants fell into one of these error-prone groups, reinforcing that older users do need better support for selecting menu items with a pen. Preliminary analysis of the performance data suggests both of our approaches were beneficial for the low hitters, but that additional techniques are needed to meet the needs of the high hitters and to address the challenge of supporting both groups in a single interface.

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

General Terms: Design, Experimentation, Human Factors

Additional Key Words and Phrases: Pen-based target acquisition, menu design, aging, older users, interaction techniques

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).

Authors' addresses: K. Moffatt and J. McGrenere, Department of Computer Science, University of British Columbia, 201-2366 Main Mall, Vancouver, BC, Canada, V6T 1Z4; email: {kmoffatt, joanna}@cs.ubc.ca.

http://doi.acm.org/10.1145/1525840.1525843.

ACM Reference Format:
Moffatt, K. and McGrenere, J. 2009. Exploring methods to improve pen-based menu selection for younger and older adults. ACM Trans. Access. Comput. 2, 1, Article 3 (May 2009), 34 pages. DOI = 10.1145/1525840.1525843. http://doi.acm.org/10.1145/1525840.1525843.

1. INTRODUCTION

Technology is increasingly being promoted as a means of addressing cognitive and sensory impairments and enabling individuals to live more independently (e.g., Hawkey et al. [2005], Lee and Dey [2007], Massimi et al. [2007], Moffatt et al. [2004], Mynatt et al. [2000], Rowe et al. [2007], Wu et al. [2007]). Pen-based technologies such as Personal Digital Assistants (PDAs) and Tablet PCs are appealing platforms for these endeavors because they are small, mobile, and powerful. They also allow users to take full advantage of their hand-eye coordination skills, in a familiar form of interaction [Greenstein 1997]. Relative to the mouse, pen-based technologies have been shown to be particularly beneficial for older adults [Charness et al. 2004; 1995], suggesting promise for a wide range of users. However, many of the target users for cognitive and sensory assistive technologies will have associated motor impairments impeding their ability to interact with small devices. For example, in our work designing mobile technology for cognitively impaired individuals [Moffatt et al. 2004], we informally observed many struggling with target acquisition using a stylus (e.g., selecting an icon).

These observations motivated us to gain a better understanding of the challenges inherent to pen interaction, and to thereby build improved pen-based interfaces. Research to date has been biased towards the needs of young-healthy adults, who can more easily adapt to different techniques. Many parameters, including a user's sensory and motor ability, are likely to affect target acquisition and manipulation skill. A broader perspective can be gained by examining a range of abilities. A second bias is that research has tended to focus on designing novel techniques that expand the interaction capabilities of the interface over supporting the basic target acquisition needs of a wide range of users. Thus, despite considerable research aimed at developing improved target acquisition techniques for the pen (e.g., Accot and Zhai [2002], Hourcade and Berkel [2008], Mizobuchi and Yasumura [2004], Ren and Moriya [2000]), point and tap (i.e., selection based on the location of the tap up) remains the de facto standard.

A third bias in research on pen interaction is one of speed over error reduction. Error rates are often low in interaction technique studies,[1] which has led most research to focus on improving speed, sometimes at the cost of accuracy. We argue that it is essential to address errors because they carry a high cost for recovery and can be overly frustrating, especially for older users [Birdi and Zapf 1997; Czaja and Sharit 1998; Rabbitt 2002]. Time-savings mostly benefit expert users by offering small additive savings. Error reduction mostly helps those users who, like many older adults, are easily confused and discouraged

[1]Fitts' type studies typically aim to have an error rate of 4% [MacKenzie 1992].

by errors. This is often lost in laboratory studies, as it is impossible to encapsulate the true cost of recovery time even when penalties are included.[2] It is also not clear that rates observed in laboratory settings truly reflect real-world frequencies: Small changes in task instruction can have a large impact on accuracy [Zhai et al. 2004].

In a previous baseline study [Moffatt and McGrenere 2007], we began to counter these biases by gathering information on the underlying causes of target acquisition difficulty across the adult lifespan. We identified three primary sources of target acquisition difficulty: (1) *slipping*, landing on the desired target, but unintentionally slipping off before lifting the pen; (2) *drifting*, accidentally hovering over (and thus triggering) an adjacent menu; and (3) *missing just below*, erroneously selecting the top edge (i.e., the top 10% or 2 pixels) of the menu item immediately below the target item.

The research reported in this article seeks to address one of these difficulties, namely, missing just below. We describe missing just below in detail in Section 3. Briefly, it occurs in a menu selection task when a user's tap distribution is downwardly shifted, resulting in frequent erroneous selection of the top edge of the item below the target item, and infrequent selection of the corresponding top edge region of the target item itself. This coupling of frequent erroneous selection with infrequent valid selection is important as it suggests we may be able to identify and address missing just below errors with only minor adjustments to the standard point and tap interaction. Minor unobtrusive adjustments are preferable to radical new techniques as they do not require the user to learn a new interaction. Rather they attempt to match existing user behavior.

With this in mind, we designed and developed two possible approaches for addressing missing just below. In the *reassigned edge* approach, input on the top edge of a menu item results in selection of the item above. This approach effectively shifts the target region of each menu item down (in motor space). In the *deactivated edge* approach, input on the top edge is ignored. This approach effectively shrinks the height of each item (in motor space), and adds an invisible menu separator between items. Figure 1 illustrates both of these experimental approaches relative to a traditional menu.

The existence of a downward shift in the tap distribution implies a disparity between where the user is aiming and the center of the menu item. Thus, the idea behind the reassigned edge approach is to reduce this inconsistency by matching the target bounds to the user's actions. We predict that most users would not notice the small shift, but that those who make missing just below

[2]Although most modern programs offer extensive undo functionalities, these facilities do not necessarily address all costs associated with making an error. Selecting the wrong menu item can have a high cost; for example, even though the effects of selecting the wrong program from the Windows Start menu can be easily reversed (by closing the undesired program and reselecting the correct one), the user must first wait for the undesired program to load, which can be time consuming. Moreover, undo facilities cannot always be guaranteed. Web sites commonly use menu-like navigation layouts. Selecting the wrong item on a Web page can be costly: It can result in the loss of work (e.g., Web forms), or it can cause the browser to navigate away from a page that was costly to load (e.g., streamed video).

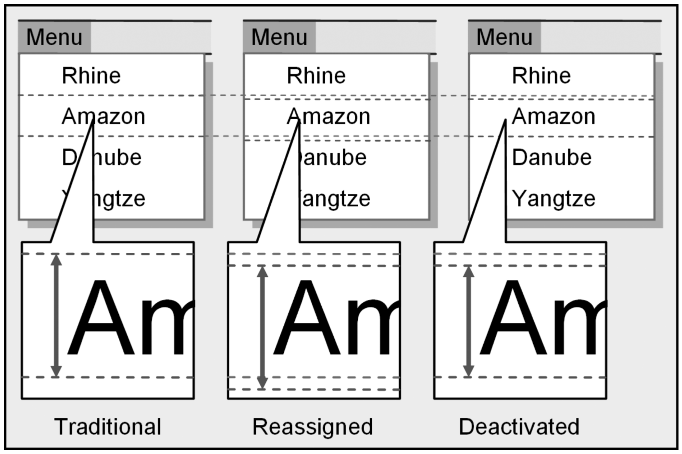

Fig. 1. In a traditional menu, the active target region (shown by hashed red lines) is centered on the text (left); in a reassigned edge menu, it is shifted down by 10% of its height (center); and in a deactivated edge menu, it is reduced by 10%, such that the top region becomes an invisible menu separator (right).

errors would benefit from fewer errors. Its disadvantage is that it turns a small number of would-be-correct selections into errors (i.e., those on the top edge of the target item itself). The deactivated edge approach is different in that it does not introduce any new errors. However, it introduces a performance penalty on all top edge selections because these selections are ignored. The user must continue until a true selection is made. Some users might not notice when their tap does not register, move on, and subsequently have to go back to try again. This may particularly affect older users as they are less able to adapt to changing task requirements [Heath et al. 1999]. The effectiveness of the deactivated edge approach thus hinges on the relationship between the cost of retapping versus the cost of correcting erroneous selections.

This distinction is subtle but important. Classically, Fitts' Law studies would define both of these user inputs as errors. We, on the other hand, have separated them because they have considerably different costs. The cost of tapping on an inactive region (including deactivated pixels, menu separators, and other inactive or blank regions) is accounted for by the extra time needed to complete the selection. In contrast, true errors (i.e., selections on undesired features) require one or more corrective steps for recovery. This cost is not captured by speed, and furthermore cannot be adequately accounted for in a laboratory environment as it is highly dependent on the real-world task, and thus is varied and difficult to predict.

The primary contributions of this work are the results of a laboratory study examining the effectiveness of our reassigned and deactivated edge approaches with younger (19–30) and older (66–81) adults. Overall, only the deactivated edge approach showed any promise. Further analysis of our data revealed that individual differences played a large role in our results and identified a new source of selection difficulty. Specifically, we observed two error-prone groups of users: the low hitters, who, like participants in the baseline study, made missing just below errors, and the high hitters, who, in contrast, had difficulty with errors on the item above. All but one of the older participants fell into one of these error-prone groups, reinforcing the idea that older users do need better support for selecting menu items with a pen. Prelimnary analysis of the performance data suggests both of our approaches were beneficial for the low hitters, but that additional techniques are needed to meet the needs of the high hitters and to address the challenge of supporting both groups in a single interface.

The remainder of this article is organized as follows. In the next section, we summarize related work. Section 3 describes missing just below in more detail and discusses relevant findings from the baseline study. Sections 4–6 describe the experiment conducted to evaluate the reassigned and deactivated edge approaches. The final sections of the article discuss the results of this investigation and identify areas for future research.

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. In this section we also describe the small body of literature that has specifically examined pen-based interaction in the context of aging. We then describe general target acquisition research. The research in this space is quite large and diverse; thus, we focus specifically on that work which has examined decoupling the motor and visual spaces, as it is most closely related to our research. Finally, for completeness, we review research aimed at improving menu interaction. Most of this research has focused on improving speed rather than accuracy, and thus has limited applicability to our work.

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. That research can be helpful in understanding why we might expect older adults to have more difficulty with missing just below errors than younger adults. In particular, research has found that older adults use different strategies concerning the speed-accuracy trade-offs involved in movement control. They tend to be more conservative, and make more corrective submovements [Walker et al. 1997]. Older adults have also been found to cover less distance with their primary movement [Ketcham and Stelmach 2004], to make many more submovements en route [Keates and Trewin 2005], to make less smooth movements [Yan 2000], and to have difficulty staying on the target while clicking [Smith et al. 1999]. In addition, slower selection speeds have been attributed to lower peak velocities [Keates and Trewin 2005; Ketcham and Stelmach 2004], longer deceleration phases [Ketcham and Stelmach 2004], and more pauses while homing in on the target [Keates and Trewin 2005].

Very little work has examined the use of pen-based systems with older adults. Charness et al. [2004; 1995] performed an age-related comparison of the mouse and the light-pen, and found that the pen outperformed the mouse for all ages and that it reduced the performance gap between ages, but that the mouse was rated as being more acceptable and easier to use (across ages). However, this work was done with a light-pen on a monitor, which required the pen to be held up, a position that is unnatural for many users. Modern Tablet PCs are designed to be more comfortable, and thus should result in higher satisfaction. More recently, [Hourcade and Berkel 2008] compared two pen-based selection techniques, tapping and touching (selection if the pen touches the target at any time before tap up), across three adult age groups. They found that for the smallest target size examined (3.8 mm), the oldest group was more accurate using touch. However, one limitation of the touch technique is that it requires that only one target be "touched" during the selection. When targets are directly adjacent, it degrades to tap, and therefore has limited applicability to a menu task.

2.2 Improving General Target Acquisition

Some general target acquisition research has examined decoupling the motor and visual spaces and is relevant to our research, specifically the reassigned edge approach, which also decouples these spaces. Semantic Pointing decouples the motor and visual size of targets, such that the motor size depends on the target's importance for interaction, and the visual size depends on the amount of visual information conveyed [Blanch et al. 2004]. In other words, targets that are likely to be selected are bigger physically (but not visually) than surrounding targets that are less likely to be selected. A similar technique increases the physical size of all targets, by shrinking the size of adjacent inactive space [Worden et al. 1997]; Object Pointing takes this to the extreme, eliminating inactive space altogether [Guiard et al. 2004]. A major difference between the aforementioned body of work and our research is that the techniques just described were all specifically designed for mouse interaction and depend on the ability to adjust the control-display ratio of the cursor. Because the mouse is an indirect input device, techniques can manipulate this ratio between mouse and cursor movement to produce a wide range of interactions. The direct mapping between the pen and the cursor considerably limits the options available in this design space.

One technique in this area was designed for direct touch interaction. Offset Cursor moves the cursor above the user's finger, so as to address the problem of finger occlusion in touch interfaces [Potter et al. 1988; see also Vogel and Baudisch 2007, for a recent extension]. While this technique shares some obvious similarities to our reassigned edge approach, one major difference is that Offset Cursor requires the user to adapt to the technique. Specifically, the user is expected to learn to aim below their target and then use the on-screen cursor to fine-tune the selection. In contrast, we are trying to match the interaction to the user's existing behavior.

Finally, research has also explored the use of alternative input modalities, such as device orientation, for handheld interaction (e.g., Oakley and O'Modhrain [2005], Rekimoto [1996]). These approaches are beyond the scope of our current investigation, but provide additional avenues for future research.

2.3 Improving Menu Interaction

Many researchers have investigated methods for improving menu interaction, but most have focused on younger users, and as a result have tended to focus on making selections faster, rather than more accurate. One common approach has been to develop techniques that reduce the distance the user must travel to reach items. Fish-eye menus minimize the physical size of items, and use fish-eye visualization to dynamically increase their visual size when the cursor is near. Because this makes selection difficult, a focus lock mode is used to make a segment of the menu act as a normal menu [Bederson 2000]. Pie Menus instead minimize travel distance by arranging items circularly around the cursor [Callahan et al. 1988; Kurtenbach and Buxton 1993; Venolia and Neiberg 1994]. As a final example, Split Menus reduce average targeting time by placing the most frequently used items at the top of the menu [Sears and Shneiderman 1994]. None of these techniques specifically addresses accuracy.

Other research has focused on increasing the size of target items. Larger targets are easier to accurately select; thus, we would expect that this research would be more likely to have applicability to missing just below. Morphing Menus progressively grow frequently selected menu items over time, allowing them to borrow motor (and visual) space from neighboring infrequently used items [Cockburn et al. 2007]. This technique should reduce errors (including missing just below errors) for the frequently used items, but it would do so at the expense of infrequently used items, leaving its overall value unclear. A similar approach in this area uses the relationship between size and distance described by Fitts' Law [Fitts 1954] to "Fittsize" menu items [Walker and Smelcer 1990]. In this approach, menu items that are farther from the initial position of the cursor (i.e., from the menu head) are larger than those nearby. This effectively equalizes the index of difficulty across items. This approach was shown to improve speed when evaluated with university students. However, it did not improve accuracy. Finally, Bubbling Menus use gestures to switch between two modes, a regular menu mode and a mode that uses a Bubble Cursor [Grossman and Balakrishnan 2005] to enable quick selection of a predicted set of items [Tsandilas and Schraefel 2007]. Though this technique did improve selection time when the prediction accuracy was high, it was found to reduce selection accuracy and to be difficult to master, suggesting it is likely a poor match for individuals who have difficulty with missing just below.

Another focus has been on improving the selection of submenu items in cascading menus. Although that research targets a different problem than the one we aim to address, we note there are some similarities between the difficulties associated with cascading menus and missing just below. A common difficulty with cascading menus is that users have a tendency to angle down from the target item while traveling to the submenu. Often the downward component of this motion is sufficient to cause the user to move onto the item below, and thus lose the submenu. While the underlying cause of this downward motion is likely different from that of pen-based missing just below errors, there may be overlap in potential solutions. Cockburn and Gin [2006] grew the activation area of each cascading item downward over noncascading items, allowing users to move more directly to the target item of the submenu. This approach shares some similarity to our reassigned edge approach in that both techniques allow target items to borrow activation space from nontarget items, albeit for different purposes. Other research has attempted to break apart the horizontal and vertical movement components, and use that information to improve submenu navigation. For example, target stickiness and force-fields have been used to support horizontal movement towards the submenu [Ahlström 2005; Ahlström et al. 2006]. Another example uses horizontal movement to the right to open submenus, and movement to the left to close them; thus once open, a submenu is “locked on” until an explicit movement closes it [Kobayashi and Igarashi 2003]. These approaches are not well suited to missing just below because, in contrast to the cascading menu problem, missing just below is not related to a bidirectional movement task.

3. MISSING JUST BELOW: RESULTS FROM THE BASELINE STUDY

Full details of the baseline study can be found in Moffatt and McGrenere [2007]. To summarize, we performed a controlled laboratory evaluation of pen-based target acquisition across two tasks (Fitts’ tapping task and menu selection) and three adult age groups (18–54, 55–69, and 70–85), with 36 participants in total. We included two tasks to gain a better understanding of how task might affect targeting ability, especially in terms of accuracy. We involved users from three different age groups to help us understand both the general shortcomings of pen-based interaction, and those unique to older users. In this section we describe those results which pertain specifically to missing just below and provide additional relevant details to what were published in our earlier paper.

Missing just below occurs in a menu selection task when a user’s tap distribution is downwardly shifted, resulting in frequent erroneous selection of the top edge of the item below the target item and infrequent selection of the corresponding region of the target item itself. In the baseline study, we saw a general tendency towards this behavior across all participants, as shown by Figure 2. Of the total 5184 trials across all 36 participants (144 trials each), only 4 selections were made on the top edge of the target item, while 56 selections were made on the corresponding region of the item below.[3] Thus, a selection on the top edge of a menu item was 14 times more likely to be intended for the item above the selected item than the selected item itself. In total, missing just below accounted for 41% (56/135) of the errors in the menu task, and affected a substantial subset (20/36) of individuals.

There was no indication that missing just below decreased over the course of the study or was affected by age. We reanalyzed the baseline study data, running 4x3 (block of 36 trials x age group) repeated-measures ANOVAs on the

[3] In other words, 44 misses and 12 slips. In Moffatt and McGrenere [2007], we considered these errors separately. We combine them here as we believe our approaches apply to them equally.

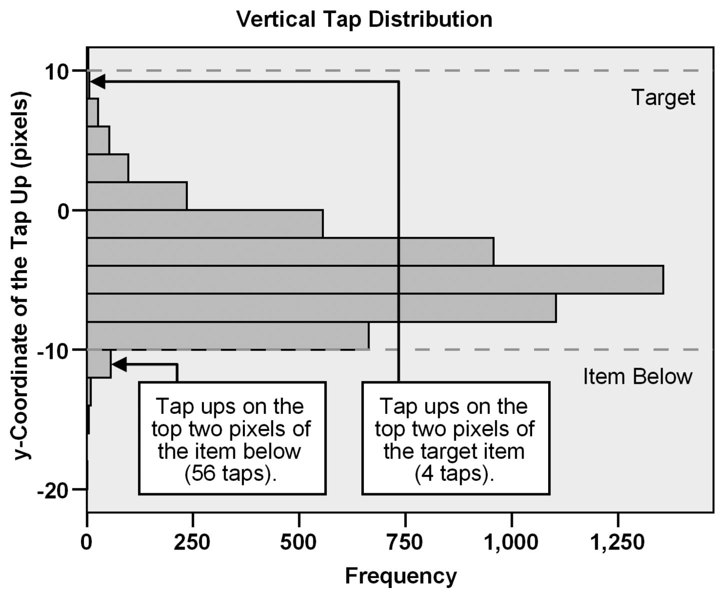

Fig. 2. Histogram of the vertical tap distribution in the baseline study (for taps on the target item and the item below), relative to the center of the target. (Bin size = 2, N = 5127). An additional 57 selections fell outside the range shown. These included errors such as wrong menu selections, randomly wrong item selections, and slips off the menu head. There was otherwise no dominant pattern to these 57 erroneous selections.

mean y-coordinate of the selection and on total missing just below errors, and found no significant differences. Figure 3 provides an age breakdown of the tap distributions and demonstrates the similarities across age: All three age groups showed a similar shift in the peak of the distribution (approximately 5 pixels below the center of the target item), and made a similar number of missing just below errors. However, the two older age groups did have somewhat wider and flatter distributions, reflecting more variability in their selections; thus, a more sensitive study design may have detected a relationship with age.

Error rates were lower in the menu condition than we expected. Of 5184 trials, there were 135 errors, yielding an error rate of 2.6%. This is less than the 4% typically expected in a Fitts'-like experiment, and as we were including a much broader age range, we expected error rates to be higher. One possible explanation for the low error rates is that participants may have been overly focused on accuracy. Though instructed to balance speed and accuracy equally (i.e., to move as quickly and as accurately as possible), many seemed to aim for 100% accuracy, becoming visibly frustrated by errors, but unconcerned with speed. Moreover, to accommodate both the tapping and the menu tasks in the baseline study, the menu task was relatively short. The small number of trials observed may have combined with an accuracy bias to result in the particularly low error rates observed, limiting our power to detect age-related differences. Thus, although we did not find a relationship between age and missing just

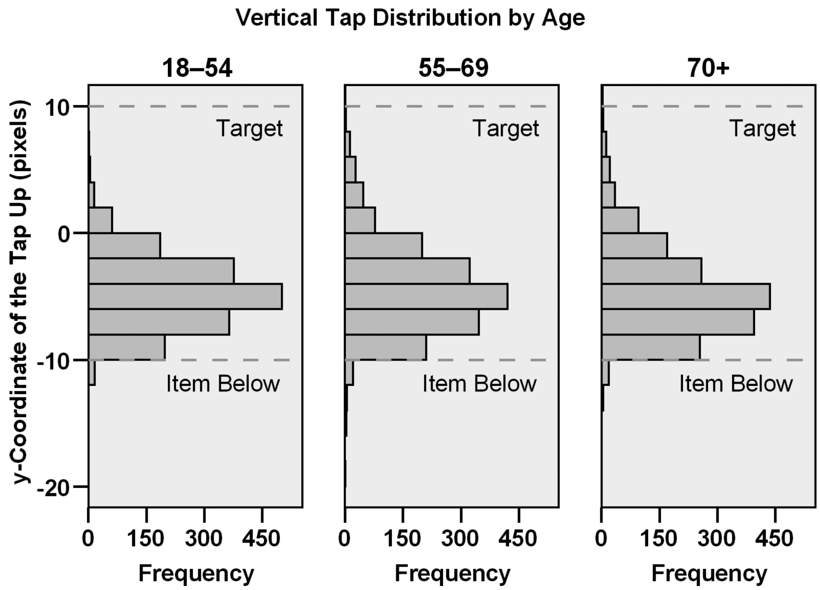

Fig. 3. Histogram of the vertical tap distribution in the baseline study by age (for taps on the target item and the item below), relative to the center of the target. (Bin size = 2, N = 5127). An additional 57 selections fell outside the range shown.

below in the baseline study, we have chosen to maintain it as a factor in the current investigation.

4. EXPERIMENTAL METHODOLOGY

In this section we describe the experimental methodology for our current study, a controlled laboratory experiment with younger and older adults to compare the effectiveness of our two experimental approaches, relative to each other and a control condition. While much of the methodology is similar to that of the baseline study, there are some notable differences, which we describe in Section 4.11.

4.1 Menu Conditions

The three menu conditions are defined next, relative to their handling of input on the top edge (i.e., the top 2 pixels or the top 10%) of a menu item.

Reassigned. Input on the top edge of an item results in selection of the item above.

Deactivated. Input on the top edge is ignored, but the menu does not close.

Traditional. Input on the top edge of an item results in selection of that item.

Reassigned and Deactivated were the experimental conditions and Traditional was used as a control.

4.2 Participants

We recruited 12 participants from each age group, for a total of 24 participants.

Younger. Aged 19–30 (M = 24, 7 female)
Older. Aged 66–81 (M = 73, 6 female)

The younger participants were recruited through advertisements posted on campus, and were paid $15 for approximately 1.5 hours of participation. The older participants were recruited through a local community group. As older participants typically take longer to complete the same task, they were paid $20 for approximately 2 hours of participation. No participants from the baseline study were included. Participants were right-handed, free of diagnosed motor impairments to their right hand, 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. None had any previous Tablet PC experience, and none owned a PDA, but one reported having tried a friend's PDA. The younger participants did, however, have greater general computer experience (in terms of frequency of use, breadth of applications used, and self-rating) than the older participants. All participants were screened using the North American Adult Reading Test [Spreen and Strauss 1998] to ensure sufficient English fluency to follow our instructions; all participants met our minimum criteria for participation.

4.3 Motor Skill

Because motor skill is known to be one of the main factors accounting for age-related differences in targeting ability [Smith et al. 1999], we administered four standardized tests to gather data about our participants' motor abilities.

To measure perceptual speed, we used the Digit Symbol Substitution Test [Wechsler 1981]. In this test, the digits 1–9 are each paired with a simple symbol. The participant is presented with a sheet of numbers and asked to fill in as many of the symbols as they can in a set amount of time (we used one minute). Participants are not expected to memorize the symbols, but are provided with a legend for reference.

As a measure of motor-coordination, we used the right-hand component of the Purdue Pegboard test [Tiffin and Asher 1948]. This test involves the sequential insertion of small pins into a pegboard (similar to a cribbage board). Participants have 30 seconds to place as many pins as they can, using their right hand only. We repeated the test three times and computed the average score.

For reaction time, we used a simple reaction time test [Wilkinson and Allison 1989]. Simple reaction time is the time required to respond to a stimuli. We asked participants to press a button as soon as a green light appeared on the computer screen. Five stimuli were presented (after a random delay) and an average reaction time was calculated.

Finally, to measure hand steadiness, we used a Nine-Hole Steadiness tester [Haaland and Harrington 1998]. In this test, participants are asked to move a metal tipped stylus through nine progressively smaller holes without touching the edge. For each diameter of hole, we recorded whether the participant inserted the stylus without touching. Participants repeated the nine-hole sequence three times. As a final score, we used the number of holes successfully completed on at least two of the three trials.

4.4 Apparatus

We used the same experimental setup as in the baseline study. All experimental conditions were run on a Fujitsu LifeBook T3010D Tablet PC with a 1.4 GHz Pentium M processor and 768MB RAM, running the Windows XP Tablet PC Edition operating system. The display was 12.1 inches large, with a resolution of 1024 x 768. The standard inductive pen that came prepackaged 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 (based on previous pilot studies) of approximately 35 degrees from horizontal. We chose this setup, as it is difficult to see the screen when it is flat, and holding the tablet can be strenuous for older users. Future work could examine the additional difficulties that arise when holding the device. Participants were encouraged to adjust the position of their chair and the placement of the stand for comfort and most participants made these adjustments. They were further encouraged to rest their hand on the screen to reduce arm fatigue.

4.5 Design

The experiment used a 2x3 factorial design with age (Younger, Older) as a between-subjects factor, and interface condition (Reassigned, Deactivated, Traditional) as a within-subjects factor. Interface was chosen to be a within-subjects factor because this increases the power of the design and allows for comparative comments on the interfaces. To minimize learning effects, we fully counterbalanced the order of presentation of the interfaces. To ensure that familiarity with the menu contents did not bias the results, a different menu scheme was randomly generated for each condition as described in Section 4.7.

4.6 Procedure

Participants started the study by completing the motor tests and a brief questionnaire about their background and computer experience. They then completed the first 8 steps of the built-in tutorial "Get Going with the Tablet PC". (Steps 9–17 concern text input and were not relevant to the study.) In this tutorial, participants were introduced to using the pen as an input device. Most importantly, they were informed of the following features: (1) they can rest their hand on the screen during input, (2) the computer tracks the pen both when it is touching the screen and when it is slightly above it, and (3) an on screen cursor provides feedback of the current cursor location. The tutorial also provided participants with an opportunity to practice using the pen for input. Once participants finished the tutorial, the tablet was calibrated to each participant using the built-in Windows XP (Tablet PC Edition) calibration utility. This utility presents four crosshair targets, one in each corner of the screen. The user taps on these targets, and based on the location of these taps, the system calibrates itself to the user. The main purpose of this calibration is to account for parallax; that is, the apparent displacement of the cursor caused by the small separation between the input sensors and the surface of the screen.

Participants then completed the menu conditions. Participants were told that they were going to be using three different interfaces, but not how they differed. They were instructed to use the programs naturally, and assured that the differences between the programs would be discussed at the end of the experiment. To enable comparative comments on the interfaces without biasing the results, each interface was masked with a neutral-sounding name. All spontaneous verbal comments made by participants throughout the experiment were documented by the researcher. After each condition, each participant was asked specifically to reflect on that condition; these comments were made verbally and documented by the researcher to allow participants to rest their arms. Between conditions, participants also completed short verbal distracter tasks. Between the first and second condition participants completed the North American Adult Reading Test [Spreen and Strauss 1998] and between the second and third condition they completed the FAS test of verbal fluency [Nussbaum 1998]. These two tasks were chosen to engage participants mentally, but not physically. Finally, at the end of the study, participants were asked to rank the conditions, and encouraged to make additional comments. These were also made verbally and documented by the researcher. 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.

4.7 Task

Figure 4 illustrates the experimental interface. All three conditions had the same visual appearance. For each trial, a menu item was displayed across the top of the screen. Participants were instructed to select that item from the menu as quickly as possible while remaining accurate. The system advanced to the next item when the participant selected a menu item, regardless of correctness. It did not advance if the participant selected a deactivated region (or any nonmenu-item component of any of the interfaces; e.g., a menu separator or the interface background). A soft clicking sound provided feedback for correct selections, and a louder beep alerted participants to selection errors (i.e., selection of a nontarget menu item); there was no auditory feedback on deactivated regions, or on any other parts of the interface.

For each condition, participants completed a short 12-trial practice block followed by six 36-trial blocks with an enforced 45-second break between blocks.

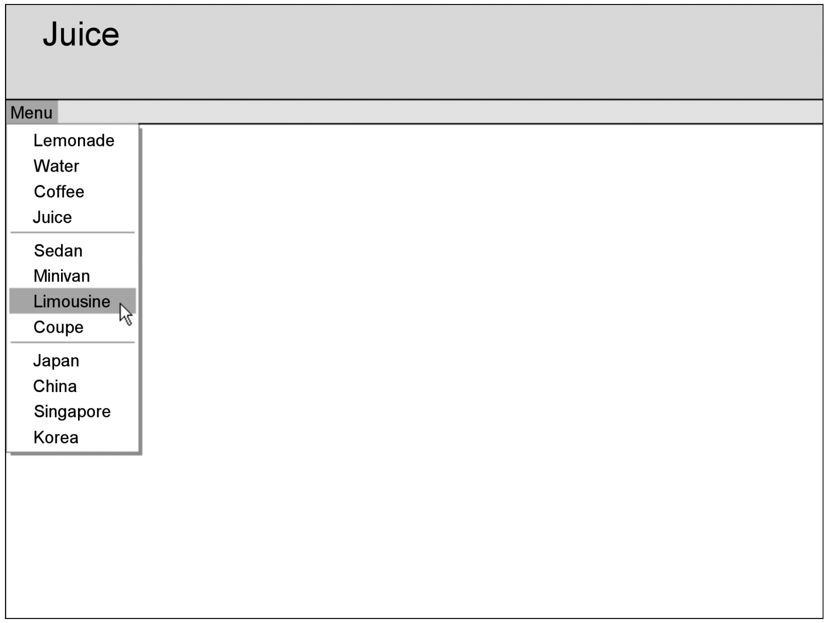

Fig. 4. Screenshot of the upper-left corner of the experimental interface. All three interfaces had the same visual appearance. Not shown: In the lower-left corner, a counter displayed the number of trials remaining in the block.

Each block consisted of a randomly ordered selection sequence from a single 12-item menu. Each item was selected three times in each block (once in each practice block). Each menu item was 20 pixels (4.8 mm) high, and each menu separator was 5 pixels (1.2 mm) high. In total, each participant completed 36 trials x 6 blocks x 3 interfaces for a total of 648 trials (excluding the 12 practice trials in each condition). One participant only completed 4 blocks for each interface, for a total of 432 trials. Where necessary, his data is scaled.

Menu contents remained constant within each condition, but changed between conditions. Each menu contained three groups of four semantically related items. These schemes were randomly generated for each participant using the approach presented by Cockburn et al. [2007]. Specifically, three 4-item semantic groups were randomly selected from a collection of such groups. The items were then randomly ordered within that group, and no group was reused in any other condition. (See Figure 4 for an example of a generated menu.) By randomly generating menu contents, we reduced the impact of the particular menu contents on our measures and prevented a confound between content and interface. Moreover, semantic groupings provide structure in the menu, making them more like real menus. The semantic groups were separated visually by menu separators. For Reassigned, the top edge of each separator was assigned to the item above it so that all menu items behaved consistently. For Deactivated, no additional changes were required as menu separators are by default inactive.

4.8 Duration

The experiment was designed to fit into a single 120-minute session. All participants finished in between 75 and 120 minutes, with the older users requiring, on average, more time. Although 120-minute sessions may at first appear inappropriately long, these sessions comprised multiple different activities and included regular breaks. Specifically, verbal distracter tasks were inserted between conditions and breaks were inserted between blocks of the same condition to allow participants to rest their arms. Previous research has shown that when these techniques are used (activity switching and short breaks), sessions of this length do not result in fatigue, even for older adults [Uttl et al. 2000]. On average, the younger adults spent approximately 57 seconds per block (17 minutes total tapping time across all three conditions), while the older adults spent just over 134 seconds per block (40 minutes total).

4.9 Speed Accuracy Trade-Off

We introduced a monetary incentive and a graphical feedback mechanism to encourage participants to perform both quickly and accurately. For the incentive, an additional $10 was awarded to the top 1/3 performers in each age group. The 1/3 ratio was chosen to encourage participants to believe they had a reasonable chance of receiving the incentive. Performance was calculated as the number of correct selections divided by the time taken to complete all selections. Shown in Figure 5, the graphical feedback was presented during the breaks between blocks to ensure participants understood the performance measure used for the incentive and to allow them to accurately gauge their performance on both speed and accuracy.

4.10 Measures

Because Reassigned turns a small number of would-be-correct selections into errors, we need to consider the impact of both the missing just below errors it prevents (i.e., those selections on the top edge of the item *below* the target), and the errors it introduces (i.e., those selections on the top edge of the *target* itself). Thus, to compare our interfaces we computed a net benefit score. For Reassigned, the net benefit was equal to the number of missing just below errors prevented minus the number of errors introduced. For example, if a participant selected the top edge of the item below the target 12 times in Reassigned, and the top edge of the target 5 times, then their net benefit for Reassigned would be 7. For Traditional, the net benefit was instead equal to the number of correct selections along the top edge of the target minus the number of incorrect selections on the top edge of the item below (i.e., the missing just below errors). For Deactivated, nothing happened when the top edge of a target was selected. Thus, for it, we considered the final outcome of the trial: The net benefit was equal to the number of *correct* trials which included input on the top edge of an item minus the number of *error* trials which included input on the top edge of an item. These definitions are summarized in Table I.

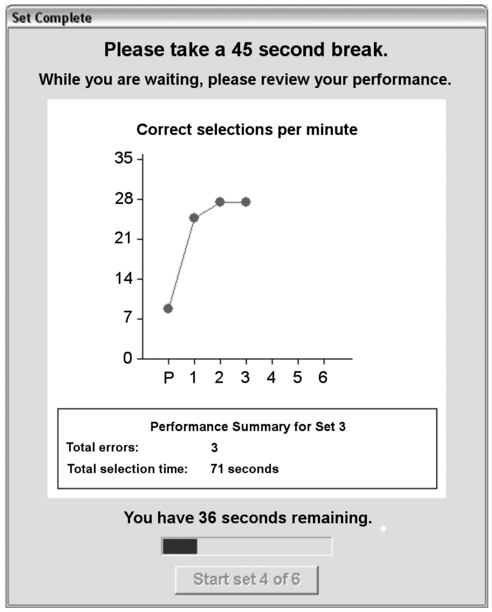

Fig. 5. The performance feedback presented between blocks, which provided the rate of correct selections for each block completed (including the practice block) and a summary of the errors and time for the most recent block. In this example, the user has just completed block 3.

Table I. Definition of Net Benefit for Each Interface

Interface	Net Benefit		
Traditional	Selections on the top edge of the *target item*	minus	Selections on the top edge of the *item below*
Reassigned	Selections on the top edge of the *item below*	minus	Selections on the top edge of the *target item*
Deactivated	*Correct* trials following selection of the top edge of any item	minus	*Error* trials following selection of the top edge of any item

For Deactivated, recall that nothing happens when the top edge is selected; thus, observation of the subsequent selection is necessary.

We recognize that for Deactivated, the measure of net benefit is imperfect. It does not fully capture the cost associated with selecting a deactivated pixel, and may overestimate the true benefit. For example, in terms of overestimation, if a correct selection is made after a selection on the top edge of the target item itself (the deactivated region), this is counted positively in the net benefit calculation though the interface actually interfered with the menu selection. To partially address these limitations, we provide a breakdown of the selections that contributed to the net benefit in the Deactivated condition. We additionally measure the average time and the average number of taps required to select an item.

Many of our older participants struggled with opening the menu, at times requiring several attempts before succeeding. This difficulty was different from missing just below and independent of our interfaces. When we included opening the menu in our time and taps measure, the variability among participants was so great that it hindered interpretation of the results. Thus, for clarity and to allow us to focus uniquely on the contribution of our interfaces to the results, we excluded from these measures the time and taps required to open the menu itself. Specifically, time and taps were measured from the point the menu was last opened until the final selection was made. Because tapping on a deactivated region does not close the menu, we do not expect any difference among our interfaces in terms of time spent opening menus as no treatment was applied to the menu heads in any condition.

4.11 Comparison to the Baseline Study Design

The basic task setup of the current investigation closely mirrors that used for the menu task of the baseline study [Moffatt and McGrenere 2007]. In this section, we highlight the key differences. As mentioned previously, we observed a bias towards accuracy in the baseline study. To encourage participants to perform both quickly and accurately, we introduced a performance feedback mechanism and a monetary incentive. At the individual trial level, we also added auditory feedback to correct selections because the sudden beep of an incorrect selection seemed to be startling in some cases in the baseline study. By increasing the regularity of the auditory feedback, we hoped to reduce its disruption while maintaining its information content. To increase statistical power, we increased the number of trials in each condition. In order to keep the total study duration reasonable, increasing the number of trials required changing the task from a discrete task, where users return to a home button in the center of the screen between trials, to a continuous task, where items are selected in an uninterrupted stream [Soukoreff and MacKenzie 2004].

We also modified the age groups to simplify the design based on our age findings in the baseline study. Specifically, in the baseline study we divided the older end of spectrum into two categories (55–70 and 70+) and for the younger group included the entire spectrum from 18–55. These groupings were based on age-related changes that occur in cognition [Craik and Salthouse 1992], 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. However, we found few significant differences between the two older groups in the target acquisition tasks and no differences between them on the motor tests. Thus, for this study we chose to include only two groupings representing the younger (18–30) and older (65+) ends of the spectrum.

Menu contents were randomly generated for each condition in this study. In the baseline study, a single scheme was used for all participants as there was only one menu condition. Moreover, participants selected items from three menus in the baseline study, but from only one in the current study. Drifting, one of the difficulties uncovered in the baseline study, involved the interaction between menus. Thus, we included only a single menu in this study so as to prevent drifting from interfering with our results.

5. HYPOTHESES

We had the following hypotheses for this experiment:

H1. Both Deactivated and Reassigned will have higher net benefit than Traditional, but Deactivated will have higher net benefit than Reassigned. This hypothesis is based on data from our baseline study. Additionally, because Reassigned turns a small number of would-be-correct selections into errors (i.e., those on the top edge of the target item itself), we expect it to be slightly less effective than Deactivated.

H2. Deactivated will require more taps and take longer overall. Because Deactivated ignores all selections along the top edge of an item, we predict it will require longer selection time and more taps to select, relative to Traditional and Reassigned.

H3. Both age groups will benefit from the experimental approaches, but the older users will benefit more so. Research has indicated that older adults move less smoothly [Yan 2000], and have more difficulty staying on targets [Smith et al. 1999]. So although the baseline study suggests missing just below affects all users, we expect that with a more sensitive study design, we will find that it is a larger problem for older adults.

H4. Both experimental approaches will be preferred to Traditional, and Reassigned will be preferred to Deactivated, especially by the older participants. We expect that the predicted performance benefits will result in an overall preference for our experimental interfaces. We further predict that Reassigned will be preferred over Deactivated, because although it is expected to be slightly less effective in terms of net benefit, we predict many will find retapping frustrating.

6. RESULTS

For each of our main performance measures (Net Benefit, Taps to Select, and Selection Time) we performed a two-way repeated measures ANOVA (Age x Interface). Bonferroni corrections were used for all post hoc pairwise comparisons. Significant heterogeneity of error variance was an issue in many of our analyses of the effect of age group (with the older group showing more variability). This is not surprising; it has been previously found that individual variability increases with age [Gregor et al. 2002]. For these analyses, we used a Welch's ANOVA, which is robust against unequal error variances. In all our repeated measures analyses (except trial time), sphericity was an issue; thus, Greenhouse-Geisser adjustments were used. We defined outliers as scores more than two standard deviations from the mean. Analyses where outliers have been removed are noted. For completeness, we did a preliminary analysis which included presentation order as a between-subjects factor. As expected, there were no significant main or interaction effects for the presentation order, giving us confidence that counterbalancing the interfaces sufficiently accounted for any learning or fatigue effects.

Analysis of the motor tests confirmed overall motor decline with age: Univariate ANOVAs revealed a significant main effect of age on each of the

Table II. Results of the Motor Tests, by Age (N = 24)

Motor Test	Mean (SD)		Significance
	Younger	Older	
Simple Reaction Time (ms)	275.8 (37.53)	455.8 (189.76)	$p < .001^*$
Purdue Pegboard (# pins)	16.1 (1.82)	12.5 (1.64)	$p < .001$
Digit Symbol Substitution (# subs.)	54.2 (5.01)	25.6 (6.02)	$p < .001$
Hole-type Steadiness Test (# holes)	5.0 (0.95)	3.83 (0.72)	$p < .01$

(*) denotes Welch's ANOVA.

Table III. Breakdown of the net benefit results for Deactivated

Age Group	Item Type of Top Edge Selection	Mean (SD) Correct Trial	Error Trial	Total
Younger	Target	1.00 (2.30)	0.00 (0.00)	1.00 (2.30)
	Non-target	2.08 (2.50)	0.00 (0.00)	2.08 (2.50)
Older	Target	4.91 (4.48)	0.73 (1.01)	5.64 (5.18)
	Non-target	6.91 (6.04)	0.14 (0.45)	7.05 (6.40)

For each age group, the average number of trials involving selection of the top edge of an item is provided, broken down by whether or not the selection was on the top edge of the target item, and whether or not the subsequent (nontop edge) selection was correct (i.e., on the target item) (N = 23).

motor tests. These results are summarized in Table II. Interestingly, the older participants performed significantly worse on the digit symbol substitution test than comparably aged individuals in the baseline study,[4] perhaps foreshadowing the differences found between the results of the two studies. However, there were no differences between the studies for the Purdue pegboard test (and no comparable data was available for the steadiness or reaction time tests).

6.1 Net Benefit

A two-way repeated measures ANOVA (Age x Interface) for net benefit, excluding one outlier (older), revealed a significant main effect of interface ($F_{1.24,26.09} = 6.653$, $p = .011$, $\eta^2 = .241$), but no main effect of age ($p = .20$). Post hoc pairwise comparisons further revealed that Deactivated had a higher net benefit than both Traditional ($p = .003$), and Reassigned ($p = .001$). No significant difference was found between Reassigned and Traditional ($p = 1.000$).

A breakdown of the net benefit for Deactivated is provided in Table III and summarized as follows. Older participants selected the top edge of the target item 5.64 times on average, and the top edge of a nontarget item 7.05 times. Following selection of a deactivated top edge, older participants selected the correct target item 87% and 98% of the time (following selection of the top edge of target and nontarget items, respectively). Younger participants selected the top edge of the target item 1.00 time on average and the top edge of a nontarget

[4]In the baseline study, 17 participants were aged 65+, and 7 were aged 19–30. We performed a Univariate ANOVA (Study x Age Group) on the digit symbol substitution test scores and found a significant interaction between study and age group ($F_{1,44} = 10.058$, $p = .003$, $\eta^2 = .186$). Post hoc pairwise comparisons further revealed that the older participants in the baseline study outperformed those in the current study ($p = .001$). Not surprisingly, there was also a main effect of age group ($F_{1,44} = 10.058$, $p = .003$, $\eta^2 = .186$). Mean scores for the older group: Baseline study: $M = 34.20$ ($SD = 6.4$); Current study: $M = 25.58$ ($SD = 6.0$).

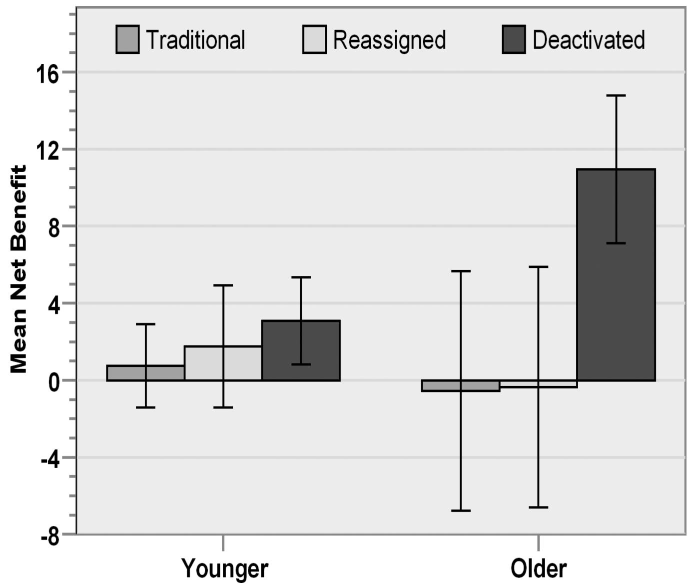

Fig. 6. Average net benefit for 216 trials, by interface and age group (N = 23). Error bars represent 95% confidence intervals. (Note: A higher score in this graph represents better performance of the interface.)

item 2.08 times on average. Subsequent to selecting the top edge of an item, the younger participants always went on to select the correct item.

Thus, our predictions in hypothesis H1 were partially supported. Contrary to hypothesis H1, Reassigned did not provide any performance benefit. Consistent with H1, Deactivated did result in a significantly higher net benefit, but selections on the top edge of the target item made a substantial contribution to the positive net benefit observed for Deactivated, overestimating its true benefit. Interestingly, individual participant scores were polarized: Participants either made selections on the top edge of the target or on the top edge of the edge item below, but no one made marked use of both edges, suggesting that individual differences were at play. We explore this idea further in Section 6.5.

Our analysis also revealed a trend suggesting an interaction between interface and age ($F_{1.24,26.09}$ = 3.48, p = .065, η^2 = .142). Figure 6 illustrates this interaction and suggests that Deactivated may have disproportionately helped the older group, as predicted by H3. Further investigation with a larger sample is needed to confirm this preliminary evidence.

6.2 Speed and Taps to Select

A two-way ANOVA (Age x Interface) on taps to select, excluding one outlier (older), revealed a significant main effect of interface ($F_{1.22,25.56}$ = 9.58, p = .003, η^2 = .313), and significant interaction between age and interface ($F_{1.22,25.56}$ = 6.40, p = .014, η^2 = .234). Post hoc pairwise comparisons further revealed that, for the older group, Deactivated required more taps than both Traditional ($p < .001$), and Reassigned (p = .004), as predicted by H2. Figure 7

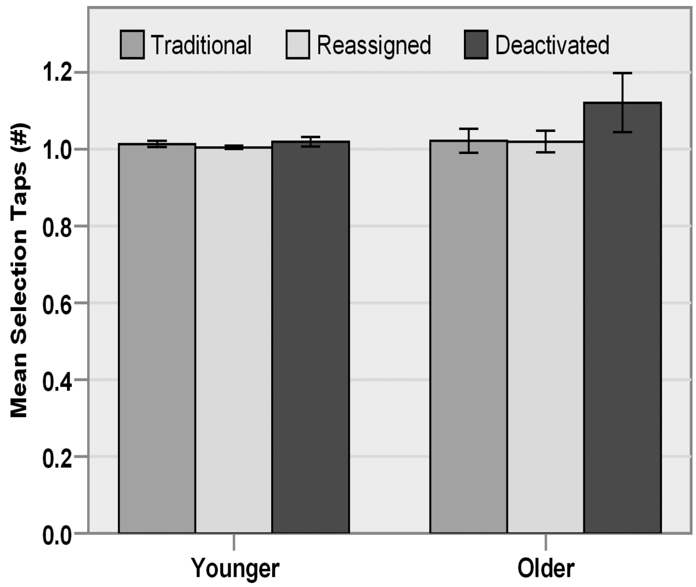

Fig. 7. Average taps needed to select an item, by interface and age group ($N = 23$). Error bars represent 95% confidence intervals.

shows the average number of taps required to make a selection for each interface, by age. There was also a significant main effect of age ($F_{1,10.66} = 7.78$, $p = .018$, $\eta^2 = .288$, Welch's ANOVA), indicating the older group required more taps to make a selection than the younger group.

Contrary to hypothesis H2, a two-way ANOVA (Age x Interface) on selection time, excluding one outlier (younger), revealed neither a significant main effect of interface ($p = .39$) nor an interaction between interface and age ($p = .46$), suggesting the cost of the deactivated condition may not have been as great as we had predicted. As expected, there was a significant main effect of age, $F_{1,11.16} = 34.52$, $p < .001$, $\eta^2 = .600$, Welch's ANOVA), indicating that older participants were generally slower than their younger counterparts. In Traditional, Reassigned, and Deactivated, respectively, younger participants took on average 980 ms, 972 ms, and 986 ms ($SD = 120, 83, 128$), while older participants took on average 2436 ms, 2290 ms, and 2502 ms ($SD = 1052, 852, 767$).

6.3 Subjective Responses

Many participants reported difficulty completing the ranked questionnaire at the end of the study. This difficulty may have been caused by the fact that the differences between the conditions were subtle. Many participants, in particular those in the older group, did make insightful comments on the interfaces, so we instead provide a descriptive account of those comments. The top portion of Table IV provides a summary of the number of participants who commented positively and negatively for each interface by age group. These counts are

Table IV. Breakdown of Comments Made by Participants, by Age Group and by Tap Distribution Group, Reported as # Positive (# Negative) (N = 24)

		Deactivated	Reassigned	Traditional
Age group	Younger	2 (3)	4 (2)	2 (2)
	Older	1 (6)	5 (1)	2 (0)
	Total	3 (9)	9 (3)	4 (2)
Tap distribution group	Low hitters	1 (3)	6 (0)	1 (1)
	Neutrals	2 (0)	1 (2)	1 (1)
	High hitters	0 (6)	2 (1)	2 (0)
	Total	3 (9)	9 (3)	4 (2)

based on comments made throughout the study sessions in reference to one of the interfaces. No neutral comments were made.

Despite its positive performance results, there was a strong negative reaction to Deactivated. Nine participants commented negatively on it, while only 3 made positive comments. In contrast, 9 commented positively and 3 negatively on Reassigned, and 4 commented positively and 2 negatively on Traditional. The older participants were responsible for most of the polarity between Deactivated and Reassigned: 6 commented negatively on Deactivated (versus 1 positively), while 5 commented positively on Reassigned (versus 1 negatively).

For Deactivated, the negative comments reflected confusion and disruption. As one participant from the older group put it, “[It] really throws you off when you have to click more than once.” Others were less specific, making comments such as “[Deactivated] seems to be a little more awkward,” and “[With Deactivated, it] was harder to make selections.” Both of these comments were made by participants from the older group. Other negative comments reflected misconceptions over why taps were not being recognized by the system. One common interpretation was that more force or longer contact was required. For example, one older participant reported, “This one seems to need you to press harder,” while a younger participant speculated, “I think you need to hold [the pen] for quite a while [with Deactivated].” Positive comments on Deactivated were less specific; for example, one older participant simply stated that he liked Deactivated “better”, but could not further qualify his preference.

Negative comments on Reassigned reflected an awareness of the discrepancy between motor and visual space in that condition. For example, one older participant asked, “Why does this keep happening, I see I have got it right but then it tells me I’m wrong.” In contrast, the positive comments reflected a sense that things were somehow easier. One older participant described it as, “[Reassigned] was a bit easier. I seemed to be able to manipulate it a bit better.” Another reported, “I thought I slipped off but the computer didn’t think so.”

6.4 Summary of the Main Results

All four of our hypotheses were partially supported by the data.

H1. Both Deactivated and Reassigned will have higher net benefit than Traditional, but Deactivated will have higher net benefit than Reassigned.

Contrary to our hypothesis, Reassigned did not result in a performance benefit over Traditional. Consistent with our hypothesis, Deactivated resulted in a higher net benefit than both Traditional and Reassigned; however, selections on the top edge of the target item itself substantially contributed to the positive net benefit of Deactivated, suggesting the measure of net benefit overestimated its true benefit. Further research is needed to better understand this result.

H2. Deactivated will require more taps and take longer overall. For the older group, Deactivated increased the number of taps required to make a selection; however, there was no evidence that it increased the overall trial time, suggesting that the cost may not be as large as predicted.

H3. Both age groups will benefit from the experimental approaches, but the older users will benefit more so. Preliminary trends suggest that the older group may have benefited more from Deactivated than the younger group, however further research is necessary to confirm this indication.

H4. Both experimental approaches will be preferred to Traditional, and Reassigned will be preferred to Deactivated, especially by the older participants. Many participants found the comparative rankings difficult, and thus, those results were not informative. The subjective comments made by participants throughout the experiment did provide some interesting insight into user preferences: Although Reassigned did not show a performance benefit, a number of participants commented favorably on it, and despite the positive performance result for Deactivated, it received a number of negative comments, especially from the older group.

6.5 Individual Differences: Understanding the Performance Results

We performed a secondary analysis to provide insight into the unsuccessful performance results for Reassigned and to better understand the breakdown of net benefit for Deactivated. Across all participants, Reassigned prevented a total of 68 missing just below errors, but introduced 50 new errors; overall, it performed no better than Traditional. Deactivated did result in a significantly higher net benefit, but for the older and younger participants, respectively, 44% and 32% of selections on the deactivated top edge were on the target item, suggesting Deactivated is less effective than indicated by its net benefit score.

To determine if individual differences played a role in these results, we examined each participant's vertical tap distribution across all conditions. Visual analysis suggested three distinct types of users. A K-means cluster analysis on the mean y-coordinate of all errors on the item above and below the target (using data from all conditions) confirmed our visual analysis and identified three clusters (with an observed significance of $p < .001$).

The *low hitters* followed the distribution observed in the baseline study: Their distribution was shifted down, and they tended to select the top edge of the item below the target, seldom selecting the top edge of the target itself. The *high hitters* displayed a somewhat diametric distribution. Their distribution was skewed upwards, and their errors tended to be on the item above the target. They rarely selected the top edge of the item below. These tendencies

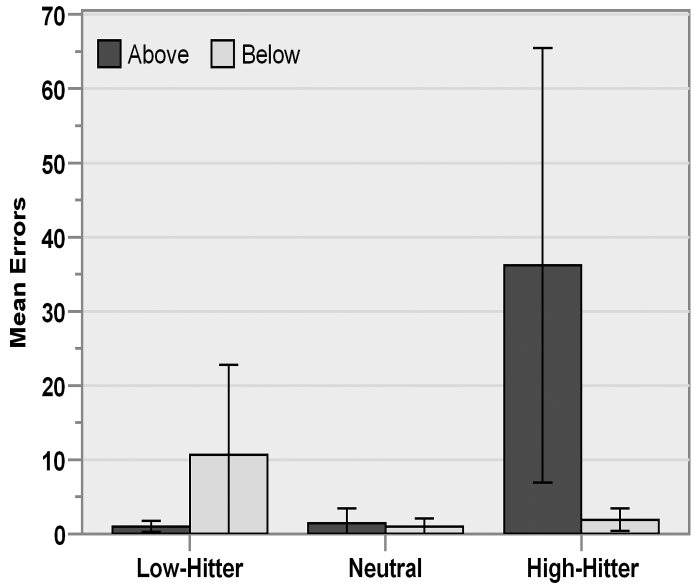

Fig. 8. Errors above and below the target item (across all interfaces), by tap distribution group (N = 24). Error bars represent 95% confidence intervals.

were relatively strong: The low hitters all had at least twice as many selections on the item below than on the item above, and likewise the high hitters all had at least twice as many selections on the item above than on the item below. There were 7 low hitters (4 young), and 10 high hitters (2 young), accounting for 17 of the 24 participants in the study. While both the high and the low hitters made substantial and skewed use of either the item above or the item below the target, the remaining 7 participants, the *neutrals*, seldom used the top edge of any item and showed no strong tendency for either. We would thus expect these individuals to be neither hindered nor helped by the experimental interfaces as they would have experienced very little difference between them. Not surprisingly, most of the participants who were classified as neutral were from the younger group (6/7). To summarize, the tap distribution groups were as follows.

Low Hitters. 4 older, 3 younger, 7 total.
Neutrals. 6 older, 1 younger, 7 total.
High Hitters. 8 older, 2 younger, 10 total.

Figure 8 highlights the contrasting error patterns of the groups and Figure 9 shows each group's tap distribution across all three interfaces. (There were no differences in the tap distributions across the three interface conditions; i.e., participants did not change their tapping behavior in response to the interfaces). We would especially like to highlight the high-hitter group's marked use of the bottom half of the item above the target, indicating that

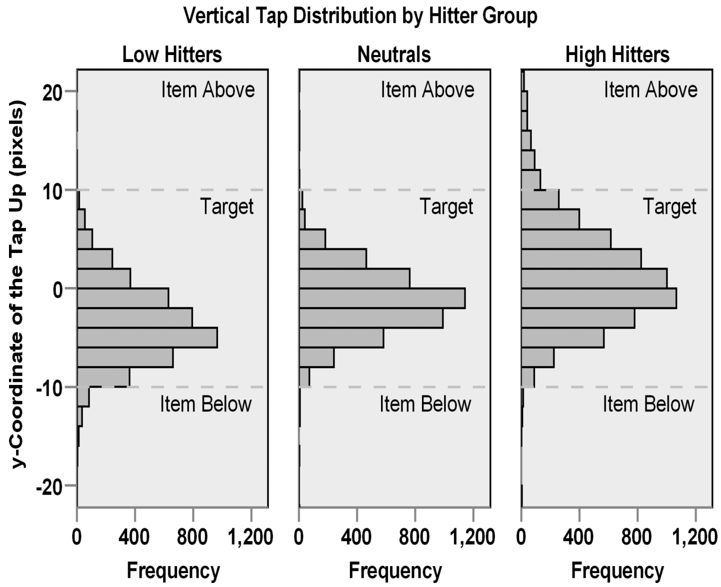

Fig. 9. Histograms of the vertical tap distributions (for taps on the target item and the lower/upper half of the item above/below) for each hitter group and across all interfaces, relative to the center of the target item. (Bin size = 2.)

although they do not make missing just below errors, they do have interaction difficulties.

In consideration of these differences, we reexamined net benefit blocking on these three groups of users. These results need to be considered as very preliminary indications only, as these groups were not identified a priori and were not controlled for in the experiment. Though the spread of participants across each of the groups is reasonably balanced, presentation order was not counterbalanced across them. A two-way repeated measures ANOVA (Distribution Group x Interface) for net benefit revealed a significant main effect of interface ($F_{1.51,31.68} = 7.55$, $p = .004$, $\eta^2 = .264$), a significant main effect of distribution group ($F_{2,10.44} = 6.56$, $p = .014$, $\eta^2 = 0.283$), and a significant interaction between interface and distribution group, ($F_{3.02,31.68} = 8.219$, $p < .001$, $\eta^2 = .439$). Post hoc pairwise comparisons further revealed that for the low hitters, both Deactivated ($p = .001$), and Reassigned ($p = .016$), had a significantly higher net benefit than Traditional, but for the high hitters, Reassigned had a significantly lower net benefit than both Traditional ($p = .028$), and Deactivated ($p < .001$). Figure 10 illustrates this interaction between interface and distribution group. In particular, it contrasts the positive effect of Reassigned for the low hitters, against the negative effect it had for the high hitters.

These results also help explain the net benefit score for Deactivated. Figure 9 shows that the high hitters were mostly selecting the top edge of the target. Thus, although they saw a positive net benefit for Deactivated (as shown in Figure 10), this was mostly due to selections on the top edge of the

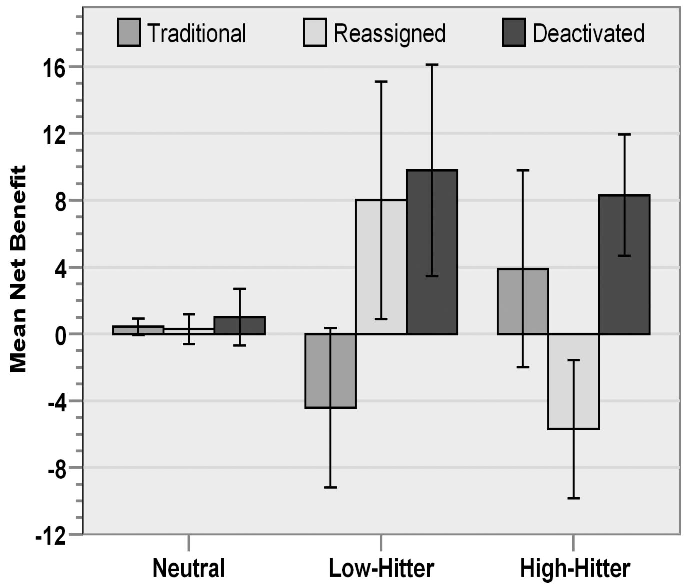

Fig. 10. Average net benefit for 216 trials, by interface and tap distribution group (N = 24). Error bars represent 95% confidence intervals.

target item. In contrast, for the low hitters, Deactivated had a positive net benefit due to selections on the top edge of the item below (i.e., missing just below errors).

Finally, these individual differences help explain the seemingly contradictory subjective responses observed. The bottom section of Table IV provides a breakdown of the positive and negative comments made for each interface by tap distribution group. Notably, 6 of the 9 negative comments made on Deactivated were made by high hitters, and 6 of the 9 positive comments made on Reassigned were made by low hitters. This response pattern suggests that although Deactivated did not impede the high hitters in terms of increasing their selection errors, they were aware of the cost of unregistered taps and that this cost was not being offset by any benefit. Most (5/6) of the high hitters who complained about Deactivated had at least one other condition before it, and thus were commenting from a reference point of having already experienced a condition without deactivated areas. In contrast, although the low hitters were also incurring the cost of retapping, their much lower number of negative responses suggests that this cost may have been balanced by the benefit of fewer selection errors.

7. DISCUSSION

The primary goal of this research was to evaluate two potential approaches for addressing missing just below errors. Only the deactivated edge approach, where input on the top edge of menu items was ignored, showed a performance improvement, though for some users this benefit was inflated due to selections on the top edge of the target. Promisingly, the cost of having to reenter ignored input was not as large as we had thought it might be. Though the deactivated edge approach required significantly more taps to make a selection for the older group, this did not translate into an increase in the selection time. However, many participants did not like the deactivated edge approach and found it confusing when selections were ignored. Thus, for the deactivated edge approach to be a viable technique, refinements are needed to make its functionality clearer; in Section 8, we discuss possible refinements to explore in the future. The reassigned edge approach, for which input on the top edge of an item resulted in selection of the item above, did reduce missing just below errors, but the number of errors it introduced (on the top edge of the target itself) negated any potential benefit.

A secondary goal for this work was to further examine the role of age in missing just below. Consistent with the baseline study, we did not see any indication that missing just below disproportionately affects older users, though there was some preliminary evidence that the older users were disproportionately helped by the deactivated edge approach. However, including older users in this research was important. Our analysis of individual selection patterns identified two error-prone groups of users: the low hitters, who, like participants in the baseline study, made missing just below errors, and the high hitters, who, in contrast, had difficulty with errors on the item above. Most of the high hitters were from the older group (8/10), and their needs would likely not have been identified had we conducted this study with younger users only. Moreover, all but one of the older users fell into one of the low- or high-hitter groups, and the older users required more taps and took longer to make a selection than the younger participants. These findings reinforce that older users do need better support for selecting menu items with a pen. This work presents a first step, but it only addresses the needs of the low hitters.

The existence of these two diametric error-prone groups makes developing general interaction techniques to assist users more challenging. Indeed, one of our main motivations in performing this work was that the results from the baseline study suggested missing just below errors could be addressed for those who make them without hindering those who do not. The results of this study indicate that a single predetermined solution will not likely meet the needs of all users. Additional techniques will likely be needed to detect the user's distribution before custom functionality can be offered. Some researchers have begun to examine methods for detecting real-world pointing problems [Hurst et al. 2008], but more work is needed to make this a viable approach to offering customized support. In the context of low- and high-hitting behavior, one approach might be to track the occurrence of programs and dialog boxes that are closed immediately after being opened from a menu, and whether a subsequent selection was made on either the item above or below. This might be an effective indication of difficulty selecting the correct item, and a way of predicting the type of difficulty the user is experiencing.

It is interesting that we saw no evidence of high-hitting behavior in the baseline study, while two-thirds of the older adults in the current study were high hitters. One possibility is that individuals in the two studies were different. Indeed, the older adults in this study scored significantly lower on the digit symbol substitution test than similarly aged individuals in the baseline study. Another possible explanation is that the continuous menu-selection task used in this study encouraged participants to initiate upward movement towards the menu head (to start the next trial) before fully completing the item selection (of the current trial). In contrast, the discrete task used in the baseline study required participants to move towards the center of the screen (down and right) after a selection, which may have encouraged missing just below.

In a real-world setting it is impossible to predict where the user will go after making a menu selection. In many instances they will move towards a dialog box (initiated by the menu selection), likely in the center of the screen. However, many other configurations are possible (e.g., the user may have multiple windows open, or be working in multiple applications), suggesting that in a real-world task we might expect to see an even wider range of behavior. Nonetheless, as some of our participants did demonstrate missing just below behavior despite the continuous task used, we believe that some users may have a downward tendency regardless of the task context. Perhaps the most important implication of this interstudy variability is that it highlights the need for increased replication in human-computer interaction research. We believe this may be especially important in research with older or disabled populations, as the high variability in these populations may make the outcomes of such studies especially sensitive.

8. FUTURE WORK

One area for future work is developing a better understanding of what factors influence the different error types observed in this study. Understanding what causes a user to be a low or high hitter may shed light onto how these different user types can be identified and supported. Scanning direction and hand occlusion of the contents are two possible reasons for why errors may occur, and individual differences in hand posture may account for the opposing (low versus high) categories. We have not investigated these possibilities experimentally; however, we believe they are promising avenues for further exploration as each would have unique implications for design. If scanning direction were to prove to be a factor, we would expect that menus anchored from below (such as the Windows Start menu) might impact the selection patterns. If hand occlusion was found to be a factor, it would suggest that wider menus, with more whitespace to the right (or to the left, for left-handed users) would ease errors for both low and high hitters without increasing the distance to any of the items in the menu.

Parallax is also a possible explanation, though the results of our baseline study suggest otherwise in that missing just below was menu specific; we did not find similar results for the tapping task. Furthermore, in the current study, we calibrated the tablet to each participant, which should have addressed any parallax. It is worth noting, however, that the built-in calibration utility used may not be sufficient for older adults. Though four calibration points may be

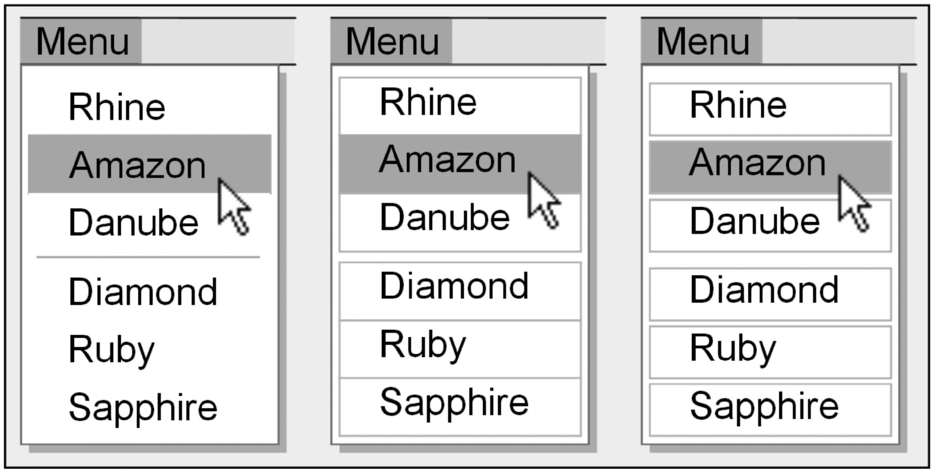

Fig. 11. Proposed visual presentations for the reassigned edge (center) and deactivated edge (right) approaches as compared to a regular menu (left). For the reassigned edge approach, the label is shifted up within the target area. For both approaches, the target boundaries are clearly marked by borders.

sufficient for detecting an offset in a consistent able-bodied individual, for a user with more movement variability it may be insufficient.

Another avenue for further work is to improve the visual appearance of our approaches. The deactivated edge approach, in particular, was not liked by the participants. One likely factor contributing to this response is confusion over what was happening when taps were ignored. For the evaluation, we did not explain to participants the differences between the interfaces because we did not want them actively trying to adapt. We additionally maintained a consistent visual representation across the interfaces so that visual appearance would not be a confounding factor. Thus, it is possible that a better understanding of why taps are being ignored, coupled with better visual and auditory feedback, may improve subjective opinion of the deactivated edge condition. In addition, some of the comments made with respect to the deactivated edge condition indicate that some participants need better feedback informing them when contact has been made with the screen. This feedback could be provided by adding a pen-down visual response, an auditory response for all pen-down actions, a tip-switch [Buxton 1990] to the physical pen to give it a clicking feel, or a combination of these approaches.

In general, the visual representations of both the deactivated and the reassigned edge conditions could be improved. For the deactivated edge approach, the target boundaries could be better delineated to make it clear where the target is active and where it is not. For the reassigned edge technique, a better approach may be to shift the label up rather than shifting the motor region down. Figure 11 demonstrates how these ideas might be achieved in an interface. Further research could investigate whether and how much these modifications would increase the effectiveness and reception of the approaches.

9. CONCLUSION

To date, pen-based target acquisition research has mostly focused on the needs of younger users by improving efficiency and designing novel techniques which extend the capabilities of the interface. In contrast, addressing basic target acquisition difficulties has not received very much attention. We argue that it is essential to address errors because they carry a high cost for recovery and are overly frustrating, especially for older users. Time-savings mostly benefit expert users by offering small additive savings. Error reduction mostly helps those users who, like many older adults, are easily confused and discouraged by errors. In this article we sought to address missing just below, a difficulty some users experience when selecting menu items with a pen-based input device. Missing just below occurs when a user's tap distribution is downwardly shifted, resulting in frequent erroneous selection of the top edge of the item below the target item, and infrequent selection of the corresponding top edge region of the target item itself.

We designed and developed two possible approaches for addressing missing just below. In the *reassigned edge* approach, input on the top edge of a menu item resulted in selection of the item above, while in the *deactivated edge* approach, input on the top edge was ignored. An evaluation comparing the effectiveness of these two approaches relative to a standard menu revealed that only the deactivated edge approach showed an overall benefit in terms of addressing missing just below. The reassigned edge approach did reduce missing just below errors, but, the number of errors it introduced (on the top edge of the target itself) negated this benefit. Further analysis of our data revealed that individual differences played a large role in our performance results. In contrast to the baseline study, many of our participants did not make missing just below errors. Instead, we found that there were three distinct user types: The low hitters, the high hitters, and the neutrals. Only the low hitters made missing just below errors, and for them, preliminary analysis suggests both of our approaches were beneficial. The high hitters, in contrast, had difficulty with errors on the item above the target. The reassigned edge approach further impeded them, and though the deactivated edge approach did not introduce errors for the high hitters, it did not help them either. The neutrals rarely made any errors, and thus were neither helped nor hindered by our interfaces.

Though our interfaces were not as effective overall as we had hoped, this research highlights the need for better support, especially for older users. All but one of the older users fell into one of the error-prone groups. Pen-based devices are increasingly being used in the development of assistive technology for a variety of age-related impairments. For this to be a viable approach, the accessibility of these devices needs to be improved, with a focus on reducing errors, and ensuring adequate undo facilities for correcting errors when they do occur. In this article, we examined one source of error and found some preliminary evidence for how this error can be addressed for the low hitters. However, additional research is needed to consider the practical implications of deploying these techniques in real-world applications, and to expand upon them to address the needs of the high hitters.

ACKNOWLEDGMENTS

The authors wish to thank L. Findlater for feedback on early drafts. We are also grateful for all the helpful comments and suggestions from the editors and anonymous reviewers.

REFERENCES

ACCOT, J. AND ZHAI, S. 2002. More than dotting the i's — Foundations for crossing-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'02)*. ACM, New York, 73–80.

AHLSTRÖM, D. 2005. Modeling and improving selection in cascading pull-down menus using Fitts' law, the steering law and force fields. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*. ACM, New York, 61–70.

AHLSTRÖM, D., HITZ, M., AND LEITNER, G. 2006. An evaluation of sticky and force enhanced targets in multi target situations. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction (NordiCHI'06)*. ACM, New York, 58–67.

BEDERSON, B. B. 2000. Fisheye menus. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST'00)*. ACM, New York, 217–225.

BIRDI, K. S. AND ZAPF, D. 1997. Age differences in reactions to errors in computer-based work. *Behav. Inf. Technol. 16*, 6, 309–319.

BLANCH, R., GUIARD, Y., AND BEAUDOUIN-LAFON, M. 2004. Semantic pointing: Improving target acquisition with control-display ratio adaptation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04)*. ACM, New York, 519–526.

BUXTON, W. 1990. A three-state model of graphical input. In *Proceedings of the 3rd IFIP International Conference on Human-Computer Interaction (INTERACT'90)*. North-Holland Publishing, Amsterdam, The Netherlands, 449–456.

CALLAHAN, J., HOPKINS, D., WEISER, M., AND SHNEIDERMAN, B. 1988. An empirical comparison of pie vs. linear menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'88)*. ACM, New York, 95–100.

CHARNESS, N., BOSMAN, E. A., AND ELLIOTT, R. G. 1995. Senior-Friendly input devices: Is the pen mightier than the mouse? In *103rd Annual Convention of the American Psychological Association Meeting*. APA, Washington, DC.

CHARNESS, N., HOLLEY, P., FEDDON, J., AND JASTREMBSKI, T. 2004. Light pen use and practice minimize age and hand performance differences in pointing tasks. *Hum. Factors 46*, 3, 373–384.

COCKBURN, A. AND GIN, A. 2006. Faster cascading menu selections with enlarged activation areas. In *Proceedings of Graphics Interface (GI'06)*. Canadian Information Processing Society, Toronto, Ontario, Canada, 65–71.

COCKBURN, A., GUTWIN, C., AND GREENBERG, S. 2007. A predictive model of menu performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, New York, 627–636.

CRAIK, F. I. M. AND SALTHOUSE, T. A., Eds. 1992. *The Handbook of Aging and Cognition*, 2nd ed. Erlbaum, Hillsdale, NJ.

CZAJA, S. AND SHARIT, J. 1998. Age differences in attitudes toward computers. *J. Gerontol. Psychol. Sci. 53*, 5, 329–340.

FITTS, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of human movement. *J. Exper. Psychol. 47*, 381–391.

GREENSTEIN, J. L. 1997. *Handbook of Human-Computer Interaction*, 2nd edition ed. Elsevier, Amsterdam, 1317–1348.

GREGOR, P., NEWELL, A. F., AND ZAJICEK, M. 2002. Designing for dynamic diversity: Interfaces for older people. In *Proceedings of the 5th International ACM SIGACCESS Conference on Computers and Accessibility (Assets'02)*. ACM, New York, 151–156.

GROSSMAN, T. AND BALAKRISHNAN, R. 2005. The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*. ACM, New York, 281–290.

GUIARD, Y., BLANCH, R., AND BEAUDOUIN-LAFON, M. 2004. Object pointing: A complement to bitmap pointing in GUIs. In *Proceedings of Graphics Interface (GI'04)*. Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 9–16.

HAALAND, K. Y. AND HARRINGTON, D. L. 1998. *Neuropsychology*. Springer, New York, 421–437.

HAWKEY, K., INKPEN, K. M., ROCKWOOD, K., MCALLISTER, M., AND SLONIM, J. 2005. Requirements gathering with Alzheimer's patients and caregivers. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility (Assets'05)*. ACM, New York, 142–149.

HEATH, M., ROY, E. A., AND L.WEIR, P. 1999. Visual-Motor integration of unexpected sensory events in young and older participants: A kinematic analysis. *Developm. Neuropsychol. 16*, 2, 197–211.

HOURCADE, J. P. AND BERKEL, T. R. 2008. Simple pen interaction performance of young and older adults using handheld computers. *Interact. Comput. 20*, 1, 166–183.

HURST, A., MANKOFF, J., AND HUDSON, S. E. 2008. Understanding pointing problems in real world computing environments. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility (Assets'08)*. ACM, New York, 43–50.

KEATES, S. AND TREWIN, S. 2005. Effect of age and Parkinson's disease on cursor positioning using a mouse. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility (Assets'05)*. ACM, New York, 68–75.

KETCHAM, C. J. AND STELMACH, G. E. 2004. *Technology for Adaptive Aging*. National Academies Press, Washington, DC, 64–92.

KOBAYASHI, M. AND IGARASHI, T. 2003. Considering the direction of cursor movement for efficient traversal of cascading menus. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST'03)*. ACM, New York, 91–94.

KURTENBACH, G. AND BUXTON, W. 1993. The limits of expert performance using hierarchic marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'93)*. ACM, New York, 482–487.

LEE, M. L. AND DEY, A. K. 2007. Providing good memory cues for people with episodic memory impairment. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (Assets'07)*. ACM, New York, 131–138.

MACKENZIE, S. I. 1992. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact. 7*, 1, 91–139.

MASSIMI, M., BAECKER, R. M., AND WU, M. 2007. Using participatory activities with seniors to critique, build, and evaluate mobile phones. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (Assets'07)*. ACM, New York, 155–162.

MIZOBUCHI, S. AND YASUMURA, M. 2004. Tapping vs. circling selections on pen-based devices: Evidence for different performance-shaping factors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04)*. ACM, New York, 607–614.

MOFFATT, K., MCGRENERE, J., PURVES, B., AND KLAWE, M. 2004. The participatory design of a sound and image enhanced daily planner for people with aphasia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04)*. ACM, New York, 407–414.

MOFFATT, K. A. AND MCGRENERE, J. 2007. Slipping and drifting: Using older users to uncover pen-based target acquisition difficulties. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (Assets'07)*. ACM, New York, 11–18.

MYNATT, E. D., ESSA, I., AND ROGERS, W. 2000. Increasing the opportunities for aging in place. In *Proceedings of the Conference on Universal Usability (CUU'00)*. ACM, New York, 65–71.

NUSSBAUM, P. D. 1998. *Neuropsychology*. Springer, New York, 83–105.

OAKLEY, I. AND O'MODHRAIN, S. 2005. Tilt to scroll: Evaluating a motion based vibrotactile mobile interface. In *Proceedings of the 1st Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'05)*. IEEE Computer Society, Los Alamitos, CA, 40–49.

POTTER, R. L., WELDON, L. J., AND SHNEIDERMAN, B. 1988. Improving the accuracy of touch screens: An experimental evaluation of three strategies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'88)*. ACM, New York, 27–32.

RABBITT, P. 2002. Consciousness is slower than you think. *Quart. J. Exper. Psychol. 55A*, 4, 1081–1092.

REKIMOTO, J. 1996. Tilting operations for small screen interfaces. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology (UIST'96)*. ACM, New York, 167–168.

REN, X. AND MORIYA, S. 2000. Improving selection performance on pen-based systems: A study of pen-based interaction for selection tasks. *ACM Trans. Comput.-Hum. Interact. 7*, 3, 384–416.

ROWE, M., LANE, S., AND PHIPPS, C. 2007. Carewatch: A home monitoring system for use in homes of persons with cognitive impairment. *Topics Geriatric Rehabil. Smart Technol. 23*, 1, 3–8.

SEARS, A. AND SHNEIDERMAN, B. 1994. Split menus: Effectively using selection frequency to organize menus. *ACM Trans. Comput.-Hum. Interact. 1*, 1, 27–51.

SMITH, M. W., SHARIT, J., AND CZAJA, S. J. 1999. Aging, motor control, and the performance of computer mouse tasks. *Hum. Factors J. Hum. Factors Ergonomics Soc. 41*, 389–396.

SOUKOREFF, R. W. AND MACKENZIE, S. I. 2004. Towards a standard for pointing device evaluation, Perspectives on 27 years of Fitts' law research in HCI. *Int. J. Hum.-Comput. Studies 61*, 751–789.

SPREEN, O. AND STRAUSS, E. 1998. *A Compendium of Neuropsychological Tests: Administration, Norms, & Commentary*, 2nd ed. Oxford University Press, New York.

TIFFIN, J. AND ASHER, E. J. 1948. The Purdue pegboard: Norms and studies of reliability and validity. *J. Appl. Psychol. 32*, 234–247.

TSANDILAS, T. AND SCHRAEFEL, M. C. 2007. Bubbling menus: A selective mechanism for accessing hierarchical drop-down menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, New York, 1195–1204.

UTTL, B., GRAF, P., AND COSENTINO, S. 2000. Exacting assessments: Do older adults fatigue more quickly? *J. Clinical Exper. Neuropsychol. 22*, 496–507.

VENOLIA, D. AND NEIBERG, F. 1994. T-cube: A fast, self-disclosing pen-based alphabet. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'94)*. ACM, New York, 265–270.

VOGEL, D. AND BAUDISCH, P. 2007. Shift: A technique for operating pen-based interfaces using touch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, New York, 657–666.

WALKER, N., PHILBIN, D. A., AND FISK, A. D. 1997. Age-Related differences in movement control: Adjusting submovement structure to optimize performance. *J. Gerontol. Psychol. Sci. 52B*, 1, 40–52.

WALKER, N. AND SMELCER, J. B. 1990. A comparison of selection time from walking and pull-down menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'90)*. ACM, New York, 221–226.

WECHSLER, D. 1981. *Wechsler Memory Scale-Revised Manual*. The Psychological Corporation, New York.

WILKINSON, R. T. AND ALLISON, S. 1989. Age and simple reaction time: Decade differences for 5,325 subjects. *J. Gerontol. Psychol. Sci. 44*, 2, 29–35.

WORDEN, A., WALKER, N., BHARAT, K., AND HUDSON, S. 1997. Making computers easier for older adults to use: Area cursors and sticky icons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'97)*. ACM, New York, 266–271.

WU, M., BAECKER, R., AND RICHARDS, B. 2007. *Universal Usability*. John Wiley & Sons, West Sussex, 317–356.

YAN, J. H. 2000. Effects of aging on linear and curvilinear aiming arm movements. *J. Exper. Aging Res. 26*, 4, 393–407.

ZHAI, S., KONG, J., AND REN, X. 2004. Speed-Accuracy tradeoff in Fitts' law tasks—On the equivalency of actual and nominal pointing precision. *Int. J. Hum.-Comput. Studies 61*, 6, 823–856.

Received May 2008; revised February 2009; accepted February 2009