Addressing Age-Related Pen-Based Target Acquisition Difficulties

by

Karyn Anne Moffatt

B.A.Sc. Computer Engineering, The University of British Columbia, 2001

M.Sc. Computer Science, The University of British Columbia, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Computer Science)

The University of British Columbia

(Vancouver)

January 2010

Abstract

Technology is increasingly being promoted as a means of addressing age-related cognitive and sensory impairments and enabling seniors to live more independently. Pen-based devices such as Personal Digital Assistants and Tablet PCs are appealing platforms for these endeavors because they are small, mobile, and powerful. Relative to the mouse, pen-based devices have been shown to be particularly beneficial for older adults. However, in terms of garnering wide-spread adoption, the mouse has historically dominated, leading researchers to focus chiefly on identifying and addressing its age- and motor-related limitations. In contrast, pen-based limitations for older users have been relatively unexplored. This thesis begins to fill that gap in the literature.

Our first experiment, an empirical evaluation of pen-based target acquisition across the adult lifespan, identified three main sources of pen-based target acquisition difficulty—missing-just-below, slipping, and drifting—and demonstrated how these difficulties vary across task situation and age. In addition, this work showed that including older adults as participants can help uncover general pen-interaction problems: the missing-just-below and drifting difficulties were evident in both younger and older users alike.

We next developed seven new target acquisition techniques to improve pen-based interaction, specifically addressing the three difficulties identified, and particularly targeting older adults. Our techniques built upon existing mouse-based techniques developed for older users and pen techniques for younger users. In total, we conducted three experiments to evaluate the seven new pen-based techniques: Reassigned and Deactivated (for missing-just-below), Tap and Glide (for drifting), and Steady, Bubble, and Steadied-Bubble (for slipping). Through these

evaluations, we established where our proposed designs were successful at reducing errors, and where further refinement is needed.

Finally, we reflected on our findings across studies to identify age-related, contextual, and technological factors which contributed to our results. These factors help illuminate the underlying reasons for pen-based targeting difficulties and shed light onto areas still needing attention. Overall, the results of this research support our main thesis that the accessibility of pen-based interfaces can be improved for older adults by first examining the sources of age-related acquisition difficulty, and then using the results of this examination to develop improved techniques.

Table of Contents

List of Tables

List of Figures

Acknowledgments

First, I would like to thank my supervisor, Dr. Joanna McGrenere, for her motivation and support. Joanna's dedication, diligence and enthusiasm for research are inspiring. She is the model of an excellent supervisor, and I feel very fortunate to have had her as mine.

The members of my supervisory committee, Dr. Peter Graf, Dr. Alan Mackworth, and Dr. Shari Trewin, have played an important dual role of critic and fan. Their astute feedback challenged me to be a better researcher, and their words of encouragement kept me moving forward. The additional members of my examining committee, Dr. Jim Little, Dr. Alan Kingstone, and Dr. Anind Dey provided many insightful comments, greatly improving the quality of this thesis. I would also like to thank Dr. Kelly Booth who, though never an official member of my committee, was always generous with his time and expertise. Undergraduate research assistants, Justine Yang, Sandra Yuen, and Natalie Forssman helped run many of the studies in this thesis; it was a pleasure working with them.

The research in this thesis was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institute of Health Research (CIHR); I am very grateful for this support. I am also especially grateful to the many participants who donated their time and energy to this project.

I have been incredibly lucky to be a member of the Interaction Design Reading Group. There are lots of really smart (and generous) people in this group, and I have benefited immensely from being part of it. They have challenged me to think more critically, and have provided insightful feedback on research, paper drafts, and practice presentations. I would particularly like to thank Tony Tang and Leah Findlater. Tony and Leah have been there since the beginning (which was a really long time ago). In addition to being generous with their smarts, they have kept me sane throughout this process. Leah shared many things with me throughout graduate school (from supervisors to offices to hotel rooms), and as such has witnessed all the ups and downs. She made the good times better and the bad times tolerable.

My family and friends have been overwhelmingly supportive of my graduate studies. I would especially like to thank my parents, Rory and Lorri Moffatt; it has been a long haul and they have been behind me every step of the way. I have been fortunate to have many relatives on whom I could count for input; my grandma and papa, Audrie and Tony Vernerey, have been particularly helpful. My siblings, Anthony and Krisi Moffatt, also deserve credit for keeping my ego in check.

Last, but certainly not least, I would like to thank my partner, Geoffrey Lefebvre. I could never have done this thesis without his selfless support and relentless encouragement. He has been lovingly patient every bump along the way. (And there were many!) It has not gone unnoticed that he always found a reason to work late when I did and was always there when I needed a mental health break. I admire his insight and tenacity, and have learned a great deal from him.

Dedication

To my parents.

Statement of Collaboration

With the exception of Chapter 6, I am the primary contributor to all aspects of this research, which was performed under the supervision of Dr. Joanna McGrenere. Under my direction, several undergraduate research assistants helped with the execution of this research. Justine Yang ran study sessions for the follow-up study presented in Chapter 3. Sandra Yuen helped with the software development, contributed feedback on the study design, and ran the study sessions for Technique Study Two (Chapter 5). Natalie Forssman ran the study sessions for Technique Study Three (Chapter 7).

The work reported in Chapter 6 is the result of a collaboration among Dr. Shari Trewin, Dr. Simeon Keates, and myself during my internship at IBM T.J. Watson during the summer of 2005. Throughout this work, wherever I took the lead, Drs. Trewin and Keates provided input in a supervisory capacity. The identification of the research problem and the initial idea for the Steady Clicks technique occurred prior to my involvement and was based on a study led by Paradise, Trewin, and Keates [82]. Dr. Trewin developed a skeleton prototype to demonstrate their idea, and with this as a starting point, I developed the core Steady Clicks functionality and took the lead on identifying and exploring the various design options described in Section 6.1.2. For the evaluation, I took the lead on designing the protocol, including the design and implementation of the study task and the design of all of the study materials. The study was conducted primarily by Dr. Trewin; however, I initiated some of the recruiting work contacting relevant organizations, and led the early evaluation sessions. The data were analyzed by Drs. Trewin and Keates. Dr. Trewin was the primary author on the co-authored paper [109]; however, Sections 6.1.2 and 6.2 are largely based on an internal report I authored.

Chapter 1

Introduction

1.1 Motivation

Technology is increasingly being promoted as a means of addressing cognitive and sensory impairments and enabling individuals to live more independently (e.g., [47, 67, 70, 77, 79, 93, 130]). Pen-based devices 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 [39]. Relative to the mouse, pen-based devices have been shown to be particularly beneficial for older adults [22, 23], suggesting promise for a wide range of users.

In contrast to touch technology (used with either a finger or passive stylus), inductive pen technology offers several advantages. It uses a dedicated pen, which reduces the likelihood of accidental selections. The use of a pen instead of a finger also makes selections more precise and reduces somewhat the problem of hand occlusion. Moreover, inductive pen technology is more powerful; it senses the angle of the pen relative to the screen, the pressure of the pen on the screen, and the location of the pen both when it is touching the screen's surface and when it is hovering in near proximity.

However, to be viable for these applications, older individuals need to be able to perform basic tasks with ease. Many of the target users for these technologies will also have associated motor impairments impeding their ability to interact with small devices. For example, in our work designing mobile technology for cognitively impaired individuals [77], we informally observed many struggling with target acquisition using a stylus.

Basic target acquisition, such as selecting small icons or menu items, is arguably the most important and prevalent form of interaction. Other interactions, such as dragging, are also important, but are less dominant. Moreover, difficulty with basic target acquisition can impede adoption. Thus, we believe that improving basic targeting is an important first step towards the development of accessible pen-based technology. These observations motivated us to gain a better understanding of the challenges inherent to pen-based target acquisition and to 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 abilities, 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 younger able-bodied users over adapting basic target acquisition techniques to support the needs of a wide range of users. Thus, despite considerable research aimed at developing improved pen-based target acquisition techniques (e.g., [1, 52, 73, 89]), point and tap—whereby selection is determined based on the location of the tap up—remains the de facto standard. To clarify, in point and tap, selection is made by (i) tapping down, (ii) possibly moving the pen, and (iii) tapping up.

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. However, it is not clear that rates observed in laboratory settings truly reflect real-world frequencies. Research has not attempted to measure real-world error rates, but it has shown that small changes in task instruction can have a large impact on the

[1] Fitts' type studies typically aim to have an error rate of approximately 4% [68]

observed accuracy [133]. This shows that error rates are not fixed, suggesting they may fluctuate depending on the task context.

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 [12, 33, 85]. 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. This is often lost in laboratory studies as it is impossible to encapsulate the true cost of recovery time even when penalties are included.

Although most modern programs offer extensive undo functionalities, these facilities do not necessarily address all costs associated with making an error. 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 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).

1.2 Thesis Goals

The high-level objective of this thesis was to increase the accessibility of pen-based technology, especially for older adults, by investigating mechanisms for assisting users to select more easily. Specifically, our goal was to answer the following two questions:

1. What types of difficulties and errors do users encounter while using a pen-based device to acquire targets? Do these difficulties vary in terms of their nature and severity with age? Do they vary over different task situations?

2. Can we build new interaction techniques—by extending and combining existing techniques for younger users and mouse interaction—to address the age-related pen-based target acquisition difficulties identified, and thereby improve pen-based interaction for older adults?

1.3 Thesis Approach and Overview

To answer these questions, we conducted a series of five controlled laboratory studies. In the Baseline Study, we collected quantitative data on pen-based target acquisition, using standard point and tap mechanisms, across a range of ages and task situations. Technique Studies One, Two, and Three build on these findings to design and evaluate interaction techniques to address the difficulties uncovered. In designing these techniques, we drew from existing techniques designed for younger users and mouse interaction, modifying and combining them in novel ways. We additionally present the Mouse Study, in which we depart slightly from our core thesis goals to present Steady Clicks, a new interaction technique designed to help prevent mouse-based slipping errors for individuals with motor-impairments. We include this study in the thesis, because it motivates one of the approaches examined in Technique Study Three.

1.3.1 Identifying Pen-Based Target Acquisition Difficulties Across the Adult Life-Span

The first goal of this research was to gather information on the underlying causes of target acquisition difficulties across the adult lifespan. To fulfill this goal, we conducted a controlled laboratory experiment (the Baseline Study) with 36 participants from three age-groups (18–54, 55–69, and 70+; 12 participants in each), and two selection tasks (multi-dimensional tapping and menu item selection). In this study, we identified three primary target acquisition difficulties: (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 directly below the target item. Slipping was specific to older users and common to both tasks, while drifting and missing-just-below were specific to the menu selection task, but affected users of all ages. An additional finding was that including older users as participants allowed us to uncover pen-interaction deficiencies that we would likely have missed otherwise. Drifting and missing-just-below were not behaviors we predicted; rather our observations of the older users during the experimental sessions prompted us to investigate them, thus revealing their general impact.

1.3.2 Design and Evaluation of Techniques to Support Pen-Based Menu Interaction

Our next two studies investigated techniques to support accurate menu interaction.

In Technique Study One, we investigated techniques to facilitate menu item selection by reducing missing-just-below errors. Missing-just-below 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 desired item 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. That is, in the reassigned edge approach input on the top edge of an item resulted in selection of the item above, whereas in the deactivated edge approach, input in this region was ignored. In a laboratory evaluation of 24 participants from two age groups (younger 18–30 and older 65+; 12 participants in each), only the deactivated edge approach showed any 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 the idea 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.

In Technique Study Two, we examined methods to prevent drifting, a menu navigation difficulty that occurs when a user accidentally hovers over an adjacent menu, causing the currently focused menu to close and the adjacent one to open. We proposed two approaches to address drifting: *Tap*, which prevents drifting by requiring an explicit tap to switch menus, and *Glide*, which uses a distance threshold to delay switching, thereby reducing the likelihood of a drift. We performed a comparative evaluation of these approaches against a control interface, with 24 participants from two age groups (younger 18–30, older 65+; 12 participants in each). Tap was effective at reducing drifts for both groups, but the reduction was greatest for the older group. Tap closed the performance gap between younger and older users such that with Tap the older group's performance was comparable to that of the younger group. In terms of preference, Tap was ranked highly by the older participants, but was not well received by the younger participants who felt it was slow (though this was not supported by the performance data). Glide surprisingly did not show any performance improvement. Additional research is needed to determine if the negative findings for Glide are a result of the particular threshold used, or reflect a fundamental flaw in the Glide approach.

1.3.3 Design and Evaluation of Techniques to Support General Pen-Based Target Acquisition

We next turned our attention to general pen-based target acquisition, paying particular attention to the slipping difficultly.

First, we present *Steady Clicks* (the Mouse Study), a mouse-based interaction technique designed to help motor impaired individuals who have difficulty holding the mouse still while clicking. Though we focus mostly on older adults in this thesis, increasing the accessibility of technology for individuals with motor impairments was part of the original motivation for this thesis research [77], and this work represents a step in that direction. Steady Clicks suppresses these slip errors by freezing the cursor during mouse clicks. Evaluation with 11 motor-impaired users found that Steady Clicks reduced errors, and in particular slipping errors. Overall task performance times were significantly improved for the six participants with the highest slip rates. Nine participants preferred Steady Clicks to the unassisted (control) condition.

In Technique Study Three, we expand upon the Steady Clicks technique to address pen-based slipping behavior. Specifically, we investigated two cursor enhancements: a *Steadied* cursor (similar to Steady Clicks), and a *Bubble* cursor (where the activation area of the cursor expands to facilitate selection), and additionally, a combined *Steadied-Bubble* cursor. We were interested in not only how well these techniques prevent slip errors, but also in their ability to reduce miss errors (where the pen both lands and lifts outside the target bounds) as these two error types account for the majority of errors observed in the tapping task of the Baseline Study. An evaluation with 12 younger (18–30) and 12 older (65+) users revealed that the effectiveness of each technique depended on the task parameters. Overall, techniques based on the Bubble cursor (i.e., Bubble and Steadied-Bubble) were most effective at reducing errors as the Bubble technique addresses both slip and miss errors. However, when targets were directly adjacent (i.e., there was no whitespace between targets) only techniques based on a Steadied cursor (i.e., Steadied and Steadied-Bubble) provided benefit (by reducing slip errors). There was no evidence of a cost associated with combining the two techniques, so the best support was offered by the combined Steadied-Bubble cursor. It prevented both miss and slip errors when targets were separated by whitespace and prevented slip errors when they were directly adjacent. Both age groups benefited from these designs, but the older participants were particularly helped by them.

1.3.4 Towards a Model of Age-Related Pen-Based Interaction Difficulty

We finally revisit the topic of pen-based interaction difficulty. Drawing on literature on aging and motor-control, we reflect on our combined findings across studies to expand on our understanding of the difficulties uncovered in the Baseline Study, and identify the relevant factors that would play a role in a model of pen-based interaction difficulty. Such a model would be useful both for informing the design of accessible pen-based techniques, and for aiding practitioners in predicting the accessibility of techniques. We conclude this chapter by outlining additional broad areas for future work, which provide interesting new avenues for investigation.

1.4 Summary of Thesis Contributions

This thesis provides three main contributions. We provide a high-level description of each contribution here and elaborate on them further in our conclusions chapter (Chapter 9).

1. We present the first in-depth examination of the underlying causes for age-related pen-based target acquisition difficulties. Through an empirical evaluation, we identified three novel pen-based difficulties—slipping, drifting, and missing-just-below—and demonstrated how they varied across task situation and age.

2. We introduce seven new pen-based target acquisition techniques to address the difficulties identified. We based our designs on techniques developed for the mouse and younger users, modifying and combining them in novel ways. We then evaluated these techniques in controlled laboratory studies to establish the ability of each to reduce errors, and to identify areas where further refinement is needed.

3. We show how systematically including age as an explicit factor can lead to improved pen-based interfaces overall. We included a range of ages in every stage of this work, and through this approach, we were able to carefully document where the needs of the older adults align with—and where they diverge from—the needs of the younger adults. Ultimately, this led to richer findings and improved interfaces for older and younger users alike.

We also claim two secondary contributions. The first is the design and evaluation of Steady Clicks, a novel mouse-based technique for supporting individuals with motor-impairments. While somewhat outside the core goals of this thesis, this work was highly influential on our subsequent pen-based technique development. The second is the identification of factors which contributed to the target acquisition difficulties we observed. These factors represent a first step towards the development of a model of age-related pen-based target acquisition difficulty.

1.5 Thesis Outline

We begin in Chapter 2 by presenting background and previous work related to aging and target acquisition. Chapter 3 describes our Baseline Study to identify pen-based target acquisition difficulties. In Chapters 4, 5, and 7 we investigate techniques for addressing each of the main difficulties identified in the Baseline Study. In Chapter 6, we present the design and evaluation of Steady Clicks, a mouse-based target acquisition technique that forms the basis for one of the techniques explored in Chapter 7. In Chapter 8, we reflect on our findings across studies and outline factors which contributed to the difficulties we observed. In this chapter, we also describe avenues for further research. Chapter 9 summarizes the thesis work, and highlights its contributions in greater detail. Finally, there are a number of appendices included at the end of the thesis, which detail the materials used in the studies.

The majority of the work presented in this thesis has already been published: Chapter 3 in the Proceedings of ACM ASSETS 2007 [74], Chapter 4 in ACM Transactions on Accessible Computing [75], Chapter 5 in the Proceedings of ACM ASSETS 2008 [78], and Chapter 6 in the Proceedings of ACM ASSETS 2006 [109]. Chapter 7 has been accepted for publication in the Proceedings of ACM CHI 2010 [76]. More details are provided in Appendix A.

Chapter 2

Background and Related Work

We begin our coverage of the literature with an overview of the motor control theories that underlie many of the interaction techniques discussed in this chapter. We then provide an overview of the effects of aging on motor skill to highlight the reasons for age-related differences in targeting ability. These differences also shed light onto why some techniques are successful with younger users but do not work well for older adults and can help predict how well techniques evaluated with younger users will perform when used by older adults. We then describe research on direct interaction, focusing on pen and touch techniques. Because indirect pointing devices, and especially the mouse, have been the subject of much more attention historically, we also review work in that area, paying special attention to techniques that may also have applicability to pen interaction. Finally, many of the studies presented in this thesis involve menu selection, so we also review research in this area.

2.1 Movement Control

Fitts' law [36, 68] is an empirical model of human movement that uses information theory to quantify a pointing task's difficulty. Although it was originally developed in the context of real-world (direct) pointing tasks, it has been subsequently shown to be applicable to computer tasks with indirect input devices [20]. The law states that performance increases monotonically as the distance to the target decreases and the width of the target increases. That is, targets that are further away or smaller take longer to accurately select than targets that are closer or bigger. Fitts quantified these ideas empirically using a reciprocal pointing task, in which participants moved a metal tipped stylus back and forth between two conductive plates [36]. By systematically varying the width of the plates and the distance between them, he derived a mathematical relationship between task difficulty (ID), movement distance (D), and target width (W). The following provides the currently preferred Shannon formulation of this relationship [68, 99] (for other formulations see [36, 121]):

$$ID = log_2(\frac{D}{W} + 1) \tag{2.1}$$

Movement time increases linearly with ID according to the following equation, where a and b are empirically derived task-dependent constants.

$$MT = a + b \times ID \tag{2.2}$$

This law has been influential to research on pointing techniques as it strongly suggests the most promising avenues for improving target acquisition are to devise methods of reducing the distance to the target or increasing the target's width.

It can also be informative to consider the prevailing motor control models that aim to explain Fitts' law. The *iterative corrections* model proposes that movement consists of a series of discrete submovements towards the target, where each moved is triggered by a feedback loop indicating that the target has not yet been attained [32, 60] . In contrast, the *impulse variability* model asserts that movement is solely governed by an initial impulse that fires the limb towards the target [97]. The currently prevailing theory [92], the *optimized initial impulse* model [72], is a hybrid of these two earlier theories and better captures the range of findings covered in the literature [92, 126]. It posits that movement is governed first by a large ballistic movement towards the goal, and then, if this initial impulse does not reach the target, a series of small corrective movements is initiated to finish the action.

2.2 Effects of Aging on Movement Control

Aging leads to declines in a wide variety of cognitive and motor processes, which have been shown to contribute to reduced performance on movement tasks. Considerable research 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. We briefly review this area of research in this section, and then come back to this topic in Chapter 8, where we consider it in the context of our findings.

Movement is impeded by changes in joint and muscle characteristics. Older adults have reduced range of motion in many joints including their wrists [21]. They also have reduced maximum force capabilities [62] and a higher noise-to-force ratio [117, 122]. Older adults experience a general loss of muscle mass, particularly in terms of the fast twitch fibers that are used to generate short bursts of speed [62]. They have reduced proprioceptive capabilities (i.e., the ability to sense how one's body segments are oriented relative to one another), which are important for producing smoothly controlled movements. Changes in vision can also have a substantial and complex impact on motor function. Among many changes, aging is associated with decreased visual acuity, less efficient visual processing, increased susceptibility to glare, and reduced sensitivity to color [95, 96]. However, it should also be noted that older adults often compensate for deficits by relying more heavily on other capabilities or knowledge [94]. This makes it difficult to tease apart the exact impact of motor loss on task performance.

Age-related cognitive changes also impact motor performance. Notably, older adults process information more slowly than younger adults [7, 123], and as task complexity increases this age gap widens [62]. For example, older adults are roughly 26% slower than younger adults on a simple reaction time test, but 30–60% slower for a more complicated choice-reaction time test (with 2–4 choices) [62].

These changes affect targeting performance in a number of ways. Older adults are slower than younger adults overall, and are disproportionately slowed by increased difficulty, both in terms of increased movement distance and decreased target size [61]. Older adults also have very different movement profiles than younger adults. They demonstrate 30–70% lower peak velocities [58, 62], cover 10–70%

less distance with their primary movement, have a 20–40% longer deceleration phases [62], and make less smooth movements towards the target [131]. Their performance is also more variable [105]: they demonstrate higher variability in their movement trajectories, end-point positions, movement durations, peak velocity, and the ratio between acceleration and deceleration [28, 34, 62]. Finally, older adults also show a bias for accuracy over speed [94, 100, 106]. They make more pauses while homing in on the target [58], and they make more corrective movements once inside the target [117].

2.3 Direct Input Target Acquisition Techniques

In this section, we present research on pen- and touch-based interaction techniques. We note that some researchers have also explored the use of alternative input modalities, such as device orientation, for interacting with handheld devices (e.g., [81, 88]). These approaches are beyond the scope of our current investigation, but provide additional avenues for future research.

2.3.1 Pen-Based Techniques

To date, very little work has examined the use of pen-based systems with older adults. Charness et al. performed an age-related comparison of the mouse and the light-pen [22, 23], 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 vertical monitor, which required the pen to be held up unnaturally. Modern Tablet PC systems are designed to be more comfortable and, thus, should result in higher satisfaction.

Additionally, Hourcade and Berkel [52] compared two pen-based selection techniques, tapping and touching (selection if the pen touches the target at any time before tap up), across age groups. They found that for the smallest target size examined (3.8 mm) the oldest group was more accurate using touch, but that many found it more tiring. One limitation of this technique is that only one target can be ‘touched’ during selection; thus, when targets are directly adjacent it degrades to tap. It is also interesting that participants found it more tiring as tapping is a form of touching. It is possible participants did not fully understand the range of interactions that would count as touching, and felt obligated to do longer crossing actions. Nonetheless, this result has implications for other techniques (including circling and crossing, which are described later in this section), which do require longer contact with the screen.

More research has been devoted to developing improved pen-based interaction techniques for younger adults. However, in terms of overall performance gains, results have been modest. Novel techniques often only slightly outperformed standard tapping, or only in specific constrained situations. The most extensive examination to date was that done by Ren and Moriya [89]. They constructed a state transition model of pen-based target selection, and using that model, derived the following six selection techniques:

Direct On. Selection on tap down of the target at the pen down location (if one exists).

Slide Touch. Selection of the first selectable item touched after pen down.

Direct Off. Selection on tap up of the target at the pen up location.

Slide Off. Selection on tap up of the last target touched during contact.

The Direct Off technique corresponds to the standard point and tap used in most off-the-shelf interfaces. The final two techniques, described next, take advantage of the hover space of a Tablet PC; that is, the space above the screen within which the tablet can sense the location of the pen without contact with the screen. For these techniques, the targets are considered to be three dimensional entities that include the hover space above them.

Space On. Selection on tap down of the target at the pen down location (if one exists). This technique is an extension of Direct On, with the addition that the target is highlighted once the pen enters its hover space, providing feedback that it is targeted for selection.

Space Touch. Selection on tap down of the target whose hover space was last touched before the pen down. This technique is an extension of Space On, relaxed such that the pen can land on inactive whitespace outside the target bounds.

Their evaluation showed that for targets smaller than 1.8 mm that Slide Touch was best in terms of speed, accuracy, and subject preference. However, their evaluation used a single target, and they cautioned that this technique would not be suitable for densely populated displays. For this they recommended either Direct On or Direct Off, but noted that both require good hand/eye coordination.

Mizobuchi and Yasumura compared tapping to circling for a multi-target selection task [73]. With circling, selection is specified by drawing a circle around the target or targets. They hypothesized that circling would be faster and more accurate than tapping, but found that it was only better in the specific situation where targets formed a cohesive group with low shape complexity. This work would indicate that for single targets, tapping is superior to circling; however, the focus of this study was limited to university student subjects with normal dexterity. Further investigation is needed to explore whether or not circling could be beneficial to other user groups.

Accot and Zhai compared tapping to crossing (which is similar to the Slide Off technique [89] described above) and found crossing was at least as fast and had similar accuracy [1]. Although not outright better than tapping, they concluded crossing is a viable interaction technique and suggested there may be special situations in which it has specific advantages, including support for elderly or motor impaired users. They did not pursue these ideas, but other researchers have since followed up on some of them. Apitz and Gumbretiere applied crossing to a drawing program in CrossY. This work showed that crossing actions can be combined to create expressive and fluid multi-action sequences and used to create a full application [5]. More recently, Gajos and Wobbrock applied crossing to a mouse-based interface and showed that for motor-impaired individuals, crossing resulted in higher throughput and more fluid interactions [127].

Other work has also investigated novel pen-based interfaces for motor-impaired individuals. Barrier pointing leverages the elevated physical edges surrounding the screen to improve pointing accuracy [38]. In that regard it shares some similarity to EdgeWrite, which uses the physical edges of a template overlaid on the screen to facilitate pen-based text entry on a PDA [128]. While techniques developed for motor-impaired individuals may have some applicability to older users, a notable difference is that all of the above work has focused on individuals with a much greater degree of motor impairment than is associated with normal aging.

A final class of techniques involves multi-action selection. For these techniques, a single, yet difficult to perform action (such as selecting a very small target), is replaced with a sequence of multiple less precise actions. The major drawback to these approaches is that the sequence itself is often complex. Zliding [86], and Pointing Lens [87] are pen-based examples of such techniques. Both provide support for zooming the target to facilitate selection. Switching between zooming and selection modes can be disruptive to the task, and while Zliding and Pointing Lens both provide methods for fluidly switching between these modes, their switching mechanisms require precise control over one of steadiness, pressure, or speed. Thus, it is uncertain how suitable they are for older users.

2.3.2 Touch-Based Techniques

Although direct touch is perhaps the most natural and intuitive form of interaction, it suffers from two major limitations: (1) occlusion of the target by the finger in the final stages of selection, and (2) ambiguity with regard to which part of the finger defines the selection point. Although pen interaction does present some occlusion challenges, the tip of the pen is much smaller than the finger, resulting in a more clearly defined selection point.

Offset cursor is an early but important effort towards addressing the above limitations [84]. With Offset cursor, an on-screen cursor is positioned just above the user's finger, upon contact with the screen. The offset between the finger and cursor is held constant so selections can be refined by sliding the finger on the screen, until the cursor is aligned with the target. While this technique did improve accuracy for small targets, indirectly positioning a cursor in this way is slow and loses many of the benefits of direct interaction. Moreover, targets along the bottom edge cannot be selected because of the upward offset.

Shift builds on the Offset cursor technique [116]. It creates a callout showing a copy of the occluded screen area (the area under the finger), augments it with a cursor and places it in a non-occluded region nearby. The user can then refine his selection by shifting his finger before lifting to confirm. Unlike Offset cursor, Shift is only invoked when necessary, thus improving performance when targets are large enough to be selected without assistance. Also, by optimally placing the callout, Shift makes it possible to select items located anywhere on the screen.

A subsequent technique, Escape, also improves upon the Offset cursor [132]. With Escape, targets are selected by gestures cued by target position and appearance. Targets are teardrop shaped and neighboring targets are uniquely oriented. To make a selection, the user lands near the target, and then gestures in the direction of the target's orientation. In an evaluation, Escape was significantly faster than Shift, and had a similar error rate. In terms of real-world applicability, there is class of applications, for which the escape technique seems feasible (e.g. marking locations of interest on a map); however, in general, constraining the shape and orientation of targets poses design challenges.

Multi-action selection techniques have also been popular for addressing the difficulties inherent in in selecting very small targets from a touchscreen. Like the pen-based Zliding and Pointing Lens approaches described above, Zoom-Pointing also provides a zooming mode to ease selection [4, 10, 44] . However, with Zoom-Pointing. the user must explicitly switch between zooming and pointing modes. In real applications (for example, in the commercial product Adobe Photoshop) this is achieved via a toolbar control. This sort of mode switching can be very disruptive to the task at hand, and is costly in terms of time, as the user must travel back and forth between the toolbar and the target region.

Other techniques have addressed this limitation by overlaying tuning widgets onto the work area. With Cross-Keys, the user taps in the approximate region of the target to deploy a crosshair augmented with four graphical arrow keys [4]. Tapping on an arrow key nudges the crosshair allowing for fine tuning of the selection point. A final tap on the crosshair itself activates selection. Although, Cross-Keys was shown in an evaluation to have low error rates, it was unusable near the screen edges (as there was not enough room to place the arrow keys), and was unpopular with users. Precision Handle [4] uses a lever metaphor such that large movements on one end, the input end, result in small precise movements on the other, the selection end. Precision Handle was shown to outperform Offset cursor in terms of speed and accuracy; however, like the other approaches, its ability to afford precision comes at the cost of simplicity.

Recently, researchers have begun to explore the use of the back of the device for touch interaction [8, 124]. So far these approaches have focused on enabling younger users to interact with increasingly smaller devices. However, it would be interesting to explore their ability to help older adults interact with more typically sized devices. Other researchers have investigated bimanual and multi-touch interaction techniques as a means of improving touch interaction [11, 15], but thus far, very little is known about the use of bimanual input for older adults, and there is some evidence that older adults have difficulty mastering interaction techniques that require both hands [6, 51].

2.4 Indirect Input Target Acquisition Techniques

To date, indirect input devices, such as the mouse, have received much more attention in the literature than direct devices such as pen and touch. In this section, we review this work, paying particular attention to those techniques that might have applicability to direct pen-based interaction.

Most research on improving mouse interaction has focused on easing cursor positioning. One technique that has shown some promise is to dynamically expand targets as the cursor approaches [24, 71]. Although an evaluation of this technique [71] showed that (younger) users benefited from the expansion even if it occurred after 90% of the distance to the target had been covered, this approach may be less promising for older users: Heath and Roy explored the effect of changing target size on younger and older participants, with an experiment where the target expanded or shrank immediately following the subject's first movement [48]. They found that unlike their younger counterparts, older individuals did not adjust their speed-accuracy strategy to the new target size but instead used their programmed response for the initial target. Moreover, accurate expansion requires the ability to predict the user's desired target early on in the interaction [71]. For pen interaction, the ability to remain outside the detectable range of the device until late in the interaction likely hinders this calculation. The variable movement patterns of older adults may further degrade the ability to make accurate predictions.

A related approach, area cursors [41, 56, 129], has been shown to have promise specifically for older adults [129]. The general idea of this technique is to replace the standard single hot spot cursor with a cursor that covers an expanded region. Thus, selection occurs if the target is covered (partially or fully) by the area cursor. In order to support selection from multiple proximate targets, several ideas have been proposed [41, 129]. Most notably, the Bubble cursor addresses this by dynamically resizing the cursor such that only one target is selectable [41]. In [41], this technique was shown to be extremely effective in assisting younger mouse users to home in on targets more easily. In Technique Study Three, the Bubble cursor is one of the techniques we investigate for pen-based interaction and older users. We note that one weakness that has limited the adoption of Bubble cursors in real-world applications is that they require knowledge of target location. This is not possible for all applications; for example, web browsers can not always determine the location of all selectable targets on a page.

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 [13]. That is, 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 [129]; Object Pointing takes this to the extreme, eliminating inactive space altogether [43]. A major difference between the work described in this paragraph 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 space.

2.5 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. Many of these techniques share some commonality with the general target acquisition techniques described in the previous section. For example, 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. This speeds up selection when the menu is very long by reducing scrolling. One problem with this approach is that it makes items very small in motor space, making selection difficult; thus, a focus lock mode is used to make a segment of the menu act as a normal menu [9]. Pie Menus instead minimize travel distance by arranging items circularly around the cursor [19, 64, 113]. As a final example, Split Menus reduce average targeting time by placing the most frequently used items at the top of the menu [98]. None of these techniques specifically address 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 older adults. Morphing Menus progressively grow frequently selected menu items over time, allowing them to borrow motor (and visual) space from neighboring infrequently used items [26]. This technique should reduce 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 [36] to 'Fittsize' menu items [118]. 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 [41] to enable quick selection of a predicted set of items [111]. 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 older individuals.

Another focus has been on improving the selection of submenu items in cascading menus. 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 we have not considered cascading menus in this research, we might expect that similar, if not greater, difficulties would exist for pen-based interaction and older adults.

Moreover, there is some similarity between the downward motion in the cascading menu problem and the missing-just-below difficulty identified in Chapter 3 and examined in Chapter 4. Though the underlying causes are likely different, it can still be informative to consider the solutions proposed for cascading menus. Cockburn and Gin [25] grew the activation area of each cascading item downward over non-cascading items allowing users to move more directly to the target item of the submenu. In Chapter 4 we present an approach to reducing missing-just-below errors that also allows target items to borrow activation space from non-target items (the reassigned edge approach), 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 [2, 3]. 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 [63]. 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.

2.6 Summary

In this chapter, we described background literature on motor control theories and the effects of aging on motor skill, and we reviewed related work on the development of interaction techniques. In particular, we reviewed three main areas of interaction technique research: mouse-based, touch-based, and pen-based. Relatively little research has investigated the development of pen-based interaction techniques, and even less has considered the uniques needs of older adults. Mouse-based interaction has historically dominated in terms of breadth of use, and accordingly it has received the most research attention. Some of the techniques developed for the mouse, including the Bubble cursor, may also be applicable to pen-based interaction. However, a large proportion of the techniques developed for the mouse depend on the ability to manipulate the control-display ratio of the cursor, which is not possible with a direct input device like the pen. With touch-based interaction, the main challenge is to enable selection of very small targets, particularly those that are smaller than the tip of the finger. Most of the techniques we reviewed did this by layering tuning functionality on top of the interaction. This add complexity to the interaction, leaving it unclear how suitable these techniques are for older adults. A common theme that emerged across all forms of input is that the vast majority of previous work on improving target acquisition has focused on the needs of younger users, and accordingly has tended to value speed over accuracy. This was particularly true of research aimed at improving menu selection.

Chapter 3

Baseline Study

Identifying Pen-Based Target Acquisition Difficulties Across the Adult Lifespan

In this chapter, we present the results of the Baseline study. This study was designed to address the first goal of this thesis, namely, to explore the types of difficulties users encounter while using a pen-based device, and to determine if these difficulties vary in terms of their nature and severity with age and task situation.

3.1 Introduction and Motivation

As discussed in Chapter 1, direct pen-based input takes full advantage of hand-eye coordination, and offers a familiar form of interaction [39]. These benefits have been shown to be particularly beneficial for older adults [22, 23]; however, this research was conducted on a light pen. Despite the advantages shown, the light pen's high cost relative to a standard mouse, coupled with the fatigue associated with using a pen on a vertical display surface (i.e., a standard monitor), left it largely unadopted. With current-day Tablet PCs and stylus-based PDAs gaining popularity, it now seems that pen input is finally in a position to succeed. However, in our own work designing mobile technology for older and motor-impaired individuals [77], we informally observed many struggling with target acquisition using a stylus. These observations motivated us to seek a better understanding of the challenges inherent to pen interaction, and to ascertain the extent to which age is a factor.

Although there has been a great deal of research aimed at developing improved target acquisition techniques, including a sizable amount directed specifically to the pen [1, 52, 73, 89], room for improvement remains: many users still experience difficulties, and standard point and tap remains the dominant technique. We note three limitations that span the majority of pen-based target acquisition research: (1) the narrow focus on young-healthy adults, who can more easily adapt to different techniques, (2) the focus on evaluation with a single, typically highly-constrained, task, and (3) the focus on designing and evaluating novel techniques over developing a deeper understanding of how users manage basic tapping.

In terms of the first limitation, there are many parameters, including a user's sensory and motor ability, that are likely to affect target acquisition and manipulation skill. Thus, a broader perspective can be gained by examining a range of users and abilities. Since aging leads not only to reduced capability, but also to greater variability [40], older adults may provide especially rich information. Such information could lead to improvements that benefit not only the older demographic but perhaps younger ones as well.

The second limitation relates to restricting the evaluation of techniques to one task. Although not exclusively used, the standard for comparing interaction techniques is a Fitts' tapping task [36, 103]. Its main advantage is that it provides well-understood measures of speed and accuracy. However, it only reflects very simple interaction with a single isolated target. Real-world applications require much more complicated forms of interaction. But more complex interactions are less well understood, harder to analyze, and often lead to less clear conclusions. Thus, we believe it is important to include multiple tasks to capture both concrete comparative measures, and complex interaction.

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

The work described in this chapter attempts to fill this niche by gathering information on the underlying causes of target acquisition difficulties. We used two tasks—a Fitts' tapping task and a menu selection task—to provide a range of interactions to examine. We involved users from three different age groups to help us understand both general shortcomings, and those issues unique to older users. Specifically, the goal of this work was threefold: (1) to perform a detailed analysis of the types of difficulties users encounter while tapping to acquire targets, (2) to determine if these difficulties vary over task situations, and (3) to determine if these difficulties vary in terms of their nature and severity with age.

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

3.2 Experimental Methodology

To address the aforementioned goals, we performed a multi-task evaluation of pen-based target acquisition across multiple age groups.

3.2.1 Apparatus

All experimental conditions were run on a Fujitsu LifeBook T3010D Tablet PC with a 1.4 GHz Pentium M processor and 768 MB RAM, running the Windows XP Tablet Edition operating system. It had a 12.1 inch (307 mm) diagonal display, with a resolution of 1024 by 768 pixels (246 mm by 184 mm; i.e., each pixel measured 0.24 mm). 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 (approximately 35 degrees from horizontal). We chose this setup, because pilot studies indicated difficulty viewing the screen when it was horizontal on the table, and holding the tablet can be strenuous. 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.

3.2.2 Participants

Thirty-six participants were recruited from three age groups (12 each).[2]

Young (19–54). Actual range: 19–53, $M = 31$, 7 female.

Pre-old (55–69). Actual range: 55–68, $M = 62$, 8 female.

Old (70+). Actual range: 73–85, $M = 76$, 9 female.

The justification for these groupings rests on the age-related changes that occur in cognition [30], 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.

Participants received $5 for each half hour of participation. The younger participants were recruited through advertisements on campus and through word-of-mouth advertisement. They completed the study in 66 to 115 minutes ($M = 78$). The older participants were recruited through word-of-mouth advertisement and a

[2]The naming of these age groups is based on accepted terminology in the aging literature [90].

local community group. The pre-old group completed the study in 73 to 109 minutes ($M = 86$) and the old group completed the study in 77 to 111 minutes ($M = 89$). All participants were right-handed, free of diagnosed motor impairments to their right hand, and had normal or corrected-to-normal eyesight.[3] To control for any biases between age and previous Tablet PC experience, we limited participation to individuals with no Tablet PC experience and no or limited PDA experience. None of our participants had previously owned a PDA, but some reported having tried a friend's, or participating in other studies involving PDAs. Furthermore, within and across each age group, participants had a wide range of computer experience, from novice to expert.

Additionally, we administered the North American Adult Reading Test [104] to help ensure participants had sufficient English fluency to follow our instructions. Based on these results, we excluded three participants from the 18–54 age group (not included in the 36 above). They were allowed to finish the study, but their data were not included in our analysis.

3.2.3 Motor Skill

Because motor skill is known to be one of the main factors accounting for age-related differences in targeting ability [101], we administered three standardized tests to gather data about our participants' motor abilities.

To measure perceptual speed, we used the digit symbol substitution test [120]. In this test, the digits 1–9 are each paired with a simple symbol. Participants were 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 were not expected to memorize the symbols, but were provided with a legend for reference.

As a measure of motor coordination, we used the Purdue pegboard test [108]. This test has four subcomponents. The first three subtests involve the sequential insertion of small pins into a pegboard (similar to a cribbage board). Participants were given 30 seconds to place as many pins as they could, using their right, left, and both hands. The fourth subtest is an assembly task: participants placed a pin in the pegboard with their right hand, a washer on the pin with their left, a collar on to

[3] Based on self-reported data.

the pin with right hand, and a final washer on top with their left. Participants were given a minute to make as many assembled pieces as they could. For all subtests, we repeated the test three times and computed the average score.

Finally, to measure hand steadiness, we used a nine-hole steadiness tester [45]. In this test, participants were asked to hold a metal tipped stylus in nine progressively smaller holes without touching the edge for 10 seconds. 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.

3.2.4 Task

To gain a better understanding of how task might affect targeting ability, we used two tasks in this study: a multidimensional Fitts' tapping task [103] and a menu selection task. The tapping task was selected because it is the gold standard for evaluating input techniques, and provides well understood measures of speed and accuracy. The menu task was selected because it provides a greater degree of realism. Additionally, it requires more cognitive effort as the user must first find the item and then acquire it.[4] We hypothesized that this additional cognitive load might disproportionately affect older users, especially in terms of accuracy.

Multidimensional Tapping Task

For the tapping task (Figure 3.1), each trial started with a single blue *start* circle in the middle of the screen. Once tapped on, the start circle immediately turned light gray, and a red *target* circle appeared. The next tap ended the trial, regardless of whether the tap successfully acquired the target or not. An audible beep provided feedback when the trial was unsuccessful. Participants were instructed to tap on the target circle as quickly as possible while remaining accurate.

[4]Cockburn, Gutwin, and Greenberg [26] showed that the selection time of a menu item is predicted by the sum of the decision/search time and the Fitts' Law pointing time for that item. In this model, decision/search time depends on the user's expertise; it is an interpolation between a linear visual search (reflective of novice performance) and Hick-Hyman decision time (reflective of expert performance).

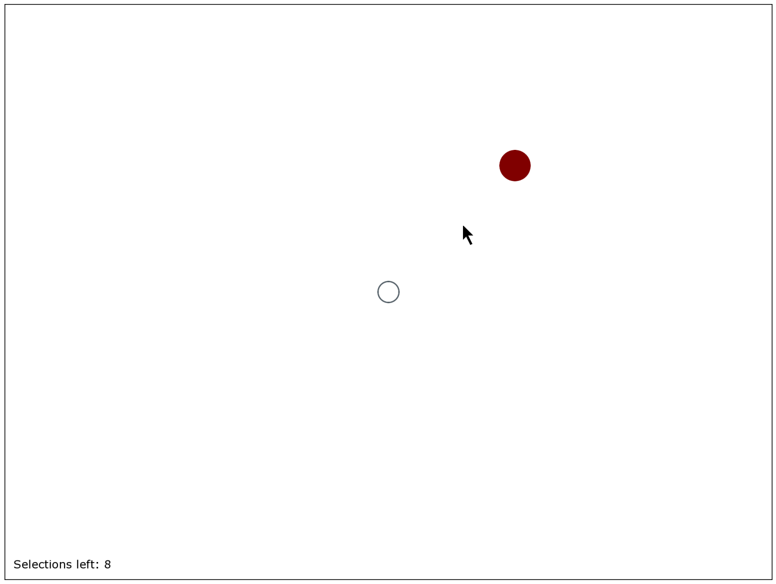

Figure 3.1: Screen shot of the multidimensional tapping task midtrial.

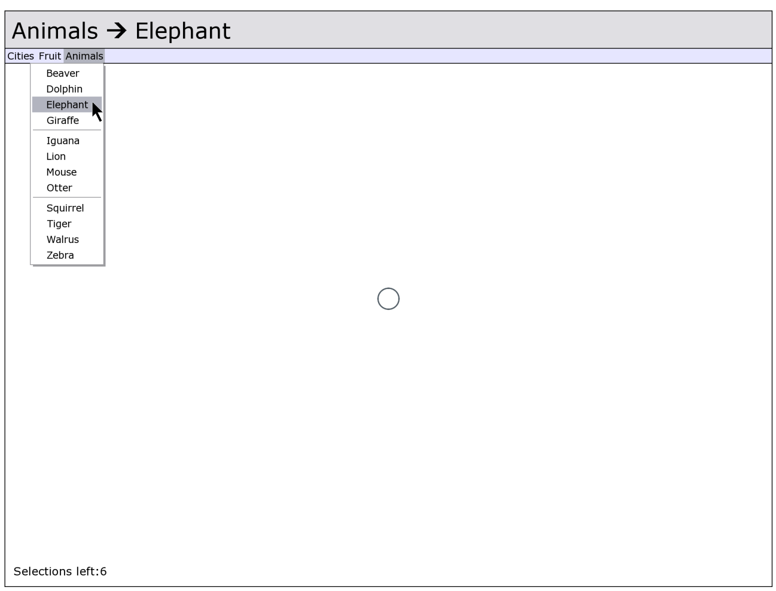

Figure 3.2: Screen shot of the menu task midtrial.

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

Menu Task

Each trial in the menu task (Figure 3.2) also began with a single blue start circle. When the participant tapped on it, it immediately turned light gray (as in the tapping task), and a prompt appeared above the menu bar indicating which menu-item pair was to be selected. The trial ended when the participant successfully selected a menu item, regardless of correctness. An audible beep provided feedback when the wrong item was selected. Again, participants were instructed to make selections as quickly as possible while remaining accurate.

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

We maintained the default Tablet PC menu layout for right-handed users; that is, menu heads were aligned with the left edge of the screen, and menus aligned with the right edge of the menu head where possible. This layout minimizes hand-occlusions relative to the normal menu layout (i.e., for non-Tablet PCs), which instead aligns the menus with the left edge of the menu head. One implication of the Tablet PC layout is that the menus can only be moved as far as the left edge of the screen. As a result, the leftmost menus tend to be in the same location, which can make it harder to notice when the wrong menu is open.

The menus otherwise operated like normal menus; however, we highlight one potentially overlooked feature. Menus open based on a tap down event occurring on the menu head, while selection of an item is based on a tap up event. As a result, it is possible to tap down on the head, drag the pen to the correct item, and lift the pen to select the item. However, use of this feature was exceptionally rare. All participants generally used two taps to make a selection.

3.2.5 Design

This experiment used a mixed design with two counterbalanced tasks (menu, tapping). Because the structure of each task was different, we present them here as two separate subdesigns. The tapping task used the following design: 3 (age groups) $\times$ 4 (blocks) $\times$ 3 (target widths) $\times$ 3 (target amplitudes) $\times$ 8 (angles). The presentation order of each combination of target width, amplitude, and angle was randomized. For the menu task, the design was: 3 (age groups) $\times$ 4 (blocks) $\times$ 3 (menus) $\times$ 12 (items). Each participant was assigned one of the six possible menu order permutations at random, and the presentation order of the menu-item pairs was randomized. Age was the only between-subjects factor. Thus, each participant completed 288 trials in the tapping task, and 144 in the menu task.

3.2.6 Procedure

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

We began with the motor tests, which were given in the order: Digit Symbol Substitution, Purdue Pegboard, and Steadiness. Next was the North American Adult Reading Test. Participants then completed the first 8 steps of "Get Going with the Tablet PC", the native tutorial that introduces new users to the Tablet PC and using the pen. (Steps 9–17 concern text input and were not relevant to the study.) 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.

After the tutorial, we presented the first task (either menu or tapping). After completing all four blocks of the first task, participants were given a brief questionnaire about their background and computer familiarity. They then completed their second task. We note that beyond the instructions given in the tutorial, participants were not instructed to use the pen in any particular manner. We explicitly wanted to observe how participants would naturally approach the task.

3.2.7 Measures

We included measures of speed and accuracy. For speed, we measured trial time as the time from the pen up action off the start circle to the pen up action that ended the trial. We included several measures of accuracy, as we were interested in not only the numbers of errors but also the types. For the tapping task, we modified for pen interaction a classification of mouse errors.[5]

Slips. The pen lands on target, but slips off before it is lifted.

Near misses. The pen lands off target, and lifts at a distance less than 50% of the target radius away from the target boundary.

Not-so-near misses. The pen lands off target, and lifts between 50% and 100% of the target radius away from the target boundary.

Other (or unclear). The pen lands off target, and lifts between 100% and 200% of the target radius away from the target boundary.

Accidental taps. The pen lands off target, and lifts at a distance greater than 200% of the target radius away from the target boundary.

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

[5]This classification was first presented in [109], a version of which appears as Chapter 6 in this thesis. However, it is based on the results of an earlier study by Keates, Trewin, and Paradise [58].

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

Slips. The pen lands on the target item or menu head, but slips off before lifting.

Correct-menu misses. Selection of an incorrect item from the correct menu.

Incorrect-menu misses. Selection of an item from an incorrect menu.

Note that we counted both slips off the target item and slips off the menu head (and onto an incorrect item) as slips. In all cases, slips off the menu head resulted in selection of the first item of the menu, which in almost all cases was not the goal target. In the rare case where this was the goal target, it is not clear whether or not the action was intentional. To be conservative, we did not count these as errors. For correct-menu misses, we further recorded the proximity to the correct item.

In both tasks, for slips we recorded the distance traveled between the pen down and up (i.e., the distance slipped).

3.2.8 Hypotheses

We had the following hypotheses for this study.

H1. Speed and accuracy will decrease as age increases.

H2. Age will impact the types of errors made.

H3. Task will impact the types of errors revealed.

3.3 Results

In this section we present our results. Unless otherwise noted, Bonferroni corrections were used for all posthoc pairwise analyses. Where Levene's test revealed significant heterogeneity of error variance, we used a Welch's ANOVA for testing the main effect and Games-Howell corrections for posthoc pairwise comparisons. Both are robust against unequal error variances. Finally, in all our repeated measures analyses, sphericity was an issue; thus, Greenhouse-Geisser corrections were used. Along with statistical significance, we report partial eta-squared (η^2), a measure of effect size, which is often more informative than statistical significance in applied human-computer interaction research [65]. To interpret this value, .01 is a small effect size, .06 is medium, and .14 is large [27].

Not surprisingly, analysis of the motor tests confirmed overall motor decline with age. Unexpectedly however, they did not detect differences between the pre-old and old groups, which foreshadows a lack of significant differences between these age groups in our analyses of the target acquisition tasks. We additionally performed some exploratory data analysis to see if the motor tests results could provide additional insight into our performance measures. This analysis was not fruitful, likely due to our small sample size.

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

3.3.1 Tapping Task

Speed decreased with age. As H1 predicted, older users were slower. (A significant effect of age was revealed by a one-way ANOVA on median trial time, $F_{2,32} = 4.26$, $p = .023$, $\eta^2 = .210$, which excluded 1 outlier from the young group). However, posthoc pairwise comparisons only detected a difference between the young and old groups ($p = .019$), with trends (as shown in Figure 3.3) suggesting a possible general slowing with age.

Everyone misses, but older users also slip. For the young, pre-old, and old groups respectively, the average overall error rates were 1.9%, 4.2%, and 6.44% ($SD = 1.8, 5.2, 7.3$). In each group, the minimum error rate was 0, and the maximums were 6.6%, 16.7%, and 22.2%, respectively. In terms of the individual error types, slips and near-misses accounted for 90% of the errors observed (with no other category accounting for more than 5% percent), so we focused our analyses on them.

For each of slips and near-misses, we performed a repeated-measures ANOVA on target width and age.[6] We found that while slipping clearly increases with age,

[6]We included target width in this analysis as previous research has found interactions between age and target size for tapping accuracy measures [52, 101].

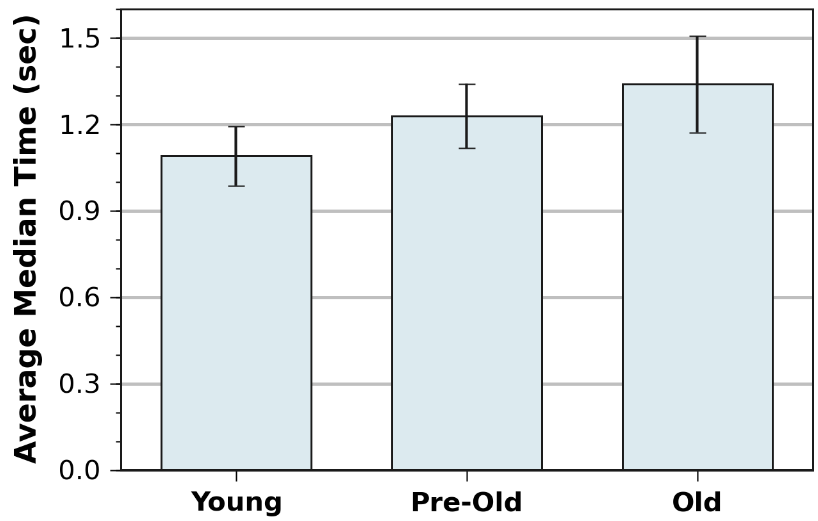

Figure 3.3: Average trial time by age group for the tapping task ($N = 35$). Error bars represent 95% confidence intervals.

near-missing remains relatively constant, as shown in Figure 3.4. (There was a main effect of age for slipping, $F_{2.0,15.9} = 3.86$, $p = .043$, $\eta^2 = .185$, but posthoc pairwise comparisons did not produce any significant results. In contrast, there was no effect of age for near-misses). As we would expect, we also found main effects of target width for both slips ($F_{0.2,39.0} = 18.3$, $p < .0001$, $\eta^2 = .357$) and near-misses ($F_{1.2,38.5} = 40.3$, $p < .0001$, $\eta^2 = .550$), indicating that both these errors increased as targets got smaller.

In addition, we also found that older users did have greater difficulty than younger users with slipping from smaller targets than larger ones, whereas slipping was infrequent for the young users across all widths. However, there was no such effect for near-misses. (There was an interaction between age and width for slipping, $F_{2.4,39.0} = 5.87$, $p = .004$, $\eta^2 = .263$, but none for near-misses.) It is also interesting to note that slips were overall relatively short. On average they were 12 pixels (2.9 mm) long (median: 9 pixels, standard deviation: 7), and while the largest slip was 49 pixels (11.8 mm), over 90% were less than 25 pixels (6.0 mm).

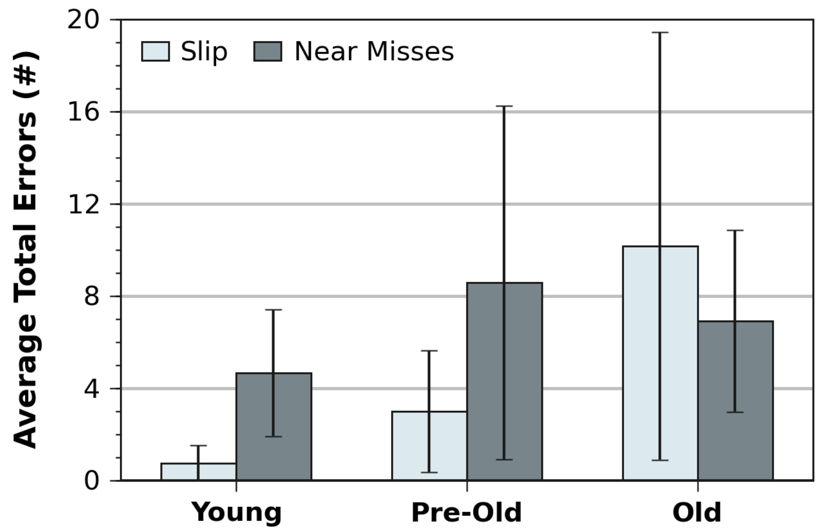

Figure 3.4: Average total number of slips and near misses for the tapping task ($N = 36$). Error bars represent 95% confidence intervals.

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

3.3.2 Menu Task

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

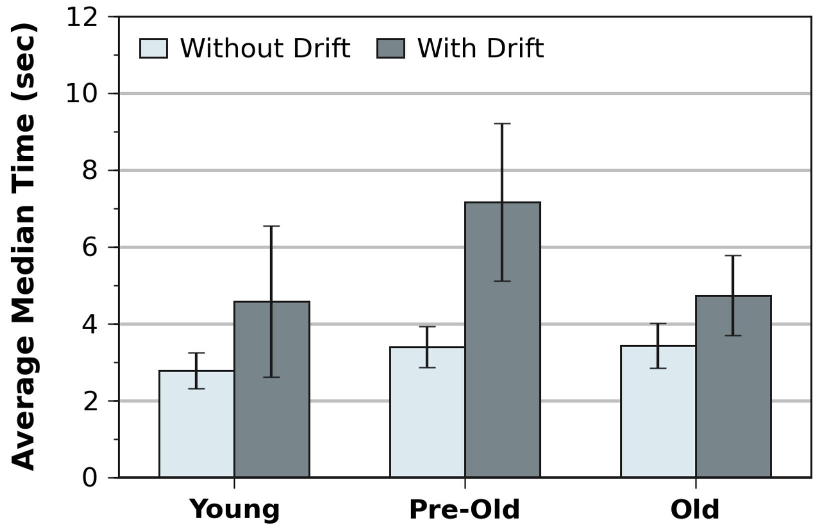

Figure 3.5: Average trial time by age group for the menu task for trials without and with drift ($N = 35$). Error bars show 95% confidence intervals.

the next menu. In consideration of this dominant behavior, we chose to consider drifting in our analysis.

Drifting impeded task performance. Participants were often confused when drifting occurred: they reported not knowing why the wrong menu was open and not being sure how to proceed. Many participants would attempt to reopen the desired menu; however, when the pen neared that menu, the hovering would trigger it to reopen. But, some participants would not notice, and would tap on it anyway, which actually resulted in it closing. Needless to say, this led to considerable confusion. Thirty-five out of 36 participants drifted at least once, and 31 responded to a drift by retapping (and thus closing) the target menu at least once. Moreover, performance was significantly impeded by drifting in terms of slower trial times, as can be seen in Figure 3.5. However, overall accuracy was not affected. (Paired t-tests on the 35 participants who drifted for both median trial time and overall accuracy revealed a significant effect on speed, $t = 5.12$, $df = 34$, $p < 0.0001$, but no effect on accuracy, $p = .16$).

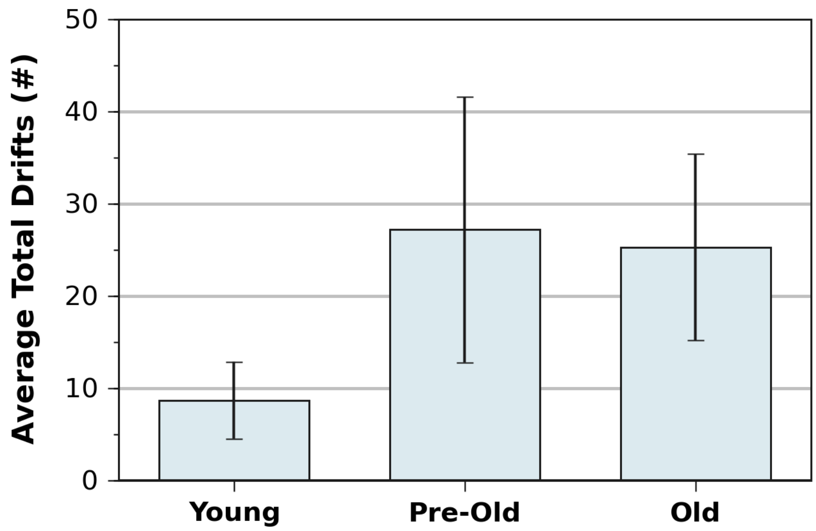

Figure 3.6: Average total number of drifts by age group ($N = 36$). Error bars represent 95% confidence intervals.

Older users drifted more. Although drifting affected all age groups, pre-old and old participants drifted more than participants from the young group. Figure 3.6 shows mean drifts by age group. (A one-way Welch's ANOVA excluding three outliers, 2 young and 1 old, revealed a significant effect of age, $F_{2,16.5} = 9.35$, $p = .002$, $\eta^2 = .221$. Posthoc pairwise comparisons further showed significant differences between the young and both the pre-old, $p = .038$, and the old, $p = .008$ groups.) We note that this further supports our hypothesis that accuracy would decrease with age (H1). Although drifting did not have an explicit affect on overall task accuracy, it does represent a difficultly in accurate interaction.

Drifting did not decrease with learning. It is also interesting to note that drifting behavior did not improve over the course of the menu task; that is, participants did not get used to the designed interaction. (A repeated-measures ANOVA on block, with age as a between-subjects factor was not significant for either the main effect of block, $p < .51$, or the interaction between block and age, $p < .45$).

Drifting aside, older users were still slower. In consideration of our findings for drifting, we performed our analysis of age on median trial time based solely on drift-free trials to determine if there was an effect independent of that caused by age-related differences in drifting behavior. Comparing only drift free trials, both the pre-old and old groups were significantly slower than the young group, which is also supported by Figure 3.5. (A one-way ANOVA on age for drift-free trials revealed a significant main effect of age, $F_{2,32} = 4.68$, $p = .017$, $\eta^2 = .226$, while posthoc pairwise comparisons revealed that the young group was significantly faster than both the pre-old group, $p = .04$, and the old group, $p = .03$.)

Error rates were low overall. Overall errors were lower in the menu condition than we expected. Of 5184 trials (36 participants $\times$ 144 trials) there were only 135 errors, yielding an overall error rate of 2.6% ($SD = 2.6$). By age group, the average error rates were 1.3, 3.6, and 2.9, for the young, pre-old, and old groups, respectively ($SD = 1.1, 5.2, 3.7$). 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. Moreover, 60 (44%) of the 135 errors were committed by 3 individuals, one in each of the three age categories.

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, 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.

In light of the low error rates, we did not attempt to make any age comparisons for errors. Instead we examined our data for general trends in the types of errors observed.

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

Missing occurred just below the target item. We further analyzed the correct-menu misses based on their proximity to the target item. Of the total 5184 trials across all 36 participants (144 trials each), only 4 selections were made on the top two pixels of the target item, while 56 selections (44 misses and 12 slips) were made on the corresponding top edge region of the item below. 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. This general tendency towards missing-just-below is demonstrated by Figure 3.7. Note that an additional 57 selections fell outside the range shown in this figure. 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.

There was no indication that missing-just-below decreased over the course of the study or was affected by age. We ran 4×3 (block of 36 trials $\times$ age group) repeated-measures ANOVAs on the mean y-coordinate of the selection and on total missing-just-below errors, and found no significant differences. Figure 3.8 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.

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

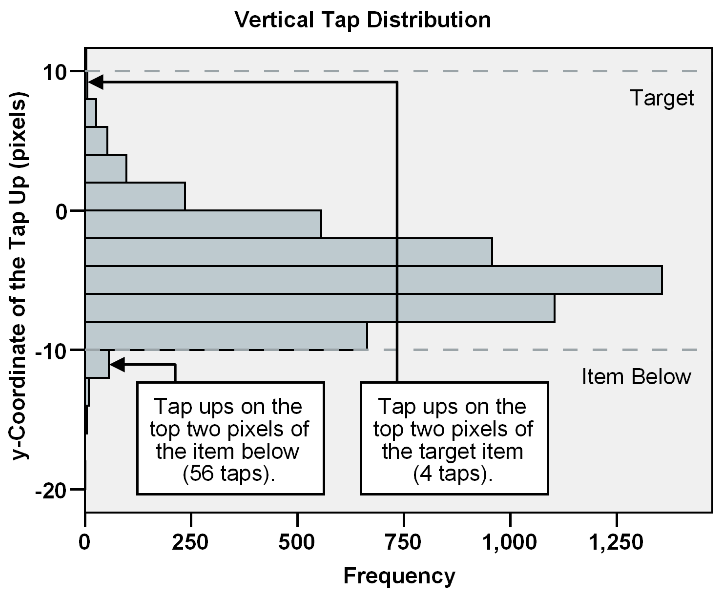

Figure 3.7: Histogram of the vertical tap distribution (for taps on the target item and the item below), relative to the center of the target. (*Bin size* = 2, $N = 5127$).

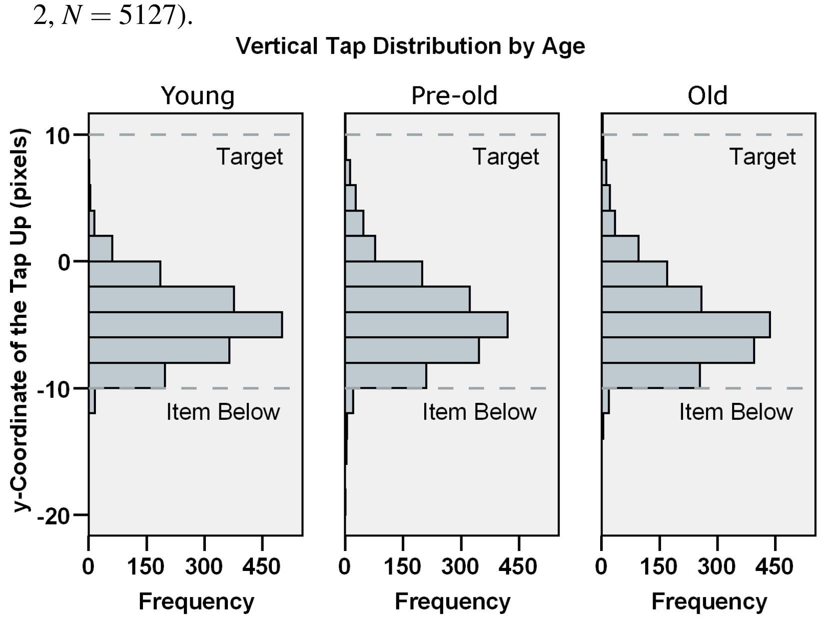

Figure 3.8: Histogram of the vertical tap distribution by age (for taps on the target item and the item below), relative to the center of the target. (*Bin size* = 2, $N = 5127$).

3.3.3 Summary of Results

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

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

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

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

3.4 Follow-Up Study

To gain a better understanding of the results for the menu task, we ran a small informal follow-up study with younger participants to explore the impact of two factors: input device (pen versus mouse) and menu orientation (top-down versus bottom-up).[7] Because of the informal nature of this study, we do not include a full description of it here. For the interested reader, the full methodology is presented in Appendix B. Here we summarize the salient results.

[7] This study was was performed by Justine Yang as part of her summer undergraduate research assistantship in 2008. We present it here because it fits with the goals of the Baseline Study; chronologically, it occurred after both Technique Studies One and Two.

As the main purpose of this study was to provide insight into the underlying reasons for missing-just-below and drifting, we did not focus on comparative performance measures such as speed and accuracy, but rather looked for differences in the tap distributions (i.e., the x, y coordinates of selections). For this analysis, all coordinates were normalized relative to the middle of the target such that a positive x-coordinate was to the right of the center and a positive y-coordinate was above the center of the target. Menu items were 20 pixels high and 100 pixels wide.

3.4.1 Menu Orientation

We compared top-down to bottom-up menus to explore whether or not the downwardly shifted tap distributions we observed in the Baseline study could be attributed to the direction in which participants scanned the menu. That is, we were looking to see if having the menu head at the bottom of the menu would result in an upwardly shifted tap distribution. However, we did not find any evidence of this. Both menu orientations resulted in distributions that were relatively symmetric across the goal item. For the top-down and bottom-up menus respectively, the mean y-coordinates were -0.54 and -0.33 pixels ($SD = 3.37$ and 3.17). This result may, however, foreshadow the results of Technique Study One (Chapter 4), which also found that many of the younger participants did not demonstrate a downwardly shifted tap distribution pattern.

3.4.2 Input Device

Our goal in comparing input devices was to explore whether or not missing-just-below and drifting were pen specific or more general. Consistent with the results for menu orientation, the vertical distributions were relatively symmetric around the center of the target. For the mouse and pen respectively, the mean y-coordinates were -0.03 and -0.01($SD = 4.45$ and 3.36).

Analysis of the x-coordinates was more fruitful. The mean x-coordinate for the mouse was -4.97 ($SD = 13.8$), while for the pen the mean was 3.94 ($SD = 12.1$). A repeated-measures ANOVA confirmed a main effect of device on mean x-coordinate ($F_{1,8} = 12.6$, $p = .008$, $\eta^2 = .611$). This finding, though preliminary, suggests (right-handed) users move further to the right when using a pen, and helps explain the drifting difficulty observed in the Baseline study.

3.5 Discussion

Our results revealed three primary target acquisition difficulties: one, slipping, which was specific to older users, and two, drifting and missing-just-below, which apply generally to all ages.

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

For drifting, the reason for our initial bias was that the effect was more pronounced in the older population. Because they moved more slowly overall, it was easier to follow their actions and catch inefficiencies. Also, they were more overt about their interactions, making comments such as, "Now what happened here?" (Upon realizing the wrong menu was open.) Or, "No. I want that one!" (Before meticulously retapping on the desired menu, causing it to close.) Younger users, on the other hand, recovered more quickly and were considerably less vocal about their experience. For missing-just-below, the effect was more subtle. Articulated confusion by the older participants over incorrect selections similarly prompted us to more closely investigate the vertical tap distributions leading to the discovery that the majority of miss errors occurred on the very topmost region of the item below.

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

3.6 Summary

In this chapter, we presented the findings of an experiment designed to gather information on the underlying causes of pen-based target acquisition difficulties, with a particular focus on how age affects targeting ability. From this examination, we identified three dominant difficulties: slipping, drifting, and missing-just-below. In the following chapters, we investigate different design possibilities for improving pen-based target acquisition and addressing the difficulties identified in this chapter. We begin with missing-just-below.

Chapter 4

Technique Study One

Methods to Reduce Missing-Just-Below

In the previous chapter, we presented a Baseline Study designed to identify age-related pen-based target acquisition difficulties. In this chapter, we begin to explore techniques to address the difficulties uncovered. We begin with missing-just-below, a menu-specific difficulty, which occurs when a user's selection pattern is downwardly shifted, resulting in frequent erroneous selections of the top edge of the item below the target item. In the Baseline Study, missing-just-below affected younger and older participants alike.

4.1 Introduction and Approach

An important and interesting characteristic of the missing-just-below behavior observed in the Baseline Study is that the frequent erroneous selections on the top edge of the item below the target were coupled with infrequent selections of the corresponding top edge of the target item itself. This suggests we may be able to identify and address missing-just-below 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 harness 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

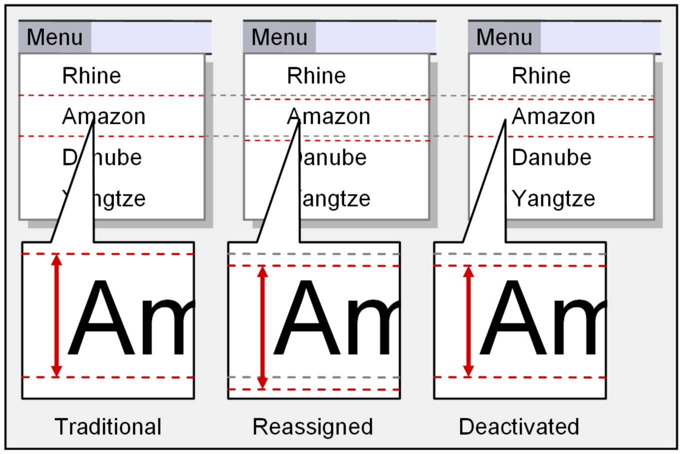

Figure 4.1: In a traditional menu, the active target region (shown by dashed lines and arrow) is centered on the text (left); in a reassigned edge menu, it is shifted down by 10% of its height (center); and in a deactivated edge menu, it is reduced by 10%, such that the top region becomes an invisible menu separator (right).

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 4.1 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 predicted that most users would not notice the small shift, but that those who make missing-just-below 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 [48]. 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.

To evaluate our reassigned and deactivated edge approaches, we conducted a laboratory study with younger and older 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. 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.

4.2 Experimental Methodology

In this section we describe the experimental methodology for 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.

4.2.1 Comparison to the Baseline Study Design

The methodology used in the current investigation is similar in many aspects to that of the Baseline Study (Chapter 3). However, there are some notable differences, which we believe play a role in our results. We, thus, begin this section with a brief overview of the differences between the two study methodologies.

We observed a bias towards accuracy in the Baseline Study; thus, to encourage participants to perform both quickly and accurately, we introduced a performance feedback mechanism and a monetary incentive (see Section 4.2.10). At the individual trial level, we also added distinct auditory feedback to correct selections as described in Section 4.2.6 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 from 144 trials in the menu task of the Baseline Study to 216 trials in each condition of the current study. 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 [103].

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 [30], notably that higher cognitive functions remain 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. These groupings are commonly used in age-comparative studies [62, 95].

Menu contents were randomly generated for each condition in this study (as described in Section 4.2.6 below). In the Baseline Study, the same menu contents were 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.

4.2.2 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 768 MB RAM, running the Windows XP Tablet PC Edition operating system. It had a 12.1 inch (307 mm) diagonal display, with a resolution of 1024 by 768 pixels (246 mm by 184 mm; i.e., each pixel measured 0.24 mm). 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 of approximately 35 degrees from horizontal.

4.2.3 Menu Conditions

The three menu conditions are defined below, 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.

4.2.4 Participants

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

Younger (19–30). Actual range: 19–29 ($M = 24$, 7 female)

Older (65+). Actual range: 66–81 ($M = 73$, 6 female)

The younger participants were recruited through advertisements posted around 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. All participants were right-handed, free of diagnosed motor impairments to their right hand, and had normal or corrected-to-normal eyesight.[8] 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.

4.2.5 Motor Skill

As in the Baseline Study (Chapter 3), we administered standardized tests to gather data about our participants' motor abilities. Our main motivation for including these tests was to ensure that our participants were consistent with their population norms. We used four tests in total. We measured perceptual speed with the digit symbol substitution test [120], and motor-coordination with the right-hand component of the Purdue pegboard test [108].[9] In this study, we also measured hand steadiness using the nine-hole steadiness tester [45]; however, we used an alternate version of the test as results in the Baseline Study were not informative. In this version, participants were not timed, but only needed to move through the holes at

[8] Based on self-reported data.

[9] In the Baseline Study we administered all four components of this test. Results were consistent across the components, so we included only the right-hand component to allow time for the reaction time test instead.

their own pace. We additionally measured simple reaction time [125]; that 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.

4.2.6 Task

Figure 4.2 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 non-menu-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 non-target menu item); there was no auditory feedback for made on the 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. 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 $\times$ 6 blocks $\times$ 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 were scaled.[10]

Menu contents remained constant within each condition, but changed between conditions. Each menu contained three groups of four semantically related items. These groupings were randomly generated for each participant using the approach presented by Cockburn et al. [26]. That is, three 4-item semantic groups were randomly selected from a collection of such groups. The items were then ran-

[10]Specifically, scaling was used for total errors and total selections (which were used to compute a net benefit score). In each case, his data were multiplied by 1.5 (648/432).

Figure 4.2: Screenshot of the evaluation task used in Technique Study One. All three interfaces had the same visual appearance.

domly ordered within that group, and no group was reused in any other condition. (See Figure 4.2 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.

4.2.7 Design

The experiment used a 2 $\times$ 3 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. To minimize the impact of learning effects on our analysis, we fully counterbalanced the order of presentation of the interfaces. To ensure that familiarity with the menu contents did not bias the results, different menu contents were randomly generated for each condition as described in Section 4.2.6.

4.2.8 Procedure

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 [112]. 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).

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", as described in Chapter 3. Once participants finished this 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 using the ISO 9241-9 independent ratings of comfort scale [35]; these questions were asked verbally, and the responses were 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 [104] and between the second and third condition they completed the FAS test of verbal fluency [80]. These two tasks were chosen to engage participants mentally, but not physically. Finally, at the end of the study, participants were encouraged to make additional comments and asked to rank the conditions in terms of overall preference, speed, accuracy, and frustration. 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.

4.2.9 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

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 any top edge	minus	*Error* trials following selection of any top edge

Table 4.1: Definition of net benefit for each interface. For deactivated, recall that nothing happens when the top edge is selected; thus, observation of the subsequent selection is necessary.

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 4.1.

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.

4.2.10 Motivation

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 third of performers in each age group. The one-third 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 4.3, 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.

4.2.11 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 [131], and have more difficulty staying on targets [101]. So although the Baseline Study suggests missing-just-below affects all users,

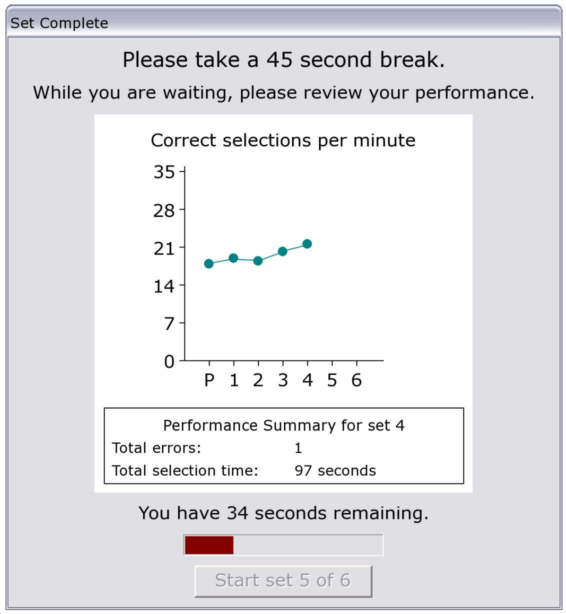

Figure 4.3: The performance feedback presented between blocks, which provided the rate of correct selections for each block completed (including the practice block) and a summary of the errors and time for the most recent block. In this example, the user has just completed block 3.

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.

4.3 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 $\times$ In-

Motor Test	Mean (*SD*) Younger	Older	Significance
Simple Reaction Time (sec)	.275 (.038)	.456 (.190)	$p < .01^*$
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)	5.00 (0.95)	3.83 (0.72)	$p < .01$

Table 4.2: Motor test results by age ($N = 24$). (*) denotes Welch's ANOVA.

terface). Bonferroni corrections were used for all posthoc pairwise comparisons. Along with statistical significance, we report partial eta-squared (η^2), a measure of effect size, which is often more informative than statistical significance in applied human-computer interaction research [65]. To interpret this value, .01 is a small effect size, .06 is medium, and .14 is large [27].

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 [40]. 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 motor tests. These results are summarized in Table 4.2. Interestingly, the older participants performed significantly worse on the digit symbol substitution test than comparably aged individuals in the Baseline Study,[11] perhaps foreshadowing the differences

[11] We regrouped the participants in the Baseline Study according to the age groups in this study: 17 were aged 65+, and 7 were aged 19–30. We performed an ANOVA (Study × Age Group) on the digit symbol substitution test scores and found a significant interaction between study and age

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 were available for the steadiness or reaction time tests.

4.3.1 Net Benefit

The overall error rates were very similar across interfaces. For the older participants the mean error rates, for control, reassigned, and deactivated, respectively, were 8.2%, 9.6%, and 10.1% ($SD = 7.6, 12.6, 11.7$); for the younger adults they were 1.2%, 1.2%, and 0.8% ($SD = 1.2, 1.0, 1.0$). For control, missing-just-below accounted for 30.8% of errors for the older group ($SD = 27.8$) and 22.2% of errors for the younger group ($SD = 35.1$).

A two-way repeated measures ANOVA (Age × Interface) for net benefit, excluding one outlier from the older group, revealed a significant main effect of interface ($F_{1.2,26.1} = 6.653$, $p = .011$, $\eta^2 = .241$), but no main effect of age ($p = .20$). Posthoc 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 ($p = 1.00$).

A breakdown of the net benefit for Deactivated is provided in Table 4.3 and summarized as follows. For the older participants, 44% of selections on the deactivated top edge were on the target item (i.e., they selected the top edge of the target item 5.64 times on average, and the top edge of a non-target item 7.05 times). For the younger participants 32% of selections on the top edge were on the target item (i.e., 1.00 versus 2.08 for target and non-target top edge selections, respectively). Following selection of a deactivated top edge, the correct target item was generally selected. For the older adults, the correct target item was selected 87% (4.91/5.64) of the time following selection of the top edge of the target, and 98% (6.91/7.05) of the time following selection of the top edge of a non-target item. Subsequent to selecting the top edge of an item, the younger participants always went on to select the correct item.

group ($F_{1,44} = 10.058$, $p = .003$, $\eta^2 = .186$). Posthoc 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 $M = 25.58$ ($SD = 6.0$).

	Average Total Trials (*SD*)			
	Younger		Older	
	Target	Non-Target	Target	Non-Target
Top Edge Selections	1.00 (2.30)	2.08 (2.50)	5.64 (5.18)	7.05 (6.40)
Subsequent Selection				
Correct	1.00 (2.30)	2.08 (2.50)	4.91 (4.48)	6.91 (6.04)
Incorrect	0.00 (0.00)	0.00 (0.00)	0.73 (1.01)	0.14 (0.45)

Table 4.3: Breakdown of the net benefit results for Deactivated. The top half of this table lists, for each age group, the the average number of trials involving selection of the top edge of target and non-target menu items. The bottom half provides a break down of this total according to the correctness of the subsequent selection ($N = 23$).

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 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 4.3.5.

Our analysis also revealed a trend suggesting an interaction between interface and age ($F_{1.2,26.1} = 3.48$, $p = .065$, $\eta^2 = .142$). Figure 4.4 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.

4.3.2 Speed and Taps to Select

A two-way ANOVA (Age $\times$ Interface) on taps to select, excluding one outlier (older), revealed a significant main effect of interface ($F_{1.2,25.6} = 9.58$, $p = .003$, $\eta^2 = .313$), and significant interaction between age and interface ($F_{1.2,25.6} = 6.40$,

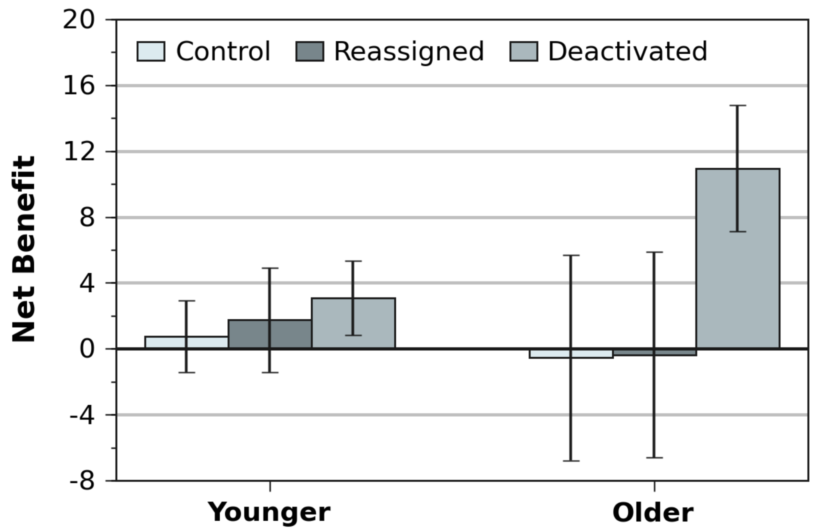

Figure 4.4: Average net benefit for 216 trials, by interface and age group ($N =$ 23). Error bars represent 95% confidence intervals. (Note: A higher score in this graph represents better performance of the interface).

$p = .014$, $\eta^2 = .234$). Posthoc 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 4.5 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 ($F_{1,10.7} = 7.78$, $p = .018$, $\eta^2 = .288$, 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 $\times$ Interface) on median 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 ($F_{1.0,11.2} = 34.52$, $p < .001$, $\eta^2 = .600$, Welch's ANOVA), indicating that older participants were generally slower than their younger counterparts. Figure 4.6 shows the average median time required to make a selection for each interface by age.

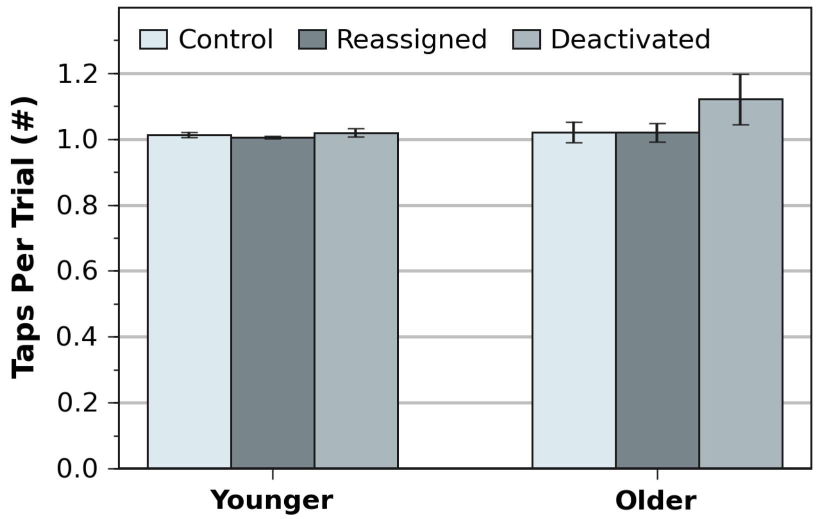

Figure 4.5: Average taps needed to select an item, by interface and age group ($N = 23$). Error bars represent 95% confidence intervals.

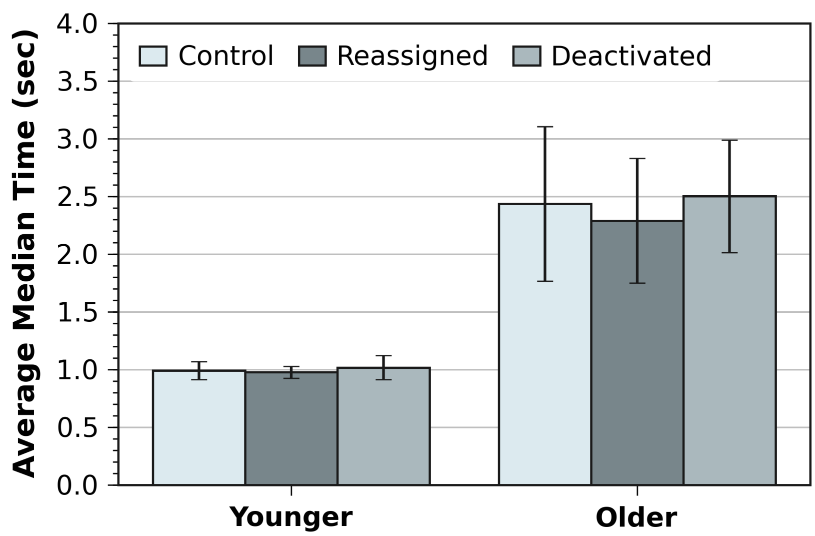

Figure 4.6: Average median time to select an item, by interface and age group ($N = 23$). Error bars represent 95% confidence intervals.

		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)
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)

Table 4.4: Breakdown of comments made by participants, by age group and by tap distribution group, reported as # Positive (# Negative) ($N = 24$).

4.3.3 Subjective Responses

Many participants reported difficulty completing the ratings at the end of each condition and 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 4.4 provides a summary of the number of participants who commented positively and negatively for each interface by age group. These counts are 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 (versus1 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."

4.3.4 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.

4.3.5 Individual Differences

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 hit-*

ters 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 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 younger), and 10 high hitters (2 younger), 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. 3 older, 4 younger, 7 total.

Neutrals. 1 older, 6 younger, 7 total.

High hitters. 8 older, 2 younger, 10 total.

Figure 4.7 highlights the contrasting error patterns of the groups and Figure 4.8 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 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 $\times$ Interface) for net

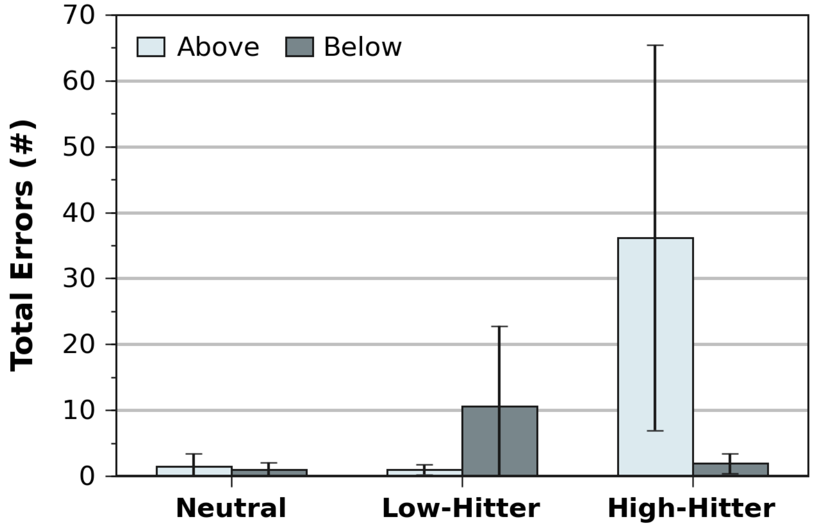

Figure 4.7: Errors above and below the target item (across all interfaces), by tap distribution group ($N = 24$). Error bars represent 95% confidence intervals.

benefit revealed a significant main effect of interface ($F_{1.5,31.7} = 7.55$, $p = .004$, $\eta^2 = .264$), a significant main effect of distribution group ($F_{2.0,10.4} = 6.56$, $p = .014$, $\eta^2 = .283$), and a significant interaction between interface and distribution group, ($F_{3.0,31.7} = 8.219$, $p < .001$, $\eta^2 = .439$). Posthoc 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 4.9 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 4.8 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 4.9), this was mostly due to selections on the top edge of the target item. In contrast, for

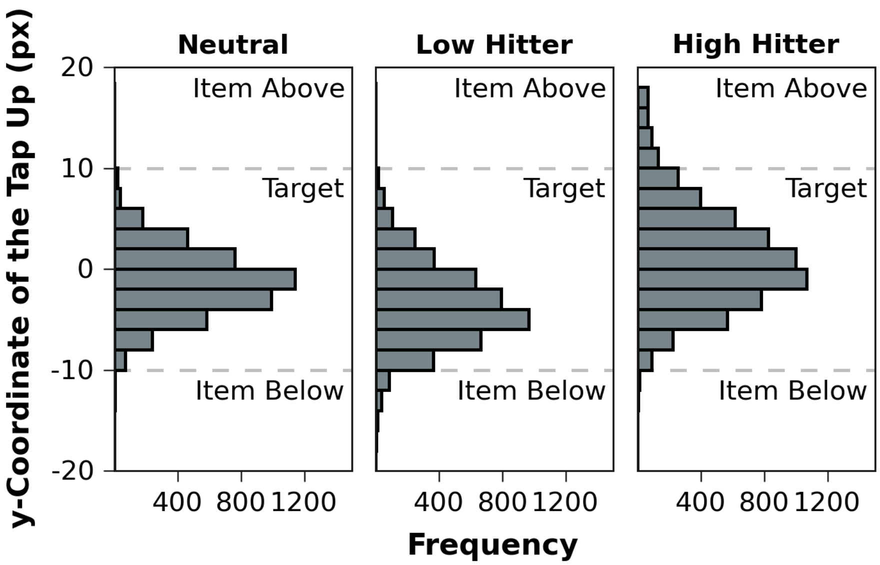

Figure 4.8: Histograms of the vertical tap distributions (for taps on the target item and the lower/upper half of the item above/below) for each hitter group and across all interfaces, relative to the center of the target item. ($Bin\,size = 2$.)

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 4.4 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

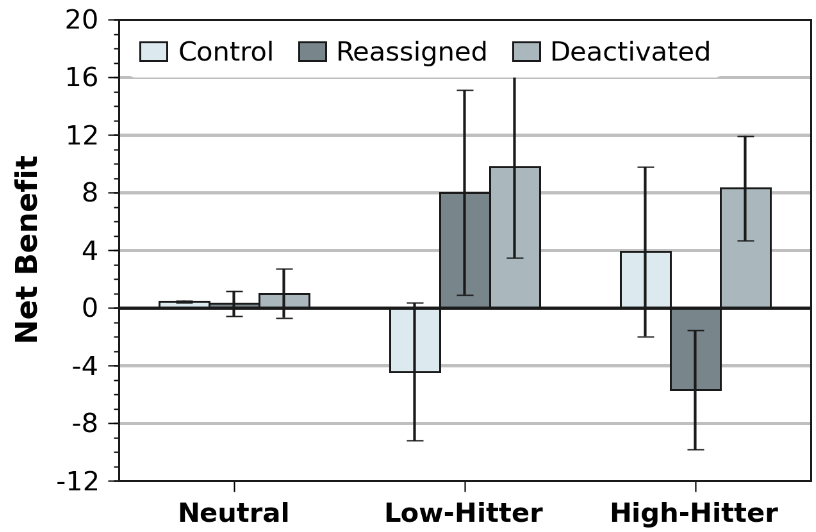

Figure 4.9: Average Net Benefit for 216 trials, by interface and tap distribution group ($N = 24$). Error bars represent 95% confidence intervals.

lower number of negative responses suggests that this cost may have been balanced by the benefit of fewer selection errors.

4.4 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 4.5, 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 the idea 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 [53], 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.

4.5 Future Work

One area for future work is developing a better understanding of what factors influence the different error types observed in this study. We explore this topic in Chapter 8. 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.

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 to the physical pen to give it a clicking feel [18], 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 4.10 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.

4.6 Summary

In this chapter, we examined methods to address missing-just-below, a difficulty that 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.

Though our interfaces were not as effective overall as we had hoped, this research highlights the need for better support, especially for older users. In contrast to the Baseline Study, many of our participants did not make missing-just-below

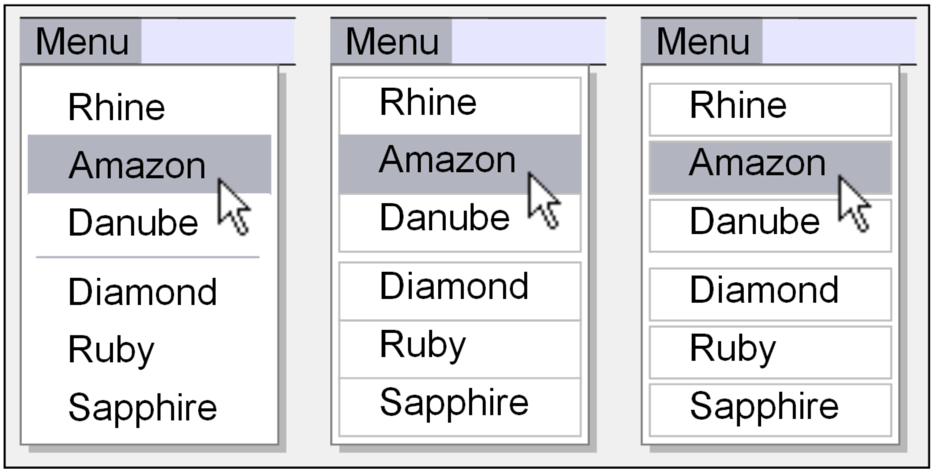

Figure 4.10: Proposed visual presentations for the reassigned edge (center) and deactivated edge (right) approaches as compared to a regular menu (left). For the reassigned edge approach, the label is shifted up within the target area. For both approaches, the target boundaries are clearly marked by borders.

errors, but instead, we found three distinct user types: the low hitters, the high hitters, and the Neutrals. All but one of the older users fell into one of the two error prone groups (the low hitters or the high hitters). When we take into account these different user types, we found preliminary evidence that our approaches were beneficial 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.

Chapter 5

Technique Study Two

Methods to Reduce Drifting

In this chapter, we continue our investigation of pen-based menu interaction difficulties, by examining methods to address the drifting. Like missing-just-below, this difficulty affected users of all ages in the Baseline study.

5.1 Introduction and Motivation

Drifting involves the unintended invocation of a menu adjacent to the one in focus. It occurs as a result of the tracking capabilities provided by inductive pens. In contrast to passive pen technology (most often associated with PDAs), inductive pen technology can sense the location of the pen both when it is touching the screen's surface and when it is hovering in proximity. When in this hovering (or tracking) state [18], pen technology behaves like a mouse when no mouse buttons are pressed; for example, the cursor's location is tracked and tool tips are dynamically displayed.

One notable use of the tracking state is to support hover-switching between menus. That is, when a menu is open, cursoring over any other menu head causes the currently open menu to close and the menu under the cursor to open, as shown in Figure 5.1. This can also be done with the mouse in the tracking state, and generally provides a quick and efficient way to scan through menus. However, with the pen, users tend to accidentally cursor (or drift) over the adjacent menu closest

Figure 5.1: Illustration of the drifting difficulty. When a menu is open, hovering the pen over a new menu causes the currently opened menu (i.e., file) to close, and the menu under the cursor (i.e., Edit) to open. Drifting is the accidental invocation of this feature.

to their hand, causing the desired menu to close, and the adjacent one to open. Individuals tend not to expect the system to respond when they are not touching the screen. Thus, although the switching interaction is consistent between mouse and pen, it feels less natural with the latter.

Occlusion further impedes the interaction and contributes to drifting. When a user's hand occludes the menu contents, it is common for the user to lift the pen and move it away to read the contents, thus increasing the likelihood of a drift occurring. Moreover, because occlusion can obscure the action, it can be difficult for users to learn the cause of drifting. We believe this limits their ability to self-correct the problem.

An additional difference between mouse and pen is that the Microsoft Windows XP Tablet PC Edition provides automatic layout modifications based on the handedness of the user. The limitations of this layout protocol make it difficult to recognize when a drift has occurred, leading to further confusion. In particular, the alignment of menus is adjusted to minimize hand occlusion; thus, for a right-handed user, each menu is moved to the left, and aligned, if possible, with the right edge of the menu head (by default a traditional mouse-based interface left-aligns the menu with its menu head). However, menus can only be moved as far as the left edge of the screen. As a result, the leftmost menus tend to be in the same location (see Figure 5.2 on page 82 for an example), reducing the visual shift that might alert the user to a drift.

In the Baseline Study, 35 out of 36 of our participants drifted at least once, and this behavior did not improve over the course of the study [74]. Moreover, although drifting did not have an impact on the overall task accuracy, it did impede performance with trial time almost doubling when drifting occurred. Finally, although drifting did affect both younger and older users, the older participants (those aged 55 and over) were disproportionately affected: they drifted significantly more often than the younger participants.

5.2 Proposed Solutions

We have developed two approaches that we predict will address drift. First, we note that in the Baseline Study, no one intentionally used hovering to switch between menus. Thus the simplest way to prevent drifting may be to turn off hover-switching, and require an explicit Tap to switch between menus. This approach, which we call *Tap*, clearly eliminates the drifting difficulty. However, the cost of this approach is not entirely clear. In our previous study, participants were all novices to pen-based interaction and were prompted to the correct menu (i.e., the task prompt for each trial provided both the menu and the item to be selected). It is possible that for expert users, or when the task requires browsing through menus for the correct item, being able to switch menus without touching the screen may prove useful.

An alternative approach, which we call *Glide*, uses a distance threshold to differentiate between accidental drifts and intentional hover-switches. In the Baseline Study, most drifts were short relative to the width of the menu head: over 80% of those drifts were fewer than 10 pixels (2.4 mm) into the adjacent menu head before the pen exited either the hover region of the tablet, or the top or bottom of the menu head. We predict that when intentionally switching to a new menu, (right-handed) users will bring the pen clear across the menu towards the right edge, so as to minimize hand occlusion and enable them to read the menu contents. Based on these two factors, we chose a threshold of 20 pixels (i.e., 4.8mm or approximately 40% of the width of the menu head). This is clearly larger than the majority of drifts in the previous study, and yet less than half way across the menu head.

We note that a time threshold may also be possible. We suspect that when drifting, users spend less time over the menu head than when they are intentionally switching to that menu. So, essentially, we could delay the menu switch by time, rather than distance. However, a time threshold is likely to require user specific customization, whereas we believe a distance threshold depends more on the width of the menu head than on individual differences in motor behavior. Thus, we chose to first investigate the simpler option, a distance threshold.

To assess the effectiveness of our proposed designs, we ran a controlled laboratory experiment to compare the Tap and Glide interfaces relative to each other and to a control interface. Tap was effective at reducing drifts, particularly for the older group. Tap was so effective for the older group that it closed the performance gap between the age groups. In terms of preference, Tap was ranked highly by the older participants, but was not well received by the younger participants who felt it was slow (though this was not supported by the performance data). Glide surprisingly did not show any performance improvement. Additional research is needed to determine if the negative findings for Glide are a result of the particular threshold used, or reflect a fundamental flaw in the Glide approach.

5.3 Experimental Methodology

In this section we describe the experimental methodology for 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. Because younger participants are easier to recruit, we decided to design our study such that we could learn from the younger group, and if necessary further refine our study before focusing on the older adults. Thus, we ran the study as two independent experiments, fully completing the younger participants before starting with the older ones. Between the experiments, we reflected on the experimental design and made a small but important adjustment to how we referred to our interfaces, as described in Section 5.3.2.

5.3.1 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 768 MB RAM, running the Windows XP Tablet PC Edition operating system. It had a 12.1 inch (307 mm) diagonal display, with a resolution of 1024 by 768 pixels (246 mm by 184 mm; i.e., each pixel measured 0.24 mm). 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 of approximately 35 degrees from horizontal.

5.3.2 Menu Conditions

We needed to provide an easy way to refer to our interfaces as we wanted participants to provide comparative feedback on all three. Initially, we named our interfaces Tap, Delay, and Slide. We used *Delay* to reflect the influence of the threshold on the menu-switching interaction. However, some participants in the younger group interpreted it to mean that the Glide condition was inherently slower than the other conditions even though this was not supported by the performance data. Though this limits our ability to interpret the subjective measures for the younger group (particularly their response to the Delay interface), we did not want it to also limit our analysis for the older group.

Thus, we renamed *Delay* to *Glide*. We also renamed the control condition from *Slide* to *Entry* to prevent confusion between Slide and Glide, and to more clearly reflect the subtle distinction that in the control condition the menu changes as soon as you enter, whereas in the Glide condition it happens at some point once you have entered. Note, we did not specifically indicate to participants at what point the menu changes for Glide because we did not want users focusing on the threshold. The following summarizes the original and revised names for each of our interfaces, and provides the short descriptions used to introduce participants to the conditions.

Tap (unchanged). The selected menu changes *when you tap* on a new menu.

Glide (was Delay). The selected menu changes *as you move* the pen across a new menu, even while not touching the screen.

Entry (was Slide). The selected menu changes *as soon as you move* the pen over a new menu, even while not touching the screen.

5.3.3 Participants

Our main focus for this investigation was on the effectiveness of our designs for older adults. However, we also wanted to ensure that the advantages of these designs for the older users would not have a corresponding negative impact on performance for younger users. Thus, we recruited 24 participants, 12 from each age group.

Younger (19–30). Actual range: 19–25 ($M = 21$, 9 female)

Older (65+). Actual range: 65–85 ($M = 72$, 8 female)

Participants received $5 for each half hour of participation. The younger participants were recruited through advertisements posted on campus, and completed the study in 60 to 90 minutes ($M = 75$). The older participants were recruited though postings in the community and word-of-mouth advertisement. They took 100 to 150 minutes to complete the study ($M = 120$). One participant in the older group (not included in the 12 above) was unable to complete the study tasks. His data were not included in our analysis.

Participants were right-handed, free of diagnosed motor impairments to their right hand, and had normal or corrected-to-normal eyesight.[12] To control for biases between age and Tablet PC experience, all were novices to pen-based computing. Within and across each age group, participants had a wide range of computer experience. Nonetheless, there were some notable differences: younger participants were more frequent computer users and used a greater number of applications; older participants were more educated and had been using computers for longer. Surprisingly, there were no differences between the groups in terms of self-rated computer expertise.

5.3.4 Motor Skill

As in the previous studies, we administered standardized tests to gather data about our participants' motor abilities and ensure our participants were consistent with the norms for their age group. In this study, we used the same tests as used in Technique Study One (Chapter 4): the digit symbol substitution test [120], the right-hand component of the Purdue pegboard test [108], the nine-hole steadiness tester [45], and a simple reaction time test [125].

5.3.5 Task

The menu task was as follows. For each menu interface, participants completed a short 12-trial practice block followed by 6 blocks of trials with an enforced 45 second break between blocks. Each block consisted of a 36-item randomly ordered selection sequence from three 12-item menus. Each item was selected once in each experimental block, and the practice block consisted of a random subset of the items. Thus, each participant completed 36 trials $\times$ 6 blocks $\times$ 3 interfaces for a total of 648 trials (excluding the 12 practice trials).

For each trial, a menu item was displayed across the top of the screen, as shown in Figure 5.2. Participants were instructed to find and select that item from the menus as quickly as possible while remaining accurate. The system advanced to the next item, only when the participant successfully selected the correct menu item. A soft clicking sound provided feedback for correct selections and a louder

[12]Based on self-reported data.

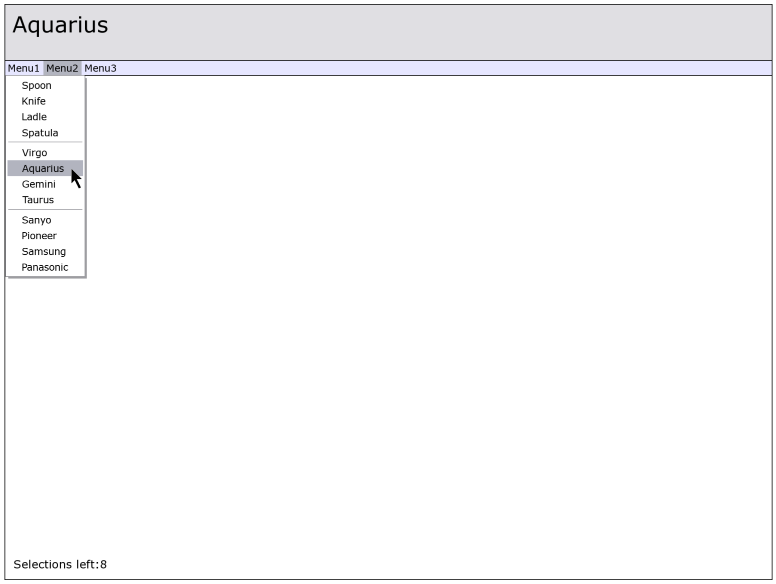

Figure 5.2: Screen shot of the evaluation task used in Technique Study Two. Notice that the open menu is aligned with the right edge of its menu head. This is the the default behavior for a right-handed user on the Tablet PC.

beep sound alerted participants to selection errors. Specifically, participants were not told which menu contained the target item. We wanted to ensure participants would need to search through the menus to find the correct item. This was done to encourage intentional use of hover-switching, thus addressing a limitation of the Baseline Study.

Menu contents remained constant within each menu 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. [26]. That is, three four-item 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 5.2 for an example of a generated menu.) Each menu item was 20 pixels (4.8 mm) high, and each menu separator was 5 pixels (1.2 mm) high.

5.3.6 Design

We used a repeated-measures design with interface condition (Tap, Glide, Control) as a within-subjects factor. We chose this design for the increased power a within-subjects methodology provides. We fully counterbalanced the presentation order of the interfaces to reduce the impact of learning effects on our measures.

As the study was run in two distinct chronological phases, and as we modified the interface names between the phases, we are careful about performing age-related statistical comparisons in our analyses of the results. However, establishing differences between younger and older users is not the main focus of this study as these differences were established in the Baseline Study. To summarize, our design consisted of two independent experiments (one per age group). Each experiment used a 3 (interfaces) $\times$ 6 (presentation orders) factorial design. Each participant was randomly assigned to one of the six possible presentation orders.

5.3.7 Procedure

The study began with the motor tests, followed by a questionnaire on background and computer experience. Participants were then introduced to the Tablet PC and completed steps 1–8 of "Get Going with the Tablet PC", as described in Chapter 3. After the tutorial, the tablet was calibrated to each participant using the built-in Windows XP (Tablet PC Edition) calibration utility.

Participants then completed the menu conditions. At the start of each condition, a short description of that condition was provided (as described above in Section 4.2). Between conditions, participants completed short verbal distracter tasks (the North American Adult Reading Test [104] and the FAS test of verbal fluency [80]). Finally, at the end of the study, participants were asked to rank the conditions, and encouraged to make additional comments. 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 they would naturally approach the task.

5.3.8 Measures

Our main goal for this study was to examine the effect of each of our interfaces on drifting. As a measure of drifting, we recorded the number of extra target menu invocations for each trial; that is, the number of times the target menu was opened in excess of the once required to complete the task. Clearly, factors other than drifting can cause additional invocations. For example, the user may miss the target item on the first pass, or accidentally close the target menu before making a selection. However, we would not expect these factors to disproportionately affect any of our conditions; thus, we can interpret differences in the numbers of extra target menu invocations needed as an indication of differences in drifting behavior.

Additionally, we wanted to ensure we captured any unforeseen effects of our designs on other aspects of performance. Thus, we also included time as a dependent measure. An implicit error penalty was included by forcing participants to correctly select the target item before continuing to the next trial. Selection errors were also recorded independently for completeness. We note that in the Baseline Study, drifting was found to significantly impact trial time. Thus, time may also be a useful indication of drifting difficulty. However, as drifting affects a relatively small number of trials and many other factors can affect trial time, we do not expect that overall task time will necessarily be sufficiently sensitive to detect difference between our conditions.

Finally, subjective data were collected after each condition using the ISO 9241-9 independent ratings of comfort scale [35]. At the end of the study, participants ranked the cursor techniques on overall preference, speed, accuracy, frustration, and initial ease of use.

5.3.9 Motivation

To motivate quick and accurate performance, a $10 incentive was awarded to the top third of performers in each age group. The one-third ratio was chosen to encourage participants to believe they had a reasonable chance of succeeding. To help participants gauge their performance, graphical feedback of performance was presented during the breaks between blocks. This feedback was the same as that shown in Figure 4.3 for Technique Study One (page 58). It consisted a graph showing 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.

5.3.10 Hypotheses

We had the following hypotheses for the experiment:

H1. Target Menu Invocations. We expect that both Tap and Glide will reduce drifting as measured by the number of extra target menu invocations.

H2. Speed and Accuracy. For speed, we predict that (1) Tap will be at least as fast as Control as the cost of having to explicitly Tap should be offset by reduced drifting, and (2) Glide will be faster than Control as it will reduce drifting without impeding hover-switching. For accuracy, we predict that there will be no differences among the interfaces.

H3. Subjective Response. We predict that both Tap and Glide will be preferred over the Control interface. Between Glide and Tap, we predict there may be some age-related difference, with the older users tending to prefer the control provided by the Tap interface, and the younger users preferring the efficiency of Glide.

5.4 Results

For each of our performance measures, we performed separate two-way repeated measures ANOVAs (Presentation order $\times$ Interface) for *each* age-group. Bonferroni corrections were used for all posthoc pairwise comparisons. Along with statistical significance, we report partial eta-squared (η^2), a measure of effect size, which is often more informative than statistical significance in applied human-computer interaction research [65]. To interpret this value, .01 is a small effect size, .06 is medium, and .14 is large [27].

Initial analysis of the data revealed a practice effect. We thus examined the data for differences between the blocks. We found that for the older group the first two blocks were significantly slower than the latter four, but that there were no significant differences among the last four blocks. For the younger group, only the first block was significantly slower than the others. Thus, in all subsequent analyses, we exclude the two blocks of each interface for the older group, and the first block for the younger group. This does not entirely eliminate the practice effect (i.e., there is still an interaction between interface and order on time, see Section 5.4.2); however, it reduces its impact.

5.4.1 Target Menu Invocations

Two-way repeated-measures ANOVAs (Presentation order × Interface) revealed a main effect of interface on the number of extra target menu invocations for both the older ($F_{2,12} = 12.0$, $p = .001$, $\eta^2 = .667$) and younger groups ($F_{2,12} = 5.94$, $p = .016$, $\eta^2 = .498$). As expected there was no interaction effect between interface and presentation order for either group. Posthoc pairwise comparisons further revealed that, for the older group, Tap resulted in significantly fewer extra target menu invocations than Control ($p = .013$). For the younger group, there was a similar trend between Tap and Control ($p = .079$).

Figure 5.3 shows the mean number of extra target menu invocations required per trial by interface and age group and shows that the older group made substantially few extra target menu invocations with Tap than with either Glide or Control. Though not significant, a similar but attenuated pattern is evident for the younger group. The figure also shows that the older participants drifted more than the younger participants (as we would expect based on the Baseline Study findings) and that the Tap interface reduced this performance gap.

Thus, hypothesis H1 was supported for Tap, but not for Glide. Consistent with our predictions, Tap reduced the number of unnecessary target menu invocations required, for both age groups. However, contrary to our predictions, Glide did not provide any significant improvement for either group.

5.4.2 Speed and Accuracy

For the older group, a two-way repeated-measures ANOVA (Presentation order × Interface) on median trial time revealed a main trend for interface ($F_{2,12} = 3.44$, $p = .066$, $\eta^2 = .364$) and a significant interaction between interface and order ($F_{10,12} = 6.55$, $p = .002$, $\eta^2 = .845$). For the younger group, the same analysis yielded no significant results. Although selection accuracy is implicitly captured by speed (see Section 5.3.8, for completeness we examined the number of times an incorrect menu item was selected. As expected, there were no significant differences. For

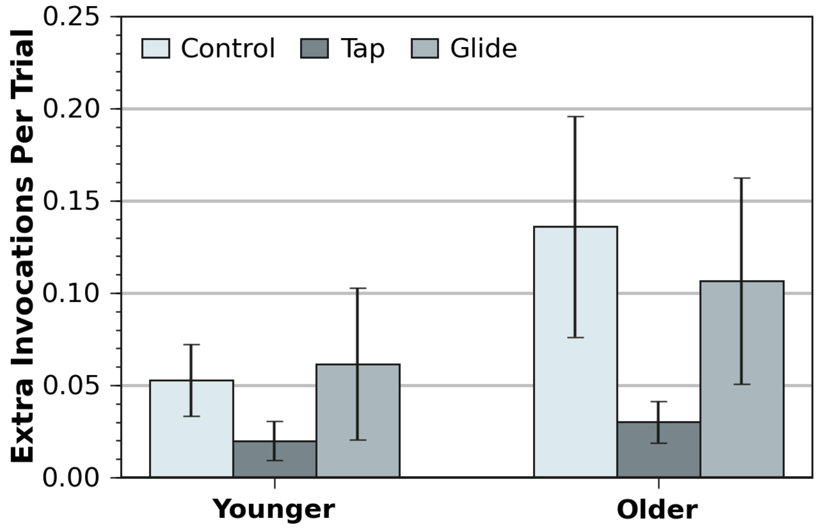

Figure 5.3: Mean number of extra target menu invocations per trial by interface and age group ($N = 24$). Error bars represent 95% confidence intervals.

the older group, the overall error rates were 1.4%, 1.4%, 1.8% for Control, Tap, Glide, respectively ($SD = 1.4, 1.4, 2.1$); and for the younger group they were 0.9%, 0.7%, 1.1% ($SD = 0.7, 0.7, 1.1$).

Figure 5.4 shows the average median trial times by interface and age group. Though it clearly highlights the large variability in this measure, it does support a trend towards the Tap interface being faster for the older age group overall. Figure 5.5 shows average median trial times for the older group by interface and presentation order, and more clearly demonstrates the interaction effect observed. To summarize, Tap was faster than Control in all three orders where Control was presented before Tap, but Control was faster than Tap only in the condition where Tap was presented first and Control last, suggesting that practice was playing a role in the overall speed improvements. Glide performed comparably to Control in all orderings except Glide-Control-Tap, where it was slower than both Tap and Control. With only 2 individuals per order and age group; individual differences likely played a role here.

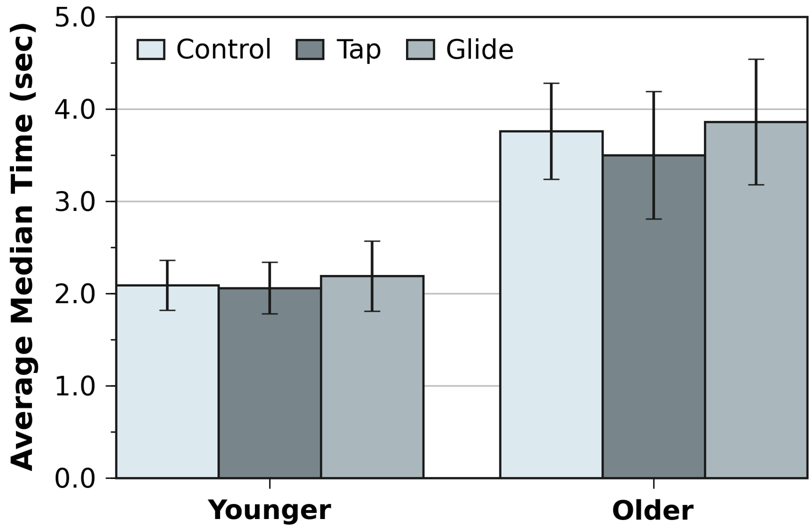

Figure 5.4: Mean trial time by interface and age group ($N = 24$). Error bars represent 95% confidence intervals.

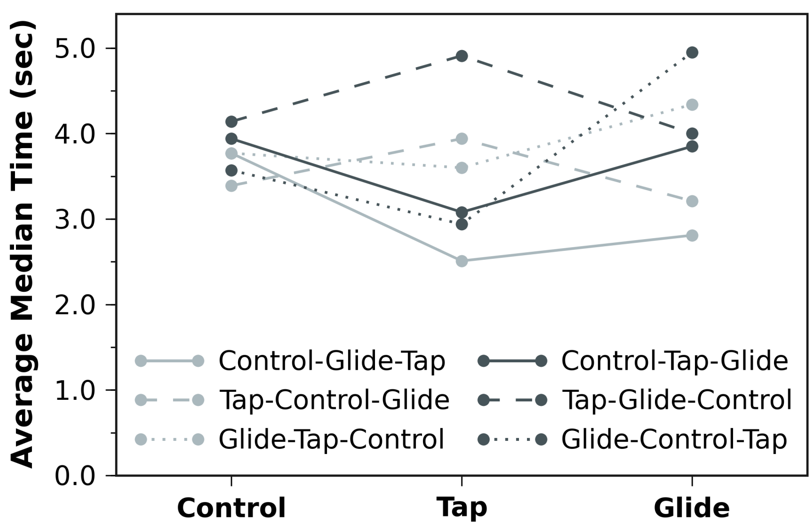

Figure 5.5: Mean trial time by interface and order for the older age group ($N = 12$). Error bars represent 95% confidence intervals.

Thus, hypothesis H2 was also supported for Tap, but not for Glide. Consistent with our predictions for Tap in hypothesis H2, Tap was not slower than Control. Moreover, a trend suggests there may in fact be a small speed benefit for the Tap interface for the older group. However, further research is necessary to confirm this trend, as there was also a significant interaction between interface and order. In contrast to our predictions for Glide in hypothesis H2 (but consistent with our negative findings for hypothesis H1) the Glide interface was not faster than either of the other two interfaces. As expected there were no differences for accuracy.

5.4.3 Age-Related Differences

Although age-related differences are not our main focus, for completeness we reexamined our variables of interest with age as between-subjects factor. Specifically, we performed a three-way repeated-measures ANOVA (Age group $\times$ Presentation order $\times$ Interface) on the time and extra target menu invocations required. Not surprisingly, there was a main effect of age on both measures (time: $F_{1,12} = 31.7$, $p < .001$, $\eta^2 = .726$; invocations, $F_{1,12} = 8.34$, $p = .014$, $\eta^2 = .410$). For extra target menu activations, there was also an interaction between age and interface ($F_{2,24} = 4.30$, $p = .025$, $\eta^2 = .264$) suggesting that the Tap interface disproportionately helped the older group. The remaining results were consistent with our separate analyses. These results should be considered preliminary due to the design limitations described in Section 5.3.6.

5.4.4 Subjective Response

After all three conditions, participants were asked to rank the interfaces according to five measures: preference, speed, accuracy, frustration, and initial ease. Table 5.1 summarizes their responses. We analyzed these results using the randomization test of goodness-of-fit [69, 102]. A chi-square test of goodness-of-fit is more commonly used for this type of analysis; however, our expected values (4) were too low for this test (which requires a minimum of 5). The randomization test uses Monte Carlo simulation to calculate the probability of the observed frequency values occurring by chance. It is robust against small sample sizes and low expected values [102].

	Number of Votes					
	Younger Group			Older Group		
Measure	Control	Tap	Glide	Control	Tap	Glide
Most Preferred	6	3	3	0	8	3
Least Preferred	1	7	4	4	3	5
Fastest	7	1	4	1	8	1
Slowest	0	7	4	3	3	6
Most Accurate	5	2	4	1	6	2
Least Accurate	5	3	3	2	1	7
Most Frustrating	2	5	4	5	1	5
Least Frustrating	7	3	1	0	8	1
Easiest Initially	4	5	3	0	7	3
Hardest Initially	3	5	4	5	2	5

Table 5.1: Summary of self-reported preferences (N=24). Recall that for the younger participants Glide was called Delay, which may have contributed to its lower ranking. Some participants did not answer all questions; thus some rows sum to less than 24. Significant rankings are denoted by a gray background.

Hypothesis H3 was partially supported. The Tap interface was well received by the older group, but there was some evidence of a negative response to it by the younger group. The older group rated Tap as the most preferred ($p = .009$), the fastest (p = 0.010), the least frustrating ($p = .003$), and the easiest initially ($p = .022$). In contrast, the younger group perceived Tap as the slowest ($p = .037$), and though not significant, it was also the least preferred by a majority of younger participants ($7/12$). In contrast, the comfort scale ratings did not reveal any differences among the interface. This might indicate that the preferences reflected in the rankings were not strong.

5.4.5 Summary

To summarize all our hypotheses were partially supported.

H1. Target Menu Invocations. Tap significantly reduced the number of extra target menu invocations used to make a selection for the younger group, and a trend suggests a reduction for the younger group as well. However, Glide did not result in fewer extra target menu invocations.

H2. Speed and Accuracy. There were no significant differences for either speed or accuracy. For accuracy, these findings were consistent with our predictions. For speed, they were consistent with our predictions for Tap, but not for Glide, which we had predicted would be faster than Control.

H3. Subjective Response. Contrary to our predictions, Tap and Glide were not consistently preferred to Control across both groups. The older adults ranked Tap as most preferred, fastest, least frustrating, and easiest initially, but the younger adults ranked it as slowest and there was some evidence that it was also least preferred for the younger group. Glide was not considered significantly different from Control on any of the subjective measures.

5.5 Discussion

In this section, we discuss our findings, focusing on their implications for design and on avenues for future research.

5.5.1 Eliminating Drifting Difficulties

Our results suggest that pen-based menus can be designed to prevent drifting and accommodate the needs of older adults, without impeding browsing. The success of the Tap interface for reducing excess target menu invocations for both age groups suggests that hover-switching could be removed entirely without compromising overall menu navigation.

However, the Tap interface was not well received by the younger group: they ranked it slowest and least preferred. What is not clear is the magnitude of this preference; further research would be needed to determine whether this is a mild or strong preference. In the meantime, a safer approach may be to support personalization and allow users to turn off hover-switching. One limitation of a personalization approach is that it is unclear whether individuals (and particularly older adults) can self assess drifting behavior. In both this study and the Baseline Study, users seemed largely unaware of why the wrong menu sometimes opened. Thus, one avenue for future investigation would be to examine whether drifting can be inferred from user input. If so, the system could preemptively present this information to the user and offer to deactivate hover-switching.

Finally, the contradictory preference results observed in this study are noteworthy as they underline the critical importance of evaluation across the lifespan. Although the Tap interface benefited both groups, had the study only focused on the younger demographic, the negative preference rankings for that group would have made interpretation difficult. Including both groups allowed us to see that the Tap interface does provide benefit, but likely the magnitude of that benefit is small enough for the younger group that personal preference dominates.

5.5.2 Study Limitations

One important limitation of this study is the null performance result for the Glide interface. One possible interpretation is that using a distance threshold is the wrong approach and it is not possible to distinguish between accidental drifts and intentional menu invocations by distance. However, we did only try one threshold. Although it was chosen based on our previous data, it is possible that a different threshold may have yielded better results. Because we designed this study such that it is not possible to directly identify drifts,[13] we cannot use our data to make preliminary predictions on other thresholds (as we did in the Baseline Study). Future research could further explore different thresholds, perhaps incorporating elements of both time and distance.

A second limitation is the significant practice effect that hindered performance in participants' first session. In retrospect, it is clear that we needed to provide more training. However, there is always a tension between maximizing training and minimizing study length. For the older adults, our study was already an average of two hours long. It is not clear we could have made it much longer.

[13]That is, we did not tell participants which menu contained the target item; thus, it is unclear whether or not an extra target menu invocation was the result of a drift. Though this limits our ability to directly identify drifts, we chose this design so that we could explore the utility of the hover-switching feature.

5.5.3 Accessible Hover-Space Interaction

Recently, there has been a growing interest in hover-space interaction as a means of increasing the functionality of pen-based systems [37, 42, 49]. For example, with Hover Widgets, users perform gestures above the surface of the screen to invoke functionality without moving away from their current work area [42]. Tracking Menus use the tracking state (of either a pen or a mouse) to enable a menu, such as a tool palette, to track the input device and remain close the user's work area [37]. One use of Stitching Gestures is to allow users to drag items (such as documents) between tablets with gestures in the hover-space [49]. Some research has also extended the capabilities of the hover space by exploring multiple layers of interaction [57, 107].

Although our work has focused on the negative potential of hover-space interaction, not all such interaction is necessarily detrimental to the performance of older adults. Many of the proposed uses of the hover space have been to provide shortcut-style access to features that can also be accessed via more standard mechanisms (e.g., with a button) [37, 42, 49], thus users need not rely on interacting within the hover space. Moreover, some hover-space research has paid careful attention to designing hover commands that cannot easily be triggered by accident. For example, Hover Widgets were designed so that their invocation gestures were both sufficiently simple to perform and sufficiently unambiguous that they would not be activated unintentionally [42]. Although it is unlikely that these interfaces will be usable by older adults, the combination of unobtrusive design and alternate access mechanisms should ensure that, at a minimum, they do not hinder performance.

5.5.4 Generalization to Other Interfaces

Finally, although we limited our investigation to menu interaction, we note that there are likely similarities between menu interaction and other interfaces (e.g. toolbars). With pen-based interaction, the combination of small widgets in close proximity, hand occlusion of the task area, and ambiguity in the interaction, may all contribute to the drifting difficulty. Proximity of the widgets is an important factor: Clearly as the distance between widgets increases, small movement away from the target is less likely to result in drifting. Similarly, if hand occlusion were not present or if users did not need to read the contents of the menu, then there would be less of a need to move away from the target area in the first place. Finally, hover-switching is ambiguous; drifting would not be a problem if the required action were not one the user was likely to perform unintentionally. Designers should consider all these factors when building pen-based interfaces to avoid introducing similar problems.

5.6 Summary

In this chapter, we presented the findings of an experiment comparing the effectiveness of two techniques designed to reduce accidental menu invocations caused by drifting; that is, unintentionally cursoring over an adjacent menu head while in the tracking state. The *Tap* interface, which turned off hover-switching and required users to make an explicit tap to switch between menus, was effective for both younger and older users in terms of reducing extra target menu invocations (an indication of drifting difficulty). However, despite the performance benefit, younger users did not like the Tap interface. The *Glide* interface, which used a distance threshold to differentiate between accidental drifts and intentional menu invocations, did not show a performance benefit. Further research is needed to examine whether this is because using a threshold is the wrong approach, or because the particular threshold used in this experiment was suboptimal.

Chapter 6

Mouse Study

Developing Steady Clicks

The previous two chapters focused on menu interaction difficulties. In the next two chapters, we turn our attention to general target acquisition difficulties. Recall that in the Baseline Study, we identified slipping as a major source of pen-based target acquisition difficulty for older adults. In this chapter, we present Steady Clicks, a mouse-based slip reduction technique designed to help motor impaired individuals click more accurately.[14] Though we focus mostly on older adults in this thesis, increasing the accessibility of technology for individuals with motor impairments was part of the original motivation for this work [77]. The work presented in this chapter is a step in that direction. Moreover, in the next chapter, we build on the technique presented here to address pen-based slipping errors for older adults.

6.1 Introduction and Motivation

Accurate pointing and clicking with a computer mouse can be a challenge for some users. Though hardware solutions to meet the needs of motor impaired individuals do exist, many people still prefer to use a standard mouse. There are a number of reasons for this preference. They may share a computer with others who use a mouse, or have used a mouse before acquiring a disability. They may find alterna-

[14]This work is the result of a collaboration among Dr. Shari Trewin, Dr. Simeon Keates, and myself during my internship at IBM T.J. Watson during the summer of 2005.

tives too expensive, or simply find that the mouse is easier to understand than other devices.

As a result, software that supports mouse users in pointing and clicking tasks is a valuable accessibility tool. While a number of researchers have investigated techniques to support easier and more accurate pointing, little work has addressed problems with clicking. This chapter describes clicking problems observed in an empirical study of older adults and people with Parkinson's Disease. It presents Steady Clicks—a clicking assistance technique intended to tackle some of the problems observed. Results from an evaluation of Steady Clicks with the target population are presented and discussed.

6.1.1 Mouse Clicking Problems

Studies of the effects of age and disability on computer use have identified that advanced age and disabilities make mouse use and movement increasingly inaccurate [17, 91, 110]. For example, Trewin and Pain [110] found that 14 of 20 participants with motor disabilities had error rates greater than 10% in a point-and-click task. Participants had difficulty positioning the cursor over small targets, and keeping the cursor over the target while clicking. Many of the participants also clicked the mouse button unintentionally before reaching the target.

In an exploratory study, Keates, Trewin and Paradise [58] examined pointing and clicking performance of 24 participants evenly split across the following groups: young adult (20–26), adult (40–59), older adult (73–82), and adult with Parkinson's disease (48–63). The results of this study indicated that older adults, and adults with Parkinson's disease, used very different movement strategies to the adults and young adults. They showed lower peak velocities and used many more submovements, with multiple pauses during the movement. These pauses were associated with movement around and through the target. Across all 3552 selections in their study, a total of 210 errors were observed, classified as follows:

Near misses (110). The mouse down position was within 50% of the target radius.

Not-so-near misses (35). The mouse down position was between 50% and 100% of the target radius.

Slips (32). The mouse button was pressed when the cursor was on the target but the cursor slipped off the target before the button was released.

Accidental clicks (14). Unintentional clicks made at a distance greater than 200% of the target radius (9), or cases where the user presses down a button, and then presses another button before releasing the first button (5).

Middle button press (2). The user pressed the wrong button.

Unclear (17). The remaining 17 errors were unclear.

The majority of errors were made by older adults (112). Perhaps surprisingly, young adults made as many errors as individuals with Parkinson's disease—35 compared with 34. All 32 of the slips, and 13 of the accidental clicks were made by the older adult and Parkinson's groups. In interviews [82], five of the six study participants with Parkinson's disease reported that accidental clicks were a problem for them. Two reported slips to be a problem. It is not clear why the error rates were so low for the Parkinson's group relative to the older adults. One possibility is that the individuals in the Parkinson's group had developed compensatory strategies, or slowed down more to accommodate their difficulties. One limitation of the study was that no visual feedback was provided when the cursor was over the target, and the visible border of the targets was not considered part of the target (i.e., selection of the border was counted as an error). The older adults in particular had difficulty seeing when the cursor was in the target. This contributed to the high number of near misses observed in the data.

As discussed in Chapter 2, a number of support techniques to aid cursor positioning have been proposed, including dynamic target expansion [71], area cursors [41, 129], 'sticky' targets [129] or the use of crossing actions instead of clicking to make selections [1, 127]. All of these techniques are designed to help a user get a cursor onto a target. If their demonstrated benefits can be shown to extend to real applications with multiple small targets in close proximity (e.g. toolbar buttons), and to users with motor impairments, they may significantly reduce near misses and not-so-near misses. Target expansion, area cursors and sticky icons may also reduce slip errors by helping to keep the cursor within the target while clicking, at least for targets that are not able to be dragged. However, none of these techniques reduce accidental clicks or wrong button presses, and none of them were specifically designed to reduce slipping. In this chapter, we propose a general assistance technique that tackles slips, accidental clicks, and wrong button presses. It could be used in conjunction with these existing techniques to address all observed sources of mouse-based target acquisition error.

6.1.2 The Steady Clicks Feature

Steady Clicks is designed to help in situations where people successfully click down on a target but slip before releasing the mouse button, accidentally click buttons while en route to the target, or click while trying to press a different button. Our software helps prevent slips by freezing the cursor at the button down location until either the button is released (causing a steadied click to occur), or the mouse is moved beyond the freeze threshold (returning it to normal operation). The freeze threshold defines the maximum Euclidean distance (currently 100 pixels) the cursor can travel away from the button down location.

Steady Clicks identifies accidental clicks based on two criteria: velocity and button status. It filters out all clicks that occur while the mouse is moving above the velocity threshold or while another mouse button is pressed. Velocity is calculated using the naive algorithm: Euclidean distance moved since previous mouse event divided by time elapsed since that event. The velocity threshold (currently 0.25 pixels per millisecond) was derived from the clicking data from the earlier exploratory study [58]. Although more accurate velocity calculations are possible, the approach chosen has the advantage that it does not need data from the future and is computationally quick to perform. This allows decisions to be made at the time the click event is received, producing a more responsive behavior.

Several design options were explored in the development of the freezing functionality of Steady Clicks, including:

Jitter. Should the cursor jitter slightly to provide visual feedback to the user that freezing is occurring?

Special Cursors. Should a special cursor be used to indicate freezing?

Jumping. When freezing releases, should the cursor jump to where it would be had freezing not occurred, or should it just release from its frozen position?

Animation. If jumping the cursor, should that jump be animated to smooth the transition?

Cursor trail. Should the unadjusted cursor location be indicated during a freeze? Should this be displayed with a trail of dots? Or should the cursor be left free but the button down location marked (e.g. with an '$\times$')?

For the evaluation prototype, we chose to implement Steady Clicks as follows. Freezing causes the cursor to jitter visually. When freezing releases, the cursor jumps to where it would have been if there had been no freezing. At the freeze threshold the cursor jumps, with no animation. There are no special cursors or other forms of visual feedback. These options were chosen based on initial exploration of the options by the researchers. The result is assistance that is unobtrusive, with minimal observable difference from standard cursor operation.

6.2 Experimental Methodology

An evaluation was performed to compare target selection using Steady Clicks to target selection without its assistance. The evaluation was done by mouse users who experience the specific issues that Steady Clicks is designed to tackle; that is, accidental clicks and slipping while clicking.

6.2.1 Apparatus

All computer tasks were presented on an IBM T41p ThinkPad running the Linux Fedora Core 3 operating system, and using an IBM 3 button optical mouse with scroll wheel. The experimental software was written in the C++ programming language using the X11 display protocol. In addition to presenting the study task, the software recorded time-stamped log files detailing each participant's cursor movements and mouse button presses. It also recorded the actions taken by the Steady Clicks feature, and the positions of the targets being clicked on.

Participant	Sex	Age	Disability	Computer Experience
P1	M	77	Parkinson’s	50 years
P2	F	27	Cerebral Palsy	11 years
P3	M	80	Parkinson’s	30–40 years
P4	M	32	Cerebral Palsy	5 years
P5	M	50	Multiple Sclerosis	2 years
P6	M	61	Neuromuscular condition	None postdisability
P7	F	54	Spinal injury	6 months
P8	F	36	Stroke	2 years
P9	F	38	Stroke	1 year
P10	F	44	Stroke	1 year
P11	M	44	Neuromuscular condition	35 years

Table 6.1: Participants in the Steady Clicks evaluation.

6.2.2 Participants

Table 6.1 describes the 11 individuals with motor impairments (5 women and 6 men) who participated in the evaluation. Participants were between 27 and 80 years old, with average and median ages of 49 and 44 respectively. Two of the participants had Parkinson’s disease, two had cerebral palsy, three had impairments resulting from a stroke, one had multiple sclerosis, another had spinal damage resulting from a gunshot accident, and two had impaired manual dexterity caused by unspecified neuromuscular conditions. Four participants used a computer daily, five used a computer several times a week, and two used a computer once a week or less often. Four had more than 10 years of computer experience. The remaining 7 had 0.5–5 years of experience. All were familiar with the standard mouse, and used that as an input device.

Participants received $50 for their involvement. Potential participants were screened in advance to identify individuals for whom the Steady Clicks feature would be most relevant. Where possible this screening was done by observing them using the mouse and asking them to use a clicking program that measured how much they slipped the mouse while clicking. A total of 18 individuals were screened in this way, and those who slipped the most while clicking, or made any accidental clicks during the screening, were included in the study. For participants

P1 and P3 this screening was not possible, and they were included based on a telephone interview.

6.2.3 Visual Acuity

A short vision test was designed to ensure participants had sufficient visual acuity to read the words used in the study task. Existing tests were considered, but precise vision data were not considered necessary. The test consisted of a series of short phrases presented to the participant in succession. As the test progressed, the font size of the phrases got smaller. The phrases were taken from popular nursery rhymes, with a few words selectively changed in each phrase to help ensure that the participant was reading and not reciting from memory. The smallest font size used in the test was 9 point; in the study the smallest font-size used was 18 pt.

6.2.4 Task

Two tasks were used in the study. In addition to the main clicking task used to evaluate Steady Clicks' ability to prevent slips and accidental clicks, a brief dragging task was included to gather preliminary feedback on Steady Clicks' potential negative impact on dragging.

Clicking task

The clicking task interface consisted of a 19-column by 30-row grid of rectangles. Each rectangle was 52 pixels wide by 22 pixels tall, and had a two to five character word printed on it in 18pt font. A screen shot of the interface is shown in Figure 6.1. All of the rectangular targets could be dragged; thus, slip errors resulted in the target being dragged the distance of the slip.

The clicking task was designed to make it clear where participants were supposed to be clicking, and when they were over the correct target, to reduce the opportunity for errors in which the participant intentionally clicked in the wrong location. For each trial, a single rectangle was selected to be the *target* item and highlighted in blue; participants were instructed to click on the blue target. Once the target item was successfully clicked on, the system automatically advanced to the next trial: the blue highlighting was removed from the previous target, and a

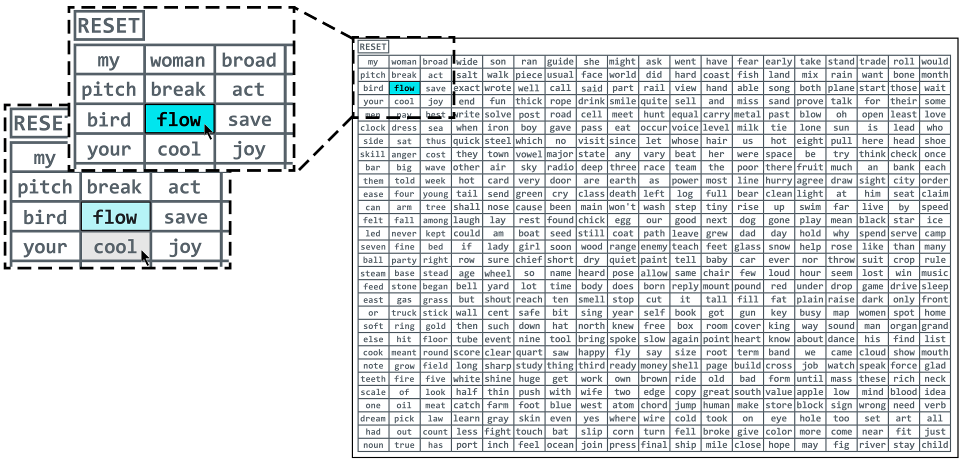

Figure 6.1: Screen shot of the evaluation task showing how the target underneath the cursor was highlighted when it was the desired target (upper magnification, cursor on 'flow') and when it was not (lower magnification, cursor on 'cool').

new target was selected and highlighted. Strong visual feedback was provided to indicate which target was under the cursor, as shown in Figure 6.1. Target items changed to a brighter blue color when the cursor was over them, and non-target items were highlighted with gray shading. The border around each target and non-target item was considered part of the target. Selection on the border resulted in selection of the corresponding item.

In order to obtain more natural clicking data, participants were asked to remember the words of the targets they clicked on. The intention was to engage the participants, so that they did not focus entirely on clicking, as they had to devote some of their attention to trying to remember the words. At the end of each block of 37 trials, participants were presented with a list of nine words, informed that three of the words had been targets in the last set, and asked to try to identify those three. A different set of nine words was used each time.

Additional realism was introduced by ensuring the user was aware of the errors they were making, and by making the penalty for errors roughly parallel the penalty inherent in a real-world web-browsing task as follows.

Drags (i.e., slips) on any item or left button clicks outside the target item. Caused the task to halt and a *reset* button was highlighted (using the same blue color as the highlighting for the target items) to indicate that clicking on it was required to restart the task. (The reset button was in the upper left corner of the screen as shown in the magnified section of Figure 6.1.) This was considered analogous to using the back button in a web browser.

Right or center mouse button clicks. Caused a window to pop-up. Clicking again hid the window. This was considered analogous to triggering a popup menu.

Dragging task

One possible disadvantage of Steady Clicks is the potential for it to interfere with dragging. A brief dragging task was included in order to get a qualitative impression of participants' initial reactions. This task employed the same rectangle grid as the clicking task. For each trial, two rectangles were highlighted and verbal instructions were given to drag one of the highlighted rectangles on top of the other, and then to press the reset button. The task consisted of five drags including drags shorter and longer than the freeze threshold.

6.2.5 Design

The experiment used a repeated-measures design with two counterbalanced conditions (with and without Steady Clicks). For each condition, participants completed two blocks of 37 trials, for a total of 74 trials with each interface. Trials were randomly generated to include a balanced number of short, medium and long movements presented at different angles from the previous target (the presumed starting cursor position). Each block presented a different set of targets and used a fresh word grid. The participants did not know what the next target position or the target words would be.

6.2.6 Procedure

Participants completed a vision test, a semi-structured interview, a clicking task performed both with and without the Steady Clicks feature, and a dragging task. The experiment was designed to fit into a single 90-minute session.

Each session began with the vision test. This was followed by the clicking tasks. When starting a new condition, participants were able to first practice on a few targets. Each block was separated by a segment of the semi-structured interview, to reduce fatigue effects, and to allow participants to provide feedback while the task was still fresh in their minds. The interviews were conducted verbally with responses being recorded by the experimenters. Again, this was done to minimize the physical effort demanded of the participants. After completing all of the clicking tasks (with and without Steady Clicks), the dragging task was presented, and a final interview gathered user feedback on dragging while using Steady Clicks.

When using Steady Clicks, participants were told "this program will keep the mouse steady while you are clicking and ignore any clicks you make while moving the mouse or clicking other mouse buttons." This information was provided to enable the participants to adjust their clicking strategies if they chose. This is also the most likely real-world scenario: An individual already familiar with clicking with a mouse tries using Steady Clicks, with some idea of what behavior to expect. When not using Steady Clicks, participants were told "this program operates as a regular mouse."

All of the participants were able to comfortably see text at the 18pt font size used in the experiment, but several found it cognitively difficult to read the words while performing the main selection task. For these participants, we read out the words on the target items as they clicked on them, to reduce the cognitive load to a more comfortable level.

6.2.7 Measures

For data analysis, the time spent in each trial was split into three states: the main target acquisition state, and two error states.

Target acquisition state. The target is available for selection, and no corrective action is needed. This is the default state.

Reset state. An error has occurred causing the task to halt. Selection of the reset button is needed to continue (i.e., to return to the target acquisition state).

Popup state. The popup window has been activated. A click is needed to clear the popup window and continue with the trial.

Finally, a semi-structured interview was used to: (1) gather information about each participant's background, computer experience, and web browsing behavior; (2) record the participants' subjective impressions and preferences during the clicking and dragging tasks; and (3) record observations made by the researchers during the session.

6.2.8 Hypotheses

We had the following hypotheses for this experiment.

H1. Participants will be able to click more accurately with Steady Clicks. Specifically, we predict that Steady Clicks will lead to fewer slips and accidental clicks.

H2. Participants will prefer clicking with Steady Clicks. We predict that with Steady Clicks participants will have to concentrate less on clicking, which will lead to higher satisfaction.

6.3 Results

In this section, we present the results of the study. We begin by examining the clicking activity in the two conditions, first in terms of errors in the target acquisition state, and then in terms of invocations of the error states. As a supplement to the error data, we additionally consider the effect of Steady Clicks on overall trial completion times and time spent in each state. Finally, we discuss the participant's subjective experiences, including their impressions of the dragging activity.

From the clicking task, with practice sessions excluded a total of 1554 trials were recorded, including 740 trials without Steady Clicks and 814 trials with Steady Clicks. The difference is due to two participants (P5 and P11) who found

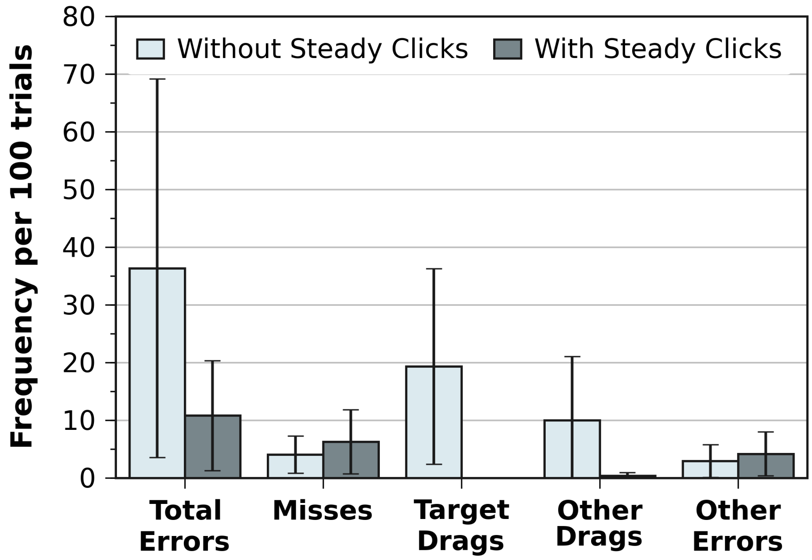

Figure 6.2: Frequency of errors per 100 trials ($N = 11$). From left to right the bars represent: total errors, misses (i.e., selections of a non-target item), drags on the target item, drags on a non-target item, and other (i.e., clicks outside the target region or blocked clicks). Error bars represent 95% confidence intervals.

the without Steady Clicks condition so tiring that they could not complete all the trials, and an additional seven trials for P2 that were lost due to a program crash. Both P5 and P11 used the without Steady Clicks condition first, then went on to successfully complete two blocks using Steady Clicks. Where necessary the data for P2, P5, and P11 were scaled.

6.3.1 Clicking Errors

Figure 6.2 shows the frequency of errors per participants, without and with Steady Clicks. To ease interpretation, we scaled the data to show the frequency per 100 trials. (Recall that there were 74 trials per condition.) In addition to the total errors indicated by the first set of bars, we provide a breakdown of the error types as follows.

Misses. A click on a non-target item.

Target Drags. A slip on the target item resulting in that item being dragged.

Other Drags. A slip on a non-target item resulting in that item being dragged.

Other Errors. All other errors.

When using Steady Clicks, participants slipped less leading to an overall reduction in errors. When Steady Clicks was not active, there was a total of 224 observed errors (across all participants). These were predominantly cases of a slip on the target item (124 instances) or a non-target item (55 cases), resulting in the item being dragged. Clicks on a non-target item (misses) accounted for only 28 cases. When Steady Clicks was active, almost all of the slips were suppressed—only a very long slip would break out of the freeze threshold and cause a drag, which happened 3 times (all on non-target items). The 88 errors with Steady Clicks consisted of 51 misses and 34 other errors. Wilcoxon signed ranked tests on each of the error types revealed a significant difference for total errors ($z = -2.67$, $p = .008$), target drags ($z = -2.81$, $p = .005$), and other drags ($z = -2.53$, $p = .012$). There was no significant difference for either misses ($p = .26$) or other errors ($p = .50$).

Wrong button presses were unexpectedly rare with Steady Clicks. Without Steady Clicks, participants pressed a wrong mouse button 27 times, causing a popup to appear. Of these, 13 were a result of two button presses that overlapped in time. Two of these had the wrong (middle or right) button going down first, while the other 11 had the intended (left) button going down first. For P5 and P11, there was a sequence of many errors in a row, where the user's finger had moved onto a position between two buttons (P11) or on the wrong button (P5) and they did not realize. Their first wrong button press caused a popup to appear, and their next press of any button cleared it. Sometimes this was the overlapping button press, so for example P11 produced and cleared the popup in a single action, and then did the same thing again, several times in a row. P5 kept pressing the scroll wheel, which caused a popup to appear and clear repeatedly. P5 and P11 account for 15 of the wrong mouse button presses, including 7 of the 13 examples of overlapping button presses. Surprisingly, wrong button presses were substantially less common when Steady Clicks was active: With Steady Clicks, only 6 wrong mouse button presses were observed. Of these, 50% were successfully blocked by Steady Clicks.

Overlapping button presses occurred 3 times, with 2 of these having the left button going down first and being blocked by Steady Clicks. Steady Clicks blocked a third wrong button press that occurred at high velocity.

The velocity filter helped block accidental clicks, but further refinement to the algorithm is needed. Twenty-eight clicks were blocked because the mouse velocity was too high prior to the mouse down event. Only 17 of these blocked clicks were genuine incorrect clicks, and these clicks were made by P10, P8 and P5. A further 11 clicks were incorrectly blocked by Steady Clicks. Nine of these errors occurred because the velocity calculation is based on the time period between two consecutive mouse events, and there were instances where this was only 4 ms or less. At such small time differences, even 1 pixel of movement caused the calculated velocity to exceed the threshold. This is an artifact of the operating system event reporting mechanism. The velocity calculation should be smoothed over a longer time period, or else movements of only 1 pixel should not be considered high velocity movements. Of these nine blocked clicks, 7 were actually over the target. Two further clicks were wrongly blocked due to a program bug (neither of these was on the target). Overall, the click blocking feature suppressed 17 high velocity errors but introduced a further 7 errors.

Steady Clicks lead to not only fewer errors, but also to better error recovery Because there were fewer errors in the Steady Clicks condition, only 6.6% of trials involved a reset operation, as opposed to 31.9% without Steady Clicks. This difference is reflected in our previous analysis of errors. Interestingly, however, the resets in the Steady Clicks condition were also performed with fewer mouse clicks per reset: 1.26 instead of 1.73. A Wilcoxon signed ranked test confirmed this difference ($z = -2.67$, $p = .008$). This finding indicates that additional errors were often made while correcting errors, and, underscores the need for error prevention techniques. We note, a similar difference was observed for clicks to clear a popup window after a wrong button press; however, analyzing this difference is not possible as the underlying behavior was vastly different between the conditions. (Recall that there were 27 wrong button presses without Steady Clicks and only 6 with Steady Clicks.)

	Average time per trial (sec)							
	Overall		Target Acq.		Reset		Popup	
	WO	W	WO	W	WO	W	WO	W
P1	3.30	3.20	3.30	3.2	0.0	0.0	0.00	0.00
P2	7.70	5.00	6.00	4.90	1.70	0.10	0.04	0.00
P3	2.80	3.30	2.70	3.10	0.10	0.20	0.00	0.04
P4	2.30	2.00	2.10	2.00	0.20	0.00	0.10	0.00
P5	55.20	10.90	43.00	9.60	10.20	0.90	2.00	0.30
P6	11.40	7.70	9.00	7.10	2.10	0.50	0.30	0.00
P7	4.60	2.20	3.90	2.20	0.70	0.00	0.00	0.00
P8	6.60	4.80	4.60	4.30	2.00	0.50	0.04	0.05
P9	6.10	3.20	4.10	3.00	2.00	0.20	0.00	0.00
P10	9.00	6.10	4.70	4.60	4.30	1.50	0.00	0.00
P11	21.00	5.30	12.60	4.80	8.30	0.50	0.10	0.00
Mean	11.80	4.90	8.70	4.40	2.90	0.40	0.20	0.04
Median	6.60	4.80	4.60	4.30	2.00	0.20	0.04	0.00

Table 6.2: Breakdown of average time per trial, overall and in each state, without (WO) and with (W) Steady Clicks. Rows highlighted in gray denote participants for whom Steady Clicks produced a significant difference in total time.

6.3.2 Movement Time

Table 6.2 gives an overview of the times taken, in seconds, with and without Steady Clicks, showing each participant's average time taken per trial, and time spent in target acquisition, reset and popup states per trial. For each of these measures, we performed a paired t-test. Ten of the participants had lower times when using Steady Clicks, but overall there was only a trend suggesting average trial times were faster with Steady Clicks ($t_{10} = 1.99$, $p = .074$, $\eta^2 = .285$). In terms of the individual states, there was no significant difference for time spent in the target acquisition state ($t_{10} = 1.68$, $p = .13$, $\eta^2 = .219$) or popup state ($t_{10} = 1.30$, $p = .22$, $\eta^2 = .145$). However, there was a significant difference for time spent in the reset state ($t_{10} = 2.58$, $p = .027$, $\eta^2 = .400$), indicating that Steady Clicks reduced the time spent in this state.

From data in Table 6.2, it is clear that individual differences played a large role in our time results. Averaging across participants can be misleading, particularly when individual differences are large [115]. We thus performed independent samples t-tests on each participant's trials. It is important to note that this type of analysis does not allow us to generalize beyond the participants in our study. However, it does provide insight into each individual's performance. Individually, there were six participants for whom the Steady Clicks condition did show a significant difference in the trial times observed: P2 ($p = .003$), P5 ($p = .005$), P7 ($p < .001$), P9 ($p < .001$) P10 ($p = .028$) and P11 ($p < .001$). These participants are highlighted with a gray background in Table 6.2.

6.3.3 Subjective Findings

On average, participants preferred working with Steady Clicks. Overall, 9 out of 11 participants preferred Steady Clicks, 1 (P8) preferred without and 1 (P3) had no preference. Eight participants felt that they worked faster with Steady Clicks, 2 (P8 and P10) reported the without condition to be fastest, and 1 (P3) reported no difference between the conditions. Seven participants felt that they had fewer errors while using Steady Clicks, 1 (P3) reported fewer errors without, and 3 (P1, P2, P7) reported no difference between the conditions.

There were no noticeable differences between conditions in the number of participants who noticed the mouse slipping while clicking (8 without and 7 with Steady Clicks), or who noticed accidental clicks (5 without and 6 with Steady Clicks). Five participants (P2, P4, P7, P9, P10) thought Steady Clicks had helped prevent them from slipping, and 3 (P8, P9, P10) thought it had filtered out unwanted clicks. No participant reported deliberate clicks being filtered out. One participant (P9) felt that Steady Clicks interfered with moving the cursor.

One participant (P5) described his clicking strategy without Steady Clicks as being to put the cursor in the corner of the target to compensate for upward slippage. He explained that he had to keep looking at his finger to see where it was positioned on the mouse. His index finger had a tendency to move to the right, sometimes causing a middle button click. He said that he could not predict, from looking at the screen, whether a click would work or not, and commented that it "takes a lot of time, so gets frustrating, because you want to be moving on. I can't feel where my finger is, so I get into position but then find my finger is not over the button." When using Steady Clicks, he commented "This is a lot easier because I'm thinking that the mouse isn't going to move. I feel a little confident [sic] in the program that I can move faster." He described that he was no longer looking down at his finger as much because he was working faster, so there was less finger slippage. He said "Now I'm not even going in the corner anymore. I'm just getting in the spot and moving on." He was initially not sure how much this was due to practice, and how much to the Steady Clicks feature. By the end of the clicking task his eyes were tiring. He expressed a clear preference for Steady Clicks, but commented that sometimes he became overconfident.

P11 reported a similar strategy of aiming at the top right of each target. She felt that this reduced slippage. She said "I have a habit of holding the button down too long. If I hold it down for half a second instead of 1 second it doesn't move so much." When using Steady Clicks she commented that it "...didn't seem to slip as much as it did earlier." P7 commented that there are "...some [mice] that when you move it jiggles and goes to another place. This one doesn't. This one stays where you want to put it."

6.3.4 Dragging Task

Participants P6 and P11 did not do the dragging task. P6 had never done dragging before. P11 ran out of time in his session.

From observation of the other participants performing long and short drags with Steady Clicks active, it was clear that Steady Clicks did interfere with dragging. Participants would start to drag, and then stop and start again because nothing seemed to have happened. However, after several attempts, without instruction, all succeeded in dragging the targets. Two participants commented that the delay before the drag would start did make dragging more difficult, but that this would not necessarily stop them from using the utility, if it was helpful in normal clicking. Three commented that they do not normally do dragging anyway. Surprisingly, 4 participants did not notice any difference at all when dragging with Steady Clicks, versus normal dragging.

6.3.5 Summary

Both of our hypotheses were supported by the data.

H1. Participants will be able to click more accurately with Steady Clicks. Supported. When working with Steady Clicks, participants made fewer errors overall. This result was mostly due to a reduction in slip errors.

H2. Participants will prefer clicking with Steady Clicks. Supported. Overall, 9 out of 11 participants preferred Steady Clicks. The majority of participants also thought they were faster and made fewer errors with Steady Clicks.

6.4 Discussion

When using Steady Clicks, participants had significantly fewer errors, and spent significantly less time in the reset state trying to correct errors. Time spent in the reset state of this task is effectively a magnification of the extent of an individual's problem, since it is caused every time an error is made, and compounded by further errors while clicking the reset button. When considered individually, significant time savings were observed for 5 individuals. The benefit of Steady Clicks for these individuals is due to the slip blocking effect. Those with the highest rates of slipping while clicking had the greatest benefit.

The click errors we observed break down very differently to those in Keates, Trewin, and Paradise [58]. Without Steady Clicks, 55% of click errors were entirely due to slipping (versus 15% previously), and 37% were misses (versus 69% previously). This is partly a reflection of the difference in participant population: This study specifically sought out participants for whom slipping was a problem. It is also partly a result of the task itself. Several participants commented that the strong visual confirmation of which target the cursor was over was very helpful, and made clicking easier than the clicking they normally do in other applications. This may have significantly reduced the number of errors that were misses. Furthermore, the targets used in the current study were larger than the smallest targets used in [58]; it was on these that the majority of misses occurred.

Our participants did not make many accidental clicks; most qualified for the study based on their slipping behavior. As a result, blocking of high velocity and overlapping clicks was less helpful for these individuals. Nonetheless, it is clear that performance could be improved by modifications to the algorithm used. The individuals themselves were aware that they made accidental clicks, and two correctly observed that Steady Clicks had filtered out unwanted clicks for them. None reported noticing when Steady Clicks mistakenly filtered correct interactions.

It is interesting that there were substantially fewer wrong button presses with Steady Clicks than without (6 versus 27). One possibility is that Steady Clicks made interaction so much easier that participants could concentrate less on clicking, and thus could pay more attention to ensuring they pressed the correct button. However, given the small size of our study, further investigation would be needed to confirm this.

Our preliminary look at dragging suggests that although Steady Clicks has a negative impact on dragging, for many users this may not be a big issue. Some users rarely do dragging. Others did not notice the effect, perhaps because they dragged quite quickly and broke the threshold before noticing the freeze. It is encouraging that users were able to complete the dragging tasks without help. Stronger visual feedback of the Steady Clicks freezing effect may be helpful.

The evaluation task required the user to take action to correct all accidental and slipped clicks, and this action was the same for every error. In real user interface tasks, accidental and slipped clicks have highly variable consequences. One slip may remain within the tolerance of the button being clicked, so that the click is successful, while another causes a folder to be dragged into another folder, or the focus to move to a different window. It is very difficult and time consuming to recover from such errors, because the user is often not sure what action they performed, or even aware that they have taken any action. The most important benefit of Steady Clicks for everyday use would be to prevent this kind of "I think I just did something but I don't know what" problem.

6.5 Summary

Previous studies showed that slipping while clicking and accidental clicks are a source of errors for older adults and individuals with disabilities when using a mouse. The Steady Clicks assistance feature suppresses these errors by freezing the cursor during clicks, preventing overlapping button presses, and suppressing clicks made while moving at a high velocity. Overall, the slip filtering feature was the most helpful. In the next chapter, we continue our investigation of slip prevention, but return to our main focus of pen-based interaction. In that investigation, we build on the Steady Clicks technique presented here.

Chapter 7

Technique Study Three

Methods to Reduce Slipping and Missing

In this chapter, we examine techniques to improve general pen-based target acquisition, and in particular to reduce slipping and missing errors. In the Baseline Study, slipping was identified as a major source of errors for the older adults. This finding was particularly interesting because slipping had not previously been identified a source of pen-based target acquisition error. Missing was less novel, but also a main source of error. Combined, missing and slipping accounted for 90% of the errors in the tapping task. Thus, in devising techniques to improve general pen-based target acquisition, we consider both slipping and missing.

7.1 Introduction and Motivation

In the tapping task of the Baseline Study, two target acquisition difficulties accounted for the majority of errors: (1) missing, landing and lifting outside the target bounds; and (2) slipping, landing inside the target bounds, but slipping out before lifting the pen. Missing was common to all age groups and remained relatively constant across age. In contrast, slipping was unique to the older participants and accounted for over half of the errors for the oldest group. We wanted to explore the feasibility of extending and combining existing techniques for younger users and mouse interaction to address the needs of older individuals using a pen. We, thus, focused on two promising mouse-based techniques: Steady Clicks and

Bubble cursor [41]. As these two techniques address different aspects of target acquisition, we also assessed the feasibility of combining them to capitalize on each of their strengths.

7.1.1 Steady Clicks

As discussed in the last chapter, Steady Clicks is a mouse-based technique designed to help in situations where the user successfully clicks down on a target but slips off before releasing. It works by freezing the cursor at the button down location until either the mouse button is released (causing a steadied click to occur), or the mouse is moved beyond the freeze threshold (returning it to normal operation). Although slipping is common to both mouse and pen interaction, with a mouse, it has generally been attributed to an inability to hold the mouse still while clicking. Tap selection does not have an analogous button clicking action, so it is not immediately clear that techniques designed to reduce slipping for the mouse will directly translate to pen interfaces. One potential barrier to using Steady Clicks with a pen is that it alters the ratio between mouse and cursor movement. The direct mapping between the cursor and the tip of the pen makes this less ideal.

7.1.2 Bubble Cursor

The Bubble cursor is a dynamic area cursor [129], in which a circular cursor grows and shrinks to capture the nearest (and only the nearest) target [41]. Evaluation of the Bubble cursor showed that it was faster and more accurate than a standard point cursor, and that its performance could be predicted by Fitts' Law by using the size of the cursor as the target's *effective width* (EW). Although this technique was not designed to address slipping, it essentially makes targets bigger in motor space, which should reduce the likelihood of a slip movement resulting in an error. Bubble cursors have not been evaluated with older adults, but the static area cursors upon which they are based have been shown to improve mouse-based pointing performance for older adults [129], suggesting promise.

7.1.3 Combining Approaches: The Steadied-Bubble

The Bubble cursor and Steady Clicks techniques each target different aspects of pointing. The Bubble cursor mostly helps ease the initial positioning of the cursor, while Steady Clicks is designed to help keep it steady once it is in place. Thus, it seems feasible to combine them into a single technique that fully captures the advantages of each. For our combined Steadied-Bubble approach, a circular Bubble cursor grows and shrinks to capture the nearest target while the pen-tip is within the hover-range of the display. Once the pen-lands on the screen, the cursor is frozen in both its location and size; that is, it is locked onto the last target captured before landing. If the pen moves beyond the freeze threshold, the Bubble cursor returns to its normal operation: the center of the area cursor tracks the tip of the pen, and the cursor grows and shrinks to capture the nearest target. Note that the freeze threshold is constant, but for any particular freeze, it may be larger or smaller than the radius of the Bubble cursor, depending on the target layout and density.

Each of the Steady Clicks, Bubble cursor and Steadied-Bubble approaches has inherent benefits and drawbacks. The Steady Clicks approach is cognitively simpler, but the mismatch between the cursor's position during freezing and the physical pen tip may be confusing to some users. Another disadvantage of this approach is that if the user misses the target on landing, it is harder to correct the selection by sliding the pen along the surface, as the user must first break the freeze threshold. An advantage of the Bubble cursor is that it could potentially address both slipping and missing. However, its effectiveness degrades as target density increases, making it least helpful when it is most needed. That is, when targets are dense, errors are more likely to activate unwanted functionality, and such errors are more costly than selections on inactive space. The combined Steadied-Bubble cursor seems most promising. The strong visual feedback provided by the Bubble cursor should help ease the mismatch caused by freezing, and overall it should offer the most support. However, it is more complicated than either technique on its own, which some users might find overwhelming.

To evaluate these tradeoffs, we conducted a laboratory experiment comparing the Bubble, Steady, and Steadied-Bubble cursors to each other and to standard point and tap, with younger and older adults. We found empirical evidence demonstrating that (1) Bubble was effective at reducing both slips and misses, but only when targets were not directly adjacent; (2) Steady was only effective at reducing slips, but its support was independent of target spacing; and (3) combining them into a single technique, Steadied-Bubble, successfully integrated the benefits of each—Steadied-Bubble prevented misses when targets were not adjacent, and slips independent of spacing. Our results demonstrate that both the Bubble cursor and Steady Clicks techniques can be successfully adapted for use in a pen-based interface, and that they are particularly helpful for older adults. They further establish that these techniques can be successfully integrated to address multiple target acquisition difficulties across a range of task contexts. Finally, we draw on these context-specific findings to establish design considerations for technique selection.

7.2 Experimental Methodology

We conducted a controlled laboratory experiment with younger and older adults to compare the individual and combined effectiveness of the Bubble cursor and Steady Clicks techniques for reducing pen-based targeting errors.

7.2.1 Apparatus

The experiment was conducted using a Wacom Cintiq 12WX direct input pen tablet, and a 2.26GHz Duo Core laptop with 2 GB of RAM and Microsoft Windows XP. A Cintiq Classic pen was used for all the computer tasks, with the barrel buttons deactivated to ensure participants did not use them unintentionally. The Cintiq senses 1024 levels of pressure and has a 307 mm (12.1 inch) diagonal display with a resolution of 1280 by 800 pixels (261 by 163 mm; i.e., each pixel measured 0.2 mm). The software was coded in Python using the Pygame SDK and the Wintab wrapper of the Python Computer Graphics Kit; it recorded all timing and error data. For the experimental tasks, the Cintiq was inclined using its built-in stand, which positioned the screen at a comfortable viewing angle (approximately 25 degrees from the horizontal). Participants were encouraged to adjust the position of their chair and the computer for comfort; most made these adjustments.

7.2.2 Pointing Techniques

We examined the following four *cursor types* (*CT*).

Control. Standard arrow cursor. No slip filtering.

Bubble. Dynamic Bubble cursor. No slip filtering.

Steady. Standard arrow cursor. Slips filtered.

Steadied-Bubble. Dynamic Bubble cursor. Slips filtered.

The Steady and Steadied-Bubble conditions filtered movement below a threshold of 60 px (12 mm). That is, while the pen remained within 60 px of its initial landing position, the cursor remained fixed at this position, and lifting the pen resulted in a selection event at the initial landing position. Once the pen moved more than 60 px away, the cursor returned to normal operation, tracking the tip of the pen; lifting resulted in selection at the lift position. The 60 px threshold was chosen to be larger than most of the slips observed in the Baseline Study. For the Bubble and Steadied-Bubble conditions, the cursor was rendered in a light semitransparent gray. We enforced a maximum cursor diameter of 100 px (20 mm) based on the suggestion by Grossman and Balakrishnan [41].

7.2.3 Participants

We recruited 12 participants from each *age group* (*AGE*).

Younger (19–30). Actual range: 19–29 ($M = 21$, 9 female)

Older (65+). Actual range: 65–85 ($M = 72$, 8 female)

The younger participants were recruited through campus postings. They received $15 for participating and completed the study in 60–80 minutes ($M = 68$ min). The older participants were recruited though community postings and word-of-mouth advertisement. They received $20 for their participation and completed the study in 75–120 minutes ($M = 88$ min). All participants were right-handed, free of diagnosed motor impairments to their right hand, and had normal or corrected-to-normal eyesight.[15] Additionally, they all had normal or corrected-to-normal eyesight.

To control for biases between age and Tablet PC experience, all were novices to pen-based computing. Within and across each age group, participants had a wide range of computer experience. Nonetheless, there were some notable differences: younger participants were more frequent users, used a greater number of applications, and were familiar with a greater number of advanced tasks. Their self-rating of expertise was also higher. However, the older participants had been using computers for longer.

7.2.4 Motor Skill

As in the previous studies, we administered standardized tests to gather data about our participants' motor abilities and ensure our participants were consistent with the norms for their age group. In this study, we used the digit symbol substitution test to measure perceptual speed [120], the right-hand component of the Purdue pegboard test to measure motor-coordination [108], and a simple reaction time test to measure reaction speed [125].

7.2.5 Task

Our task is shown in Figure 7.1, and was modeled after the task used by Grossman and Balakrishnan [41]. Participants selected a series of goal targets, which appeared at unpredictable locations. To control the inactive whitespace around the goal target, four neighboring distracter targets were equidistantly placed around the goal, in the line of and perpendicular to the axis of approach. Additional distracter targets were placed in the scene to create varying levels of overall target density

The goal target was rendered as a solid green circle, and distracter targets as gray outlined circles of the same size. Visual feedback was provided by changing the appearance of targets: (1) when the pen tip hovered over a goal target it turned a deep red (distracter targets turned a solid dark gray), and (2) when the pen tip touched a goal target it turned a brighter red with a dark red border (distracter tar-

[15] Based on self-reported data.

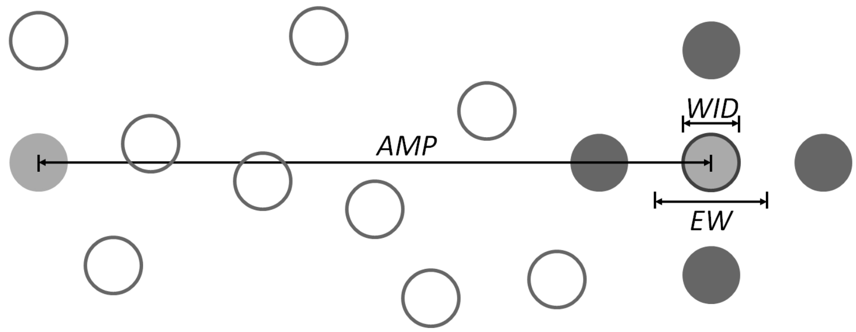

Figure 7.1: Experimental setup. Light gray denotes the previous and current goal targets, with the current goal indicated by a dark outline. The rest are distracter targets, with the 'neighboring' distracters filled in dark gray. ($DEN = 0.5$, $EWR = EW/WID = 2$.) Note the colors used in this figure are for illustrative purposes only. The colors used in the experiment are described in the text.

gets similarly turned a lighter gray with a dark gray border). We provided the latter form of feedback to help participants determine how much pressure was needed.

Consistent with Grossman and Balakrishnan's study design [41], we varied the following factors.

Target Width (*WID*) specifies the diameter of the goal target. We used three target widths for the experiment: 12, 24, and 36 pixels (2.4, 4.8, and 7.2 mm). These sizes are in line with previous studies (e.g., [41, 52, 89]), and roughly correspond to the following common widgets: the height of a text link, the size of a small toolbar icon (or the height of a menu item), and the size of a larger icon (e.g., the height of Windows Start menu items using the 'large icon' option).

Amplitude (*AMP*) is the distance to the goal target from the starting position of the trial (i.e., from the previous trial's goal). We examined three amplitudes: 256, 382, and 512 pixels (51, 76, and 102 mm) to explore a range of distances.

Effective Width Ratio (*EWR*) specifies the amount of inactive whitespace surrounding the goal target (i.e., the distance from the goal target to its neighbors). Target spacing is particularly important for Bubble (and by extension, Steadied-Bubble). Thus, we express it as the ratio between the effective target width (for Bubble and Steadied-Bubble) and the actual target width (i.e., $EWR = EW/WID$). We used three values for this factor: 1, 2, and 3. When $EWR = 1$, the goal target is directly adjacent with its neighbors.

Distracter Density (*DEN*) refers to the number of other targets on the screen. We used the same three levels as in [41]: 0, 0.5, and 1. $DEN = 0$ reflects no distracter targets (except for the four neighbors), $DEN = 0.5$, a moderate target density (see Figure 7.1), and $DEN = 1$, a high target density. (For complete details on the distracter placement, see [41].)

At the start of each new cursor condition (CT), participants were introduced to the cursor and given 10 practice trials. Participants then completed four blocks of trials with each cursor, with an enforced 45 second break between blocks. Each block consisted of 81 trials representing one of each possible combination of *WID*, *AMP*, *EWR*, and *DEN*, for a total of 324 trials per condition. The order of presentation of trials was consistent with [41] and was as follows. Each combination of *WID*, *EWR*, and *DEN* was presented in a random order. Within each of these combinations, all three levels of *AMP* were presented together (in a random order). No rationale was provided in [41] for this approach, but we suspect it was done to provide continuity between trials. Our early pilot runs of the experiment also found that having all four factors change every trial was disorientating.

7.2.6 Design

We used a $2 \times 4 \times 3 \times 3 \times 3 \times 3$ mixed factorial design, with *AGE* as a between-subjects factor, and *CT*, *WID*, *AMP*, *EWR*, and *DEN* as within-subjects factors. *CT* was a within-subjects factor to increase the power of the design. To minimize the impact of learning effects, each participant was assigned to one of the following four presentation orders, which were chosen to form a balanced Latin square.

- Control – Steady – Bubble – Steadied-Bubble
- Steady – Steadied-Bubble – Control – Bubble

- Steadied-Bubble – Bubble – Steady – Control
- Bubble – Control – Steadied-Bubble – Steady

7.2.7 Procedure

The experiment was designed to fit into a single 120-minute session. We began with the motor tests, in the order: simple reaction time, Purdue pegboard, and digit symbol substitution. Next, participants were asked to complete a brief questionnaire about their background and computer experience. They were then introduced to using the pen-based device, and shown that (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. Once participants were comfortable using the pen, the Cintiq was calibrated to each participant using the built-in calibration utility. Calibration was repeated until both the experimenter and the participant were satisfied with the alignment of the cursor and pen tip.

Participants then completed the experimental tasks. Following each condition, participants completed a short questionnaire about that condition. Between conditions, participants completed short verbal distracter tasks. These tasks were chosen to engage participants mentally, but not physically, allowing them to rest their arms. At the first break, participants completed the North American adult reading test [104]; at the second break, the FAS test of verbal fluency [80]; and at the third break, a reverse digit span test [120]. At the end of the experiment, participants were asked to rank the interfaces on a number of factors and encouraged to make additional comments.

7.2.8 Measures

For accuracy, we measured errors individually as the total number of slips and the total number of misses in each condition. We additionally included trial time as a measure, to provide an overall indication of performance. For time, the median was used to reduce the influence of outlier trials, and an implicit error penalty was used to discourage participants from overly focusing on speed. This penalty meant that participants could not advance to the next trial until they correctly completed the current trial. However, we note that while it is helpful in terms of motivating a balance between speed and accuracy, this penalty underestimates the true cost of errors as it treats errors on non-goal (distracter) targets the same as errors on inactive whitespace. In real-world tasks, selection of an unwanted feature typically requires additional corrective action. Finally, subjective data were collected after each condition using the ISO 9241-9 independent ratings of comfort scale [35]. At the end of the study, participants ranked the cursor techniques on speed, ease, frustration, and preference.

7.2.9 Motivation

To motivate quick and accurate performance, a $10 incentive was awarded to the top third of performers in each age group. The one-third ratio was chosen to encourage participants to believe they had a reasonable chance of succeeding. Top performers were those who were fastest to correctly complete the tasks. (Remember that participants could not advance until they correctly selected the goal target). To help participants gauge their performance, graphical feedback was presented during the breaks between blocks. This feedback consisted of a graph of their speed for all blocks completed with that cursor, and a text summary of their total time and errors for their most recent block.

7.2.10 Hypotheses

Our hypotheses are all relative to Control.

H1. Bubble will reduce both slips and misses, but only when the surrounding targets are not directly adjacent. That is, when $EWR = 1$, there will be no difference between Bubble and Control, but when $EWR > 1$, Bubble will result in both fewer slips and fewer misses.

H2. Independent of target spacing, Steady will reduce slips, but it will not affect misses. That is, we predict that for all EWR, Steady will result in fewer slips, but not fewer misses, than Control. Combined with H1, Bubble will reduce more errors overall than Steady, for $EWR > 1$.

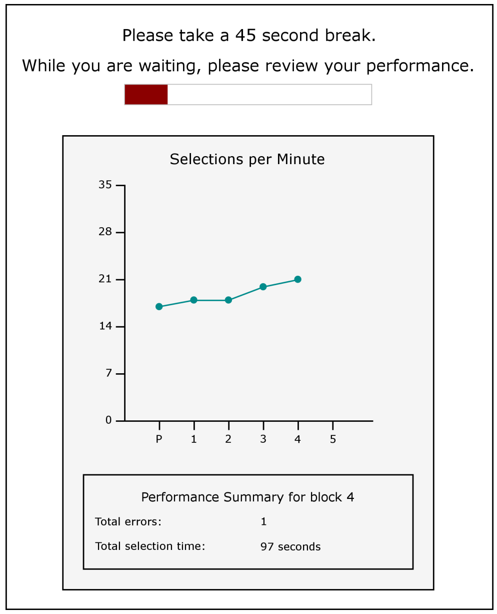

Figure 7.2: The performance feedback presented between blocks. The graph indicated speed for each block completed including the practice block (recall the errors were implicitly captured by speed) and a summary of the errors and time for the most recent block. In this example, the user has just completed block 4.

H3. Steadied-Bubble will reduce slips when targets are directly adjacent, and both slips and misses when they are not. That is, it will fully integrate the individual benefits of Bubble and Steady. Correspondingly, it will be the most effective technique at reducing errors.

H4. The experimental cursors (Bubble, Steady, and Steadied-Bubble) will reduce total errors for both age groups, but the impact will be larger for the older group as they will benefit from both slip and miss reduction. This hypothesis is based on the results of the baseline study, which suggested older users make both slip and miss errors, whereas younger users predominantly miss.

H5. A greater proportion of errors in Bubble and Steadied-Bubble will land on a distracter target. Although Bubble and Steadied-Bubble will both result in fewer errors than Control, when errors do occur, they will be more likely to hit a distracter target. For Steady, we do not predict an increase in distracter hits.

7.3 Results

For each of our main measures (trial time, misses and slips), we performed a repeated measures ANOVA. For trial time, we performed a full analysis across all of our factors of interest. For the error results, we focus our analysis on just those factors for which we have hypotheses (AGE, CT, WID, EWR), collapsing across the remaining factors. Initial analysis of the data including all factors did not suggest any main or interaction effects for AMP and DEN, and with only four trials (per each combination of all factors), the error data were too coarse-grained to measure the differences in which we were interested at this level of analysis.

In our reporting of F-statistics, where df is not an integer, we have applied a Greenhouse-Geisser adjustment for non-spherical data. All pairwise comparisons were protected against Type I error using a Bonferroni adjustment. Along with statistical significance, we report partial eta-squared (η^2), a measure of effect size. Roughly speaking, .01 is considered a small effect, .06 medium, and .14 large [27].

7.3.1 Errors

Consistent with other research [52, 89], the majority of errors in this study occurred on the smallest target ($WID = 12$). For the larger widths, error rates were low and skewed towards zero, suggesting a floor effect. Skewed data can invalidate the results of an ANOVA analysis; thus, we focus our statistical analysis on just the trials with $WID = 12$. We do note, however, that while the largest, and correspondingly the most practically significant differences, occur at $WID = 12$, a similar but highly attenuated pattern is evident for the other widths, as shown in Figures 7.3 and 7.4. In addition Table 7.1 provides a summary of the error rates for the younger and older groups.

Even after filtering out the larger target widths, there remained cases where error rates were floored. This is not entirely unfortunate; all these cases corresponded to instances where an experimental cursor substantially reduced one of the error

			M	*SD*	*Min*	*Max*
Younger	Control	Misses	7.5%	6.1%	3.1%	24.4%
		Slips	3.9%	3.0%	0.3%	11.1%
		Total	11.5%	7.4%	4.0%	28.7%
	Steady	Misses	8.5%	8.8%	2.2%	33.6%
		Slips	1.7%	2.5%	0.3%	9.0%
		Total	10.2%	9.6%	2.5%	35.2%
	Bubble	Misses	3.4%	2.3%	0.6%	8.0%
		Slips	1.9%	2.0%	0.0%	7.7%
		Total	5.2%	3.3%	1.5%	12.0%
	Steadied-Bubble	Misses	5.1%	3.7%	1.5%	14.5%
		Slips	1.1%	2.4%	0.0%	8.3%
		Total	6.2%	5.0%	1.5%	16.0%
Older	Control	Misses	16.3%	9.5%	4.3%	31.2%
		Slips	5.9%	3.4%	0.6%	11.1%
		Total	22.2%	11.7%	7.4%	41.0%
	Steady	Misses	18.2%	12.3%	6.2%	45.7%
		Slips	1.1%	0.6%	0.3%	1.9%
		Total	19.3%	12.6%	6.8%	47.5%
	Bubble	Misses	8.3%	4.4%	2.8%	16.4%
		Slips	2.3%	1.5%	0.6%	6.2%
		Total	10.6%	4.9%	4.0%	19.4%
	Steadied-Bubble	Misses	10.5%	5.6%	3.4%	19.8%
		Slips	0.4%	0.9%	0.0%	2.5%
		Total	10.9%	6.0%	3.4%	19.8%

Table 7.1: Overall error rates by AGE and CT ($N = 23$).

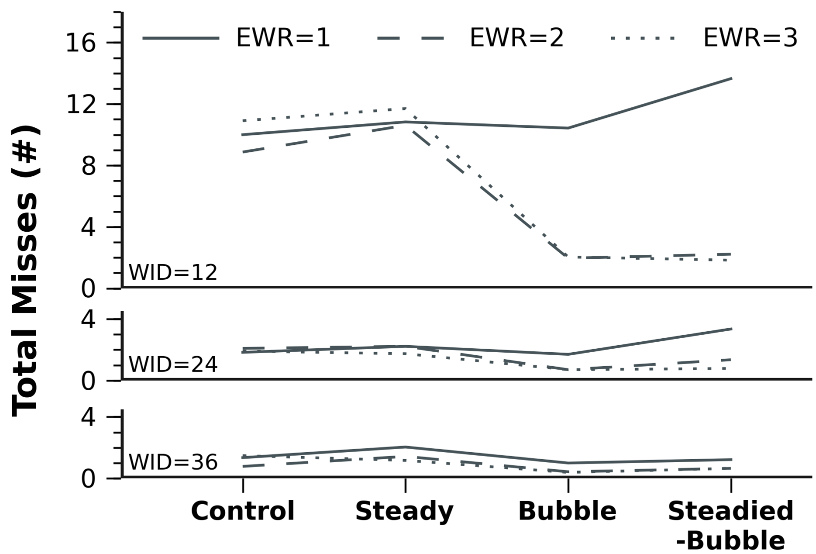

Figure 7.3: Average total misses by CT, WID, and EWR ($N = 23$, for 36 trials).

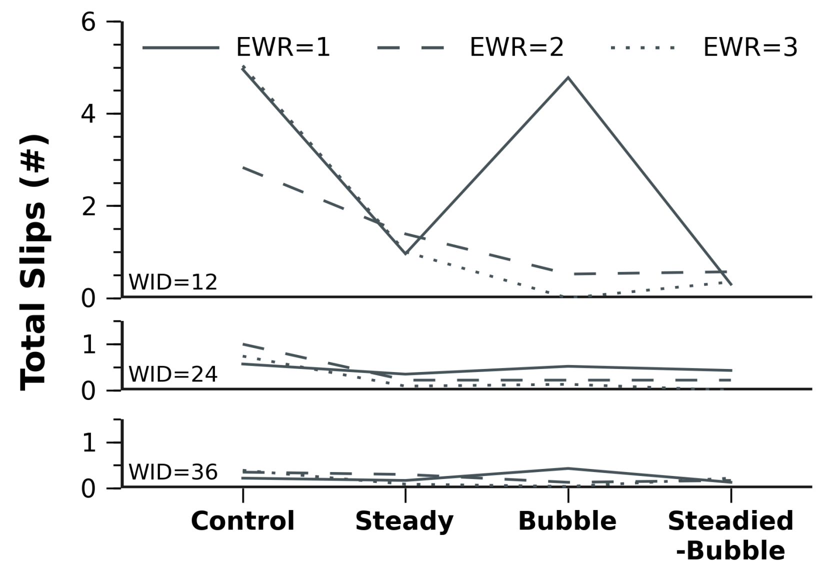

Figure 7.4: Average total slips by CT, WID, and EWR ($N = 23$, for 36 trials).

types (i.e., none of the results for Control are floored). However, to ensure that these measures do not bias the statistical results, we additionally relied on confidence intervals to aid our interpretation of the ANOVA results. Specifically, we only report those significant pairwise comparisons between cursor types from the ANOVA, where confidence interval analysis also found a significant difference.[16]

One participant in the younger group had unusually high error rates. His performance was outside the 1.5 interquartile range and more than two standard deviations from the mean. Although analysis with and without him yields the same conclusions, we exclude him to better reflect the performance of the younger group as a whole.

7.3.2 Miss Errors, *WID = 12*

Bubble and Steadied-Bubble significantly reduced misses when targets were not adjacent. There was a significant main effect of CT, and as shown in Figure 7.5, a significant interaction between CT and EWR (CT: $F_{1.9,40.1} = 8.32$, $p = .001$, $\eta^2 = .284$; $CT \times EWR$: $F_{3.9,82.0} = 11.0$, $p < .001$, $\eta^2 = .344$). Pairwise comparisons revealed that Bubble and Steadied-Bubble both resulted in significantly fewer misses than Control and Steady for $EWR = 2$ and $EWR = 3$ (all $p < .005$), but not for $EWR = 1$. Steady was not significantly different from Control for any EWR.

The older adults missed significantly more. As shown in Figure 7.6, there was a main effect of AGE ($F_{1,21} = 15.5$, $p = .001$, $\eta^2 = .425$). On average, the older group missed 2.67 times for every miss by the younger group.

There was also a main effect of EWR and a significant interaction between AGE and EWR (EWR: $F_{2,42} = 39.8$, $p < .001$, $\eta^2 = .655$; $AGE \times EWR$: $F_{2,42} = 5.24$, $p = .009$, $\eta^2 = .200$). These results simply mirror the other findings. For $EWR = 1$, both groups incurred roughly twice as many misses as they did for $EWR = 2$ and $EWR = 3$ because Bubble and Steadied-Bubble were not effective at reducing errors at $EWR = 1$. Moreover, because the older adults missed more in

[16] A confidence interval is an indication of the reliability of a measured estimate. It is more conservative than an ANOVA analysis because it does not pool variances; thus, a floor effect in one level of measurement does not affect the confidence intervals of other levels of measurement. To aid the reader, 95% confidence intervals are included as error bars in all our graphical results. Nonoverlapping error bars represent significantly different results.

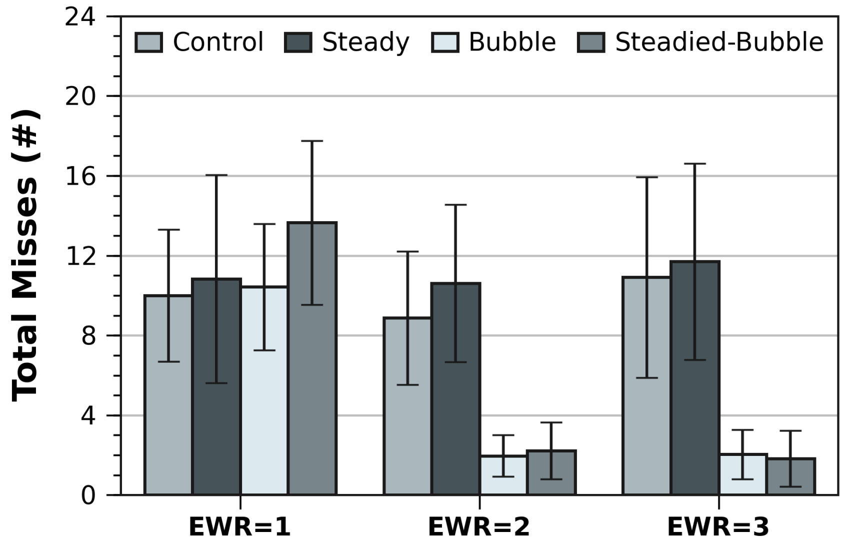

Figure 7.5: Average total misses (for 36 trials) by CT, and EWR ($WID = 12$, $N = 23$). Error bars are 95% confidence intervals.

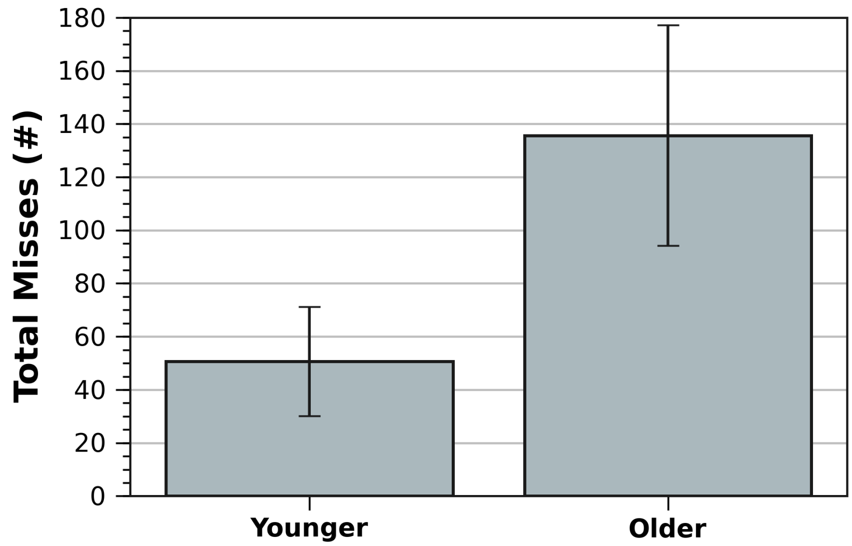

Figure 7.6: Average total misses (for 1296 trials) by AGE ($WID = 12$, $N = 23$). Error bars are 95% confidence intervals.

general, this doubling at $EWR = 1$ resulted in a greater increase for them, which explains the interaction.

7.3.3 Slip Errors, *WID = 12*

Steady and Steadied-Bubble reduced slips and their performance was consistent across target spacings. Bubble also reduced slips, but only when targets were not adjacent. There was a significant main effect of CT and a significant interaction between CT and EWR (CT: $F_{1.4,29.8} = 29.2$, $p < .001$, $\eta^2 = .582$; $CT \times EWR$: $F_{2.9,62.6} = 15.5$, $p < .001$, $\eta^2 = .425$).

Pairwise comparisons of the $CT \times EWR$ interaction (shown in Figure 7.7) revealed that Bubble resulted in significantly fewer slips than Control when $EWR = 2$ and $EWR = 3$ (both $p < .001$), but that it was not significantly different from Control at $EWR = 1$ ($p = 1.00$). Steady and Steadied-Bubble resulted in significantly fewer slips than control for all EWR (all $p < .005$), except at $EWR = 2$, where Steady and Control were not significantly different ($p = .14$). Though this last comparison is somewhat inconsistent with our hypotheses, we note that, as shown in Figure 7.7, it is mostly a reflection of lower than expected slip results for Control at $EWR = 2$.

The older adults benefited more from the experimental cursors. As shown in Figure 7.8, there was a significant interaction between $CT \times AGE$ ($F_{1.4,29.8} = 4.20$, $p = .036$, $\eta^2 = .167$). Pairwise comparisons revealed that the experimental cursors reduced the performance gap between younger and older users: For Control, the older group slipped significantly more than the younger group ($p = .05$), but there were no significant differences between the groups for any of the other interfaces. (A trend additionally suggests a difference between the two groups for Bubble, $p = .08$. This likely reflects the influence of the Bubble cursor's increased slip rate at $EWR = 1$.)

As with misses, there was a main effect of spacing (EWR: $F_{2,42} = 19.6$, $p < .001$, $\eta^2 = .482$), indicating more slips at $EWR = 1$. This was also a mirroring of the main results. Both Control and Bubble had a large number of slips at $EWR = 1$, while for $EWR = 2$ and $EWR = 3$, only Control did. Unlike misses, there was no

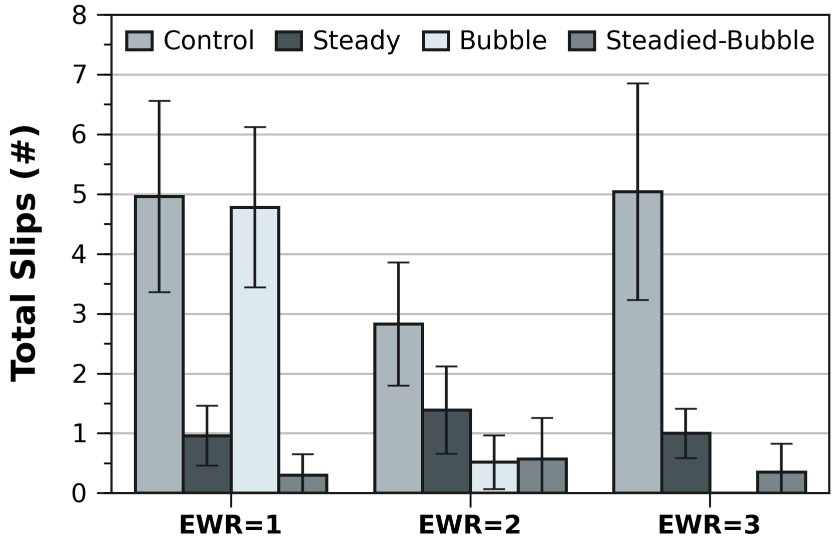

Figure 7.7: Average total slips (for 36 trials) by CT, and EWR ($WID = 12$, $N = 23$). Error bars are 95% confidence intervals.

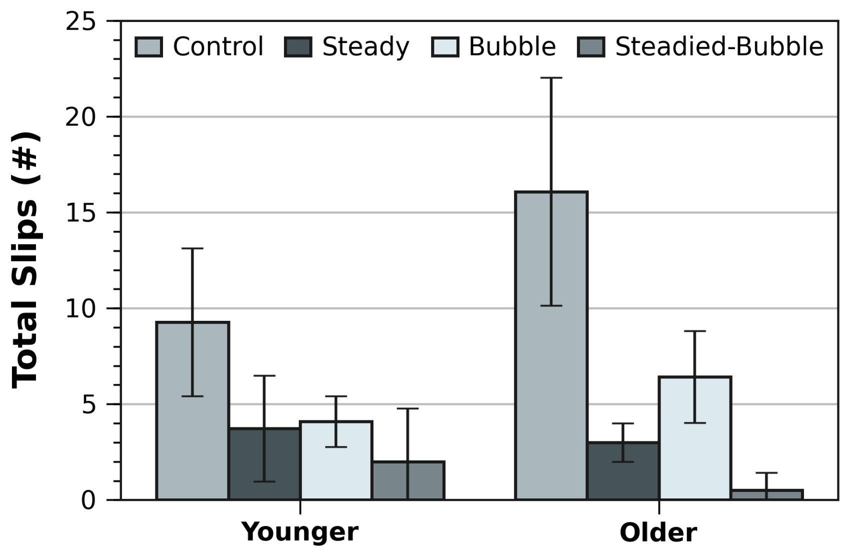

Figure 7.8: Average total slips (for 108 trials) by CT and AGE ($WID = 12$, $N = 23$). The only age difference was for Control. Error bars are 95% confidence intervals.

$EWR \times AGE$ interaction. This is most likely because, for slips, the older adults performed comparably to the younger adults for three of the four cursors.

7.3.4 Distracter Target Hits

Errors in Bubble and Steadied-Bubble were almost four times more likely to hit a distracter target. As shown in Figure 7.9, almost 100% of the errors in Bubble and Steadied-Bubble landed on a distracter target, while for Control and Steady the percentages were much lower: 28% and 23%, respectively. This difference was confirmed with an RM ANOVA on AGE and CT (main effect of CT: $F_{2.3,49.9} = 534$, $p < .001$, $\eta^2 = .960$). Pairwise comparisons confirmed the percentages were higher for Bubble and Steadied-Bubble than for Control or Steady (all $p < .001$). (There were no main or interaction effects for AGE.) Even when the average total errors are considered, there were roughly twice as many errors on a distracter target in Bubble and Steadied-Bubble (For Control, Steady, Bubble, and Steadied-Bubble, respectively: $M = 14.2, 12.0, 25.5, 27.4$; $SD = 9.6, 11.7, 15.9, 18.9$). Thus, even though there were fewer errors in Bubble and Steadied-Bubble, there were more errors on a distracter target.

7.3.5 Movement Time

Not including break times, the experimental tasks for each condition took on average 5.5 and 7.9 minutes, for the younger and older groups respectively ($SD = 0.87, 1.25$). We note that the time data were more sensitive than the error data, and thus, we were able to detect differences for the larger target widths (i.e., $WID = 24$ and $WID = 36$). Moreover, the time data were not skewed to zero, so we take full advantage of the power of our within-subjects design and do not rely on confidence intervals for this analysis.

Consistent with the results for misses, Bubble and Steadied-Bubble were significantly faster than Control, except when targets were adjacent. There was a significant $CT \times WID \times EWR$ interaction, as well as all the corresponding main and 2-way interactions ($CT \times WID \times EWR$: $F_{4.3,.81.1} = 7.27$, $p < .001$, $\eta^2 = .277$; CT: $F_{3,57} = 31.9$, $p < .001$, $\eta^2 = .627$; WID: $F_{1.0,19.4} = 291.5$, $p < .001$, $\eta^2 = .939$; EWR: $F_{2,38} = 91.0$, $p < .001$, $\eta^2 = .827$; $CT \times WID$: $F_{2.1,39.8} = 20.1$, $p < .001$,

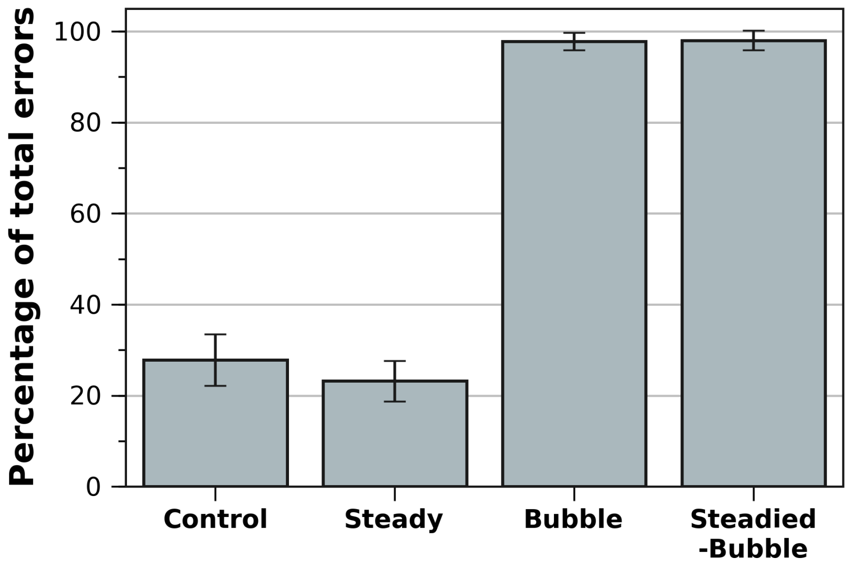

Figure 7.9: Average percentage of errors hitting a distracter target ($N = 23$). Error bars are 95% confidence intervals.

$\eta^2 = .514$; $CT \times EWR$: $F_{3.6,67.6} = 10.1$, $p < .001$, $\eta^2 = .346$; $WID \times EWR$: $F_{1.9,35.6} = 45.5$, $p < .001$, $\eta^2 = .706$).

As shown in Figure 7.10, cursors based on the Bubble cursor were faster than Control and Steady, for $EWR = 2$ and $EWR = 3$. Figure 7.11 shows this same data broken down by WID, highlighting that the gains were largest at $WID = 12$. Pairwise comparisons confirmed these differences. For $EWR = 2$ and $EWR = 3$, Bubble was significantly faster than both Steady and Control (all WID, $p < .05$) and Steadied-Bubble was significantly faster than Steady at all widths ($p < .05$) and Control at $WID = 12$ and $WID = 24$ ($p < .05$). At $EWR = 1$, none of the comparisons were significant.

The older adults were slower than the younger adults and disproportionately affected by the task factors. There was a main effect of AGE and several interactions involving AGE. In all the interactions, both age groups showed similar patterns of results, but the effects were often magnified for the older adults. For both groups, speed decreased as target width increased and amplitude increased; however, the older group was disproportionately slowed by smaller targets and larger

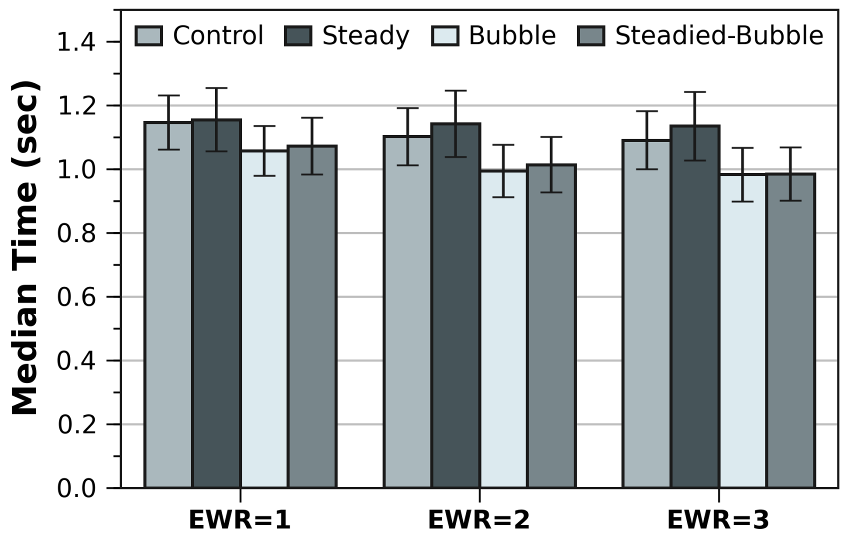

Figure 7.10: Average median trial time by CT and EWR ($N = 24$). Error bars are 95% confidence intervals.

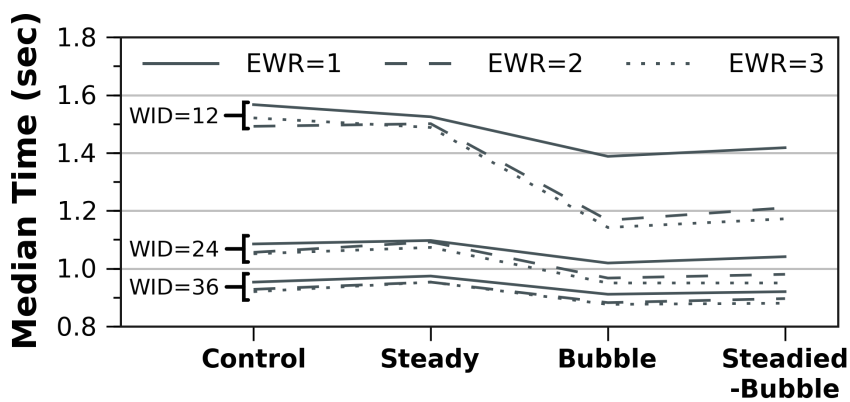

Figure 7.11: Average median trial time by CT, EWR, and WID ($N = 24$).

amplitudes. Moreover, inspection of the significant $AGE \times CT \times WID$ interaction revealed that at the smallest target size, the older adults benefited more than the younger adults from Bubble and Steadied-Bubble.[17] (AGE: $F_{1,19} = 41.5$, $p < .001$, $\eta^2 = .686$; $AGE \times CT$: $F_{3,57} = 3.87$, $p = .014$, $\eta^2 = .169$; $AGE \times WID$: $F_{1.0,19.4} = 30.0$, $p < .001$, $\eta^2 = .612$; $AGE \times AMP$: $F_{2,38} = 4.22$, $p = .022$, $\eta^2 = .182$, $AGE \times CT \times WID$: $F_{2.1,39.8} = 6.47$, $p = .003$, $\eta^2 = .254$; $AGE \times WID \times EWR$: $F_{1.9,35.6} = 4.89$, $p = .015$, $\eta^2 = .205$).

Finally, there was also a main effect of AMP ($F_{2,38} = 65.0$, $p < .001$, $\eta^2 = .774$). Pairwise tests confirmed it was typical: Speed decreased as amplitude increased (both $p < .001$).

7.3.6 Pressure

The older adults exerted significantly more pressure than the younger adults. Although not part of our main focus, we were additionally interested in exploring potential differences between older and younger adults in terms of the pressure exerted during selections. An independent t-test on AGE was run to compare the average maximum trial pressure between groups. This revealed a significant difference ($t_{22} = -3.34$,$p = .003$, $\eta^2 = .336$). Surprisingly, on average, the older adults applied approximately 50% more pressure than the younger adults, as shown in Figure 7.12.

7.3.7 Subjective Findings

Bubble and Steadied-Bubble were preferred overall. A reliability analysis confirmed high consistency among the different preference rankings (Cronbach's alpha = .912), so we collapse them into a single score for brevity. A Friedman test on the transformed rankings showed a significant main effect of CT ($\chi^2_3 = 35.27$, $p < .0001$). To understand the source of this effect, we performed pairwise comparisons using Wilcoxon Signed Rank tests and applied a Bonferroni adjustment. Bubble and Steadied-Bubble were both ranked more favorably than Steady and Control, with no differences between either Bubble and Steadied-Bubble or Steady

[17] There was also a significant $AGE \times CT \times EWR$ interaction. As in many of the other analyses, this was a mirroring of the results for interface.

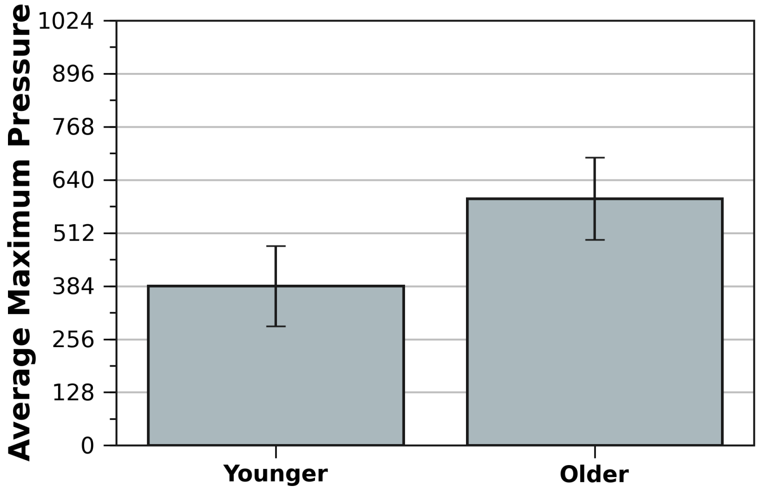

Figure 7.12: Average maximum trial pressure by *AGE* ($N = 24$). Error bars are 95% confidence intervals.

and Control (Bubble–Control: $z = -3.88$, $p < .001$; Bubble–Steady: $z = -3.85$, $p < .001$; Steadied-Bubble–Control: $z = -3.5$, $p < .005$; Steadied-Bubble–Steady: $z = -4.03$, $p < .001$).

The results of the comfort ratings were similar, but statistically less powerful. Friedman tests revealed significant main effects of *CT* on overall operation ($\chi^2_3 = 11.1$, $p < .05$), speed ($\chi^2_3 = 15.7$, $p < .005$), accuracy ($\chi^2_3 = 22.3$, $p < .0001$), mental effort required ($\chi^2_3 = 12.1$, $p < .01$), smoothness of operation ($\chi^2_3 = 20.0$, $p < .0005$), and force required ($\chi^2_3 = 10.9$, $p < .05$). Bonferroni corrected Wilcoxon Signed Rank Tests showed Steady was considered slower than Steadied-Bubble ($p < .05$), less accurate than both Bubble ($p < .05$) and Steadied-Bubble ($p < .01$), more mentally taxing than Steadied-Bubble ($p < .05$), and requiring of more force to operate than both Steadied-Bubble ($p < .05$) and Control ($p < .05$). Control was considered less smooth to operate than both Steadied-Bubble ($p < .01$) and Bubble ($p < .05$).

7.3.8 Summary

We summarize our results according to our hypotheses.

H1. Bubble will reduce both slips and misses, but only when surrounding targets are not directly adjacent. Supported.

H2. Independent of target spacing, Steady will reduce slips, but it will not affect misses. Mostly supported. Steady resulted in significantly fewer slips than Control, except at $EWR = 2$, there was no statistical difference.

H3. Steadied-Bubble will reduce slips when targets are directly adjacent, and both slips and misses when they are not. Supported.

H4. The experimental cursors (Bubble, Steady, and Steadied-Bubble) will reduce total errors for both age groups, but the impact will be larger for the older group as they will benefit from both slip and miss reduction. Partially supported. In terms of slipping, the older group benefited more from the experimental cursors than the younger group, as predicted. Statistically, there was no difference for missing; however, the older group missed significantly more than younger group. Thus, the error reduction provided would have a greater practical impact for them.

H5. A greater proportion of errors in Bubble and Steadied-Bubble will land on a distracter target. Supported.

7.4 Discussion

This study established the individual benefits of Steady Clicks and Bubble cursor for pen-based pointing with younger and older adults, and furthermore, showed that the two techniques could be successfully combined to provide the benefits of each. Our Steadied-Bubble cursor reduced misses when targets were not directly adjacent, and slips independent of spacing. Though our error analysis is limited to the smallest target size examined, we did see similar patterns for the larger target sizes. Moreover, our analysis of the movement time data did find differences for all target widths. The pattern of results for movement time was similar to the one observed for missing. Slipping did not noticeably impact movement time, which is not surprising since missing was much more common than slipping.

Though the techniques were beneficial to both age groups, they especially helped the older adults. For slipping, the experimental cursors worked so well that they reduced the performance gap between ages such that the older group's performance was no longer significantly different from that of the younger group. For missing, both groups benefited equally from the experimental cursors, but the older group missed almost three times more often. Thus techniques that reduce missing should have greater practical significance for them. We note this finding was not predicted and contrasts the results of the Baseline study, in which we did not find an effect of age for misses. One difference is that the task used in this study was more challenging—in the Baseline study, the smallest target was 38% bigger than the smallest target in this study.

Although slips were generally less frequent than misses (for both groups), slipping presents an important problem for older adults. During a slip, the pen initially lands on the target. This activates the visual feedback associated with a selection, and indicates to the user that their selection should be successful. As a result, slip errors are particularly confusing. Many older users are unaware of the cause of their difficulty, which hinders the development of self-correction strategies.

Thus, both slip and miss reduction is important for older users. The techniques we evaluated in this chapter, and particularly the Steadied-Bubble cursor addressed these two most common pen-based pointing problems. It is also important to note that in no case did any of our experimental techniques hinder the younger participants (they were either positively or neutrally affected by all our cursors). Thus, inclusion of these techniques should make it easier for older adults to interact with the same software as younger adults, reducing the need for specialized applications. Specialized applications generally try to make interaction easier by making targets bigger, often at the expense of aesthetics or features. However, they require each individual application to be adapted. Thus one major benefit of modifying the interaction technique instead of the application is that it provides older adults with enhanced access to a much broader set of applications.

As a secondary goal of this work, we examined the pressure exerted during selection. This investigation was largely motivated by earlier isolated observations (across studies) of extreme cases where an older participant would use so much pressure that the device would malfunction, leading us to wonder if others experienced some degree of pressure difficulties as well. Our finding that the older adults exerted 50% more force than the younger adults is important. Older adults often report finding pen interfaces tiring [23, 52]. Aging is known to lead to reduced maximum force capabilities [62], and at first glance, this seems to explain why they might find them more tiring. However, our results suggest that the problem is not with exerting sufficient pressure, but rather with determining how much pressure is needed. In this study, we included visual feedback indicating contact. This type of feedback is not common; thus, our results may even underestimate the extent of the problem. Devising ways of teaching older adults to use less pressure is thus an important area for further investigation.

An additional finding was that although the two techniques based on the Bubble cursor reduced errors relative to the Control cursor, when errors did occur, they were almost four times more likely to result in selection of a distracter target. In contrast, a comparable effect was not observed for the Steady cursor. This tendency for the Bubble cursor to shift errors onto unwanted functionality has not been discussed in the literature to date, and it has important practical implications as selection of an unwanted target typically requires corrective action, and thus has a much higher cost than selection of inactive whitespace. This is particularly important for older adults as they tend to find error correction more difficult. The impact of Bubble and Steadied-Bubble's higher proportion of distracter target errors is likely not reflected in our preference ratings. Though we differentiated between hits on a distracter target and hits on inactive whitespace in our analysis, from the user's perspective these errors were the same. This may have contributed to the strong preference for the Bubble and Steadied-Bubble cursors over Steady and Control.

One place where our hypotheses were not fully met is at the medium level of target spacing examined ($EWR = 2$), where Steady did not result in significantly fewer slips than Control. Inspection of the means for Steady and Control clarified that result and suggested it arose from Control performing slightly better at that spacing. Steady's performance remained relatively constant across spacings. Though the differences between levels of spacing for Control were not significant (and thus random variation is the likely explanation), it is possible that EWR=2 represents a balance between visual complexity at $EWR = 1$ and overconfidence at $EWR = 3$.

It is interesting that in our study we did not see a relationship between overall target density and performance for the Bubble cursor, while Grossman and Balakrishnan reported a negative effect of low density on performance [41]. One possible explanation is that limiting the maximum size of the Bubble cursor was effective, as they hypothesized [41]. However, it is also possible this difference reflects a deeper distinction between mouse and pen interaction. With a pen, users can remain above the detectable range of the screen until late in the interaction. While the pen is out-of-range, the cursor was hidden in all conditions; thus intermediate distracter targets did not have as much of an effect on the behavior of the bubble cursor.

Finally, though we focused on Bubble cursor and Steady Clicks, many other mouse techniques have been developed, many of which were discussed in Chapter 2. Some of these may also have applicability to pen interaction, and there may be additional opportunities to combine them, as we have done here for Steady Clicks and Bubble cursor.

7.5 Design Considerations

Our results showed clear support for the experimental cursors and illustrated that each had particular task contexts in which it performed best. Thus, we conclude this chapter, by reflecting on our findings to propose design guidelines for cursor selection.

7.5.1 Target Size and Density

We found that the biggest benefits were realized when targets were small; specifically, when they were comparable to the height of a text link. Even for the next biggest size examined (which was roughly the size of a small toolbar icon) the differences were weak and difficult to interpret. Small targets abound, and facilitating their selection is important. However, this finding does suggest some practical implications. To be effective, the Bubble cursor requires that targets are not directly adjacent. Thus, techniques based on it are particularly well suited to applications or tasks that have many small, but spaced targets. Such applications include interacting with large data visualizations, or selecting features in a drawing application.

However, small targets are often coupled with high target density, such as with word or character selection in a text editor. In these situations, the Bubble cursor is not helpful, but techniques such as Steady Clicks are. The Steadied-Bubble provides a useful balance between these factors. For example, some web pages have links tightly clustered in one area, whereas other pages have sparser links. The Steadied-Bubble supports a seamless transition between these cases, providing the best possible support for each.

7.5.2 Error Cost

Overall, the Bubble and Steadied-Bubble cursors were more effective at reducing errors than our Steady cursor. However, because the Bubble cursor technique assigns inactive whitespace to nearby targets, it results in a higher proportion of errors landing on an unwanted target. In some cases, unintended selections are easily and efficiently corrected. However, it is important to consider the cost of error correction when choosing a technique. When the cost is high, it may be better to choose a technique that has more but less costly errors.

7.5.3 Target Awareness

One notable difference between the Bubble and Steady cursors is that the former is intrinsically target-aware, that is, the system needs to know where targets are to calculate the cursor's expansion. The Steady cursor is instead target-agnostic; its functionality is independent of target location. Thus, technique selection also depends on whether or not it is practically possible or computationally feasible to track target locations.

7.5.4 Target Users

The experimental cursors differently supported the various errors types examined in this chapter. In particular, Steady only provided support for slipping, and even with the Control cursor, the younger adults demonstrated relatively little slipping. Thus, interfaces targeted exclusively to younger users may not warrant a Steady or Steadied-Bubble approach. On the other hand, both slip and miss reduction is important for older users. Importantly, none of the techniques hindered the younger participants. Thus, when targeting older individuals—or when targeting a range of users—techniques that address both error types should be adopted.

7.6 Summary

In this chapter, we presented an investigation into techniques to address missing and slipping. Towards solving these problems, we examined the feasibility of extending and combining existing techniques designed for younger users and mouse interaction, focusing our investigation on the Bubble cursor and Steady Clicks techniques. Our experimental results showed that both techniques can be adapted for use in a pen interface and that they can be successfully combined to provide greater support than either technique on its own. Through our results were especially pertinent to the older adults in our study, both ages benefited from the designs. An additional finding was that each technique's performance depended on the task context. Drawing on these contextual findings, we concluded the chapter by establishing design considerations for technique selection.

Chapter 8

Towards a Model of Age-Related Pen-Based Interaction Difficulty

In Chapter 3, we presented a Baseline Study to identify sources of pen-based target acquisition errors, and in Chapters 4, 5, and 7 we presented investigations of potential techniques to address each of the difficulties identified. In this chapter, we first review literature from kinesiology and gerontology that is particularly relevant for understanding the behaviors we saw in our studies, and we discuss technological and contextual factors that likely contribute to overall performance on pen-based target acquisition tasks.

Our goal in this chapter was to identify the core factors influencing this thesis research. We note that while some of the technological factors identified could be quantified with additional research, the age-related and contextual factors are too broad and diverse to be developed into a predictive model to simulate user behavior. Thus, modeling target acquisition at that level of granularity will remain a challenge. Rather, we present these factors as a starting point for developing a descriptive model of age-related pen-based targeting error. Such a model would be useful both for informing the design of accessible pen-based techniques, and for aiding practitioners in predicting the accessibility of techniques. We conclude this chapter by outlining additional broad areas for future work, which provide interesting new avenues for investigation.

8.1 Age-Related Factors

In this section, we draw on work from the fields of kinesiology and gerontology to place our findings in the context of research on aging and motor control. Understanding the mechanisms and age-related changes that cause older adults to find targeting more difficult is beyond the scope of this thesis. However, examining research in this area can help shed light onto our findings. In particular, we found that age-related differences in the following factors played a role in our findings: response initiation, primary movement coverage, force control, speed-accuracy tradeoff biases, and vision. We discuss each in turn.

8.1.1 Response Initiation

Response initiation, or the time between completing one step of a task and initiating the next, has been shown to contribute to longer movement times for older adults. Cousins, Corrow, Finn and Salamone found that in a finger tapping task, older adults demonstrated a longer response duration time [29]. That is, when asked to tap their finger up and down on a solid surface, the older adults took longer than younger adults to initiate an upward movement after making contact with the surface. Cousins et al. did not investigate the underlying reasons for this effect, but they noted that it could be due to a number of factors including slower muscle contraction times, slower neuronal conduction times, motivational differences, and orthopedic or medication-related differences. Although the longer response times affect many components of a tapping task, these findings are particularly relevant to the slipping difficulty we observed. A delayed response time suggests that older adults will spend more time in contact with the screen during a tap. If the pen slips while in contact with the screen regardless of age, then the slip itself will be larger for an older adult than for a younger adult: For younger adults, the movement would be small enough that it would not generally lead to an error (i.e., the pen might move while in contact with the screen but not enough to exit the target bounds), whereas for older adults, longer contact likely means larger slip distances, increasing the likelihood that this movement will result in an error.

8.1.2 Primary Movement Coverage

Older adults cover less distance with their primary movement, and as a result make many more additional submovements to reach the target. This result has implications for many mouse-based interaction techniques, particularly those that rely on the primary movement to predict the user's target (e.g., target expansion expands the size of a predicted target [71], forcing non-predicted items shrink or move away, thus hindering selection when the prediction is wrong).

In terms of pen-based interaction, there are also potential implications. One possibility is that users lift their hand from the screen for the larger primary submovement, but then rest their wrist during the final targeting. If this is true, then a shorter primary movement would place older adults farther from their goal during final targeting; to compensate, they would need to either reposition their wrist, which is time-consuming, or reach awkwardly to the target by twisting their wrist, which may help explain their higher error rates. We did not examine these factors in this thesis. With a mouse, the data stream of mouse events is sufficient for understanding the low level patterns of movement that occur on route to a target, including the path taken and the timings of individual submovements. With pen input there are two challenges to acquiring this data: (1) the movement occurs in three dimensions (rather than along a two dimensional surface as with a mouse), and (2) bounds on the tracking range of the device limits the data available. Answering the above questions requires the use of a separate motion capture system that can provide full data on the relevant six degrees of movement freedom (i.e., x, y, z position, plus pitch, yaw, and roll).

8.1.3 Force Control

In general, as the force required for a task increases, noise in the motor-control system also increases, causing accuracy to degrade. In other words, tasks performed with little force can be executed more accurately than those performed with more force. This is true independent of age; however, aging is known to lead to reduced maximum force capabilities [62] and a higher noise-to-force ratio [117, 122]. This means that accuracy degrades more quickly with increased force for older adults, which has several implications for pen-based interaction. Most notably it relates to our finding that older adults use more pressure. It suggests that using excess pressure not only leads to increased fatigue, but also decreased accuracy, underscoring the need for methods of encouraging the use of appropriate (i.e., minimal) force.

8.1.4 Biases in the Speed-Accuracy Tradeoff

Research has found that older adults use different strategies concerning the speed-accuracy tradeoff involved in movement control. They tend to be more conservative, and make more confirmatory movements once on the target, suggesting a bias towards accuracy over speed [117]. This complicates comparisons between younger and older adults by confounding strategy with capability. In terms of our results, it may help explain why in some cases we did not see a difference between the age groups in terms of accuracy, while in others we did. For example, in the tapping task of the Baseline Study we did not see an effect of age on missing, but in Technique Study Three we did. A notable difference between the two studies is that the Baseline Study task was easier (i.e., the smallest target size was 38% bigger than in Technique Study Three). We suspect that when possible, older adults will slow down to maintain a high level of accuracy. However, as the task difficulty increases, it may not be possible to slow down enough to prevent errors, rendering this strategy less effective.

8.1.5 Vision

Throughout our studies, we controlled for vision by exclusively recruiting participants with normal or corrected-to-normal vision. Nonetheless, it seems likely that older adults will be habituated to some amount of vision loss, and as a result, may overestimate their vision. Thus, some differences likely existed between the age groups, with the older adults having less visual acuity. Moreover, it has also been found that visual processing is less efficient for older adults [31, 114]. That is, the brain's ability to process visual images degrades with age, independent of the eyes' ability to perceive the scene. Thus, it may have taken the older adults longer to identify targets. This factor may have been particularly important in our studies of menu interaction as target identification was an intrinsic part of that task. For the general target acquisition studies, identification was simpler and we additionally used strong visual feedback on our targets to minimize the processing demands.

8.2 Technological Factors

In this section, we outline factors pertaining to characteristics of the device that influence the difficulties observed, including surface resistance, parallax, and physical affordances. These mostly relate to limitations of current hardware; thus, many of the factors identified here may be resolved as the technology matures.

8.2.1 Surface Resistance

Relative to paper, Tablet PC screens have low surface resistance, which can make them feel slippery and can make it hard to hold the pen still against the screen. It is likely that this lack of surface resistance contributed to the slipping error that we observed. It would be interesting to explore differing levels of resistance to see if slipping can be reduced by increasing the friction of the display. Some Tablet PCs do have more surface resistance than others; however, increasing the surface friction tends to degrade the visual appearance of the display. As such, the range of resistances available on the market is limited, making it difficult to examine this factor. Moreover, image degradation may introduce interaction difficulties of its own, especially for older adults as they are likely to have reduced vision capabilities, as noted in Section 8.1.5.

8.2.2 Screen Parallax

As shown in Figure 8.1, screen parallax is the apparent displacement of the on-screen cursor relative to the tip of the pen, caused by the small separation between the surface of the screen and the image. The impact of parallax depends on the viewing angle and the glass used in the screen (i.e., the thickness of the screen and its refractive index). To give an idea of the magnitude of parallax: a viewing angle of 20 degrees from normal results in an offset of 1 px/mm of screen thickness, and a viewing angle of 35 degrees would yield an offset of 2 px/mm.[18] LCD screen's

[18]Based on a pixel width of 0.2 mm, and assuming a refractive index of 1.517 for the glass.

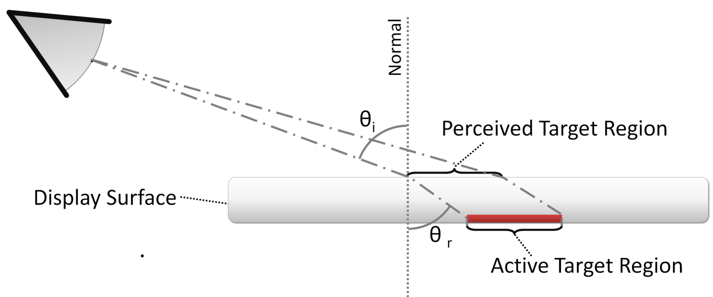

Figure 8.1: Illustration of screen parallax. When the viewing angle is oblique relative to the surface of the screen, the depth of the display hardware results in a discrepancy between the perceived location of the target and its actual location. This discrepancy depends on the viewing angle (θ_i) and the refraction through the glass (θ_r). Calibration software attempts to use a small number of user selections to infer the viewing angle and compensate for the parallax effect. (Note that the scale of this image is exaggerated for illustrative purposes; θ_i would typically be closer to normal, and the thickness of the screen would be much thinner than appears.)

are typically about 0.7 mm thick, and are at most 1.1 mm [14],[19] providing a useful estimate of the magnitude of the effect.

In modern technology, software utilities compensate for parallax by calibrating the system to the user's physical position relative to the device. However, this does not completely eliminate the effects of parallax, and we found that many of the older users had difficulty with the built-in calibration utilities.

In our studies, we tried to minimize the effects of parallax, with varying degrees of success. In the Baseline Study, the experimenter (the author of this thesis) calibrated the display once at the start of the study; this had the drawback that the calibration did not account for differences in height and posture between participants and the experimenter. To address this issue, from Technique Study One onwards, we asked participants to complete the calibration themselves. However, many participants were ineffective at calibration, even after multiple attempts, indicating that refinements are needed to make calibration utilities more accessible.

[19] We were unable to determine the thickness of the glass through the manufacturer and did not attempt to measure the thickness directly.

The Windows Tablet PC utility used in the Baseline Study and Technique Studies One and Two uses only a single calibration point in each corner of the screen to approximate the user's position. The Cintiq's utility (Technique Study Three) is even more limited; it uses only the top left and bottom right corners. These approaches might be sufficient for an able-bodied user who can consistently and accurately select the calibration target. However, for an older user with more movement variability averaging across a number of trials and positions is likely necessary. A final consideration is that younger adults may be better able to compensate for the effects of parallax in their motor planning.

8.2.3 Physical Affordances

A final factor related to the technology is the lack of tactile feedback during interaction. Pressing a physical button has an affordance that can be felt by users to assess whether their action was successful or not. The screen provides only limited tactile feedback and one or two participants in each of our studies reported difficulty knowing how hard to press on the screen. At the extreme, pressing too hard causes warping in the screen that can causes the sensors to malfunction.[20] However, even more moderate cases of excessive pressure can indirectly contribute to errors, by causing individuals to fatigue more quickly. Also, as discussed in Section 8.1.3, increased force leads to less control. Often the pen is not held perfectly perpendicular to the screen, and when this is so, excessive pressure may result in slipping. Research into providing haptic feedback through vibration [16, 50] or physical deformation of the screen [46] are emerging areas of research that may ultimately provide solutions to mitigate some of these issues. It may also be possible to compensate for the lack of physical feedback with visual or audio feedback, or by adding a tip-switch to the pen to give it a clicking feel [18, 66].

[20] In one case, a participant in the Baseline Study was pressing so hard he was causing a haloing effect around the point of contact. He remarked that he did not know why he was getting errors. He knew he was hitting the target because he could see the "circle".

8.3 Contextual Factors

In our studies, we observed instances where particular contextual factors, such as the task or work environment, appeared to play a role in our results. Context is a very broad and complex topic, and it is not our intention to produce a comprehensive list of all possible contextual factors. Instead, we highlight those factors that were particularly important in our results.

8.3.1 Task Flow

By task flow, we refer to the individual sequence of tasks. In an experimental setting, a single task is often repeated many times, and often in the same sequence, which can influence the results. For example, in the Baseline Study, we chose a discrete task sequence, (where users returned to the center of the screen between trials) and found a clear pattern of missing-just-below. In Technique Study One, we instead used a continuous sequence (where users continued directly to the top of the menu after each trial), and observed a different pattern of results: Some participants displayed a downward hitting behavior (consistent with missing-just-below), but others had the opposite tendency, hitting high on the target item and onto the item above. It is not clear why we saw this difference, and individual differences between the participants in each study may play a role. However, another possibility is that the continuous menu-selection task used in Technique Study One 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, while the discrete task used in the Baseline Study encouraged movement towards the center of the screen (down and right) after each selection. In a real-world setting it is impossible to predict where the user will go after making a menu selection, and a variety of movement patterns are likely, suggesting that in a real-world task we might expect to see an even wider range of behavior.

8.3.2 Widget Layout

The missing-just-below and drifting difficulties we observed may be closely tied to the layout or design of the interface. For example, we chose to maintain the default Tablet PC menu layout for right-handed users; that is, menu heads were aligned with the left edge of the screen, and menus aligned with the right edge of the menu head where possible. It might be possible to eliminate drifting by placing the menu heads towards the right edge of the screen. This would fully accommodate right alignment of items with the menu head, reducing occlusion of the menu contents by the hand. In the follow-on to the Baseline Study (presented in Section 3.4) we compared mouse and pen interaction and found that participants did not move as far to the right while navigating with a mouse (as reflected by the x-coordinate of their selection), suggesting this might be a promising avenue. Generally, the layout of widgets can influence the order in which users scan for functionality and the extent to which hand occlusion is a problem. In terms of missing-just-below, scanning direction and hand occlusions are both possible factors. We have not investigated their influence on missing-just-below experimentally; however, we believe they are promising avenues for further exploration as each would have unique implications for design. If scanning direction were found 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.

8.3.3 Visual Feedback During Targeting Phase

Visual feedback is often provided in computer interfaces to improve targeting accuracy. With pen-based interaction, one example is the on-screen cursor that is provided to help users overcome parallax. However, it is not clear that this approach is effective. In our studies, all participants were informed of the role of the on-screen cursor, and told to use it for targeting. However, many appeared to disregard the cursor. We believe this is because the cursor breaks the naturalness of direct interaction. Research on eye-movement during hand-eye coordination tasks has found that for direct manipulation, eye movements precede action; that is, humans focus on their intended target, not on their hand or manipulation tool [54]. This suggests that when using a pen device, users do not look at the tip of the pen during final stage of targeting but rather focus on the target itself. Thus, they may not notice that the cursor is not aligned with the target.

8.3.4 Relative Positioning

The impact of the surface resistance may also depend on how the Tablet PC is positioned relative to the user. Holding the screen on an angle may magnify difficulties caused by the lack of surface resistance, and thus increase the likelihood of slipping. However, it is difficult to work with a Tablet PC when it is lying flat horizontally, as it requires the user to lean over the screen and look down, which is a poor working posture [83]. In our studies, we placed the Tablet PC on a stand to provide extra support and position it at a more appropriate work angle [83, 119]. Considering the impact of different work postures is an important area for future work, as is examining the additional difficulties that arise when holding the device or resting it on your lap.

8.3.5 Work Environment

The environment in which the device is used can have a large influence on interaction. Clearly many environmental factors may play a role, but one example that was particularly important in this work was the effect of lighting, particularly the impact of glare from overhead lights. Glare will be disruptive and challenging for everyone, but it can be particularly detrimental to older adults [96]. We minimized glare by turning off all overhead lighting and used lamps to create a comfortable level of non-directional ambient light. However, in real-world scenarios this will not always be possible, and alternative approaches will be needed.

8.4 Additional Directions for Future Research

In the previous sections of this chapter, we reflected upon our findings and identified a number of places where focused further investigation is needed. In this section, we expand on this topic, exploring additional broad directions for future work.

8.4.1 Extension to Other User Populations

Our initial motivation for this work grew from our previous research developing handheld technology for individuals with aphasia [77]. Aphasia is an acquired cognitive disorder that is usually acquired as a result of a stroke. As such, many of the participants in that research were older with additional stroke-related motor impairments. In this thesis research, we have specifically focused on right-handed individuals free of diagnosed impairment to their hands. We did this to limit the effect of individual differences and to increase the statistical power of our studies. With the base understanding developed in this research, a next step would be to increase the scope of examination to include a wider range of motor abilities. The goal of that work would be to see how well our designs address the needs of those users and to explore what other difficulties may need to be addressed.

It would also be interesting to examine left-handed users. To a certain extent, we would expect the results of left-handed individuals to mirror those of right-handed individuals. However, left-handed individuals develop a much wider range of pen holding postures (e.g., some hook their hand while writing to avoid smearing ink with the edge of their hand), and so they may have unique needs beyond simply mirroring the techniques proposed for right-handed users. Additionally, their breadth of postures may reduce the impact of the difficulties experienced by the right-handed users for them; these workarounds may provide additional insight into potential techniques for right-handed users.

8.4.2 Long-Term Evaluation with a Real-World Application

Any controlled laboratory experiment is limited in its ecological validity. Examining our research questions in the context of a real application may reveal different results from those found in our controlled studies. When faced with a real task, users may be more motivated to develop compensation strategies. Or they may be more focused on their task and correspondingly, less focused on performing the interaction; thus, they may make more—potentially different—errors. Some researchers have begun to examine methods for detecting real-world pointing problems [53], but more work is needed to make this a viable approach. In particular, one challenge is identifying the user's intent. For example, one way of inferring

Age-related	Technological	Contextual
Response Initiation	Surface Resistance	Task Flow
Primary Movement	Parallax	Widget Layout
Force Control	Physical Affordances	Visual Feedback
Speed-Accuracy Biases		Relative Position
Vision		Work Environment

Table 8.1: Factors that contributed to the errors observed in this research. Together these form an initial set of components that might contribute to a model of age-related pen-based target acquisition difficulty.

menu selection difficulty would be to detect instances where a user selects but then quickly cancels a selection. However, it is impossible to know whether this was caused by the user accidentally selecting the wrong item, or intentionally selecting an item, but quickly identifying it as incorrect.

8.4.3 Supporting Individual Differences

In this thesis, we have focused on a one-size-fits-all approach to address pointing difficulties. While this is an important first step, users would likely benefit from a more customized approach. For example, with slipping, some users may only slip a short distance and require only minimal assistance, while others may slip further and need greater support. The Steady Clicks approach (Chapters 6 and 7) reduces slipping, but it does so at the expense of making dragging more difficult. Likely, those with greater need will be more willing to make more compromises (such as reduced dragging ability) than those who only require slight correction. Thus, being able to customize the slip threshold would allow those with better ability to maintain a more natural dragging ability, and those with greater need to trade dragging ability for increased ease in target acquisition. In addition, a customized approach is likely especially important for menu interaction given the existence of the diametric low- and high-hitter groups identified in Technique Study One. It seems unlikely that a single approach will meet the needs of these opposing groups.

8.5 Summary

In this chapter, we reflected on our combined results across studies to identify factors that contributed to the difficulties observed. It is beyond the scope of this thesis to identify how exactly each of these factors contributes to overall performance, but this discussion does allow us to interpret our results in the context of existing work on human movement, and to identify areas for future work. Table 8.1 summarizes these factors, which we have categorized along three main dimensions: age-related, technological, and contextual. Together, these serve as a starting point towards the future development of a model of age-related pen-based interaction difficulty. These factors are also important considerations for designers building pen-based interfaces for older adults. We then finished this chapter by identifying additional broad avenues for future work.

Chapter 9

Conclusions

The goal of this thesis was to increase the accessibility of pen-based technology, particularly for older adults, by investigating mechanisms for assisting users to select features more easily. Our first step in fulfilling this goal was to collect quantitative data on pen-based target acquisition difficulty—across a range of ages and task situations—using standard point and tap mechanisms. Based on the results of this study, we conducted a series of three more studies to develop and evaluate techniques for addressing the difficulties uncovered. We based our techniques on existing ones for younger users and mouse-based interaction, extending and combining them in novel ways to address the needs of older users and pen-based interaction. Finally, we also developed and evaluated a mouse-based target acquisition technique, Steady Clicks. Although, somewhat outside our core thesis goals, this work explored a method for reducing mouse-based slip errors, a difficulty we also identified for pen interaction. We expanded on this technique in our subsequent examination of pen-based slipping.

In this chapter, we summarize the contributions of this thesis, elaborating on the findings from the steps outlined above.

9.1 Primary Contributions

We claim three primary contributions for this thesis.

9.1.1 Identification of Age-Related Pen-Based Interaction Difficulties

Previous research on aging and input device selection has established that older adults perform target acquisition tasks faster and more accurately with a pen than with a mouse [22, 23]. Research has also examined the role of age and motor-control in terms of improving mouse-based interaction [58, 59, 82, 110]. However, prior to this thesis, an equivalent investigation had not been done to examine the underlying causes for age-related pen-based interaction difficulty. With the Baseline study, we began to fill this gap in the literature. Our results identified three novel pen-based target acquisition difficulties, and demonstrated how these difficulties vary across task situations and age. One strength of this work is that we included two tasks (menu selection and Fitts' tapping), enabling us to observe both simple and complex interaction.

Although identifying target acquisition difficulties was chiefly the goal of the Baseline Study, some of our subsequent studies uncovered additional target acquisition problems. Notably, in Technique Study One, we discovered a new difficulty pertaining to errors on the menu item above the target item, and in Technique Study Three, we discovered that older adults have difficulty determining how much pressure is needed for selection. These findings add to the results of the Baseline Study, and provide additional avenues for future work.

9.1.2 Design and Evaluation of New Pen-Based Techniques to Address the Difficulties Identified

Based on the findings from the Baseline study, we developed seven new target acquisition techniques, specifically addressing the three difficulties identified, and particularly targeting older adults. In designing these techniques, we built upon existing mouse-based techniques developed for younger and older users and pen techniques for younger users, extending and combining them in novel ways to address the needs of older individuals using a pen-based Tablet PC. For each of the difficulties identified, we proposed multiple potential techniques and then evaluated them with controlled laboratory experiments to compare them to each other and to a control interface. In total, we conducted three laboratory experiments to evaluate the seven new pen-based techniques: Reassigned and Deactivated (for missing-just-below), Tap and Glide (for drifting), and Steady, Bubble, and Steadied-Bubble (for slipping). Through these evaluations, we established where our proposed techniques were successful at reducing errors, and we identified areas where further refinement is needed.

9.1.3 Improvement of Pen-Based Technology Through the Systematic Inclusion of Age as a Factor

Research that considers the needs of older computers users is relatively rare. Even rarer are direct comparisons between younger and older users. Thus, an important strength of this thesis is that we systematically included age as an explicit factor in every stage of this work. We could have focused solely on the needs of older users. However, by including a range of ages, we were able to carefully document where the needs of the older adults align with—and where they diverge from—the needs of the younger adults. Specifically, an additional finding of the Baseline Study was that including older users as participants allowed us to uncover general pen-interaction deficiencies that we would likely have missed otherwise: Two of the difficulties uncovered also affected younger users. In terms of our Technique Studies, including both younger and older adults allowed us to establish places where our techniques also improved interaction for the younger adults, and moreover ensured that at a minimum our techniques did not hinder the younger adults. Thus, we believe that our approach of including a range of adult ages ultimately led to richer findings and improved interfaces for older and younger users alike.

9.2 Secondary Contributions

We also claim two secondary contributions.

9.2.1 Design and Evaluation of a Novel Mouse-Based Technique to Address Slipping

An additional contribution of this thesis was the development of Steady Clicks, a novel mouse-based interaction technique designed to help motor-impaired individuals who have difficulty holding the mouse still while clicking. While a number of researchers have investigated methods to help get the cursor to the target, this is the first work to address the problem of keeping it there once it is in place. Although, this research was somewhat outside our main thesis goals, it was important as the Steady Clicks technique formed the basis of one of the techniques we examined for addressing pen-based slipping.

9.2.2 Identification of Factors Contributing to the Difficulties Observed

Finally, we reflected on our findings across studies and outlined factors which contributed to the difficulties we observed. We based these factors on literature that has investigated the effects of aging and motor-control, and on our own observations of the technological and contextual factors that likely contributed to performance in our studies. These factors represent the first step towards the development of a model of age-related pen-based target acquisition difficulty. Such a model would be useful both for informing the design of accessible pen-based techniques, and for aiding practitioners in predicting the accessibility of existing techniques for older users.

9.3 Concluding Comments

Older adults form a growing demographic of computer users. As of 2008, 56% of Americans aged 64–72, and 31% of Americans aged 73 and over, were online [55]. Partially motivated by these statistics, technology is increasingly being promoted for older adults, particularly as a means of addressing cognitive and sensory impairments and enabling individuals to live more independently (e.g., [47, 67, 70, 77, 79, 93, 130]). Pen-based devices such as Personal Digital Assistants (PDAs) and Tablet PCs are appealing platforms for these endeavors, but despite a multitude of advantages, many older individuals find pen-based interaction challenging. 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 thesis, we have provided a first step towards identifying and addressing some of the difficulties older adults encounter when using pen-based input. Reflecting on our findings, we identified factors that contributed to our results. These factors help illuminate the underlying reasons for pen-based targeting difficulties and shed light onto areas still needing attention.

Bibliography

[1] Accot, J., & Zhai, S. (2002). More than dotting the i's—foundations for crossing-based interfaces. In *CHI'02: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 73–80). New York, NY, USA: ACM.

[2] Ahlström, D. (2005). Modeling and improving selection in cascading pull-down menus using fitts' law, the steering law and force fields. In *CHI'05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 61–70). New York, NY, USA: ACM.

[3] Ahlström, D., Hitz, M., & Leitner, G. (2006). An evaluation of sticky and force enhanced targets in multi target situations. In *NordiCHI'06: Proceedings of the 4th Nordic conference on Human-computer interaction*, (pp. 58–67). New York, NY, USA: ACM.

[4] Albinsson, P.-A., & Zhai, S. (2003). High precision touch screen interaction. In *CHI'03: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 105–112). New York, NY: ACM Press.

[5] Apitz, G., & Guimbretière, F. (2004). Crossy: A crossing-based drawing application. In *UIST'04: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, (pp. 3–12). New York, NY, USA: ACM.

[6] Apted, T., Kay, J., & Quigley, A. (2006). Tabletop sharing of digital photographs for the elderly. In *CHI'06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 781–790). New York, NY, USA: ACM.

[7] Bashore, T. R., Osman, A., & Heffley, E. F. (1989). Mental slowing in elderly persons: A cognitive psychophysiological analysis. *Psychology and aging*, *4*(2): 235–344.

[8] Baudisch, P., & Chu, G. (2009). Back-of-device interaction allows creating very small touch devices. In *CHI'09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1923–1932). New York, NY, USA: ACM.

[9] Bederson, B. B. (2000). Fisheye menus. In *UIST'00: Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, (pp. 217–225). New York, NY, USA: ACM.

[10] Bederson, B. B., & Hollan, J. D. (1994). Pad++: a zooming graphical interface for exploring alternate interface physics. In *UIST'94: Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology*, (pp. 17–26). New York, NY, USA: ACM Press.

[11] Benko, H., Wilson, A. D., & Baudisch, P. (2006). Precise selection techniques for multi-touch screens. In *CHI'06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1263–1272). New York, NY, USA: ACM.

[12] Birdi, K. S., & Zapf, D. (1997). Age differences in reactions to errors in computer-based work. *Behaviour & Information Technology*, *16*(6): 309–319.

[13] Blanch, R., Guiard, Y., & Beaudouin-Lafon, M. (2004). Semantic pointing: improving target acquisition with control-display ratio adaptation. In *CHI'04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 519–526). New York, NY, USA: ACM.

[14] Blevins, J. D., Novak, R. A., & Shay, G. C. (2007). US patent 2007/00622219 A1: Methods of fabricating flat glass with low levels of warp.

[15] Brandl, P., Forlines, C., Wigdor, D., Haller, M., & Shen, C. (2008). Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *AVI'08: Proceedings of the Working Conference on Advanced Visual Interfaces*, (pp. 154–161). New York, NY, USA: ACM.

[16] Brewster, S., Chohan, F., & Brown, L. (2007). Tactile feedback for mobile interactions. In *CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 159–162). New York, NY, USA: ACM.

[17] Brownlow, N., Shein, F., Thomas, D., Milner, M., & Parnes, P. (1989). Direct manipulation: Problems with pointing devices. In *Resna '89: Proceedings of the 12th Annual RESNA Annual Conference*, (pp. 246–247). Washington, DC: Resna Press.

[18] Buxton, W. (1990). A three-state model of graphical input. In *INTERACT'90: Proceedings of the Third IFIP International Conference on Human-Computer Interaction*, (pp. 449–456). Amsterdam, The Netherlands: North-Holland Publishing Co.

[19] Callahan, J., Hopkins, D., Weiser, M., & Shneiderman, B. (1988). An empirical comparison of pie vs. linear menus. In *CHI'88: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 95–100). New York, NY, USA: ACM.

[20] Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for the text selection on a crt. *Ergonomics*, *21*(8): 601–613.

[21] Chaparro, A., Rogers, M., Fernandez, J., Bohan, M., Choi, S. D., & Stumpfhauser, L. (2000). Range of motion of the wrist: Implications for designing computer input devices for the elderly. *Disability & Rehabilitation*, *22*(13-14): 633–637.

[22] Charness, N., Bosman, E. A., & 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*. Washington, DC, USA: American Psychological Association.

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

[24] Cockburn, A., & Firth, A. (2003). Improving the acquisition of small targets. In *People and Computers XVII: Proceedings of the 2003 British Computer Society Conference on Human-Computer Interaction*, (pp. 181–196).

[25] Cockburn, A., & Gin, A. (2006). Faster cascading menu selections with enlarged activation areas. In *GI'06: Proceedings of Graphics Interface 2006*, (pp. 65–71). Toronto, Ont., Canada, Canada: Canadian Information Processing Society.

[26] Cockburn, A., Gutwin, C., & Greenberg, S. (2007). A predictive model of menu performance. In *CHI'07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 627–636). New York, NY, USA: ACM.

[27] Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum.

[28] Cooke, J., Brown, S., & Cunningham, D. (1989). Kinematics of arm movements in elderly humans. *Neurobiology of Aging*, *10*(2): 159–165.

[29] Cousins, M. S., Corrow, C., Finn, M., & Salamone, J. D. (1998). Temporal measures of human finger tapping: Effects of age. *Pharmacology, biochemistry, and behavior*, *59*(2): 445–449.

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

[31] Cremer, R., & Zeef, E. J. (1987). What kind of noise increases with age? *Journal of Gerontology*, *42*(5): 515–518.

[32] Crossman, E. R., & Goodeve, P. J. (1963). Feedback control of hand-movement and Fitts' Law. (Reprint of the paper presented at the meeting of the Experimental Psychology Society, Oxford, July 1963). *The Quarterly Journal of Experimental Psychology Section A, 1983*, *35*(Pt 2): 251–278.

[33] Czaja, S., & Sharit, J. (1998). Age differences in attitudes toward computers. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *53*(5): P329–P340.

[34] Darling, W., Cooke, J., & Brown, S. (1989). Control of simple arm movements in elderly humans. *Neurobiology of Aging*, *10*(2): 149–157.

[35] Douglas, S. A., Kirkpatrick, A. E., & MacKenzie, I. S. (1999). Testing pointing device performance and user assessment with the iso 9241, part 9 standard. In *CHI'99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 215–222). New York, NY, USA: ACM.

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

[37] Fitzmaurice, G., Khan, A., Pieké, R., Buxton, B., & Kurtenbach, G. (2003). Tracking menus. In *UIST'03: Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, (pp. 71–79). New York, NY, USA: ACM.

[38] Froehlich, J., Wobbrock, J. O., & Kane, S. K. (2007). Barrier pointing: Using physical edges to assist target acquisition on mobile device touch screens. In *ASSETS'07: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 19–26). New York, NY, USA: ACM.

[39] Greenstein, J. L. (1997). Pointing devices. In M. Helander, T. Landauer, & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction*, 2nd ed., (pp. 1317–1348). Amsterdam: Elsevier.

[40] Gregor, P., Newell, A. F., & Zajicek, M. (2002). Designing for dynamic diversity: interfaces for older people. In *ASSETS'02: Proceedings of the 5th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 151–156). New York, NY, USA: ACM.

[41] Grossman, T., & Balakrishnan, R. (2005). The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In *CHI'05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 281–290). New York, NY, USA: ACM.

[42] Grossman, T., Hinckley, K., Baudisch, P., Agrawala, M., & Balakrishnan, R. (2006). Hover widgets: Using the tracking state to extend the capabilities of pen-operated devices. In *CHI'06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 861–870). New York, NY, USA: ACM.

[43] Guiard, Y., Blanch, R., & Beaudouin-Lafon, M. (2004). Object pointing: A complement to bitmap pointing in GUIs. In *GI'04: Proceedings of Graphics Interface 2004*, (pp. 9–16). Canadian Human-Computer Communications Society.

[44] Guiard, Y., Bourgeois, F., Mottet, D., & Beaudouin-Lafon, M. (2001). Beyond the 10-bit barrier: Fitts' law in multi-scale electronic worlds. In *People and Computers XV - Interaction without Frontiers: Joint Proceedings of HCI 2001 and IHM 2001*, (pp. 573–587). Springer.

[45] Haaland, K. Y., & Harrington, D. L. (1998). Neuropsychological assessment of motor skills. In G. Goldstein, P. D. Nussbaum, & S. R. Beers (Eds.), *Neuropsychology*, (pp. 421–437). New York, NY, USA: Springer.

[46] Harrison, C., & Hudson, S. E. (2009). Providing dynamically changeable physical buttons on a visual display. In *CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 299–308). New York, NY, USA: ACM.

[47] Hawkey, K., Inkpen, K. M., Rockwood, K., McAllister, M., & Slonim, J. (2005). Requirements gathering with alzheimer's patients and caregivers. In *ASSETS'05: Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 142–149). New York, NY, USA: ACM.

[48] Heath, M., Roy, E. A., & L.Weir, P. (1999). Visual-motor integration of unexpected sensory events in young and older participants: A kinematic analysis. *Developmental Neuropsychology*, *16*(2): 197–211.

[49] Hinckley, K., Ramos, G., Guimbretiere, F., Baudisch, P., & Smith, M. (2004). Stitching: Pen gestures that span multiple displays. In *AVI'04: Proceedings of the Working Conference on Advanced Visual Interfaces*, (pp. 23–31). New York, NY, USA: ACM.

[50] Hoggan, E., Brewster, S. A., & Johnston, J. (2008). Investigating the effectiveness of tactile feedback for mobile touchscreens. In *CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1573–1582). New York, NY, USA: ACM.

[51] Hollinworth, N. (2009). Improving computer interaction for older adults. *ACM SIGACCESS Accessibility and Computing Newsletter*, *93*: 11–17.

[52] Hourcade, J. P., & Berkel, T. R. (2008). Simple pen interaction performance of young and older adults using handheld computers. *Interacting with Computers*, *20*(1): 166–183.

[53] Hurst, A., Mankoff, J., & Hudson, S. E. (2008). Understanding pointing problems in real world computing environments. In *ASSETS'08: Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 43–50). New York, NY, USA: ACM.

[54] Johansson, R. S., Westling, G., Backstrom, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *Journal of Neuroscience*, *21*(17): 6917–6932.

[55] Jones, S., & Fox, S. (2009). *Generations online in 2009*. Pew Internet and American Life Project, Washington, DC. Retrieved 09/19/09 from http://www.pewinternet.org/Reports/2009/Generations-Online-in-2009.

[56] Kabbash, P., & Buxton, W. A. S. (1995). The “prince” technique: Fitts' law and selection using area cursors. In *CHI'95: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 273–279). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co.

[57] Kattinakere, R. S., Grossman, T., & Subramanian, S. (2007). Modeling steering within above-the-surface interaction layers. In *CHI'07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 317–326). New York, NY, USA: ACM.

[58] Keates, S., & Trewin, S. (2005). Effect of age and parkinson's disease on cursor positioning using a mouse. In *ASSETS'05: Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 68–75). New York, NY, USA: ACM.

[59] Keates, S., Trewin, S., & Paradise, J. (2005). Using pointing devices: Quantifying differences across user groups. In *Proceedings of the 3rd International Conference on Universal Access in Human-Computer Interaction*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

[60] Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, *70*(61): 387–403.

[61] Ketcham, C. J., Seidler, R. D., van Gemmert, A. W. A., & Stelmach, G. E. (2002). Age-related kinematic differences as influenced by task difficulty, target size, and movement amplitude. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(1): P54–P64.

[62] Ketcham, C. J., & Stelmach, G. E. (2004). Movement control in the older adult. In R. W. Pew & S. B. V. Hemel (Eds.), *Technology for Adaptive Aging*, (pp. 64–92). Washington, DC, USA: National Academies Press.

[63] Kobayashi, M., & Igarashi, T. (2003). Considering the direction of cursor movement for efficient traversal of cascading menus. In *UIST'03: Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, (pp. 91–94). New York, NY, USA: ACM.

[64] Kurtenbach, G., & Buxton, W. (1993). The limits of expert performance using hierarchic marking menus. In *CHI'93: Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems, (pp. 482–487). New York, NY, USA: ACM.

[65] Landauer, T. (1997). Behavioral research methods in human-computer interaction. In M. Helander, T. Landauer, & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction*, 2nd ed., (pp. 203–227). Amsterdam: Elsevier.

[66] Lee, J. C., Dietz, P. H., Leigh, D., Yerazunis, W. S., & Hudson, S. E. (2004). Haptic pen: a tactile feedback stylus for touch screens. In *UIST '04: Proceedings of the 17th annual Annual ACM Symposium on User Interface Software and Technology*, (pp. 291–294). New York, NY, USA: ACM.

[67] Lee, M. L., & Dey, A. K. (2007). Providing good memory cues for people with episodic memory impairment. In *ASSETS'07: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 131–138). New York, NY, USA: ACM.

[68] MacKenzie, S. I. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, *7*(1): 91–139.

[69] Manly, B. F. J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 2nd ed. New York, NY: Chapman & Hall.

[70] Massimi, M., Baecker, R. M., & Wu, M. (2007). Using participatory activities with seniors to critique, build, and evaluate mobile phones. In *ASSETS'07: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 155–162). New York, NY, USA: ACM.

[71] McGuffin, M. J., & Balakrishnan, R. (2005). Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Transactions on Computer-Human Interaction*, *12*(4): 388–422.

[72] Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. (1988). Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological Review*, *95*(3): 340–370.

[73] Mizobuchi, S., & Yasumura, M. (2004). Tapping vs. circling selections on pen-based devices: Evidence for different performance-shaping factors. In *CHI'04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 607–614). New York, NY, USA: ACM.

[74] Moffatt, K., & McGrenere, J. (2007). Slipping and drifting: Using older users to uncover pen-based target acquisition difficulties. In *ASSETS'07: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 11–18). New York, NY, USA: ACM.

[75] Moffatt, K., & McGrenere, J. (2009). Exploring methods to improve pen-based menu selection for younger and older adults. *ACM Transactions on Accessible Computing*, *2*(1): 3:1–3:34.

[76] Moffatt, K., & McGrenere, J. (2010). Steadied-bubbles: Combining techniques to address pen-based pointing errors for younger and older adults. In *CHI'10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM (to appear).

[77] Moffatt, K., McGrenere, J., Purves, B., & Klawe, M. (2004). The participatory design of a sound and image enhanced daily planner for people with aphasia. In *CHI'04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 407–414). New York, NY, USA: ACM.

[78] Moffatt, K., Yuen, S., & McGrenere, J. (2008). Hover or tap? Supporting pen-based menu navigation for older adults. In *ASSETS'08: Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 51–58). New York, NY, USA: ACM.

[79] Mynatt, E. D., Essa, I., & Rogers, W. (2000). Increasing the opportunities for aging in place. In *CUU'00: Proceedings of the 2000 Conference on Universal Usability*, (pp. 65–71). New York, NY, USA: ACM.

[80] Nussbaum, P. D. (1998). Neuropsychological assessment of the elderly. In G. Goldstein, P. D. Nussbaum, & S. R. Beers (Eds.), *Neuropsychology*, (pp. 83–105). New York, NY: Springer.

[81] Oakley, I., & O'Modhrain, S. (2005). Tilt to scroll: Evaluating a motion based vibrotactile mobile interface. In *WHC'05: First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, (pp. 40–49). Los Alamitos, CA, USA: IEEE Computer Society.

[82] Paradise, J., Keates, S., & Trewin, S. (2005). Using pointing devices: Difficulties encountered and strategies employed. In *Proceedings of the 3rd International Conference on Universal Access in Human-Computer Interaction*. Mahwah, New Jersey, USA: Lawrence Erlbaum Associates.

[83] Pheasant, S., & Haslegrave, C. M. (2005). *Bodyspace: Anthropometry, Ergonomics and the Design of Work*. 3rd ed. CRC Press.

[84] Potter, R. L., Weldon, L. J., & Shneiderman, B. (1988). Improving the accuracy of touch screens: An experimental evaluation of three strategies. In *CHI'88: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 27–32). New York, NY, USA: ACM.

[85] Rabbitt, P. (2002). Consciousness is slower than you think. *Quarterly Journal of Experimental Psychology*, *55A*(4): 1081–1092.

[86] Ramos, G., & Balakrishnan, R. (2005). Zliding: fluid zooming and sliding for high precision parameter manipulation. In *UIST'05: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*, (pp. 143–152). New York, NY, USA: ACM Press.

[87] Ramos, G., Cockburn, A., Balakrishnan, R., & Beaudouin-Lafon, M. (2007). Pointing lenses: facilitating stylus input through visual-and motor-space magnification. In *CHI'07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 757–766). New York, NY, USA: ACM Press.

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

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

[90] Rimkus, A., Melinchok, M. D., McEvoy, K., & Yeager, A. K. (Eds.) (2005). *Thesaurus of Aging Terminology*. Eighth ed. AARP.

[91] Riviere, C. N., & Thakor, N. V. (1996). Effects of age and disability on tracking tasks with a computer mouse: accuracy and linearity. *Journal of Rehabilitation Research and Development*, *33*(1): 6–15.

[92] Rosenbaum, D. A. (2009). *Human motor control*. 2nd ed. Academic Press.

[93] Rowe, M., Lane, S., & Phipps, C. (2007). Carewatch: A home monitoring system for use in homes of persons with cognitive impairment. *Topics in Geriatric Rehabilitation: Smart Technology*, *23*(1): 3–8.

[94] Salthouse, T. A. (1988). Cognitive aspects of motor functioning. *Annals of the New York Academy of Sciences*, *515*(1): 33–41.

[95] Schaie, K. W. (2004). Cognitive aging. In R. W. Pew & S. B. V. Hemel (Eds.), *Technology for Adaptive Aging*, (pp. 43–63). Washington, DC, USA: National Academies Press.

[96] Schieber, F. (2003). Human factors and aging: Identifying and compensating for age-related deficits in sensory and cognitive function. In N. Charness & K. W. Schaie (Eds.), *Impact of Technology on Successful Aging*, (pp. 42–85). New York: Springer.

[97] Schmidt, R. A., Zelaznik, H., Hawkins, B., Frank, J. S., & Quinn, J. T. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review*, *47*(5): 415–451.

[98] Sears, A., & Shneiderman, B. (1994). Split menus: Effectively using selection frequency to organize menus. *ACM Transactions on Computer-Human Interaction*, *1*(1): 27–51.

[99] Shannon, C. E., Weaver, W., & Shannon (1998). *The Mathematical Theory of Communication*. University of Illinois Press.

[100] Smith, G. A., & Brewer, N. (1995). Slowness and age: Speed-accuracy mechanisms. *Psychology and aging*, *10*(2): 238–247.

[101] Smith, M. W., Sharit, J., & Czaja, S. J. (1999). Aging, motor control, and the performance of computer mouse tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *41*: 389–396.

[102] Sokal, R. R., & Rohlf, F. J. (1994). *Biometry: The Principles and Practices of Statistics in Biological Research*. 3rd ed. New York, NY: W. H. Freeman.

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

[104] Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, & commentary*. 2nd ed. New York, NY: Oxford University Press.

[105] Sprott, R. L. (1988). Age-related variability. *Annals of the New York Academy of Sciences*, *515*(1): 121–123.

[106] Strayer, D. L., Wickens, C. D., & Braune, R. (1987). Adult age differences in the speed and capacity of information processing: 2. An electrophysiological approach. *Psychology and aging*, *2*(2): 99–110.

[107] Subramanian, S., Aliakseyeu, D., & Lucero, A. (2006). Multi-layer interaction for digital tables. In *UIST'06: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, (pp. 269–272). New York, NY, USA: ACM.

[108] Tiffin, J., & Asher, E. J. (1948). The purdue pegboard: Norms and studies of reliability and validity. *Journal of Applied Psychology*, *32*: 234–247.

[109] Trewin, S., Keates, S., & Moffatt, K. (2006). Developing steady clicks: A method of cursor assistance for people with motor impairments. In *ASSETS'06: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 26–33). New York, NY, USA: ACM.

[110] Trewin, S., & Pain, H. (1999). Keyboard and mouse errors due to motor disabilities. *International Journal of Human-Computer Studies*, *50*(2): 109–144.

[111] Tsandilas, T., & Schraefel, M. C. (2007). Bubbling menus: A selective mechanism for accessing hierarchical drop-down menus. In *CHI'07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1195–1204). New York, NY, USA: ACM.

[112] Uttl, B., Graf, P., & Cosentino, S. (2000). Exacting assessments: Do older adults fatigue more quickly? *Journal of Clinical and Experimental Neuropsychology*, *22*: 496–507.

[113] Venolia, D., & Neiberg, F. (1994). T-cube: A fast, self-disclosing pen-based alphabet. In *CHI'94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 265–270). New York, NY, USA: ACM.

[114] Verrillo, R. T., & Verrillo, V. (1985). Sensory and perceptual performance. In N. Charness (Ed.), *Aging and Human Performance*, (pp. 1–46). Chichester, New York: Wiley.

[115] Vicente, K. J., & Torenvliet, G. L. (2000). The earth is spherical ($p < 0.05$): Alternative methods of statistical inference. *Theoretical Issues in Ergonomics Science*, *1*(3): 248–271.

[116] Vogel, D., & Baudisch, P. (2007). Shift: A technique for operating pen-based interfaces using touch. In *CHI'07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 657–666). New York, NY, USA: ACM.

[117] Walker, N., Philbin, D. A., & Fisk, A. D. (1997). Age-related differences in movement control: Adjusting submovement structure to optimize performance. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *52*(1): P40–P52.

[118] Walker, N., & Smelcer, J. B. (1990). A comparison of selection time from walking and pull-down menus. In *CHI'90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 221–226). New York, NY, USA: ACM.

[119] de Wall, M., van Riel, M. P., Snijders, C. J., & van Wingerden, J. P. (1991). The effect on sitting posture of a desk with a 10 degree inclination for reading and writing. *Ergonomics*, *34*(5): 575–584.

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

[121] Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, *3*(3): 189–230.

[122] Welford, A. T. (1981). Signal, noise, performance, and age. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *23*(1): 97–109.

[123] Welford, A. T. (1988). Reaction time, speed of performance, and age. *Annals of the New York Academy of Sciences*, *515*(1): 1–17.

[124] Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., & Shen, C. (2007). Lucid touch: a see-through mobile device. In *UIST'07: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, (pp. 269–278). New York, NY, USA: ACM.

[125] Wilkinson, R. T., & Allison, S. (1989). Age and simple reaction time: Decade differences for 5,325 subjects. *Journals of Gerontology*, *44*(2): P29–P35.

[126] Wing, A. M. (1983). Crossman and Goodeve (1963): Twenty years on. *The Quarterly Journal of Experimental Psychology Section A*, *35*(Pt 2): 245–249.

[127] Wobbrock, J. O., & Gajos, K. Z. (2008). Goal crossing with mice and trackballs for people with motor impairments: Performance, submovements, and design directions. *ACM Transactions on Accessible Computing*, *1*(1): 4:1–4:37.

[128] Wobbrock, J. O., Myers, B. A., & Kembel, J. A. (2003). Edgewrite: a stylus-based text entry method designed for high accuracy and stability of motion. In *UIST'03: Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, (pp. 61–70). New York, NY, USA: ACM.

[129] Worden, A., Walker, N., Bharat, K., & Hudson, S. (1997). Making computers easier for older adults to use: Area cursors and sticky icons. In *CHI'97: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 266–271). New York, NY, USA: ACM.

[130] Wu, M., Baecker, R., & Richards, B. (2007). Designing a cognitive aid for and with people who have anterograde amnesia. In J. Lazar (Ed.), *Universal Usability*, (pp. 317–356). West Sussex, England: John Wiley & Sons.

[131] Yan, J. H. (2000). Effects of aging on linear and curvilinear aiming arm movements. *Journal of Experimental Aging Research*, *26*(4): 393–407.

[132] Yatani, K., Partridge, K., Bern, M., & Newman, M. W. (2008). Escape: A target selection technique using visually-cued gestures. In *CHI'08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 285–294). New York, NY, USA: ACM.

[133] Zhai, S., Kong, J., & Ren, X. (2004). Speed-accuracy tradeoff in fitts' law tasks—on the equivalency of actual and nominal pointing precision. *International Journal of Human-Computer Studies*, *61*(6): 823–856.

Appendix A

List of Publications

Large parts of Chapters 3 to 7 are updated versions of published papers or submitted manuscripts. Chapters 3, 4, 5 and 7 were written primarily by me with feedback from all co-authors. Chapter 6 was co-authored as described in the Statement of Collaboration (page xviii).

- An earlier version of parts of Chapter 3 was published as: Moffatt, K., and McGrenere, J. (2007). Slipping and drifting: Using older users to uncover pen-based target acquisition difficulties. In *Assets'07: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, 11–18. ©ACM, 2007. http://doi.acm.org/10.1145/1296843.1296848. It received an award for best student paper. This award is given to the top-rated paper which has a student as the first author.

- A large portion of Chapter 4 was published as: Moffatt, K., and McGrenere, J. (2009). Exploring methods to improve pen-based menu selection for younger and older adults. *ACM Transactions on Accessible Computing, 2(1)*: 3:1–3:34. ©ACM, 2009. http://doi.acm.org/10.1145/1525840.1525843.

- An earlier version of parts of Chapter 5 was published as: Moffatt, K., Yuen, S., and McGrenere, J. (2008). Hover or tap? Supporting pen-based menu navigation for older adults. In *Assets'08: Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, 51–58. ©ACM, 2008. http://doi.acm.org/10.1145/1414471.1414483.

- An earlier version of Chapter 6 was published as: Trewin, S., Keates, S., and Moffatt, K. (2006). Developing Steady Clicks: A method of cursor assistance for people with motor impairments. In *Assets'06: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, 26–33. ©ACM, 2006. http://doi.acm.org/10.1145/1168987.1168993.

- Chapter 7 has been accepted for publication as: Moffatt, K. and McGrenere, J. (2010) Steadied-Bubbles: Combining techniques to address pen-based pointing errors for younger and older adults. To appear in CHI'10: Proceedings of the 28th ACM conference on Human factors in computing systems.

Appendix B

Follow-Up to the Baseline Study: Experimental Methodology

This study was conducted by Justine Yang as part of her NSERC undergraduate student research assistantship (Summer 2008). The main goal of this study was to provide Justine with experience conducting user studies. Gaining additional insight into the results of the menu task of the Baseline study was a secondary goal. This study explored the impact of two factors: input device (pen versus mouse) and menu orientation (top-down versus bottom up). Because this study was informal in nature, we included only younger adults. We note however, that the two main difficulties encountered in the menu task of the Baseline study (missing-just-below and drifting) affected younger and older users alike.

B.1 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 768 MB 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. For the mouse condition, a standard two-button optical mouse was connected to the tablet. 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 of approximately 35 degrees from horizontal.

B.2 Participants

Eighteen participants aged 19–30 were recruited through campus postings to participate in the study. Participants received $10 for approximately one hour of participation. Participants were right-handed and free of diagnosed motor impairments to their right hand. They all had normal or corrected-to-normal eyesight, and all were novices to pen-based computing.

B.3 Design

This study was run in two parts, one focusing on menu orientation, and the other, on input device. Within each part, interface condition was a repeated-measures factor. That is, half the participants were in the menu orientation group and worked with both the top-down and bottom-up menus. The other half were in the input device group and worked with both the mouse and pen devices. The two parts were analyzed separately.

B.4 Procedure

The study began by introducing participants to the Tablet PC. Participants then completed steps 1–8 of "Get Going with the Tablet PC", as described in Chapter 3. After the tutorial, the tablet was calibrated to each participant using the built-in Windows XP (Tablet PC Edition) calibration utility. Participants then completed the menu conditions. Following each condition participants completed a questionnaire about that condition. Between conditions, participants completed short verbal distracter tasks and a questionnaire on their background and computer experience. Finally, at the end of the study, participants were asked to rank the conditions on a number of factors and encouraged to make additional comments.

B.5 Task

The menu task was largely based on the task used for Technique Study One (Chapter 4). For each menu interface, participants completed a short 24-trial practice block followed by 6 blocks of 48 trials with an enforced 45 second break between blocks. Each block consisted of a 48-item randomly ordered selection sequence from one 24-item menu. (Each item was selected twice in each experimental block and once in the practice block.) Thus, each participant completed 48 trials x 6 blocks x 2 interfaces for a total of 576 trials (excluding the 24 practice trials per interface).

For each trial, a menu item was displayed across the top of the screen. Participants were instructed to find and select that item from the menus as quickly as possible while remaining accurate. The system advanced to the next item, only when the participant successfully selected the correct menu item. A soft clicking sound provided feedback for correct selections and a louder beep sound alerted participants to selection errors.

Menu contents remained constant within each menu 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. [26]. That is, six four-item 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 the secod condition. Each menu item was 20 pixels (4.8 mm) high, and each menu separator was 5 pixels (1.2 mm) high. All items were 100 pixels (24 mm) wide.

B.6 Measures

Our main goal for this study was to provide insight in to the underlying reasons for missing-just-below and drifting. Thus, we did not focus on comparative performance measures such as speed and accuracy, but rather looked for differences in the individual tap distributions. Thus, our main measures for this study, were the x, y coordinates of selections. These were normalized to the center of the target menu item.

Subjective data were collected after each condition using the ISO 9241-9 independent ratings of comfort scale [35]. At the end of the study, a questionnaire asked for comparative rankings of the cursors on speed, accuracy, frustration, and preference.

B.7 Motivation

To motivate quick and accurate performance, a $10 incentive was awarded to the top third of performers in each age group. The one-third ratio was chosen to encourage participants to believe they had a reasonable chance of succeeding. To help participants gauge their performance, graphical feedback of performance was presented during the breaks between blocks. This feedback was the same as that shown in Figure 4.3 for Technique Study One.

Appendix C

Computer Task Instructions

In this appendix, we list the instructions given to participants for the main computer tasks used in the studies.

C.1 Baseline Study

> For this task you are going to use the pen to select (*circular targets on the screen* / *items from a menu*). The task is divided into 4 blocks of (*72* / *36*) selections, with a 1 minute rest between blocks. Each trial will begin with a blue outlined circle in the in the center of the screen. Tap on it to start the trial and reveal the (*target* / *prompt*).
>
> **Menu task only:** You should note that the menus are organized by category, and that the items within the menu are organized alphabetically. After every four items there is a separator.
>
> When you tap on (*the screen* / *a menu item*) the trial will end regardless of whether or not you were successful. If you were successful, nothing happens, the trial just ends. If you missed a beep will sound, but don't try to fix you error, just continue on with the next trial
>
> You should be trying to move as quickly as possible while remaining accurate.
>
> One thing to note is that if you need to take a quick break—to scratch your nose or re-adjust your position—it's best if you can do it between

trials when the blue circle is visible, but if you can't wait, don't worry, just continue on as soon as you can. Also, in the lower-left corner of the screen there is counter to let you know how many trials are remaining before your next break.

C.2 Follow-Up to the Baseline Study

C.2.1 Session I

Input device group: For the first session you will be using a mouse (or pen) to complete the menu selections, for the second session you will be using a pen (or a mouse).

Menu orientation group: For the first session the menu will be across the top of the screen and the contents will open down. In the next session, the menu will be across the bottom of the screen and the menu will open up (and vice versa).

Both groups: For the computer sessions, you are going to use two different interfaces to complete a series of menu selections. As I mentioned previously, between the two interfaces we will take a break and do some other non-computer tasks. To help you get familiarized with the interface and the menu contents, each computer session will first take you through a practice session which you should use to familiarize with the contents of the menu items. These menu items are different in each computer session, but remain the same throughout a single session. For each session, there is one menu containing a total of 24 items divided into 6 groups of 4 related items. [show]. After this practice session, you will complete 6 sets of 48 menu selections, with a 45 second break between the sets. During the 45-second breaks, you will get feedback on your performance.

This is a screenshot of the feedback that you will get. The graph shows your rate of successful trials. So it reflects both your speed [point to y axis] and accuracy [point to the number of errors]. We want you to try to make this as high as possible. [point to the line]. This requires finding a balance between speed and accuracy. If you slow down enough, you can be perfectly accurate, but as you go faster you will make more mistakes and then you lose time fixing those mistakes. So the trick is to find the balance. At the bottom you can see how many errors you made and how long you took in the last set.

So, now, let's move on to the actual task that you will be performing. Each trial will begin with a word printed across the top of the screen. [Demonstrate.] Your task is to select that word from the menu. These menus work just like with a mouse: you can either tap once on the word menu and then again on the target word, or tap down on the menu and then drag to the target word. If successful, you will hear a click sound like this. [Demo Hit.] If you miss, you will hear a louder beep like this. [Demo Miss.] Also, if you miss, the program will prompt the same word again until you get it right. In the lower-left corner of the screen, there is a counter to let you know how many trials are remaining before your next break. [point.] Most importantly, you should be trying to move as quickly and accurately as possible. Remember that we are giving bonuses to the top 1/3 individuals who successfully completed the most trials the quickest. So you may begin. If you have any question, please ask.

C.2.2 Session II

This is the second computer session. Just a little reminder that the menu contents are now different from the previous set, so you should use the practice session to try to learn the menu contents. And also, please remember that you should be trying to move as quickly and accurately as possible. If you have questions, please ask

C.3 Technique Study One (Missing-Just-Below)

C.3.1 Session I

For the computer sessions you are going to use three different menu programs: Cosmos, Dahlia, and Rose. Between each programs we will take a break from the computer and do some other tasks. You don't need to worry about how the three menu programs differ. You should just use them naturally and after we will discuss which you thought worked best.

Each computer session will begin with a practice session which you should use to familiarize yourself with the contents of the menu as they will be different for each session. The menu contains 12 items consisting of three groups of four related items. After the practice session, you will complete six sets of 36 selections, with a 45 second break between sets.

During these breaks, you will get feedback on your performance. The feedback looks like this `[show feedback]`. The graph shows your rate of successful trials. So it reflects both your speed and accuracy. We want you to try to make this as high as possible. This requires finding a balance between speed and accuracy. If you slow down enough you can be perfectly accurate, and if you make lots of errors you can probably go faster, but between the two there is a balance. At the bottom you can see how many errors you made and how long you took in the last set.

OK, so lets turn to the task. Each trial will begin with a word printed across the top of screen. Your task is to select that word from the menu. These menus work just like with a mouse: you can either tap once on the word menu and then again on the target word, or tap down on the menu and then drag to the target word. If successful, you will hear a click sound like this `[Demo Hit]`. If you miss, you will hear a louder beep like this `[Demo Miss]`. However, the trial ends regardless of whether or not you were successful, so don't try to fix errors, just continue on with the next item. In the lower-left corner of the screen there is counter to let you know how many trials are remaining before your next break.

Most importantly, you should be trying to move as quickly and accurately as possible. Remember that we are giving bonuses to the top 1/3 individuals who successfully completed the most trials the quickest.

C.3.2 Sessions II and III

This is the (*second* / *third*) computer session, remember that the menu contents change between sessions, you should use this first practice set to try to learn the menu contents. And most importantly, please remember that you should be trying to move as quickly and accurately as possible. If you have any questions please ask.

C.4 Technique Study Two (Drifting)

C.4.1 Session I

For the computer sessions you are going to use three different menu programs: Entry-switch, Tap-switch, and Glide-switch. You should just use them naturally and after we will discuss which you thought worked best. Between each program we will take a break from the computer and do some other tasks. Each computer session will begin with a practice session, which you should use to familiarize yourself with the contents of the menu as they will be different for each session.

Each menu contains 12 items consisting of three groups of four related items `[Demo]`. After the practice session, you will complete six sets of 36 selections, with a 45 second break between sets. During these breaks, you will get feedback on your performance. The feedback looks like this. The graph shows your rate of successful trials. It reflects both your speed and accuracy. We want you to try to make this as high as possible.

This requires finding a balance between speed and accuracy. If you slow down enough you can be perfectly accurate, and if you make lots of errors you can probably go faster, but between the two there is a balance. At the bottom you can see how many errors you made and how long you took in the last set. Errors give an inherent time penalty since it takes longer to correct your mistake.

So for the task, each trial will begin with a word printed across the top of screen. Your task is to select that word from the menu. These menus work just like with a mouse. If successful, you will hear a click sound like this `[Demo Hit]`. If you miss, you will hear a louder beep like this `[Demo Miss]`. The trial ends when you select the correct item, so try to fix your errors as quickly as possible. In the lower-left corner of the screen there is counter to let you know how many trials are remaining before your next break.

Most importantly, you should be trying to move as quickly and accurately as possible. Remember that we are giving bonuses to the top 1/3 individuals who successfully completed the most trials the quickest.

`[Insert method-specific instructions here.]`

C.4.2 Session II and III

This is the (*second* / *third*) computer session.

`[Insert method-specific instructions here].`

Remember that the contents change between sessions—you should use this first practice set to try to learn the menu contents. And most importantly, please remember that you should be trying to move as quickly and accurately as possible. If you have any questions please ask.

C.4.3 Method-Specific

Entry: In this condition, the selected menu changes as soon as you move the pen over a new menu, even if the pen isn't touching the screen. Note that tapping on an open menu will close that menu. You can also tap anywhere outside the menu area to close an open menu.

Tap: In this condition, the selected menu changes when you tap on a new menu. Note that tapping on an open menu will close that menu. You can also tap anywhere outside the menu area to close an open menu.

Glide: In this condition, the selected menu changes when you move the pen over a new menu, even if the pen isn't touching the screen, but there is a very short delay—it doesn't happen right away. Note that tapping on an open menu will close that menu. You can also tap anywhere outside the menu area to close an open menu.

C.5 Technique Study Three (Slipping)

C.5.1 General

You will be using four different computer programs today. Each of these is different in how it determines what a selection is. We are interested in knowing which ones make it easier for you to make selections and which ones make it harder. We will be collecting data on your speed and accuracy as well as asking you for your opinion. Remember that we are evaluating the programs and not you.

For all of the programs, your task is to repeated select the bright green target. It will appear in different locations on the screen, and in different sizes `[demonstrate this by tapping through a few trials].` the light gray circles, are distracter targets; you should avoid selecting them. When the goal target is active it turns to dark red, and when a distracter target is active it turns to dark gray. If you touch the screen while the goal target is activated, it turns a brighter red, like this. And distracter targets turn lighter like this. As soon as you correctly select the goal target, the trial will automatically advance so you should try to keep going without taking breaks during the block. If the goal target doesn't change, then you haven't correctly selected it yet, and you should keep trying.

For each program, you will select a series of targets on the screen. You will start with a short practice period so that you can get familiar with that program, and then you will complete a series of four longer blocks. Each block contains 81 selections, and takes 1–2 minutes. Between each block you will get a 30 second break to rest your arm. During these breaks you will also get feedback on your performance. We want you to try to make the selections as quickly and as accurately as you can. Remember that we are giving a bonus to the top 1/3 individuals who are quickest to successfully complete the trials .

Ok, that's all the general task instructions, do you have any questions so far?

`[Insert method-specific instructions here.]`

After first practice session: Great you are done practicing. Before you begin, I want to go over this performance feedback. You will see this after each of the blocks. This `[point to the graph]`, shows your speed in selections per minute for each of the blocks. This is the information we use to determine the top performers. You should try to get it as high as possible. At the bottom `[point to the box]`, there is some additional information for the most recent block. It tells you how many errors you made, and your total time on that block.

C.5.2 Method-Specific

Control: For this program, the cursor is an arrow. While you are touching the screen, this cursor moves with the pen. This might help you make final adjustments, but it's important the pen is on the target just before you lift off the screen.

Steady: For this program, the cursor is an arrow. While you are touching the screen the cursor doesn't move at first. This might help you stay on the target while tapping, but it's important the pen accurately lands on the target.

Bubble: For this program, the cursor is a semi-transparent circle. While you are touching the screen the cursor moves with the pen. This might help you make final adjustments, but it's important the pen is on the target before you lift off the screen.

Bubble-Steady: For this program, the cursor is a semi-transparent circle. While you are touching the screen the cursor doesn't move at first. This might help you stay on the target while tapping, but it's important the pen accurately lands on the target.

Appendix D

Neuropsychological and Motor Test Instructions

D.1 Purdue Pegboard Test

This test was used in the Baseline Study and all three Technique Studies. In the Baseline Study, we administered all four subcomponents of this test: right hand, left hand, both hands, and assembly. In the subsequent Technique Studies, we administered only the right-hand component of this test.

D.1.1 Introduction

> This is a test to see how quickly and accurately you can work with your hands. The test consists of 4 subcomponents, and we will do the test 3 times. The first time through, I will show you what to do and then you will have an opportunity to practice. After that, I will remind you which task you are to do and ask you if remember, if you don't remember let me know and I will demonstrate again. At all times make sure you understand exactly what to do before we begin. At the start of each task I'd like you to place your hands face-down on the table on either side of the board like this. Any questions?

D.1.2 Right / Left Hand

For the right (left) hand task, pick up one pin at a time with your right (left) hand from the right (left) hand cup place it in the right hand column, starting at the top. If you drop a pin, do not stop to pick it up. Simply continue by picking another pin out of the cup. Continue placing pins until I say stop. You have 30 seconds.

D.1.3 Both Hands

For this part of the test, you will use both hands at the same time. Pick up a pin from the right hand cup with your right hand, and at the same time pick up a pin from the left hand cup with your left hand. Then place the pins down the rows, beginning with the top row. Continue placing pins until I say stop. You have 30 seconds.

D.1.4 Assembly

For this part of the test you are going to use both hands for an assembly. Pick up one pin from the right hand cup with your right hand. While you are placing it in the top hole of the right hand row, pick up a washer with your left hand. As soon as the pin has been placed, drop the washer over the pin and pick up a collar with your right hand. While the collar is being dropped over the pin, pick up another washer with your left hand and drop it over the collar. While the final washer for the first assembly is being placed with your left hand, start the second assembly immediately by picking up another pin with your right hand. Continue placing assemblies until I say stop. You have 1 minute.

D.2 Digit Symbol Substitution Test

This test was used in the Baseline Study and all three Technique Studies. In the Baseline Study, participants had 90 seconds (as specified in the instructions below) to complete this test. In the Techniques Studies, this was reduced to 60 seconds.

For this test, numbers are paired with symbols as shown in this key along the top. Your task is to fill in as many symbols as you can in 90 seconds. You are free to refer to the key as often as you like throughout the test, but you must fill in the boxes in order starting at the top-left and moving left then down.

D.3 Steadiness Test

Two versions of this test are specified below. The first was used in the Baseline Study. It was not informative; thus, we administered an alternate version in Techniques Studies One and Two. We did not administer this test in Technique Study Three.

D.3.1 Version I (Baseline Study)

This is a test to measure the steadiness of your hand. There are eight holes, gradually diminishing in size. Your task is to hold the stylus in the hole without touching the sides for 10 seconds. You don't need to put the whole pen into the whole, just the tip. If you touch the sides you will hear a beep. At the end of 10 seconds I will say Stop. Between holes, you will have 10 seconds to rest. After 8 seconds of rest, I will say Ready, and at 10 seconds I'll say Start. We will start with the largest hole moving across then down. We will do the whole test 3 times; Between tests you will get 30 seconds rest. After all eight holes, you will get about 30 seconds rest. Any questions?

D.3.2 Version II (Technique Studies One and Two)

This is a test to measure the steadiness of your hand. There are eight holes, gradually diminishing in size. Your task is to move the stylus in and out of the holes starting with largest and moving to the smallest. You don't need to put the whole pen into the whole, just the tip. If you touch the sides you will hear a beep. You cannot support your arm with your other hand, the table, or anything else. I will be recording which holes you touch. We will do the test three times.

D.4 Simple Reaction Time Test

A simple reaction time test was included in each of the three Technique Studies. Two versions of this test were used: The Online Reaction Time Test[21] and Sheep Dash.[22] The first version was used in Techniques Studies One and Two, and the second, in Technique Study Three. Both provided comparable data; however, we felt the latter version was more engaging. Both tests were available online and could be downloaded for offline use.

D.4.1 Version I (Techniques Studies One and Two)

> This test measures your reaction time. To begin, press the space bar. This will turn the stop light from yellow to red. After a random amount of time, the stoplight will turn green. Your task is to tap on the space bar again as quickly as you can. We will repeat the whole test five times.

D.4.2 Version II (Techniques Study Three)

> This test measures your reaction time. Your task is to click on the button at the bottom of the screen whenever you see a sheep leaving the flock. There are five sheep to stop, and at the end there will be a summary of your response times.

D.5 North American Adult Reading Test

This test was administered in the Baseline Study and all three Technique Studies. After the Baseline Study. only the first 30 words of this test were used. (In the Baseline Study, all 61 words were included.) We used the data in this study to identify participants with insufficient English competency to complete the tasks. It also served as one of the distracter task between conditions.

[21]The Online Reaction Time Test, ©2002 by Jim Allen, http://www.getyourwebsitehere.com/jswb/rttest01.html

[22]Sheep Dash, ©2009 BBC, http://www.bbc.co.uk/science/humanbody/sleep/sheep/reaction_version5.swf

> I want you to read slowly down this list of words starting here `[point to the word` *`debt`*`]` and continuing down this column and on to the next. I must warn you that there are many words that you probably wont recognize; in fact most people don't, so just guess at any you don't know, Ok? Go ahead.

D.6 FAS Test

This test was used as a distracter task between conditions in the Technique Studies.

> For this test, I am going to give you a letter, and you are to name as many words you can beginning with that letter for 30 seconds. For example, if I gave you the letter B, you could list 'Banana', 'Bread', etc. We will be doing three letters in total. As soon as I give you the letter you may begin.

D.7 Reverse Digit Span Test

This test was used as a distracter task between conditions in Technique Study Three. An additional distracter task was needed for this study because it had four conditions. (Technique Studies One and Two only had three conditions each.)

> This time we are going to do a test of working memory—the ability to hold on to information while doing something with it. The test is simple. On each trial, I will read aloud a list of numbers. Your job is to listen carefully, and then at the end of the list, to recall the numbers in the reverse order. So, if I say the numbers 2, 9 and 5, your job would be to say them backwards as: 5, 9, 2. I'll gesture like this at the end of each sequence, so that you know it's your turn `[Gesture with hand]`.

Appendix E

Background and Computer Experience Questionnaire

Background and Computer Experience Questionnaire

Instructions: Please answer the following questions to the best of your ability. If you have any questions, please do not hesitate to ask the researcher for help.

Part I: Personal Information

1. What is your gender? ☐ Male ☐ Female

2. What is your age? ______________

3. At your primary residence (the place you call home), what is the average age-group of other adults living with you?

   ☐ Not applicable, I live alone.

____ 18–24	____ 35–39	____ 50–54	____ 65–69	____ 80–84
____ 25–29	____ 40–44	____ 55–59	____ 70–74	____ 85+
____ 30–34	____ 60–64	____ 45–49	____ 75–79	

4. Please indicate the highest level of education you have obtained. Where space is provided, please specify degree/program.

   ☐ Some high school
   ☐ Completed high school
   ☐ Some post-secondary education: ________________________
   ☐ Completed community college: ________________________
   ☐ Completed undergraduate degree: ________________________
   ☐ Some graduate or professional school: ________________________
   ☐ Completed post-graduate degree: ________________________
   ☐ Other, please specify: ________________________

5. What is your primary job or profession? (What do you do for a living?) Please select the most appropriate alternative.

☐ Professional–full time	☐ Retired–professional
☐ Professional–part time	☐ Retired–skilled trade worker
☐ Skilled trade worker–full time	☐ Student–full time
☐ Skilled trade worker–part time	☐ Student–part time
☐ Home maker	☐ Other, please specify: __________

Part II: Computer Experience

> Please note that for the purposes of this questionaire, the term computer refers to any of the following: desktop, laptop/notebook, tablet, handheld (i.e., PDA, PalmPilot, etc.)

1. When did you first use a computer:
   ☐ pre-1960 ☐ 1970–1974 ☐ 1985–1989 ☐ 2000–2004
   ☐ 1960–1964 ☐ 1975–1979 ☐ 1990–1994 ☐ 2005–pres.
   ☐ 1965–1969 ☐ 1980–1984 ☐ 1995–1999 ☐ Never

2. What kinds of computers have you used: *Tick all that apply*
   ☐ PC (Windows) ☐ Unix ☐ Tablet
   ☐ PC (Linux) ☐ Laptop/Notebook ☐ Not sure
   ☐ Mac/Apple ☐ Handheld (PDA/PalmPilot . . .) ☐ Other

3. Do you use a computer for work? (either at home or work).
   ☐ Yes ☐ No ☐ N/A
   If yes, on an average day, approximately how many hours do you spend using a computer for this? ________________ hours.

4. Do you use a computer for leisure or personal tasks? (at home or work).
   ☐ Yes ☐ No ☐ N/A
   If yes, on an average day, approximately how many hours do you spend using a computer for this? ________________ hours.

5. How familiar are you with the following types of computer applications?

	Unfamiliar		Familiar	
	Completely	Mostly	Somewhat	Very
Word processor (e.g., MS Word)	☐	☐	☐	☐
Email (e.g., Outlook, Eudora, Hotmail)	☐	☐	☐	☐
Web Browser (e.g., Netscape, Firefox, IE)	☐	☐	☐	☐
Spreadsheet (e.g., Excel, Lotus 1-2-3)	☐	☐	☐	☐
Graphics (e.g., Adobe, Corel Draw)	☐	☐	☐	☐
Presentation software (e.g., Powerpoint)	☐	☐	☐	☐
Database (e.g., MySql, Oracle)	☐	☐	☐	☐
Music/Video (e.g, iTunes, Quicktime)	☐	☐	☐	☐
Computer games	☐	☐	☐	☐
Other, please specify: _______	☐	☐	☐	☐

6. Which of the following have you done with a computer?
   Tick all that apply.
   - ☐ I have customized options or preferences within an application.
   - ☐ I have made a purchase online.
   - ☐ I have installed a computer application.
   - ☐ I have installed an operating system.
   - ☐ I have formatted a hard drive.
   - ☐ I have added a new external device (e.g., printer, scanner, camera).
   - ☐ I have added memory.
   - ☐ I have added an internal device (e.g., hard-drive, internal CD-ROM).

7. How would you characterize yourself in terms of computer knowledge?
   ☐ Basic knowledge ☐ Moderate knowledge ☐ Extensive knowledge

8. Have you ever attended a computer course?
   ☐ Yes ☐ No If yes, please specify: ____________________

9. Is there any other relevant information about your use of computers that you would like to note here?
   ____
   ____
   ____

Appendix F

Final Comparative Questionnaires

The Technique Studies each included a final questionnaire, which asked participants to rank the interfaces used in the experiment. This appendix lists the questionnaires used in each study.

Computer Interface Comparison (Technique Study One)

Please answer the following questions to the best of your ability. To refresh your memory, you used three different menu programs today in the following order:
(1) ______________, (2) ______________, and (3) ______________.

1. Did you notice any difference between any of the programs?
   ☐ Yes, I could tell all three were different.
   ☐ Yes, but I only observed a difference between the first and second.
   ☐ Yes, but I only observed a difference between the first and third.
   ☐ Yes, but I only observed a difference between the second and third.
   ☐ No, they all seemed the same to me.

   If you answered yes to question 1, please answer the following questions.

2. Which menu program did you prefer overall? Please put a 1 beside that program and a 2 beside the next preferred program.
   ___ First ___ Second ___ Third
   Comments: ____

3. With which menu program did you feel you were the fastest? Please put a 1 beside that program and a 2 beside the program with which you were the second most fastest.
   ___ First ___ Second ___ Third
   Comments: ____

4. With which menu program did you feel you made the fewest errors (incorrect menu selections)? Please put a 1 beside that program and a 2 beside the program with which you felt you made the second fewest errors.
   ___ First ___ Second ___ Third
   Comments: ____

5. With which menu program did you feel the most frustrated? Please put a 1 beside that program and a 2 beside the second most frustrating program.
   ___ First ___ Second ___ Third
   Comments: ____

Computer Interface Comparison (Technique Study Two)

Please answer the following questions to the best of your ability. To refresh your memory, you used three different menu programs today in the following order: (1) _________________, (2) _________________, and (3) _________________.

1. Which menu program did you prefer overall? Which did you least prefer?
   Most preferred ____
   Least preferred ____
   Comments: ____
2. With which menu program did you feel you were the fastest? With which were you the slowest?
   Fastest ____
   Slowest ____
   Comments: ____
3. With which menu program did you feel you made the most errors (incorrect menu selections)? With which did you make the fewest?
   Most errors ____
   Fewest errors ____
   Comments: ____
4. With which menu program did you feel the most frustrated? With which were you the least frustrated?
   Most frustrating ____
   Least frustrating ____
   Comments: ____
5. Which menu program did you initially find the easiest to use? Which was the hardest initially?
   Easiest initially ____
   Hardest initially ____
   Comments: ____

Computer Interface Comparison (Technique Study Three)

Please answer the following questions to the best of your ability.

1. You used 4 interfaces today. Please order them for each of the following, using the labels 1–4 for first used to fourth used.

   Fastest | _____ | _____ | _____ | _____ | Slowest

   Why? ______________________________

   Easiest | _____ | _____ | _____ | _____ | Most difficult

   Why? ______________________________

   Least frustrating | _____ | _____ | _____ | _____ | Most frustrating

   Why? ______________________________

   Most preferred | _____ | _____ | _____ | _____ | Least preferred

   Why? ______________________________

2. Did any of the interfaces seem similar to you?

   ☐ Yes ☐ No

   If yes, which ones? ______________________________

   Why? ______________________________

3. Do you have any additional comments?

   ____

   ____

Appendix G

Mouse Study Materials

This appendix provides the instructions used throughout the Mouse Study as well as the semi-structured interview questions. The instructions and questions are interspersed according to the order they were presented in the study.

Introduction

The study session began with an overview of study and an interview on mouse use.

> Ok, let's begin. First I'd like to remind you that any data that we collect will be fully anonymous.
>
> We will be alternating asking you some questions about you and your computer use with having you do activities on the computer. Most of our time today is going to be spent clicking, but we are going to do a little bit of dragging at the end. To give you an overview:
>
> - We will be doing 2 clicking sessions, each with a different mouse program.
> - For each session, first I'll show you how it works, and then let you practice.
> - Then we will do 2 sets of 37 targets.
> - In between the 2 sets, we will take a break and I will ask you some questions.

- Between the 2 sessions, we will also take a break. I will ask you some questions and you will have an opportunity to rest.

Any questions?

We are going to begin with an interview about your mouse use.

Interview Part 1: Mouse Setup

1. Do you normally use a mouse pad?
   ☐ Yes ☐ No ☐ Unsure

   (a) If yes, does it have a wrist pad for support?
   ☐ Yes ☐ No ☐ Unsure

2. Do you use your left or right hand to move the mouse? ☐ Left ☐ Right

3. Are you left-handed or right-handed?
   ☐ Left ☐ Right

4. Do you experience any problems when using a mouse?
   ☐ Yes ☐ No ☐ Unsure

   (a) What kinds of problems and how often do you experience them?
      i. Accidental clicks
      ☐ Yes ☐ No Freq _________ Cause _________
      ii. Slips
      ☐ Yes ☐ No Freq _________ Cause _________
      iii. Wrong button
      ☐ Yes ☐ No Freq _________ Cause _________
      iv. Fatigue
      ☐ Yes ☐ No Freq _________ Cause _________
      v. Other ______ Freq _________ Cause _________

5. Do you have a general strategy for using the mouse?
   ☐ Yes ☐ No ☐ Unsure

   (a) If yes, could you please describe those strategies?

   ____

   ____

   ____

   (b) For example:
      i. Do you use keyboard alternatives?
         ☐ Yes ☐ No ☐ Unsure
      ii. Do you Switch hands?
         ☐ Yes ☐ No ☐ Unsure
      iii. Do you brace your hand/fingers against the table?
         ☐ Yes ☐ No ☐ Unsure
      iv. Why did you develop these strategies?

         ____

         ____

         ____

Vision Test

The vision test consisted of a series of short phrases presented to the participant in succession. As the test progressed, the font size of the phrases got smaller. Participants were instructed to read the phrases aloud, and to inform us when the text got too small to read comfortably, as follows.

> We are going to ask you to read a series of short phrases. This is to give us an idea of your vision and how easy it is for you to see to the computer screen.
>
> **If not wearing glasses;** If you normally wear glasses when using the computer, please wear them for this and all computer activities we do today.
>
> Please read the text aloud, and let us know as soon as the text becomes too small to read comfortably.

We took the phrases from popular nursery rhymes, and selectively changed a few words in each phrase to help us ensure that the participant was reading and not reciting from memory. The following lists each phrase according to the font size at which we presented it.

24pt. Baa, baa, black sheep. Have you any fur?

20pt. Hickory, dickory, dick, the mouse ran up the tree.

18pt. Yankee Doodle came to town riding on a donkey.

16pt. Georgie Porgie, pudding and pie, kissed the girls and made them smile.

14pt. Hush, tiny baby, don't say a thing.

12pt. Birds of a letter flock together.

11pt. The ants go marching three by three.

10pt. A diller, a dollar, a ten o'clock bother.

9pt. Goosey, goosey, gander, whither shall I wonder?

Interview Part 2: Vision Test

1. Smallest font size read: _______

2. Smallest font size reported comfortable: _______

Computer Session 1, Set 1 Instructions

Now we are going to move on to our first computer session.

With steady clicks: We are going to begin with the steady click program. This program will keep the mouse steady while you are clicking and ignore any clicks you make while moving the mouse or clicking other mouse buttons.

Without steady clicks: We are going to begin with the basic click program. This program operates as a regular mouse. Please use it as you would normally use a mouse.

Each time we begin a new set this is what the screen will look like. To begin, click on the 'START' button, which is highlighted in blue. [Demo starting.]

Your task is to click on the blue target. Please try to avoid clicking on the other targets, but try to go as quickly as possible. As soon as you click on the target, it will turn white and the next target will turn blue. [Demo the task.]

Note that if you click any of the other targets, or drag any of the targets the task will freeze, and you will have to click the *reset* button before you can continue. [Demo reset.]

Also, if you click on the center or right mouse buttons, a pop-up window will appear telling you to use the left button. Again the task will freeze. To get rid of the pop-up and continue the task, click anywhere on the screen. [Demo popup]

You will notice that each target has a word written on it. We would like you to try to remember the words of the blue targets. You might find it helpful to read them out loud as you go.

Go ahead and try a few to practice. [Let the participant try about four.]

Great, do you have any questions? [Answer any questions.]

Let's begin the first set now.

Remember we are testing the design of the mouse not you.

If you become tired at any point during this study, please let us know so you can take a break or stop.

Interview Part 3: Computer Session 1 Mini-Interview

1. Did you notice the mouse slipping at all while you were clicking?
   ☐ Yes ☐ No ☐ Unsure

(a) If yes, how many times? _______

(b) Did this ever cause you to miss a target?

☐ Yes ☐ No ☐ Unsure

i. If yes, how many times? _______

2. Did you notice any accidental clicks?

☐ Yes ☐ No ☐ Unsure

(a) If yes, how many times? _______

3. Any comments about this session?

____

____

____

Interview Part 4: Web Browser Use

1. What kind of websites do you visit?

[For answers given unprompted, indicate order]

i.	___ Do you shop online?	☐ Yes	☐ No	Frequency ___
ii.	___ Do you play games online?	☐ Yes	☐ No	Frequency ___
iii.	___ Do you use web-based email?	☐ Yes	☐ No	Frequency ___
iv.	___ Do you read the news online?	☐ Yes	☐ No	Frequency ___
v.	___ Other: _________________	☐ Yes	☐ No	Frequency ___

2. When web browsing, how do you scroll down the page?
   i. ___ Do you use the up and down arrows on the keyboard?
      ☐ Yes ☐ No ☐ Unsure Frequency ___
   ii. ___ Do you use the page up and page down keys?
      ☐ Yes ☐ No ☐ Unsure Frequency ___
   iii. ___ Do you use a scroll wheel on the mouse?
      ☐ Yes ☐ No ☐ Unsure Frequency ___
   iv. ___ Do you drag the scroll bar with the mouse?
   v. ___ Do you click on the up and down arrows on the scroll bar?
      ☐ Yes ☐ No ☐ Unsure Frequency ___
   vi. ___ Do you click on the scroll bar itself?
      ☐ Yes ☐ No ☐ Unsure Frequency ___
   vii. ___ Other: ______________ Frequency ___

3. When web browsing, do you ever use the middle mouse button?
   ☐ Yes ☐ No ☐ Unsure

   (a) For what? ____

4. When web browsing, do you ever use the RIGHT mouse button?
   ☐ Yes ☐ No ☐ Unsure

   (a) For what? ____

Computer Session 1, Set 2 Instructions

Ok, we are now ready to do the second set.

With steady clicks: We will again be using the steady click program. Remember, this program keeps the mouse steady while you are clicking and ignores any clicks you make while moving the mouse or clicking other mouse buttons.

Without steady clicks: We will again be using the basic click program. This is the program that operates as a regular mouse. Please use it as you would normally use a mouse.

Remember that we would like you to try to go as quickly as possible and to try to remember the words of the blue targets. You might find it helpful to read them out loud as you go.

Interview Part 5: Condition One Follow-up

.

1. Did you notice the mouse slipping at all while you were clicking?
   ☐ Yes ☐ No ☐ Unsure
   (a) If yes, how many times? _______
   (b) Did this ever cause you to miss a target?
       ☐ Yes ☐ No ☐ Unsure
       i. If yes, how many times? _______
2. Did you notice any accidental clicks?
   ☐ Yes ☐ No ☐ Unsure
   (a) If yes, how many times? _______
3. How did this compare to your normal mouse use?
   ☐ Better ☐ Worse ☐ No difference
   (a) Did you notice any differences? ___________________________
   (b) How did you feel about them? ___________________________
4. On a scale from 1-7, where 1 is easiest possible and 7 is most difficult, how easy was it for you to click on the targets using this program?
   easy 1 2 3 4 5 6 7 difficult
5. Using this same scale, how easy is it for you to click on targets normally?
   easy 1 2 3 4 5 6 7 difficult

The following questions are for the with steady clicks condition only.

1. Did the steady clicks ever help prevent you from slipping off a target?

   ☐ Yes ☐ No ☐ Unsure

   (a) If yes, how many times? _______

2. Did the steady clicks ever get in the way of moving the cursor?

   ☐ Yes ☐ No ☐ Unsure

3. How would you rate the level of slip support provided?

   ☐ Too much ☐ Too little ☐ Just right

4. Did you notice any clicks being filtered?

   ☐ Yes ☐ No ☐ Unsure

   (a) Were any of these clicks that you wanted?

      ☐ Yes ☐ No ☐ Unsure

      i. How many? _______

   (b) Were any of these clicks that you didn't want?

      ☐ Yes ☐ No ☐ Unsure

      i. How many? _______

5. How would you rate the level of filtering?

   ☐ Too many ☐ Too few ☐ Just right

6. Would you use steady clicks if it were available on your computer?

   ☐ Yes ☐ No ☐ Unsure

   (a) Why or why not? _________________________________

Computer Session 2, Set 1 Instructions

Now we are going to move on to our second computer session. We will be switching to the other mouse program now, but the task remains the same

With steady clicks: We are going to begin with the steady click program. This program will keep the mouse steady while you are clicking and ignore any clicks you make while moving the mouse or clicking other mouse buttons.

Without steady clicks: We are going to begin with the basic click program. This program operates as a regular mouse. Please use it as you would normally use a mouse.

Note that if you click any of the other targets, or drag any of the targets the task will freeze, and you will have to click the *reset* button before you can continue. [Demo reset.]

Please remember to try to avoid clicking on the other targets, but try to go as quickly as possible. As soon as you click on the target, it will turn white and the next target will turn blue. [Demo the task.]

Remember that if you click any of the other targets or drag any of the targets the task will freeze, and you will have to click the reset button, before you can continue [Demo reset.]

Also, remember that if you click on the center or right mouse buttons, a pop-up window will appear telling you to use the left button and the task will freeze. To get rid of the pop-up, click anywhere on the screen [Demo popup.]

We would again like you to try to remember the words of the blue targets. You might find it helpful to read them out loud as you go.

You will notice that each target has a word written on it. We would like you to try to remember the words of the blue targets. You might find it helpful to read them out loud as you go.

Go ahead and try a few to practice. [Let the participant try about four.]

Great, do you have any questions? [Answer any questions.]

Let's begin the first set now.

Remember we are testing the design of the mouse not you.

If you become tired at any point during this study, please let us know so you can take a break or stop.

Interview Part 6: Computer Session 2 Mini-Interview

The same questions were used as for the mini-interview of Computer Session 1 (see Interview Part 3, page 211).

Interview Part 7: Basic Information

1. What is your age: _______
2. What is your Occupation: ____________________
3. Do you use a computer regularly? ☐ Yes ☐ No
   (a) How often? _____________
   (b) About how long at a time? _____________
4. About how long have you been using computers regularly? _______
5. What was your primary reason to start using computers? ________________
6. Have you ever had any computer training? ☐ Yes ☐ No
   (a) If yes, what kind? _________________________________
7. (If relevant) What is the nature of your disability? How does it affect your ability to use a mouse? _______________________________________

Computer Session 2, Set 2 Instructions

These instructions were the same as for Computer Session 1, Set 1 (see page 210).

Interview Part 8: Condition Two Follow-up

The same questions were used as for the follow-up to Computer Session 1 (see Interview Part 5, page 211).

Interview Part 9: Comparative Follow up

1. Did you prefer to use the mouse with or without steady click?
   ☐ With ☐ Without

2. Could rate that preference on a scale from 1-7, where '1' is no difference and '7' is strongly prefer?
   no difference 1 2 3 4 5 6 7 strongly prefer

3. In your opinion, was it faster with or without steady click?
   ☐ With ☐ Without

4. Could rate that preference on a scale from 1-7, where '1' is no difference and '7' is strongly prefer?
   no difference 1 2 3 4 5 6 7 strongly prefer

5. Do you think there were there fewer errors with or without steady click?
   ☐ With ☐ Without

6. How many fewer? _______ Errors with: _______ Errors without: _______

Computer Session Dragging Instructions

Great, we are now going to move on to dragging. We think the steady click program that we tried out today might help some people with clicking, but it also has an effect on dragging. We want you to try out dragging so that we can get your feedback about it.

For each drag, I'm going to ask you to drag one word on top of another. Both words will be highlighted in blue so that you can find them. Your drags don't need to be perfectly aligned, just drag the word so that it is mostly covering the other word.

Ok any questions?

Press *start* to begin

1. Can I get you to drag *about* onto to *best*? (and press reset when done)

2. …*tell* onto *step*? (and press reset…)
3. …*visit* onto *no*? (and press reset…)
4. …*down* onto *happy*? (and press reset…)
5. …and *be* onto *would*? (and press reset…)

Interview Part 10: Dragging Follow up

1. On a scale from 1-7, where 1 is easiest possible and 7 is most difficult, how easy was it to drag a long distance?
   easy 1 2 3 4 5 6 7 difficult
2. And on this same scale, how easy was it to drag a short distance?
   easy 1 2 3 4 5 6 7 difficult
3. How did this compare to your normal mouse use? ☐ Harder ☐ Easier
   (a) Would this prevent you from using steady clicks for web browsing? ☐ Yes ☐ No
   (b) Are there tasks for which you wouldn't want to use steady clicks? ☐ Yes ☐ No
      i. Which ones? ____

Interview Part 11: Closing

1. Do you have any questions or comments at this point?
   ____
   ____
   ____

Appendix H

UBC Research Ethics Board Certificates

The University of British Columbia
Office of Research Services and Administration
Behavioural Research Ethics Board

Certificate of Approval

PRINCIPAL INVESTIGATOR	DEPARTMENT	NUMBER
McGrenere, J	Computer Science	**B05-1017**

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT
UBC Campus ,

CO-INVESTIGATORS:
Moffatt, Karyn, Computer Science

SPONSORING AGENCIES
Canadian Institutes of Health Research

TITLE :
Evaluation of Handheld Technology Usability

APPROVAL DATE	TERM (YEARS)	DOCUMENTS INCLUDED IN THIS APPROVAL
JAN 2 3 2006	1	Jan. 16, 2006, Consent form / Questionnaires / Nov. 10, 2005, Advertisement

CERTIFICATION:

The application for ethical review of the above-named project has been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

[redacted]

Approved on behalf of the Behavioural Research Ethics Board
by one of the following:
Dr. Peter Suedfeld, Chair,
Dr. Susan Rowley, Associate Chair
Dr. Jim Rupert, Associate Chair
Dr. Arminee Kazanjian, Associate Chair

This Certificate of Approval is valid for the above term provided there is no change in the experimental procedures

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road,
Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL- MINIMAL RISK RENEWAL

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC BREB NUMBER:
Joanna McGrenere	UBC/Science/Computer Science	H05-81017

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Point Grey Site

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):
Karyn Moffatt

SPONSORING AGENCIES:
Canadian Institutes of Health Research - "Evaluation of Handheld Technology Usability"

PROJECT TITLE:
Evaluation of Handheld Technology Usability

EXPIRY DATE OF THIS APPROVAL: February 5, 2008

APPROVAL DATE: February 5, 2007

The Annual Renewal for Study have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board and signed electronically by one of the following:

Dr. Peter Suedfeld, Chair
Dr. Jim Rupert, Associate Chair
Dr. Arminee Kazanjian, Associate Chair
Dr. M. Judith Lynam, Associate Chair

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road,
Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL - MINIMAL RISK AMENDMENT

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC BREB NUMBER:
Joanna McGrenere	UBC/Science/Computer Science	H05-81017

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Point Grey Site

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):
Karyn Moffatt

SPONSORING AGENCIES:
Canadian Institutes of Health Research (CIHR) - "Evaluation of Handheld Technology Usability"

PROJECT TITLE:
Evaluation of Handheld Technology Usability

Expiry Date - Approval of an amendment does not change the expiry date on the current UBC BREB approval of this study. An application for renewal is required on or before: February 5, 2008

AMENDMENT(S):	AMENDMENT APPROVAL DATE: June 26, 2007

Document Name	Version	Date
Consent Forms:		
Main study consent	003	June 21, 2007

The amendment(s) and the document(s) listed above have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board and signed electronically by one of the following:

Dr. Peter Suedfeld, Chair
Dr. Jim Rupert, Associate Chair
Dr. Arminee Kazanjian, Associate Chair
Dr. M. Judith Lynam, Associate Chair
Dr. Laurie Ford, Associate Chair

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road,
Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL - MINIMAL RISK AMENDMENT

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC BREB NUMBER:
Joanna McGrenere	UBC/Science/Computer Science	H05-81017

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Vancouver (excludes UBC Hospital)

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):
Karyn Moffatt

SPONSORING AGENCIES:
Canadian Institutes of Health Research (CIHR) - "Evaluation of Handheld Technology Usability"

PROJECT TITLE:
Evaluation of Handheld Technology Usability

Expiry Date - Approval of an amendment does not change the expiry date on the current UBC BREB approval of this study. An application for renewal is required on or before: February 5, 2008

AMENDMENT(S):	AMENDMENT APPROVAL DATE: October 19, 2007

Document Name	Version	Date
Consent Forms:		
Main study consent	004	October 16, 2007

The amendment(s) and the document(s) listed above have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road,
Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL- MINIMAL RISK RENEWAL

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC BREB NUMBER:
Joanna McGrenere	UBC/Science/Computer Science	H05-81017

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Vancouver (excludes UBC Hospital)

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):
Karyn Moffatt

SPONSORING AGENCIES:
Canadian Institutes of Health Research (CIHR) - "Evaluation of Handheld Technology Usability"

PROJECT TITLE:
Evaluation of Handheld Technology Usability

EXPIRY DATE OF THIS APPROVAL: January 8, 2009

APPROVAL DATE: January 8, 2008

The Annual Renewal for Study have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road,
Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL - MINIMAL RISK AMENDMENT

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC BREB NUMBER:
Joanna McGrenere	UBC/Science/Computer Science	H05-81017

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Vancouver (excludes UBC Hospital)

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):
Karyn Moffatt

SPONSORING AGENCIES:
Canadian Institutes of Health Research (CIHR) - "Evaluation of Handheld Technology Usability"

PROJECT TITLE:
Evaluation of Handheld Technology Usability

Expiry Date - Approval of an amendment does not change the expiry date on the current UBC BREB approval of this study. An application for renewal is required on or before: January 8, 2009

AMENDMENT(S):

AMENDMENT APPROVAL DATE:
May 27, 2008

Document Name	Version	Date
Consent Forms:		
Main study consent	005	May 27, 2008

The amendment(s) and the document(s) listed above have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects

Approval is issued on behalf of the Behavioural Research Ethics Board
Dr. M. Judith Lynam, Chair
Dr. Ken Craig, Chair
Dr. Jim Rupert, Associate Chair
Dr. Laurie Ford, Associate Chair
Dr. Daniel Salhani, Associate Chair
Dr. Anita Ho, Associate Chair

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road,
Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL- MINIMAL RISK RENEWAL

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC BREB NUMBER:
Joanna McGrenere	UBC/Science/Computer Science	H05-81017

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Vancouver (excludes UBC Hospital)

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):
Karyn Moffatt

SPONSORING AGENCIES:
Canadian Institutes of Health Research (CIHR) - "Evaluation of Handheld Technology Usability"

PROJECT TITLE:
Evaluation of Handheld Technology Usability

EXPIRY DATE OF THIS APPROVAL: November 24, 2009

APPROVAL DATE: November 24, 2008

The Annual Renewal for Study have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board

Dr. M. Judith Lynam, Chair
Dr. Ken Craig, Chair
Dr. Jim Rupert, Associate Chair
Dr. Laurie Ford, Associate Chair
Dr. Daniel Salhani, Associate Chair
Dr. Anita Ho, Associate Chair

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road,
Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL - MINIMAL RISK AMENDMENT

PRINCIPAL INVESTIGATOR:	**DEPARTMENT:**	**UBC BREB NUMBER:**
Joanna McGrenere	UBC/Science/Computer Science	H05-81017

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Vancouver (excludes UBC Hospital)

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):
Karyn Moffatt

SPONSORING AGENCIES:
Canadian Institutes of Health Research (CIHR) - "Evaluation of Handheld Technology Usability"

PROJECT TITLE:
Evaluation of Handheld Technology Usability

Expiry Date - Approval of an amendment does not change the expiry date on the current UBC BREB approval of this study. An application for renewal is required on or before: November 24, 2009

AMENDMENT(S):	**AMENDMENT APPROVAL DATE:** **May 11, 2009**

Document Name	Version	Date
Consent Forms:		
Main study consent	006	April 22, 2009

The amendment(s) and the document(s) listed above have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board and signed electronically by one of the following:

Dr. M. Judith Lynam, Chair
Dr. Ken Craig, Chair
Dr. Jim Rupert, Associate Chair
Dr. Laurie Ford, Associate Chair
Dr. Anita Ho, Associate Chair