Scrolling in Radiology Image Stacks

Multimodal Annotations and Diversifying Control Mobility

by

Louise Carolyn Oram

Bachelor of Human Kinetics, UBC, 2007
Bachelor of Computer Science, UBC, 2011

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

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
(Computer Science)

The University of British Columbia
(Vancouver)

October 2013

Abstract

Advances in image acquisition technology mean that radiologists today must examine thousands of images to make a diagnosis. However, the physical interactions performed to view these images are repetitive and not specialized to the task. Additionally, automatic and/or radiologist-generated annotations may impact how radiologists scroll through image stacks as they review areas of interest. We analyzed manual aspects of this work by observing and/or interviewing 19 radiologists; stack scrolling dominated the resulting task examples.

We used a simplified stack seeded with correct or incorrect annotations in our experiment on lay users. The experiment investigated the impact of four scrolling techniques: traditional scrollwheel, click+drag, sliding-touch and tilting to access rate control. We also examined the effect of visual vs. haptic annotation cues' on scrolling dynamics, detection accuracy and subjective factors. Scrollwheel was the fastest scrolling technique overall for our lay participants. Combined visual and haptic annotation highlights increased the speed of target-finding in comparison to either modality alone.

Multimodal annotations may be useful in radiology image interpretation; users are heavily visually loaded, and there is background noise in the hospital environment. From interviews with radiologists, we see that they are receptive to a mouse that they can use to map different movements to interactions with images as an alternative to the standard mouse usually provided with their workstation.

Preface

Initial observation of radiologists occurred as a part of a class project (CS 543), in collaboration with Jeremy Kooyman and Florin Gheorghe (both Masters students from the department of Electrical Engineering). We obtained permission from Dr Bruce Forster to observe and talk to radiologists at the UBC hospital.

During a prototyping attempt, Evelyn Tsai worked on implementing gesture recognition for a particular type of touch surface; this work was done as an engineering report for EECE 597.

The study described in Chapter 6 of this thesis, as well as the questionnaire administered to radiologists in Chapter 4, were conducted with the approval of the University of British Columbia (UBC) Behavioural Research Ethics Board (BREB) and the Vancouver Coastal Health Research Institute (VCHRI), under certificate number H12-01672.

The majority of the work and writing in this thesis, at the time of this writing, is under review as a conference paper. The co-authors are my supervisors Dr Karon MacLean, Dr Philippe Kruchten and a radiology collaborator from the UBC hospital Dr Bruce Forster.

Table of contents

List of tables

List of figures

Glossary

FSR Force Sensing Resistor

UBC University of British Columbia

VCHRI Vancouver Coastal Health Research Institute

MRI Magnetic Resonance Imaging

CT Computed Tomography

PACS Picture Archiving and Communication System, workstation, software, and network for image storage and retrieval according to industry standards

DICOM Digital Imaging and Communications in Medicine, a standard for handling, storing, printing and transmitting information in medical imaging

ANOVA Analysis of Variance, a set of statistical techniques to identify sources of variability between groups

CAD Computer Aided Detection or Computer Automated Detection; also Computer Aided Diagnosis: where the radiologist uses the detected areas to help them reach a diagnosis

POLYMORPH a low melting-temperature polymer, used to create imprecise plastic parts for prototypes

Acknowledgments

Firstly I would like to acknowledge my supervisors Dr. Karon MacLean and Dr. Philippe Kruchten, I am grateful that I got to work with both of them. Karon has been supportive of my research since I started as an undergraduate in her lab, and I would not have completed this Masters were it not for her. Philippe has been a stabilizing influence throughout my Masters.

I am grateful to Dr. Joanna McGrenere, who was the second reader for this thesis, and was always helpful and encouraging when I spoke with her. I would also like to thank Cliff Edwards, my contact at McKesson, for his boundless positivity.

Thank you to all of the radiologists who let me observe them and ask questions; I would not have been able to do this research without their time and input. Specifically, I would like to thank Dr. Bruce Forster, my collaborator at UBC hospital, who was supportive of my research and an endless source of knowledge on radiology.

I am thankful for the support and help of my many labmates, particularly Jessica and Juliette as we have been together since our initial research experience as undergraduates.

Finally, I thank my partner Thomas for helping to keep me sane, and my parents and brother for their support. Also I am thankful to my friends who listened to me ramble on about my research, specifically Meghan who proofread my thesis and Rachel who gave statistics advice.

Funding for this research was provided by SurfNet (from NSERC).

Chapter 1

Introduction

To utilize the detailed information provided by today's high-resolution image capture technologies, such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), radiologists must examine ever-larger image sets. It is not uncommon for multi-trauma CT scans or coronary CT angiograms to have data sets of 4000 images [5]. Diagnosis entails a complex, time-pressured visual search task, where target conspicuity, background clutter and other attentional factors can influence the radiologist's ability to detect anomalies [5], and radiologists are put at substantial risk of repetitive strain injury [16].

Radiology images are currently viewed as single 2D slices [5] [6], arranged in a stack through which the user scrolls depthwise. The main interaction tool, a scrollwheel mouse, has not changed since 1995[1]. In contrast to singleton x-ray images, MRI and CT image stacks are continuous media streams. Efficient perusal demands fluid, controllable interaction akin to video scrubbing [25], as has been demonstrated with a haptic scrollwheel [30]. A conventional mouse's restricted input mobility (x-y hand movement and finger-level scrollwheel movement) has limited ability to support this variation and level of control.

The daunting scope of the image-viewing task makes it a candidate for semi automation, for instance Computer Aided Detection (CAD) of anomalies in images [13]. Such algorithms are tuned to find all real anomalies (true positives), at the cost of substantial rates of false positives, which radiologists must then distinguish. CAD automatically creates marks in the stack (referred to by radiologists as 'CAD marks'), so the radiologist can review the algorithm detections. The accuracy and number of CAD identified objects varies. Since it takes 5-7 seconds for a radiologist to re-evaluate a CAD-identified nodule [28], there is a cost to the potential time and accuracy gains from CAD. Similar issues exist for stack metadata generated from other sources. One important example are those created by other radiologists (known as 'annotations') in redundant procedures and peer reviews or training reviews. For instance, in a recent situation where a doubt as to quality of care arose, a review of a radiologist's work was conducted[2].

---

[1] http://en.wikipedia.org/wiki/Scroll_wheel

[2] http://www.cbc.ca/news/health/toronto-radiologist-s-3-500-ct-scans-mammograms-reviewed-1.1701023

The use of stack metadata, in particular CAD marks, can affect detection accuracy [2] [13]. Context bias is a concern when the radiologists' diagnostic sensitivity depends on expected prevalence of a given anomaly [15]; as well as, automation bias where CAD misses particular cancer types. Learned dependency could also lead the user to miss important findings.

How might alternative annotation presentation affect bias? CAD data is currently presented as visual highlights, which may be more likely than another modality to influence what the radiologist sees at perceptual, attentional and strategic levels. Integrated with care, haptic highlights might avoid an identified risk of degrading the decision process through simple sensory overload [24]: since the visual system is already heavily loaded and the hospital environment is noisy.

Neither adding a specialized device nor compromising familiar mouse functions are likely to be accepted by radiologists. They heavily use other manual tools, such as keyboards and dictaphones, and oscillate swiftly between graphical user interface (GUI) pointing and stack strolling. Proprietary data systems enforce device standards. A conventional two-dimensional (x-y) computer mouse is best for pointing [16]; its ease of use and familiarity make it favoured relative to alternative input devices in this setting (e.g. [29]).

1.1 Contributions

In this work, we aim to streamline the radiology image-scrolling task. We investigate whether alternatives in the user's input mobility can improve stack navigation. By **mobility**, we mean the ways in which the user's hand can move, as it interacts with the input device – and in particular, finger/hand movements for scrolling control. We also explore how the modality of annotation display impacts signal detection patterns; as well as the degree and kind of change in input device radiologists will tolerate, and indeed welcome.

After analyzing 19 radiologists' work via observation and/or interviews, we prototyped augmentations to the standard mouse (Figure 5.1 and Figure 6.7) which we hypothesized could support (a) more efficient image scrolling (with more fluid interaction) and (b) attentionally improve annotation display (using the haptic modality). We obtained qualitative feedback from our radiologists on these prototypes and the interactive techniques they support; and examined the impact of interaction and display on detection rates in a controlled, abstracted study with non-radiologists. We contribute

- a set of verified task examples that capture the most important radiology image interactions
- prototypes representing a set of novel scrolling inputs
- an abstracted task paradigm suitable for screening scrolling-type and annotation-modality candidates on lay users

- quantitative data on detection accuracy and subjective reactions to scrolling-type and annotation-modality
- a proposal for how haptics can increase the effectiveness of reviewing annotated image stacks

1.2 Thesis Overview

This thesis is divided into 8 chapters. Chapter 2 contains an overview of the related work. Chapter 3 goes over the initial observation of radiologists and some early prototype attempts. Chapter 4 discusses the creation and validation of 6 radiology task examples. Chapter 5 explains the prototypes used in the experiment and the design space they cover. Chapter 6 outlines the abstracted task, design and protocol used for the experiment and presents the hypotheses. Chapter 6 continues with the results of the experiment and discuss these based on the hypotheses. Chapter 7 discusses informal feedback from radiologists after they were presented with a more cohesive prototype. Chapter 8 ends the thesis with a discussion, including ideas for future work.

Chapter 2

Related Work

This section discusses the literature that is most important to the work in this thesis. Radiology specific literature, as well as human computer interaction literature, is discussed.

2.1 The Radiologist's Work Environment and Constraints

To view images, radiologists use two or three high-resolution LCD monitors; a conventional scroll-wheel mouse for stack navigation and GUI navigation; and keyboard and dictaphone to transcribe diagnoses (Figure 2.1). Data is provided via a Picture Archiving and Communication System (PACS)[1] which consists of a workstation, software and network for image storage and retrieval according to industry standards. PACS are sourced by health authorities as major capital investments from a small number of medical imaging vendors, and have proprietary elements.

2.1.1 Fatigue, Errors and Ergonomics

Errors in diagnosis are of concern for any medical system, and can be caused by fatigue and errors in perception. Fatigue occurs over the workday and has been seen to cause a small drop in detection of fractures in radiographic images between early and late in the workday [21]. Subjective radiologist feedback related this performance drop to increased physical discomfort, sleepiness, lack of energy and oculomotor strain; the eyestrain effect is worse for CT images [33]. Three causes of false-negative type perception errors have been observed [22]: never fully fixating on the object, not fixating long enough to recognize the object and fixating but not recognizing or consciously dismissing the object.

Radiologists also experience a high rate of repetitive stress injuries from prolonged use of the workstation [8]; the mouse was not designed for continuous use and, therefore, can cause repetitive stress injuries [16]. Innovations allowing images to be read more quickly without reducing accuracy or thoroughness, while improving ergonomics at the same time, would increase efficiency and reduce costs.

---

[1] http://en.wikipedia.org/wiki/Picture_archiving_and_communication_system

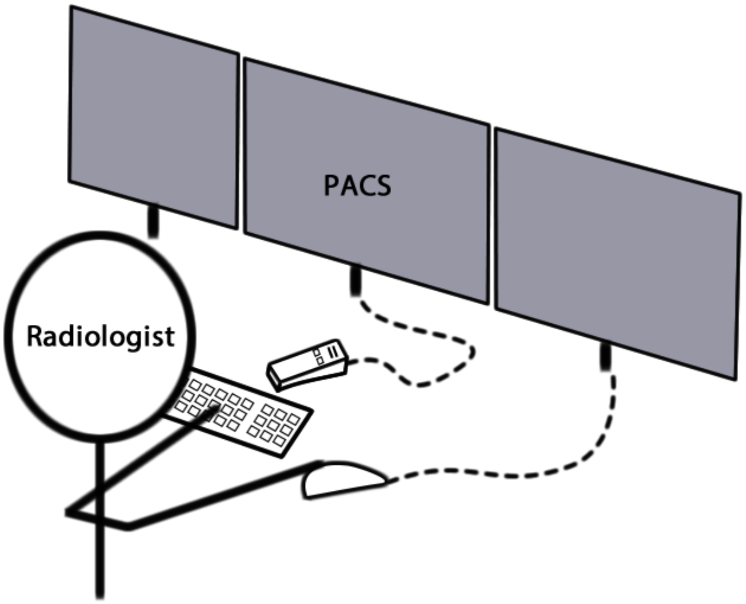


Figure 2.1: Sketch showing a radiologist, their displays, keyboard, mouse, and dictaphone.

2.2 Viewing Images by Scrolling

Scrolling is integral to image review. CT image consumption is faster when viewed in a stack rather than as tiles, where multiple images are visible at once, probably due to eased perception of 3D structures [26]. Radiologists must scroll at different speeds, stop and reverse to compare or examine locations. They are trained to review specific anatomical structures, and make successive passes focusing on each of these in turn.

PACS workstations typically support two scrolling techniques: scrollwheel or click+drag. **Scrollwheel** employs the wheel on the mouse, pairing each detent on the wheel to movement up/down an image in the stack. In **Click+drag** the user clicks on the left mouse button then drags the mouse; movement in the y-direction (along the table surface) maps to moving up/down in the stack. Both employ position control; scrolling distance is proportional to the distance moved by the mouse or angle traversed by scrollwheel. Atkins et al. compared scrollwheel and click+drag techniques to a jogwheel (a rate control device, where scrolling rate is proportional to the displacement of the sprung wheel off its central position), and found that most radiologists preferred the more familiar position control even though some were faster with rate control [6]. Relative movement rates were generally fastest for the wheel/click+drag combination, slowest with wheel alone and in between for jogwheel [6]. Sherbondy et al. used a tablet and stylus for scrolling, and found that position was faster than rate control for finding a target in a CT stack [29].

Fitts' Law models scrolling for trajectories where the target location is known, for techniques including rate control, scrollwheel and wheel with acceleration [18]. This model follows a logarithmic function dependent on distance to the target. When visual target search is required (e.g. reviewing a CAD-marked stack, where the marked slice is not visible until it has been reached), scrolling time depends linearly on distance to target [4]. In this context *'distance'* refers to the number of slices that must be traversed. In other words, the user does not know which slice, if any, will hold the target, and cannot generate a ballistic trajectory to optimize speed of approach.

2.3 Beyond the Mouse, and Direct-Touch Sensing

Multi-touch sensing has become a ubiquitous manual control, and is explored in many interaction contexts outside radiology. In an early mouse example, Hinkley et al. explored touch sensing near the scrollwheel, and found it a useful discrete scroll alternative to the wheel, for instance: tapping to page up/down [17]. Villar et al. considered multi-touch in five diverse mouse form factors [34]. They found it could extend the control degrees of freedom and support different input modes [34], mitigating the need to switch between devices. They advised locating touch-sensed areas in easy reach of one hand posture, and visibly cueing their location.

Flying mice and other tracked devices can be lifted above the table surface. Direct mapping to a 3D space makes them intuitive, and easy to learn [38]; fatigue in maintaining cursor persistence makes them unsuitable for radiology interaction.

A pen and tablet solution showed shorter times for a radiology task of outlining a region of interest relative to a mouse [12]. However, switching between different devices may hinder radiologists' workflow, as they still use mouse interactions for most other tasks. Another alternative, direct-touch, reduces the need for device switching but has occlusion issues [35] and fatigue from unsupported hands [36].

Other desk-supported variants have diversified interaction. The Rockin' Mouse adds a fourth degree of freedom; it is faster than a normal mouse in 3D [7], but scrolling was not studied. Many other control movements (i.e., extended mobility, as per our definition) could be used with a mouse-like device, but have not been explored in the radiology setting.

2.4 Haptic Feedback in Support of Scrolling

Akamatsu et al. found that for a pointing task with a mouse, tactile feedback (a pin pushing up into the fingerpad when the mouse is over the target) was quickest, and no feedback was slowest for final positioning times [1]. Levesque et al. saw variable friction feedback speed up target selection on a touch screen [23]. Oakley et al. tested an interaction where the user tilted a mobile device to scroll.

This interaction was augmented so that the user felt a vibrotactile buzz when they transitioned to the next item on the list. Vibrotactile feedback lowered task completion time, and position control was faster than rate control [27].

These results suggest that haptic feedback on possible targets will give modest performance gains even if the system does not know where the user is heading (scrolling time linearly dependent on the distance to target). The prevalence of detents on the mouse's scrollwheel in the radiology setting ([5], [6]), where radiologists have some choice as to the commodity mouse they plug into their workstation, indicate radiologists may be receptive to this.

Another form of input or output is pressure. Kim et al. presented an inflatable mouse that could give haptic feedback by inflating/shrinking quickly. Users tried scrolling with pressure, but found that the physical interaction (a combination of clicking and pressing) did not work well with the current form factor, and tasks that required high pressure caused fatigue [19]. It was suggested that this is because clicking requires stabilization from the other fingers, which affects the pressure input. On a hard surfaced mouse with Force Sensing Resistor (FSR)'s, it was found that users could control a limited number of pressure levels with one pressure sensitive location [9].

2.5 Computer Aided Detection (CAD)

Most CAD research focuses on validating that CAD information, provided as visual image annotations, improves radiologist detection sensitivity and/or speed [13]. Computer aided diagnosis (occurring after the algorithm is run for detection) is generally defined as a diagnosis made by a physician who takes into account the computer output given, based on quantitative analysis of radiological images [13].

However, annotations overlaid on the stack affect what radiologists see. Even when biased towards finding everything, CAD misses around 20% [13] and leads to automation bias. Radiologists attending to annotated areas are more likely to miss artifacts not found by the CAD. Alberdi et al. found a lower detection rate for users given CAD information in comparison to those who were not; here, the largest difference was seen in cancers not found by CAD. They hypothesized a bias effect, where users calibrate to the expected prevalence of cancers and expected proportion of cancers missed by CAD in the current data set [2]. Additionally, a criticism of many CAD studies is that the data sets used contain an unrealistic proportion of cancers, and radiologists know this [2]. We have not seen studies that modified how CAD annotations are displayed; yet this may help mitigate the detection bias that CAD produces.

Rubin et al. [28] saw that CAD had a significantly higher sensitivity to finding lesions missed by a first human reader in comparison to a second human reader. However, this comparison posits unrealistically that the user of the CAD annotations would accept all true positives and reject all false positive CAD detections.

Multiple studies (see review article: [13]) have shown that the computer output improves radiologists diagnostic accuracy. For example, there is a high false positive rate when looking for malignant pulmonary nodules in high resolution CT; when the radiologists were presented with computer feedback their performance improved. It was hypothesized, from analyzing the responses, that this is because the radiologists were able to maintain their firm correct decisions and correct their initial mistakes using CAD [13].

In low-dose CT images, a CAD scheme detected 83% of lung nodule cancers (in stacks with on average 1-2 nodules), with an average of 5.8 false positives per scan [13]. Another scheme (run on different scans, containing some potentially more subtle cancers) detected 80%, with 2.7 false positives per scan.

In our experiment, we therefore manipulate annotation display assuming a detection ratio of 80% to align with current CAD performance.

Chapter 3

Observation and Prototyping

Initial observation of radiologists took place at University of British Columbia (UBC) hospital.

3.1 Observation and Rapid Prototyping

We observed radiologists to understand the interactions they performed on images with the PACS. Figure 3.1 illustrates some conceptual relationships between common PACS interactions we observed; it segregates the interactions into distinct or fluid interactions, and further specifies which are multi-image (vs single-image). At this stage we observed 10 radiologists, for a total of around 18 hours of observation as a part of a class project [20]. We watched over the radiologists' shoulders to see what occurred on screen, as well as watch their hand movements on the mouse. We took notes on what we observed, and asked questions about their preferences and ways they performed interactions (taking care to ask these at a time that did not disrupt their workflow). From the observation and questions, we gained an understanding of the interactions that radiologists performed.

Then we made simple rapid prototypes in which a single idea for interaction was presented per prototype. Our general approach in these prototypes was to increase the mobility of the radiologist's input device. An initial physical prototype can be seen in Figure 3.2. This prototype was created to fit under the hand like a mouse, but to have two perpendicular surfaces (hypothetical touch surfaces) and be able to rock back-and-forth and side-to-side. Blog postings created for the class can be seen at: http://interactiondesignmusings.tumblr.com/.

Later prototypes interacted with the computer via a Java program that could open a Digital Imaging and Communications in Medicine (DICOM) file format image stack and scroll through it as well as several other interactions, including window/level (a.k.a contrast/brightness). After several iterations of discussing the prototypes with radiologists we created a prototype that integrated the viable ideas into one prototype. This prototype rocked both back-and-forth and side-to-side in a silicone base, had a 2D slider joystick on the side that was set to adjust brightness/contrast, and had an ipod touch on the top for gesture interactions [20]. Figure 3.3 shows the prototype attached to a laptop, and ongoing discussion with radiologists about the interactions.

In this project, we implemented simple interaction ideas: using an accelerometer to switch the

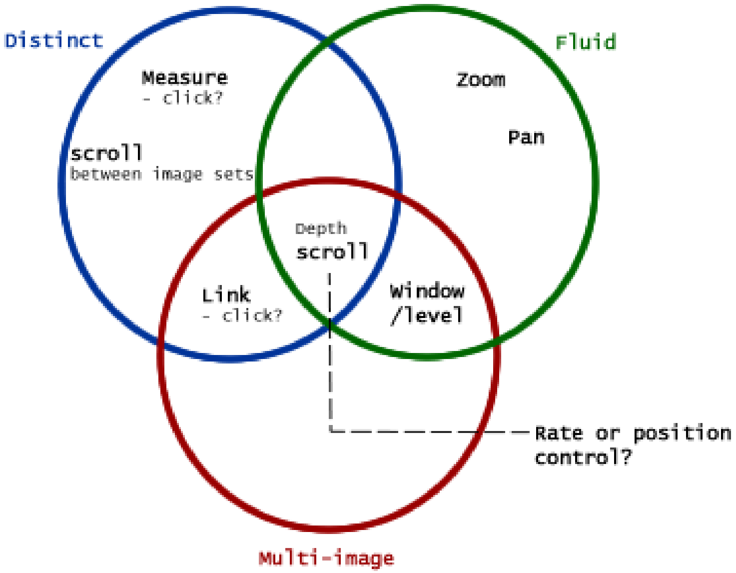


Figure 3.1: Sketch showing how several of the PACS interactions can be categorized.

Figure 3.2: Early physical prototype: created out of POLYMORPH, knitted wool and stuffing.

plane in which the image is being viewed; using pressure sensed via FSRs in an early tilting prototype; and using a 2D sliding joystick to map to contrast/brightness. From presenting these prototypes to radiologists we started to learn which might be the most promising avenues for exploration, as well as how receptive the radiologists were to the different interaction techniques. However, we still needed to validate which interactions would actually speed up their workflow.

3.2 Subsequent Prototyping

Wanting to move forward with the idea of touch surfaces on a mouse-like device, but dissatisfied with the responsiveness of the more easily available touch surfaces (such as the Nintendo DS), or


[redacted]

Figure 3.3: Showing our final prototype (attached to a laptop) to radiologists at UBC hospital.

their form factors (flat), we investigated alternative touch surfaces. Generally, we felt that there was little available that was easy to prototype with.

3.2.1 Force Sensing Resistor Matrix

We obtained an FSR matrix from Sensitronics. An FSR matrix is a type of touch surface, and is normally integrated into much larger (and often industrial) systems. An FSR matrix is essentially many FSR's in parallel; 10 by 16 lines make the matrix grid. One set of these lines is attached to digital output pins, allowing their power to be switched on/off; the other set of lines is attached to analog input pins allowing their value to be read. This was initially scanned from an easily available Arduino Uno. However, the Arduino Uno was slow because it did not have enough analog inputs and therefore required multiplexers. Then we tried a Beaglebone; it runs Linux and could potentially run gesture recognition and send only the recognized gestures over the serial port instead of the raw data. However, the way that the Beaglebone accesses the input/output pins proved slow. Finally, we settled on an Arduino Due because of its many analog input/output pins (Figure 3.4), and the speed at which it can read/write to them.

In order to scan for fingers touching the FSR surface, one digital output at at time is switched on (in order to power that line). The analog inputs are then individually scanned to see if any power came through. If there is a reading above 0 on an analog pin then there is pressure (touch) at the crossover point between the two lines. This process is done repeatedly, and as fast as possible, across the whole board. See Appendix A.4.1 for the code that runs on the Arduino Due for this process.

Initially, we worked with Java program to parse the Arduino serial stream. Because of the choice of the language Java, we had problems with the speed and consistency of polling the serial port so

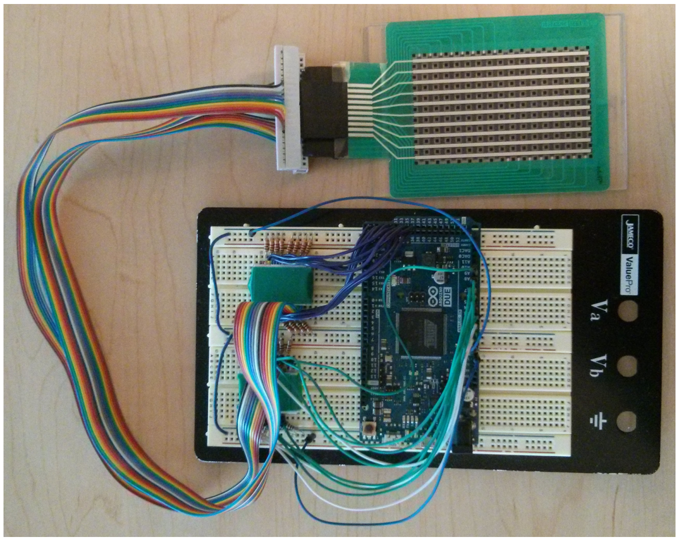


Figure 3.4: FSR matrix setup shows the ribbon cables connecting the FSR matrix to the breadboard where simple circuitry (primarily pulldown resistors) then connects it to the Arduino Due digital or analog pins.

subsequently moved to C++. Full featured gesture recognition for the FSR matrix proved difficult to implement, likely because of the difficulty in parsing/separating the finger positions [32]. We used an open-source gesture recognition algorithm (the 1$ recognizer [37]) with some success. It could recognize gestures such as clockwise circle, counterclockwise circle and line swipes. The scope and performance were not sufficient to encourage us to continue with this technology at the time.

Creating a stable system with the FSR matrix and gesture recognition would have taken too much time, and creating a product easily usable for prototyping was far beyond the scope of this thesis. Additionally, we did not have the expertise needed to solve the problem of identifying and tracking separate fingers.

If this setup had been stable, it would have allowed for a curved (in one axis) surface where we could setup the gestures we thought would be useful to radiologists. Giving the radiologists access to gestures would increase diversify the mobility of their input device. The plan was to put the FSR matrix on a mouse-like device. However, it was not stable enough for experimentation, so this was never used in a full prototype.

3.3 Next Steps

From the above we gained an understanding of the interactions radiologists perform, and the some information about the pain points in those interactions. We saw how the radiologists reacted to numerous simple prototypes, and found that they were generally interested in new ideas for interaction methods, often coming up with their own ideas. We learned that it is difficult to parse finger positions, and that if we needed a stable touch surface we would likely need to use something fully pre-fabricated (with both the hardware and software already integrated).

The observation and prototyping described here had elements of participatory design, in which we listened to the radiologists ideas and came back with modified prototypes. However, it was not structured in documenting the radiologist interactions. Moving forward we wanted to formalize our understanding of the interactions that radiologists perform.

Chapter 4

Task Analysis

We analyzed the physical interactive elements of the radiologists' workflow in a two-staged process. The initial observation, as described in the previous chapter, was continued until we reached a point of saturation in which we were largely seeing the same types of interactions and tasks again. We then chose the most promising subset of tasks to support, and validated them through a questionnaire.

We wanted to break down the basic components and goals of radiology image interaction tasks, then confirm that these tasks were important to radiologists. The goal of analyzing the physical interactive elements was to understand if it would be beneficial to diversify the mobility of the input device. We wanted to gain an understanding of what movements were most repetitive, and what image manipulations mapped to what movements.

4.1 Task Example Creation

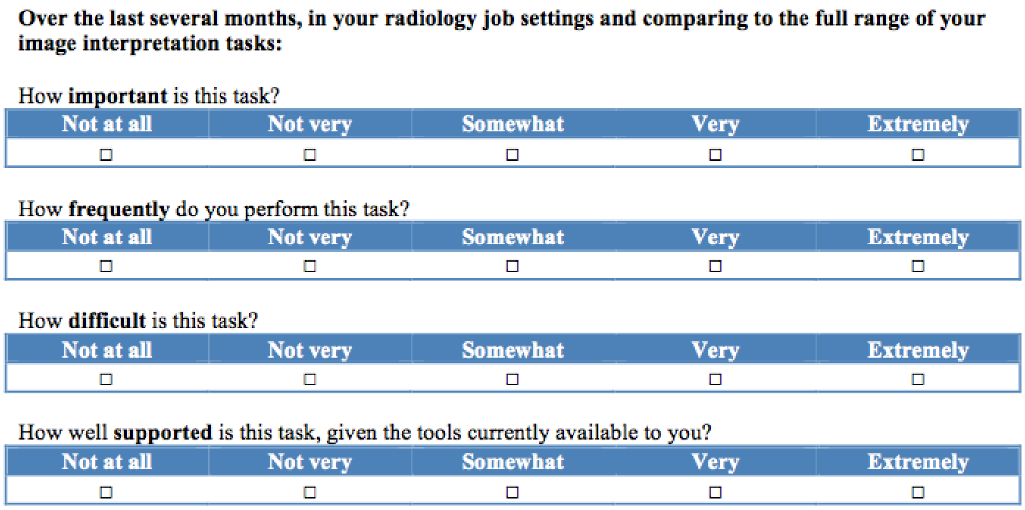
We informally observed and interviewed 12 radiologists (1 female) within a variety of work settings, over 1 to 3 sessions in blocks of 30 minutes to 1 hour. The total amount of time spent observing was around 35 hours (including the observation described in the previous chapter). The radiologists had many suggestions for PACS software improvements, a topic out of our scope, as well as for physical image interaction. We noticed some disparities between observation and self-report in activities (e.g. percentage of scrollwheel vs. click+drag use), which may point to subjective importance. We captured this domain-expert input in a set of task examples, which are described in Table 4.1.

4.2 Task Example Validation

To verify that our task examples faithfully represented the most important elements of radiology image interaction, we took these task examples more formally to a set of 10 radiologists (8 male; including 3 from the original 12), recruited by email from hospital administration and word of mouth. Our participants had experienced a variety of work settings in Vancouver, including an academic hospital radiology department, private lab and city hospital emergency room. They had also experienced many professional roles throughout their careers, such as interventional radiology,

Table 4.1: The 6 task examples, as used in questionnaires given to radiologists.

1. Identifying or finding a specific piece of anatomy The radiologist looks for an object or area of interest in one anatomical plane, looking through several slices to find and properly identify it. If unsure, or things are unusual, then they may look at the area in another plane (or several other planes if they are available). They can cross-reference a point between different planes, to see the location in other planes. Additionally, they may adjust the window/level to get better contrast between the object and its surrounds.
2. Defining the edge / size of something The radiologist may want to know the size of an object, or if it is encroaching on the area of other anatomy. Window/level may be used to get better contrast of the object to its surrounds. After looking at the object in several planes, they choose a specific image, or multiple images, to outline, circle, or measure the diameter of the object.
3. Tracking / connecting objects The radiologist follows a part of the anatomy through several slices to check for abnormalities. The radiologist moves back and forth through the image slices while watching the area of interest. If they feel they have missed something, or loose track of the object they may slow down and watch more carefully for a subset of the image slices. This is repeated as many times as needed for different anatomical parts, usually by organ system but sometimes by area (such as in the brain).
4. Comparing two images (old and new) The goal is to look for interval change: differences between the sets of image. Do new objects appear, have old objects enlarged? The radiologist brings up both sets of diagnostic images and looks at the same plane and area in each image side by side. They scroll back and forth in each set of images, comparing the areas of interest (can link the two images so they scroll together, but the slices may not land at exactly the same spots). They may re-measure objects that were found in the first diagnostic to see if they have changed in size.
5. Identifying the makeup of something The radiologist may want to know what something abnormal is composed of. They look at the item in several planes, and see the attenuation of the item. They may adjust the window/level to get the best contrast with the surrounds, or to see colour differences within the object. To know the density of the item from the imaging they can select part or all of it and see the density number.
6. Getting a second opinion If the radiologist is unsure of something, less familiar with it, or finds something unusual, they may ask the opinion of another radiologist. Another option is to look up papers on the topic to help confirm the diagnosis or learn about more nuanced aspects they cannot remember off the top of their head.



Over the last several months, in your radiology job settings and comparing to the full range of your image interpretation tasks:

How **important** is this task?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

How **frequently** do you perform this task?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

How **difficult** is this task?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

How well **supported** is this task, given the tools currently available to you?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

Figure 4.1: The questionnaire used for each task example.

diagnostics, neuroradiology etc. On-job experience ranged from 0 to 31 years (avg. 12.7). All were familiar with touch devices and owned and/or often used one. Six reported ergonomic issues from extended PACS use, including shoulder pain, eyestrain and repetitive use of the scrollwheel.

In 15-minute sessions at their workplace, volunteers read the task examples, answered a questionnaire, and were interviewed. They were given a $10 gift certificate to Starbucks in appreciation for their time.

Questionnaire: Four 5-point Likert scale questions (Figure 4.1), repeated for each task example, asked how important, frequent, difficult and well-supported that task was. Figure 4.2 summarizes questionnaire responses. The x-axis of the stacked bar charts is the number of radiologists (N = 10), while the y-axis is the task example number. These results indicate that the respondents considered most task examples were important and frequently performed and, to a lesser extent, they were felt to be difficult and well-supported.

We voice-recorded discussion of a set of open-ended questions. Participants were asked to identify the following:

- Missing tasks which they find important, frequent or difficult.
- Tasks which are well or poorly supported by PACS they have used (many had experience with different PACS brands).
- Mouse interactions they found tiring or repetitive.
- Any issues with repetitive-strain injuries.

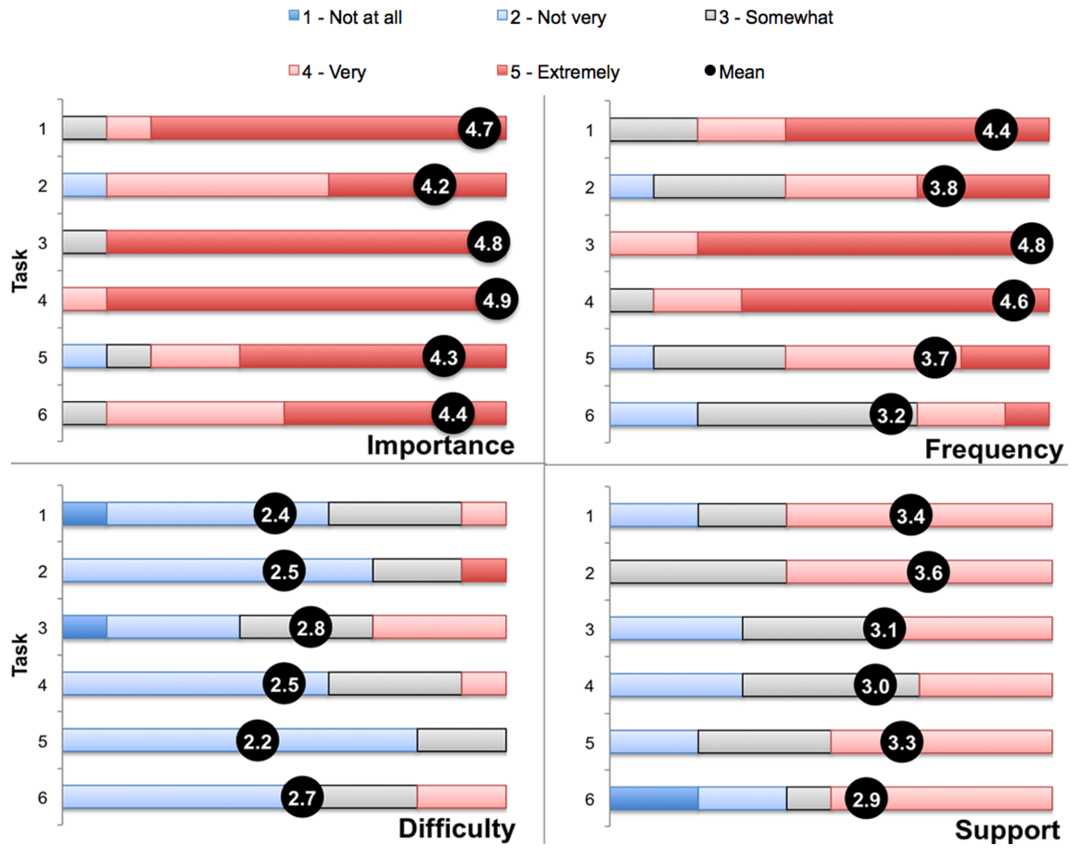


Figure 4.2: A stacked bar chart of Likert responses, with the task example # on the Y-axis (N=10).

The open-ended questions for the beginning and the end of the interview can be found in Appendix A.3.1.

4.2.1 Importance

Each of the 6 tasks was rated as very or extremely important by at least 8 out of 10 participants. Participant P1 summarized that *"they all seem extremely important to me"*. Participants either said no important tasks had been overlooked (2 participants), or gave examples of very specialized or specific tasks (8 participants).

4.2.2 Frequency and Repetitiveness

Tasks 1 to 5 were labeled very or extremely frequent by 6 to 10 participants, with Tasks 3 and 4 rated highest. Task 6 was less frequent, but of high importance. We note that area of specialization is likely to play a role in these assessments.

Participants verbally identified the most repetitive task as scrolling: *"When you are looking at [a] CT that has 350 images in it, and you are looking at every image, that takes a lot of scrolling up and down"* (P7). Participant P6 noted that scrolling is very mouse-intensive and, therefore, a way to end up with an injury. P6 then suggested having a way *"to scroll through a large amount of data sets with minimal hand motion."*

On scrolling and speed, P2 commented: *"I use scroll-wheel way more often than the drag stuff"*. When asked if it was hard to go fast enough, P2 replied, *"Yeah... but it's too hard to go slow enough with the click+drag... something in between, so if you had a dual function?"*

4.2.3 Difficulty

Generally, task difficulty arose from diagnostic complexity, for instance *"when there is complex anatomy, complex disease processes"* (P7); or ambiguity: *"to know what is normal, or what is in the range of normal, or where is starts to be abnormal or pathologic"* (P8).

Discussion resolved the potential ambiguity of responses indicating both low-difficulty and low-support (see Figure 4.2): radiologists have figured out ways to perform necessary tasks, accommodating non-optimal support, and no longer find them difficult; but still wish for better support.

4.2.4 Device Interaction

Participants suggested a range of device improvements, with many relating to functional specificity: *"I would prefer to have more buttons, with less functionality per button"* (P5). This radiologist is suggesting adding extra buttons on the mouse in order to increase functional specificity. P1 mentioned speed interfering with functional mapping; he is a rapid clicker, and double clicking was mapped to a function that he does not usually mean to invoke. P2 had even considered adding his own accessory to the PACS workstation: *"At one point I was considering getting a gaming accessory pad... so you could mouse or move over to the pad."* In general, these quotes support the idea of increasing the mobility of the input device.

4.2.5 Other Findings

Two participants brought up distraction issues, in both positive and negative ways. P1 said he had more repetitive strain problems at home because he sat in one position longer, while *"at work we are interrupted a lot, which in one sense is irritating, but in another sense is good."*

3D image interaction came up, with P4 saying that other than for 'flying' through the colon *"we generally think that 3D is overrated."* P7 mentioned that while they did not generally use it directly in diagnostics, it was a helpful alternate view.

4.3 Summary

Scrolling is an essential and frequent part of radiology interaction: validated tasks 1 to 5 require scrolling (as seen above in table 4.1), and radiologists confirmed their frequency and importance. Discussion confirmed both centrality of scrolling in routine activities, and the need and possibility for improvement of image interaction via device and/or software.

The most crucial challenges in current scrolling technology identified were reducing repetitive movements, more easily varying the speed of scrolling and mapping more functionality to the device.

Many parts of the PACS system are somewhat personalizable: e.g. *"I can't imagine using PACS without having my custom way of looking at it"* (P4). Radiologists were generally receptive to the idea that the input device could be more personalizable, for instance the ability to access pre-set scrolling speeds.

There are many directions we could explore to improve radiology interaction, which leaves us with a large possible design space. As a result, we decided going forward to refine the scope of the radiology interactions being examined. Narrowing down to scrolling interactions was a logical step as it was prevalent in the task examples described in this chapter. Optimizing scrolling through a stack is a sizeable problem in itself, and we chose to focus on it as our top priority.

Chapter 5

Design Space and Prototypes

We identified a scrolling-input-mobility design space to explore for possible improvements to challenges we saw in stack scrolling. This space includes current baseline methods and adds diversity in input control (Table 5.1, also illustrated in Figure 6.7). We populated this design space with three exploratory prototypes (one representing two enhancements of control mobility), constructed by modifying existing mice (as shown in Figure 5.1). These prototypes were used in the experiment described in Chapter 6, to compare the different interaction mobilities.

A second research question, and subsequent aspect of our design space, involves using tactile cues to display CAD marks or other annotations. As we wished to explore interactions between tactile CAD display and scrolling mobility type, we installed tactile displays in all of our prototypes.

5.1 Vibrotactile Annotation Display

A pager motor generated a vibrotactile buzz in all prototypes, perceptible in all hand positions observed. In piloting, we determined that a cue duration of 200 ms (pager motor supplied 3V, 200 Hz), delivered at the annotated image, provided a good compromise between perceptibility and intrusiveness. In the case of fast scrolling ($<$10 images/sec) we advanced delivery by one image, so that the cue started 1-image in advance, so that the majority of the buzz was felt on the image.

5.2 Devices

5.2.1 Touch

An Apple Magic Mouse, characterized by a curved multi-touch surface, was modified by adhering a pager motor to the underside of the touch surface, adding around 1 cm to its height. The multi-touch surface was of interest because custom gestures (extra button-mapped functions requested by radiologists) could be mapped onto it, but this ability was not tested here.

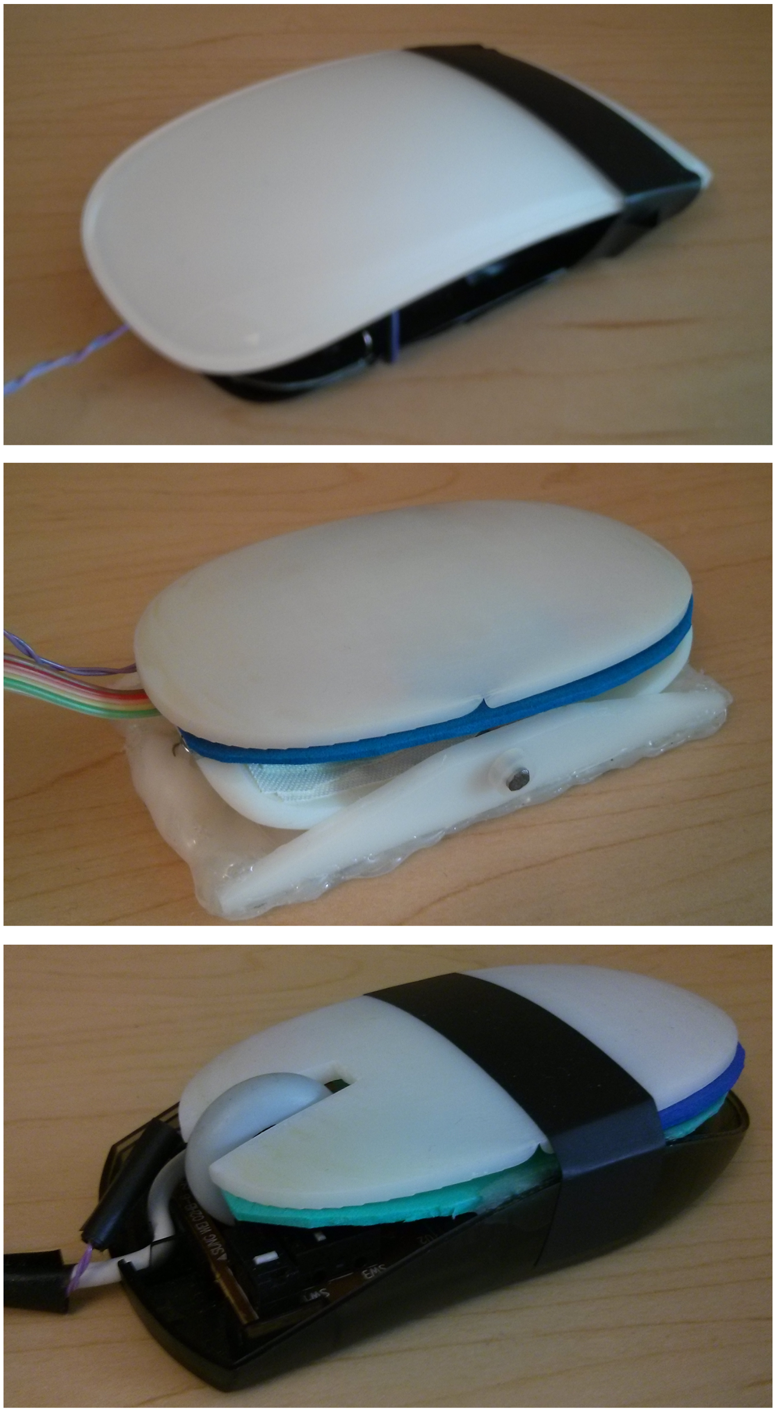


Figure 5.1: Image of the three prototypes used in the study: Touch, Tilt, Wheel/Click+Drag (both in one prototype).

Table 5.1: Describes each prototype based on its scrolling input mobility; prototypes pictured in Figure 5.1.

Type	***Prototype Name*: Motion Description**
Wheel scrolling	*Wheel (Baseline):* Traditional scrollwheel mouse functionality.
Dragging of whole mouse	*Click+drag (Baseline):* Traditional dragging and pointing functionality (combinable with Wheel).
Sliding on mouse surface	*Touch:* Sensing of a finger sliding on a smooth surface, as in current mobile touch screens; multi-touch can map gestures to specialized functions.
Rocking	*Tilt:* Maps forward/back rocking to scroll up/down; also uses rate rather than position control.

5.2.2 Tilt

A curved top surface with profile matching the Magic Mouse was 3D-printed and a pager motor placed on its underside. Springs at either end achieved stable centering of a curved bottom surface. An accelerometer, read by an Arduino Uno, detected its tilt angle which, configured for rate control, controlled rate of movement through the stack.

5.2.3 Wheel / Click+Drag

To provide baseline comparisons at a comparable level of prototype polish, we replaced the top of a traditional mouse with a 3D-printed surface identical to Tilt's but with a slot for the scrollwheel, and attached a pager motor to the underside of this surface.

5.3 Connectivity

Prototypes communicated with a custom image-viewing program (written in C++) on a control laptop via an Arduino microprocessor (Uno or Micro). This program commanded a vibration via USB-2, and received input from existing x-y, scrollwheel and multi-touch mouse channels and tilt's custom accelerometer.

Chapter 6

Experiment: Simplified Stack Scrolling

We conducted a study to compare usability of our 4 prototypes (representing points in the scrolling input design space, i.e. control mobility), as well as the impact of both scrolling type and annotation modality on the human viewer's detection performance in the face of imperfect annotation (false positives and true positives: faked CAD). As it was not feasible to access professional radiologists for this kind of input, we abstracted the stack-scrolling task to test on lay users. In constructing annotation modality conditions, we aimed to hold perceptual salience constant.

We chose to focus on the very specific sub-task of finding a region of interest within an image stack, because this is the most basic element of stack interaction. For this subtask we wished to:

- Compare scrolling methods with respect to stack scrolling performance and subjective factors.
- Compare annotation modalities with respect to impact on task performance, including both speed (where we seek speed-ups) and errors (where we do not want to increase error rates or introduce bias).

6.1 Abstracted Task

We abstracted the task of a trained radiologist scrolling through a lung CT image stack while looking for potentially cancerous nodules (a case of Task Example 1, from Chapter 4). In real stacks, lung images exhibit a bronchial tree: bronchi tubes feeding into smaller tubes called bronchioles. The alveoli sacks at the ends of this tree can look similar to, but have slightly different characteristics than, cancerous nodules (Figure 6.1).

To mimic this task in an easy-to-learn way, we placed small greyscale rectangles throughout a 60-image stack containing a uniform black field. The task was to find the true target (Table 6.1), of which there would be exactly one per stack, among 50 distractor noise rectangles which were distributed throughout the stack's images. The subject would then click one of 4 buttons on a numeric keypad, indicating the image quadrant where the target was seen. In pilots, we adjusted task difficulty (varying distractor shape, frequency and contrast) to the settings described here in Table 6.1. These supported an error rate of around 10%, which is about 1/2 to 1/3 better than the 20-

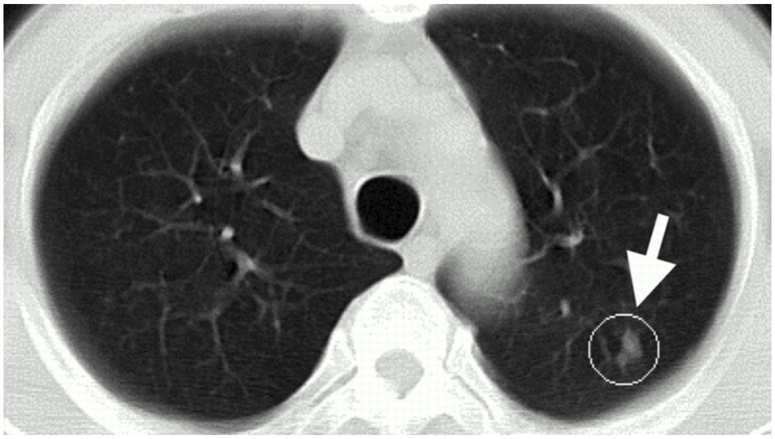


Figure 6.1: Images of lungs, with potential lung nodules detected (from: Armato S.G. et al. Radiology 225: 685-692, 2002).

Table 6.1: Abstracted task parameters used in experiment.

Object	Shape/size	Colour
True target	1 perfect square: 5-10px/side	medium grey
Distractors	50 rectangles: sizes randomly chosen between 4-12 px/side with aspect ratio exactly 40% smaller or larger than the true target	significantly lighter or darker grey
Distractor target	1 almost square: aspect ratio of 16% smaller or larger than the true target	slightly lighter or darker grey

30% documented for radiologists [22]. However, we expected to have slightly better performance as our task was cognitively easier, and we expected the lay users recruited to be less conscientious than the grad students from the department used as pilots.

For tractable analysis, we constrained the target's location to stack index of: 20, 30, 40, 50. In pilots, participants did not appear to learn target locations, as they continued to make errors at a uniform rate. This was confirmed in our study results.

A single stack consisted of one setting of scrolling input and feedback, and a single target located in an image at one of its four possible indices. For each combination of scrolling input and feedback, participants saw an initial learning example plus 20 test stacks. These stacks were presented in a random order. These comprised 5 replicates with a highlight at one of each of the 4 target indices: 4 where the true target (represented as a perfect square) was highlighted (16 stacks

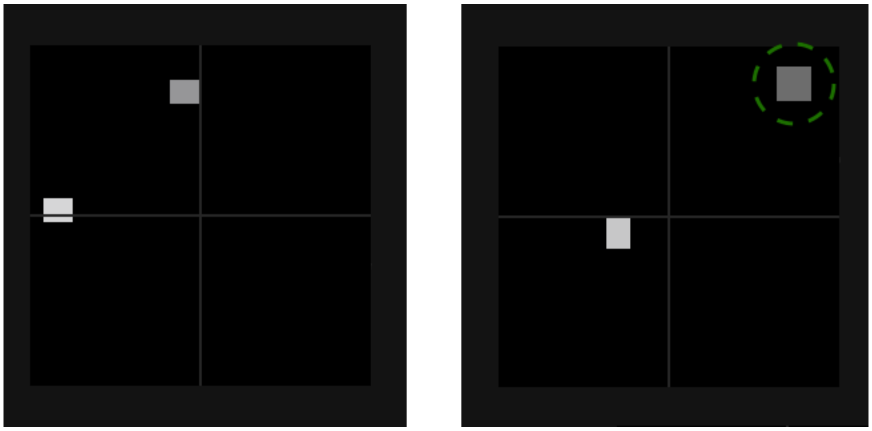


Figure 6.2: Two images of the abstracted task; a visual annotation is seen on the right image. Image size: 256 x 256 px, rendered at around 8 cm/side on a monitor.

Table 6.2: Annotation modality study factor

Visual	Dashed green circle around the target (Figure 6.2), visible when the image itself appeared.
Haptic	A 200 ms buzz (pager motor) as user approached annotated image. The buzz started one image before the annotated image if the user was scrolling rapidly (majority of the buzz felt on image), and on the image if the user was scrolling slowly.
Combined	Combination of Visual and Haptic.

total); one where a distractor target was highlighted, and the true target was located further along in the stack (4 stacks total). This ratio of true-positive highlighting and false-positive highlighting (80% to 20%) approximates current published performance of CAD algorithms [13].

The distractor target (closer to a perfect square than the distractor rectangles) always appeared before the true target, with an advance randomly selected within image 5 in the stack, and 5 images before the true target.

6.2 Experiment Design

We used a Latin square to produce 4 orderings for scrolling type (Touch, Tilt, Wheel, Click+Drag) and 3 for annotation modality (Visual, Haptic, Combined). The latter were blocked within scrolling input to minimize device switching, for a total of 12 orderings of the 12 condition combinations experienced by each participant.

We measured task completion time (from start of scrolling to keypad button click) and accuracy. A correct response comprised identification of both the correct image and quadrant of the true target.

12 lay participants (1 per ordering) were recruited via campus posters and emails to department mailing lists, and compensated $15 for 1.5 hours of their time.

6.3 Protocol

An experiment session took up to 1.5 hours. Participants were seated in a quiet room, asked to complete a demographic questionnaire and instructed to complete the task quickly and accurately. For each new combination of device and annotation modality they were given an exemplar stack to practice and the opportunity to ask questions. They then carried out target-searches on 12 sets (one for each condition combination) of 20 stacks, while listening to white noise through noise cancelling headphones to mitigate any auditory vibration feedback as the background noise at a hospital likely would.

Between each set of 20 stacks, participants answered a questionnaire about their scrolling accuracy, frustration, confidence and attentional needs for those trials; how easy it was to notice the annotations and how helpful they were. Upon completion, they were asked to rank their preference on device and annotation (see Appendix A.2.2 for the full questionnaires).

6.4 Hypotheses

We made the following five hypotheses about the quantitative and qualitative differences in the devices and annotation modality:

H1: Wheel and Touch will afford no difference in accuracy, because they both allow clutching through the images.

H2: Click+Drag and Tilt will be fastest in approaching an area, but perform poorly in finer adjustments.

H3: Combined (Haptic+Visual) annotations will afford faster detection than either alone.

H4: There will be no effect of annotation modality on error rates.

H5: Combined (Haptic+Visual) annotations will be preferred subjectively.

We also sought subjective input that would elucidate ergonomic factors, but this data was not collected in a form amenable to statistical testing.

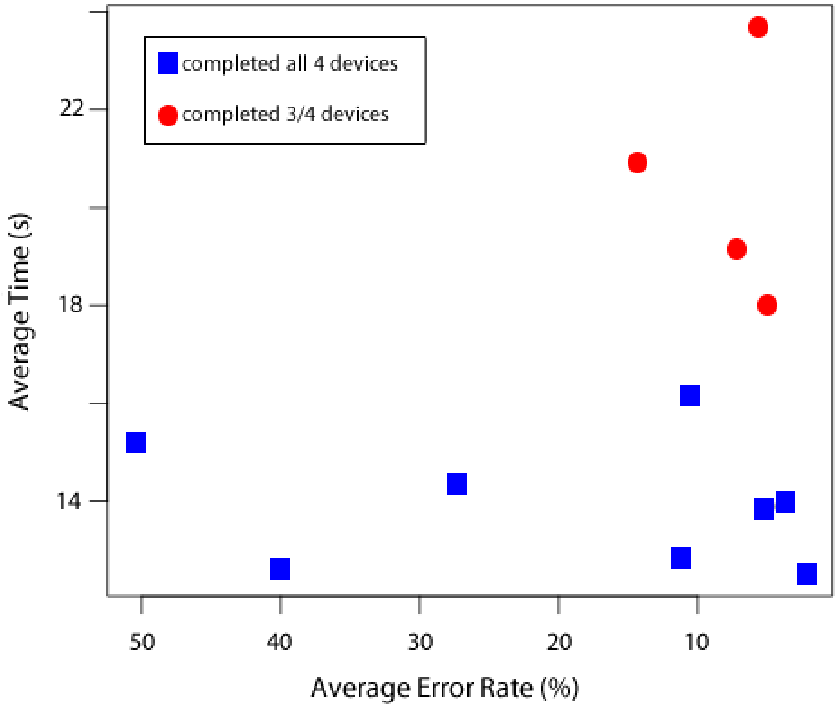


Figure 6.3: Plot showing the average time time for a task vs the average error rate, for the first 75% of the tasks (the ones completed by all subjects).

6.5 Results

We replaced one subject who completed less than half the trials, as the subject kept double clicking and did not seem to fully understand the task. We replaced a second subject who had an error rate over 50% and did not complete the last scrolling type, as the subject was both slow and inaccurate.

9 out of 12 final subjects had error rates <20%, and 3 in the range of 30-50%. Five of the 12 subjects did not complete the final 3 sets (last device) due to a time restriction of 1.5 hours; they completed 75% of the study (3/4 of the devices). Figure 6.3 uses the first 75% of the data that all the subjects completed, and shows the average time for the stack scrolling task vs error rate. The subjects who did not complete the final scrolling device were on average slower but all had low error rates (all <20%); thus the loss of their (slower) trials biases the data towards faster subjects. Some of these faster subjects were faster at the cost of higher error rates (all of those with error rates >30% completed the 4 scroll types).

6.5.1 Task Completion Time

Completion time exhibited a broad and heavily skewed distribution: targets were placed at different distances from the start point. Participants varied in the care that they took, with trials tending to go long if they did not find the square in the first pass. Conventional models like Analysis of Variance (ANOVA) and General Linear Modelling (GLM) require normality. ANOVA can also only treat whether or not they got the trial correct as a variable, whereas a Cox model can use whether or not the trial was correct to censor the data. Further, completion time and accuracy were not fully independent since with enough time a correct target could always be found in our abstracted task.

Censoring the data refers to a statistical situation wherein only partial information is known about a data item, e.g. that up to time x, the user had not completed the task [14]. In data censoring the calculation for the regression coefficients is modified, based on the assumption that it will take the user more than time x. The legitimacy of some trials comes into question if we do not censor the times by whether or not the subject found the target correctly. Short response times in which the participant was apathetic and chose a non-target image would skew results, but if censoring is used it essentially removes this data by only taking it as partial information.

We therefore used a proportional hazards model (a Cox regression [3] [11]) for completion time, which assumes that if given more time users could answer correctly. Non-error trials have all the information needed; error trials have partial information (we only know they did not find it up to a certain time). This model uses observations about each task completion time (T_{comp}), and the time is censored by whether or not a subject got the trial correct:

$$T_{comp} = P + S + A + Ti + Th + N^2 + (Th \times A) + (S \times A) \text{ [Eq. 1]}$$

where model parameters are Participant (P), Scrolling input condition (S), Annotation modality condition (A), Target index (Ti), Target highlighted (Th), and trial Number (N).

The hazard rate is the likelihood that at a given time the user will find the target. We can calculate this rate from the Cox regression and plot it as a survival curve, which shows the likelihood a task would be completed at a certain time (Figure 6.4). Most tasks are completed within 40 seconds, and it is apparent that combined annotations (Haptic+Visual) make it more likely that the task is completed earlier.

The Cox regression delivered the following statistically significant results for completion time ($p < .05$):

- S: Wheel scrolling was faster than the baseline Click+Drag (Z=2.48, p=0.013), and Tilt was slower than Click+Drag (Z=-2.47, p=0.014).
- A: Combined annotations were faster than Visual (Z=2.59, p=0.0096), and Haptic was slower than Visual (Z=-2.82, p=0.0048).

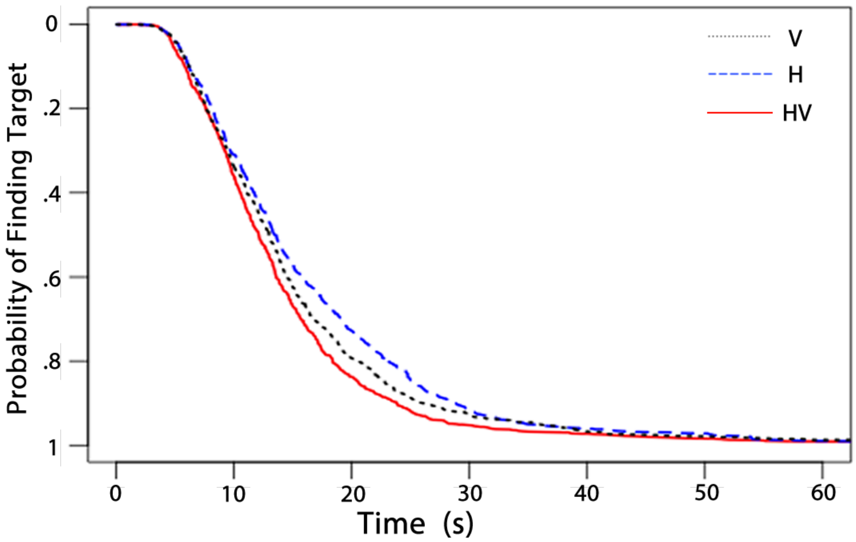


Figure 6.4: Survival likelihood (Cox regression) vs. projected completion time. V=Visual, H=Haptic, HV=Combined.

- Th: For false alarms (false-target highlighted) trials were slower (Z=-2.27, p=0.023).
- Th x A: Combined annotation was slower for false alarms than for true positives (Z=-3.30, p=0.00096).

Regarding individual variance and task validation,

- P: Participants varied widely in completion time (T_{comp} SD: 12913ms). E.g. P10 was faster than P1 (Z=5.29, p<0.0001), and P8 slower than P1 (Z=-2.82, p=0.0048).
- N2: Trial number reaching significance (Z=-5.48, p<0.0001) indicates T_{comp} fit a t2 distribution: earlier trials were slower, middle trials fastest, and later trials slower again. This suggests learning followed by boredom.
- Ti: The shortest target index distances (20) had faster trials than the two longest (40; Z=-5.46, p<0.0001) and (50: Z=-8.11, p<0.0001).

Approach analysis: The user's motion dynamics are not visually guided search, since the target's location is not known. Rather than ballistic motion with overshoot, we anticipate that motion equilibrates to either a relatively smooth scan rate, or proceeds faster and slower as suspected targets are examined. To get a sense of the motion dynamics as a function of scrolling method and annotation modality, we defined T_{app} as the period of time a user proceeded forward measured from the trial's start to a first direction reversal. To reduce noise, trajectories shorter than 10 images were removed from this analysis.

A GLM (generalized linear model) was used for analysis so as to model approach time linearly based on the distance moved, scrolling device and annotation modality. Haptic had slower approaches than Visual (t=2.46, p=0.0014). Wheel, Touch and Tilt had slower approaches than Click+Drag (t=3.88, 5.02, 8.30, all $p<0.0001$), but there was less data for Click+Drag following short-trajectory removal; we conjecture that its motion was jerkier.

6.5.2 Accuracy

To analyze trial accuracy (right/wrong) we used a GLM (with binomial distribution) with the same parameters as for T_{comp} (Eq. 1). As mentioned in the protocol section, subjects were instructed to complete the task quickly and accurately. Significant results ($p<.05$) are as follows.

- P: Participants varied widely in accuracy (average 17% error rate, min 2%, max 55%). E.g. P11 had significantly fewer errors than P1 (Z=8.90, $p<0.0001$).
- N2 (Z=-2.28, p=0.023): there is likely a learning then boredom a effect (consistently with T_{comp}).

6.5.3 Questionnaire Results

Ten participants preferred Combined annotation modalities; one preferred Haptic, and one Visual. Likert scale responses were analyzed using a proportional odds logistic regression, accounting for scale ordering along with experiment factors (scrolling input, annotation modality). This indicated ($p<0.05$):

- Wheel was deemed the most accurate device (Z=-4.79, $p<0.0001$) with Touch the runner-up (Z=-1.97, p=0.0493). Users had more confidence in Wheel (Z=-4.45, $p<0.0001$) and felt they required less attention (Z=3.03, p=0.0025).
- Wheel was rated the least frustrating (Z=4.79, $p<0.0001$), with Touch 2nd least frustrating (Z= 3.07, p=0.0021).
- Combined (haptic and visual) annotation was most noticeable (Z=3.27, p=0.0011), as well as most helpful (Z=2.34 p=0.0191).

There were generally more positive responses for Wheel in comparison to the other scrolling inputs. Combined annotation received higher ratings than either alone.

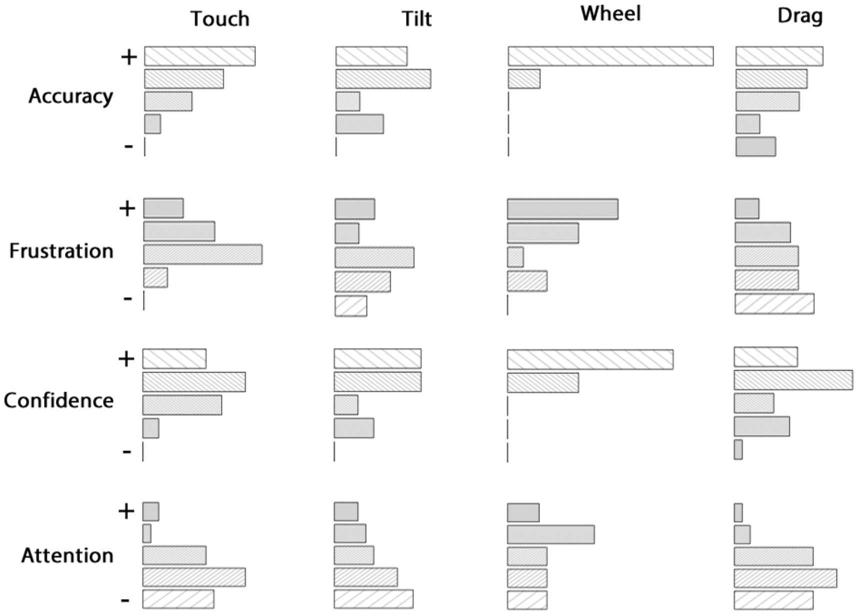


Figure 6.5: Summary of questionnaire results by device, for: Accuracy, Frustration, Confidence and Attention (N=12).

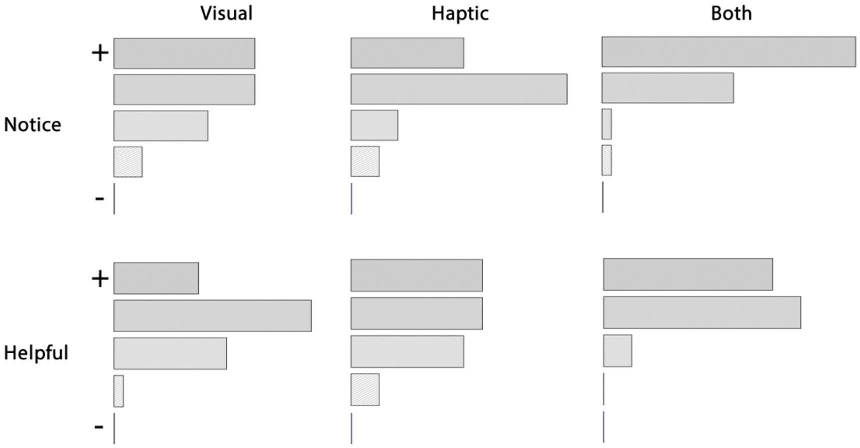


Figure 6.6: Summary of questionnaire results by annotation modality, for: Noticeability and Helpfulness (N=12).

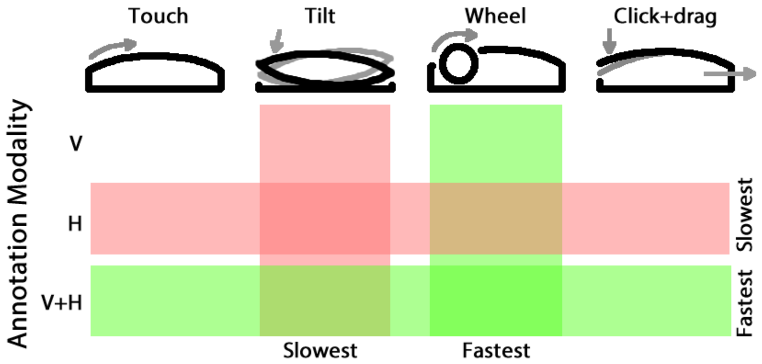


Figure 6.7: Summary of design space and results for speed.

Table 6.3: Hypotheses restated with outcomes, for discussion.

Accepted	
H3	Combined (Haptic+Visual) annotations will afford faster detection than either alone.
H4	There will be no effect of annotation modality on error rates.
H5	Combined (Haptic+Visual) annotations will be preferred subjectively.
Partially supported	
H1	Wheel and Touch will afford no difference in accuracy, because they both allow clutching through the images.
H2	Click+Drag and Tilt will be fastest in approaching an area, but perform poorly in finer adjustments.

6.6 Discussion

Here we discuss the results based on the hypotheses presented above, their outcomes are summarized in Table 6.3. These results are visualized graphically in Figure 6.7.

6.6.1 Hypotheses

Value of Haptic Feedback:

H3: Combined (Haptic+Visual) will afford faster detection than either alone - **Accepted**

Results from our non-expert, abstracted study suggests that for a task similar to image-stack scrolling, multimodal annotations are most noticeable, most helpful and improved detection times. Haptic annotation was slower than Visual. We can infer performance relative to no annotation from the cases where the true target was not annotated (distractor target highlighted); having just haptic or visual annotations showed no significant differences in speed, but multimodal annotation slowed the user relative to when the target was correctly highlighted. Overall, using both types of annotation together was still fastest.

A possible explanation for this, in addition to simple redundancy of cueing, is that each modality provided slightly different speed-related benefits. Visual annotations told the user exactly where in the image the target was; haptic may allow faster motor responses. Combined annotations benefited from both.

The timing advance of the Haptic annotations here was devised to match the performance benefit of the Visual annotation as closely possible. However, in future, the haptic annotation could be given earlier, allowing the user to slow down pre-emptively and search more carefully through the next few images. In this abstracted case, the context in which the perfect square was found does not matter, but in the case of a radiologist, tweaking the timing of the feedback may help them view and understand the context of the potential anomaly.

An important emerging source of annotations is other radiologists. Trainees must have their diagnoses checked by a board-certified radiologist, and are required to provide annotations in key images for the second radiologist to review, which haptic feedback can quickly identify in a larger image stack. Also, there is widespread pressure within diagnostic imaging [31], and medicine as a whole, to increase peer review activities as a quality assurance measure.

H5: Combined (Haptic+Visual) will be preferred subjectively - **Accepted**

10 out of 12 of the participants ranked the Combined annotations as preferred, over either alone. Additionally, it was generally seen as more helpful and noticeable (Figure 6.6).

Effect on Decisions:

H4: There will be no effect of annotation modality on error rates - **Accepted**

There was no difference in error rates between the different annotation modalities. Annotation modality did not apparently affect the lay users' ability to make a decision, as it did not impact accuracy. Overall, having Combined annotation sped users up and they showed a preference towards it, in comparison to Visual alone.

Experiment participants had the same accuracy for trials annotated correctly and incorrectly, as the Th (Target highlighted) term of the GLM model for Accuracy was not significant. However, they made slower detections in trials containing false positives (the Th term reached significance in the Cox model for task completion time).

Scrolling Type:

H1: Wheel and Touch will afford no difference in accuracy, because they both allow clutching through the images - **Partially supported**

No device emerged as the most accurate, but subjectively Wheel was felt to be the most accurate, with Touch next.

H2: Click+Drag and Tilt will be fastest in approach, but perform poorly in finer adjustments - **Partially supported**

Click+drag was fastest for approaching an area. Tilt was slowest in task completion time, so appears to be weaker for finer adjustments for the implementation we tested; however, it was also the least familiar to users, and had the least refined implementation (the others being minor revisions of commercial products).

6.6.2 Summary

The traditional and most familiar device (Scrollwheel) had the fastest task completion times for lay users, and was preferred. In most metrics, sliding-touch scrolling (Touch) was ranked second. However, Click+drag had faster initial approach (even if it was to the wrong area). This, along with familiarity, is likely why Scrollwheel and Click+drag work well together in the radiology environment.

This was a first attempt at using Tilt in a stack scrolling task, and it did not fare well by most metrics. The prototype was the least ‘slick’ of the devices. The most familiar scrolling devices (Scrollwheel, Touch) generally fared the best. Tilt appears to be a good option for integrating rate control into a mouse form factor, and further refinement and integration into true mouse form (where other scrolling options are available) is worth attempting.

Critique of Methodology: We tried to balance the length of the experiment with the need for more trials for better quantitative data. There was a boredom effect, but by including trial number as a factor in our statistics we hope to have minimized its effect.

Chapter 7

Follow up with Radiologists

To confirm and elaborate on the convergence of our prototypes and study findings, we returned to our radiologists to elicit further feedback.

7.1 Modified Prototype

We combined the best performing features found in the scrolling input and annotation types evaluated in the experiment in Chapter 6, as well as the features we felt still needed to be further explored, to create a prototype that worked as a conventional mouse. A feature that still needs further exploration is tilting to scroll, as it clearly suffered from a less polished prototype; but could, if better designed, have potential for the kind of integrated, higher control scrolling that radiologists need. This prototype had the added abilities to (a) touch-scroll, and (b) tilt backwards to access rate control scrolling (with a switch to control direction). We began with a Microsoft Wedge mouse, added a bump off the back created with acPolymorph to be able to rock and sensed tilt with a potentiometer (an accelerometer would confound translation with tilt). An Arduino relayed mouse signals, and a pager motor was installed underneath the surface (Figure 7.1).

7.2 Method

We took the modified prototype to the workplaces of 3 radiologists (2 previously interviewed from the task analysis in Chapter 4, one new), and demonstrated its movement and haptic feedback in the context of our abstracted test task. Additionally, we informally discussed its potential usefulness with them.

7.3 Highlights

Given existing customizability of PACS setups, radiologists reiterated their receptivity to the idea of a personalizable mouse. Their preferred speed of scrolling is highly personal and varies depending on the type of stack; rate control could have several preset speeds (potentially controlled via a slider on the side of the mouse). *"The goal should be to customize the mouse in a perfect world once, and*

Figure 7.1: Images of the modified prototype.

then to not have to fool with it after that" (P1).

Emergency radiologist P2 stated *"The way that I look at a large data set study is I fly through it once and get a birds eye view... I want to exclude any immediately life-threatening conditions."* Further, in a diagnosis he needed to access multiple stacks, and felt the haptic feedback would help re-orient him when switching between them. He also indicated aesthetic appreciation: *"The haptic feedback I love"*, *"The haptic feedback is really cool, even just going through this short series."*

A general summary of P2's feelings on the use of haptic feedback can be seen in this quote: *"You are interrupted, you might pull down the the case, pick up the case again, and by the time you return to this very complex case that you are going through you may not really remember... but having something like this that even is just... as you are scrolling through it buzzes on the annotations, to me is pretty helpful."*

Sometimes radiologists need to re-read other radiologist's image sets to ensure quality of care, such as when working with trainees. The haptic annotations could help speed this review: *"You mark up the image in a peer review, and then I go through it to check whoever's work, and I can find immediately what they were looking at - that is valuable"* (P1). Also, *"being able to make peer review economic timewise and easy for the radiologist because it is really important, and thats another possible application of something like this haptic technology"* (P1).

P3 noted there might be *"a temptation to go really fast"*, and worried that the haptic cues would encourage this, resulting in missing anomalies. However, he further mused that it would be useful for very large data sets, such as the lungs. He generally felt that *"You have a problem and you are trying to find a solution to the problem, and here we have a potential solution to many problems."*

Unsurprising was some mention of potential integration issues: *"Many of our workflows are so refined over the years... because we are just used to going through data sets in a certain way"* (P2).

Chapter 8

Discussion, Conclusions and Future Work

The goal of our research was to investigate how best to physically interact with radiology images, and to thoughtfully apply human computer interaction ideas to this setting.

8.1 Combining Scrolling Input Methods

Radiologists were interested in reducing the repetitive movements associated with the mouse that occur often with scrolling, for instance clutching with the mouse wheel. This encourages us to continue to refine our Tilt implementation and test it following longer learning, as a rate control approach, while continuing to support other functionality.

Multi-touch would also allow many more potential improvements in radiology image interaction, via the mapping of gestures to different tools that could reduce the need for modal interaction with PACS workstations. Radiologists are generally a technology savvy group, and most had much familiarity with multitouch interfaces; in our discussions with them, they seemed generally receptive to the idea of using gestures with PACS.

8.2 Validity of Abstracted Task + Lay Users

How are our lay subjects like/unlike radiologists? On average our lay users' error rates were similar to those seen in radiologists. We would expect radiologists to show more homogeneity (less variance) in error rate because of their training and the expected level of care in the health system, and because studies show consistent error rates of 20 to 30% [22]. Our lay users likely varied more, ranging from less than 2% error to 55% error (average 17% error rate). We would expect professionals to have fewer slow outliers, and less inter-person variability if they were to complete our abstract stack scrolling task, or any radiology task for that matter.

We must, therefore, take care in generalizing to radiologists. We saw little effect on error rate in the study with lay users, but there may be effects for radiologists. Future validation includes a compacted study to look at the effect of annotation modality on errors for trained radiologists.

8.3 Recommendations for Moving Forward

We believe radiologists would benefit from a more specialized interaction device, particularly since, depending on their specialization, they spend full 8 hour workdays interacting with images at the PACS workstation. Scrolling was our main avenue of exploration in this thesis, however, there are many other ways the interactions could be streamlined. One particular avenue is to work to reduce the amount of mode switching needed to access different PACS tools.

Based on our experiences, the solution should have the following properties:

- Diversification of available input movements on the mouse. This will allow PACS interaction that is less subject to modal constraints and workflow interruptions, and will potentially reduce repetitive movements. For instance, it may be useful to continue to explore the intuitive use of touch interactions so that they can be mapped to PACS tools (6.5.3 Scrolling Type), particularly since with more familiarity it may become more preferred.
- Haptic feedback in the mouse. This can at least be used to display radiologist and CAD annotations (6.5.1 Value of Haptic Feedback), but may also have other useful applications.
- The ability to map different PACS tools to different interactions on the mouse.
- The ability to tailor the device to individual differences, such as the difference in preferred scrolling speeds.

Literature and our discussion with radiologists highlights individual differences in interaction preferences, particularly scrolling speeds. Radiologists were receptive to the idea of a more personalizable mouse (as seen in section 4.2 Task Example Validation), so there should be an interface where they can set up which PACS tools map to which physical interactions on the mouse.

8.3.1 The Next Prototype Step

In order to achieve the properties discussed above, we propose a mouse with many more features than a traditional mouse. The solution should have the following specifications (see Figure 8.1 for one possible way they could be mapped to the mouse):

- Touch surface
- Scroll wheel
- Ability to slide on table surface
- Ability to left/right click
- Tilting (with notches)
- Pager motor

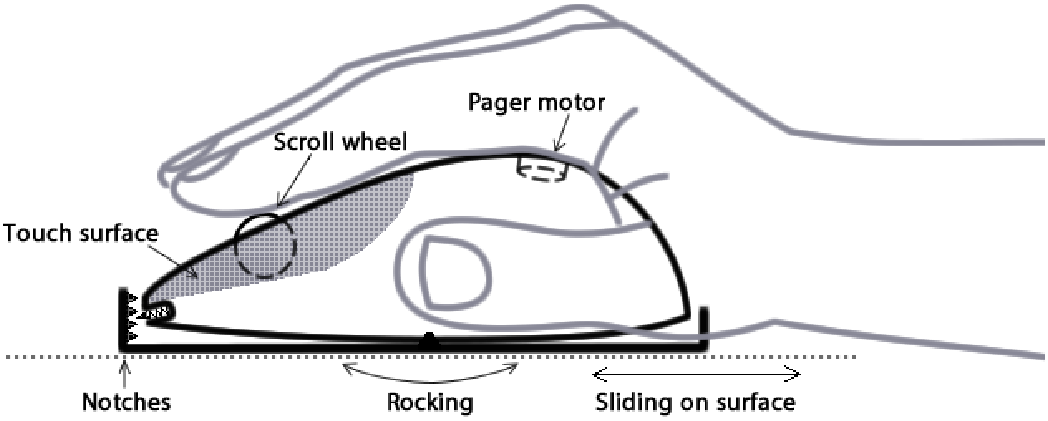


Figure 8.1: Schematic diagram of a potential next prototype. A rocking body is contained in a base that can slide along the table surface with traditional x-y mouse functionality. The primary function of the base, in this schematic design, is to house the rocking mechanism, including mechanical detent generation.

The touch surface would allow for gestures to be used, potentially to access some of the more frequently used PACS tools. The scroll wheel is needed because keeping it retains the favourite, or at least most familiar, scrolling method. The ability to slide on the surface and left/right click are mouse functions that are needed for pointing, selecting and other basic desktop interactions. The rocking would allow for easy access to scrolling via rate control, and allow for continued exploration of this interaction, as well as to monitor the learning curve. The notches would create detents when the mouse was rocked up or down. The detents would add a level of tactile awareness as to how far it has been rocked, similar to the notches on a scroll wheel. The pager motor would be used to give haptic feedback on annotations, and also allow us to continue to explore other uses of haptic feedback in the radiology setting.

Additional items that could be included on the mouse are a notched slider for different scrolling speed presets, and pressure sensitivity in the touch surface for more nuanced interactions. The slider could be on the side of the mouse, in order to be accessed by the thumb.

8.3.2 Obstacles to Adoption

One of the main obstacles to adoption is the learning curve that radiologists must go through when they are given a new interaction device; as was seen in the experiment (Chapter 6) in this thesis, the fastest and preferred device is often the one the user is most familiar with. However, if used over a longer period this preference may change, as the user may become more skilled at using another interaction method. Tilt, for instance, was one of the slower scrolling methods, but was also the least familiar.

Radiologists are highly trained individuals, and they have put a lot of time into their training. In order to adopt anything new they need to see a clear benefit to it, as it may require some time to adjust their workflow. The need to maximize their throughput is paramount to saving tax payer dollars, as they are one of the highest paid medical professionals in the US making on average over $200,000 per year[1]. As such, even a small speed up in image interaction could have a large monetary effect.

PACS vendors are also an obstacle to adoption, as they may be reluctant to try to implement something too different than the status-quo for fear of alienating customers. Additionally, PACS software is large and complicated so integrating a new device may be time consuming and therefore costly. The hospitals are also invested in a certain brand and version of PACS; the hospital administration may not be interested in transitioning to a new system because of the setup costs, unless there are major gains to be had.

8.3.3 Technical Implementation Issues to Solve

Standard computer mice send signals to the computer, but do not receive any information back. To create haptic feedback, we need to be able to send a signal from the computer to the mouse; this backwards path does exist, and can be done with peripheral devices such as an Arduino Micro. Using a specialized mouse driver, a commercial mouse the 'ifeel'[2], could create haptic feedback. However, it never achieved much popularity; this may have been because most software did not make use of the haptic feedback, and that the feedback created an obvious audible noise above 100 Hz [10]. We believe one of the main reasons it failed was because it was introduced as a technological innovation, and there was no knowledge base or guidance as to how to incorporate it into interaction design. It was therefore not useful, and unsophisticated attempts to incorporate it proved to be more annoying than helpful.

For the communication from the mouse to the computer it uses 'mouse events'. For instance, the different versions of scrolling send the same 'scroll event', but the frequency at which this is sent depends on how they are programmed. Adapting these mouse events to work for a mouse with more interaction options should be relatively simple.

---

[1] http://www.healthcare-salaries.com/physicians/radiologist-salary

[2] http://www.logitech.com/en-roeu/press/press-releases/1183

8.4 Future Work

The improved prototype described above could be used to run a similar study to the one presented in this thesis. However, in our study we required our lay users to use each scrolling input type separately. A more realistic scenario is for the user to access them all in a seamless manner: some methods are better for scanning the stack, others for fine adjustments, and yet others for other GUI uses. While the prototype proposed above likely would still need further refinement, it could be used to test the speed that users can scroll when they have multiple interaction methods available to them, and compare it to the speeds seen in our experiment where they used each separately.

We initially avoided performing structured studies with radiologist, opting for a more informal approach; however with a more stable and refined prototype we could give it to radiologists to use at their workstations. Longitudinal studies are needed to examine how a device integrates into a radiologist's everyday workflow. One idea would be to use a diary study to monitor the adoption of the mouse.

The effectiveness of the haptic feedback could be increased by personalization, to accommodate individual differences in reaction times. One could create a program that logs the reaction to the haptic cue, and adjusts the timing of the feedback based on this. Other types of haptic cues might improve attentionally on the simple buzz we used, such as a vibration fading in upon approaching a region of interest.

8.5 Conclusions

We analyzed radiologists' work and found a high prevalence of scrolling, poorly supported by traditional scrolling input devices with negative ergonomic and productivity implications that can be expected to grow in the future. The radiologists we interviewed were very interested in seeing improvements to their working tools, and some had experimented with this on their own.

We compared 4 scrolling input motions, 2 of them traditional. The scrollwheel emerged as the fastest in our study with lay users, and confirms our early observation that augmentation of established tools should be explored rather than replacement. However, novel input methods (e.g a tilt or rocking motion associated with rate control scrolling) were potentially handicapped by their newness and less optimized implementation. Rate control has been seen to be a fast stack scrolling method for radiologists [6], and we believe there is a possibility they would be more accepting of it if it was integrated into the familiar mouse form factor where other scroll methods are also available. Because the scrollwheel has known ergonomic issues from excessive repetitive movement, alternate methods still need to be explored.

Click+drag had the fastest forward movement speeds when doing a first pass through the stack of images, for 8 out of 9 radiologists [6]. This study used a task with artificially stimuli and all the radiologists were 100% accurate. However, the effect of interaction technique on errors should be further explored, as the speeds the technique allows the user to move may affect perception. When talking to radiologists about using a rate control device they mentioned wanting to be able to have different individual speed presets. It would be interesting to investigate what scrolling speeds are used in different aspects of stack interaction (e.g. first pass, careful back/forth).

In the emerging practice of incorporating annotations (from CAD or other radiologists) into radiologists' workflow, we have shown that multimodal cues are a promising approach, showing task speedup without error degradation, for a task abstracted on non-experts. Radiologists are heavily visually loaded, and may benefit from information provided through a less loaded modality, even when redundant.

Because of the many factors (economic, training) making the radiologist work environment highly change-resistant, introduction of new input devices must be undertaken with care. Our informal approach of presenting prototypes and eliciting the radiologists' feedback, which revealed enthusiasm for change, seems a promising avenue for this.

Bibliography

[1] M. Akamatsu, I. S. MacKenzie, and T. Hasbroucq. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics*, 38:816–827, 1995.

[2] E. Alberdi, A. Povyakalo, L. Strigini, and P. Ayton. Effects of incorrect computer-aided detection (cad) output on human decision-making in mammography. *Academic radiology*, 11 (8):909–918, 2004.

[3] P. K. Andersen and R. D. Gill. Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, 10(4):1100–1120, 1982.

[4] T. H. Andersen. A simple movement time model for scrolling. In *CHI05 extended abstracts on Human factors in computing systems*, pages 1180–1183, 2005.

[5] K. P. Andriole, J. M. Wolfe, R. Khorasani, S. T. Treves, D. J. Getty, F. L. Jacobson, M. L. Steigner, J. J. Pan, A. Sitek, and S. E. Seltzer. Optimizing analysis, visualization, and navigation of large image data sets: one 5000-section CT scan can ruin your whole day. *Radiology*, 259(2):346–362, 2011.

[6] M. S. Atkins, J. Fernquist, A. E. Kirkpatrick, and B. B. Forster. Evaluating Interaction Techniques for Stack Mode Viewing. *Journal of digital imaging*, 22:369–382, 2009.

[7] R. Balakrishnan, T. Baude, G. Kurtenbach, G. Fitzmaurice, and R. S. East. The Rockin Mouse : Integral 3D Manipulation on a Plane. *In Proceedings of CHI 97*, pages 311–318, 1997.

[8] P. M. Boiselle, D. Levine, P. J. Horwich, L. Barbaras, D. Siegal, K. Shillue, and D. Affeln. Repetitive stress symptoms in radiology: prevalence and response to ergonomic interventions. *Journal of the American College of Radiology*, 5(8):919–923, 2008.

[9] J. Cechanowicz, P. Irani, and S. Subramanian. Augmenting the mouse with pressure sensitive input. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1385–1394. ACM, 2007.

[10] A. Chan, K. MacLean, and J. McGrenere. Designing haptic icons to support collaborative turn-taking. *International Journal of Human-Computer Studies*, 66(5):333–355, 2008.

[11] D. R. Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B Methodological*, 34:187–220, 1972.

[12] A. Dix, J. Finlay, G. D. Abowd, and R. Beale. Human-Computer Interaction in Radiotherapy Target Volume Delineation: A Prospective, Multi-institutional Comparison of User Input Devices. *Journal of digital imaging*, 24:794–803, 2010.

[13] K. Doi. Current status and future potential of computer-aided diagnosis in medical imaging. *The British journal of radiology*, 78:S3–S19, 2005.

[14] B. Efron. The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, 72(359):557–565, Sept. 1977.

[15] T. K. P. Egglin. Context Bias: A Problem in Diagnostic Radiology. *JAMA: Journal of the American Medical Association*, 276(21):1752, Dec. 1996.

[16] N. Goyal, N. Jain, and V. Rachapalli. Ergonomics in radiology. *Clinical Radiology*, 64(2): 119–126, 2009.

[17] K. Hinckley and M. Sinclair. Touch-sensing input devices. *Proceedings of the SIGCHI conference on Human factors in computing systems CHI 99*, pages:223–230, 1999.

[18] K. Hinckley, E. Cutrell, S. Bathiche, and T. Muss. Quantitative analysis of scrolling techniques. *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world changing ourselves CHI 02*, 4:65, 2002.

[19] S. Kim, H. Kim, B. Lee, T.-J. Nam, and W. Lee. Inflatable mouse: volume-adjustable mouse with air-pressure-sensitive input and haptic feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 211–224. ACM, 2008.

[20] J. Kooyman, F. Gheorghe, and L. Oram. User input devices for radiology: Iterative prototypes, and design framework. Created as a final report for class project in CS543, UBC Vancouver, 2012.

[21] E. Krupinski, K. Berbaum, and R. Caldwell. Long radiology workdays reduce detection and accommodation accuracy. *College of Radiology*, 7(9):698–704, 2010.

[22] H. L. Kundel, C. F. Nodine, and D. Carmody. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative radiology*, 13(3):175–181, 1978.

[23] V. Levesque, L. Oram, K. MacLean, A. Cockburn, N. D. Marchuk, D. Johnson, J. E. Colgate, and M. A. Peshkin. Enhancing physicality in touch interaction with programmable friction. In *Proc. CHI'11*, pages 2481–2490. ACM Press, 2011.

[24] D. J. Manning, A. Gale, and E. A. Krupinski. Perception research in medical imaging. *The British journal of radiology*, 78(932):683–5, 2005.

[25] J. Matejka, T. Grossman, and G. Fitzmaurice. Swifter: improved online video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '13*, page 1159. ACM Press, 2013.

[26] A. G. Mathie and N. H. Strickland. Interpretation of ct scans with pacs image display in stack mode. *Radiology*, 203(1):207–209, 1997.

[27] I. Oakley, J. Ängeslevä, S. Hughes, and S. OModhrain. Tilt and feel: Scrolling with vibrotactile display. *EuroHaptics 2004*, pages 316–323, 2004.

[28] G. D. Rubin, J. K. Lyo, D. S. Paik, A. J. Sherbondy, L. C. Chow, A. N. Leung, R. Mindelzun, P. K. Schraedley-Desmond, S. E. Zinck, D. P. Naidich, and S. Napel. Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. *Radiology*, 234:274–283, 2005.

[29] A. J. Sherbondy, D. Holmlund, G. D. Rubin, P. K. Schraedley, K. Winograd, and T. Napel. Alternative Input Devices for Efficient Navigation of Large CT Angiography Data Sets. *Radiology*, 234:391–398, 2005.

[30] S. S. Snibbe, K. E. MacLean, R. Shaw, J. Roderick, W. L. Verplank, and M. Scheeff. Haptic techniques for media control. *Proceedings of ACM symposium on User interface software and technology UIST 01*, 3:199, 2001.

[31] J. O. Swanson, M. M. Thapa, R. S. Iyer, R. K. Otto, and E. Weinberger. Optimizing peer review: A year of experience after instituting a real-time comment-enhanced program at a children's hospital. *AJR. American journal of roentgenology*, 198(5):1121–5, 2012.

[32] E. Tsai. Force-sensing resistor matrix touch surface gesture recognition api. Created as a final engineering report EECE597, UBC Vancouver, 2013.

[33] T. Vertinsky and B. Forster. Prevalence of eye strain among radiologists: influence of viewing variables on symptoms. *American Journal of Roentgenology*, 184(2):681–686, 2005.

[34] N. Villar, S. Izadi, D. Rosenfeld, H. Benko, J. Helmes, J. Westhues, S. Hodges, E. Ofek, A. Butler, X. Cao, and Et Al. Mouse 2.0: multi-touch meets the mouse. In *Proceedings of ACM symposium on User interface software and technology*, volume 9, pages 33–42, 2009.

[35] D. Vogel and P. Baudisch. Shift. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, page 657. ACM Press, 2007.

[36] F. Wang and X. Ren. Empirical evaluation for finger input properties in multi-touch interaction. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, page 1063, New York, New York, USA, Apr. 2009. ACM Press.

[37] J. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 159–168. ACM, 2007.

[38] S. Zhai. User performance in relation to 3D input device design. *ACM SIGGRAPH Computer Graphics*, 32:50–54, 1998.

Appendix A

Supporting Materials

A.1 Consent Forms

Below you will find the consent forms used when interviewing the radiologists, as well as the generic consent form for the study with lay participants.

A.1.1 Radiologist Consent Form

PARTICIPANT'S COPY
CONSENT FORM

Department of Computer Science
2366 Main Mall
Vancouver, B.C. Canada V6T 1Z4
tel: (604) 822-3061
fax: (604) 822-4231

Project Title: Active-touch Form Factors to Support Radiologist Interactions with Volumetric Data

(UBC Ethics #H12-01672)

Principal Investigators: Dr. Karon MacLean; Associate Professor; Dept of Computer Science;
Student Investigator: Louise Oram; M.Sc. Candidate; Dept of Computer Science;

The purpose of this project is to evaluate the effectiveness of different prototype touch devices for interaction with volumetric data. The first session is an interview where you will be asked to look over a set of example tasks and rank them by different metrics, as well as give open-ended feedback on your interactions with the PACS system and mouse. The second session will involve trying basic diagnostic tasks using a prototype device and to provide feedback on your experiences during or immediately interacting with the device. You will also be asked to provide general demographic information such as your age and familiarity with touch screens.

REIMBURSEMENT:	Coffee shop gift card
TIME COMMITMENT:	2 × 30 minute session
CONFIDENTIALITY:	*You will not be identified by name in any study reports. Data gathered from this experiment will be stored in a secure Computer Science account accessible only to the experimenters.*

You understand that the experimenter will ANSWER ANY QUESTIONS you have about the instructions or the procedures of this study. After participating, the experimenter will answer any other questions you have about this study.

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without jeopardy. Your signature below indicates that you have received a copy of this consent form for your own records, and consent to participate in this study.

If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Info Line in the UBC Office of Research Services at 604-822-8598.

Version 1.0 / June 12, 2012 / Page 1 of 1

RESEARCHER'S COPY
CONSENT FORM

Department of Computer Science
2366 Main Mall
Vancouver, B.C. Canada V6T 1Z4
tel: (604) 822-3061
fax: (604) 822-4231

Project Title: Active-touch Form Factors to Support Radiologist Interactions with Volumetric Data
(UBC Ethics #H12-01672)

Principal Investigators: Dr. Karon MacLean; Associate Professor; Dept of Computer Science;
Student Investigator: Louise Oram; M.Sc. Candidate; Dept of Computer Science;

The purpose of this project is to evaluate the effectiveness of different prototype touch devices for interaction with volumetric data. The first session is an interview where you will be asked to look over a set of example tasks and rank them by different metrics, as well as give open-ended feedback on your interactions with the PACS system and mouse. The second session will involve trying basic diagnostic tasks using a prototype device and to provide feedback on your experiences during or immediately interacting with the device. You will also be asked to provide general demographic information such as your age and familiarity with touch screens.

REIMBURSEMENT:	Coffee shop gift card
TIME COMMITMENT:	2 × 30 minute session
CONFIDENTIALITY:	*You will not be identified by name in any study reports. Data gathered from this experiment will be stored in a secure Computer Science account accessible only to the experimenters.*

You understand that the experimenter will ANSWER ANY QUESTIONS you have about the instructions or the procedures of this study. After participating, the experimenter will answer any other questions you have about this study.

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without jeopardy. Your signature below indicates that you have received a copy of this consent form for your own records, and consent to participate in this study.

If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Info Line in the UBC Office of Research Services at 604-822-8598.

You hereby CONSENT to participate and acknowledge RECEIPT of a copy of the consent form:

Printed Name__________________________Date____________Signature_______________________

Version 1.0 / June 12, 2012 / Page 1 of 1

A.1.2 Generic Consent Form

PARTICIPANT'S COPY
CONSENT FORM

Department of Computer Science
2366 Main Mall
Vancouver, B.C. Canada V6T 1Z4
tel: (604) 822-3061
fax: (604) 822-4231

Project Title: Active-touch Form Factors to Support Radiologist Interactions with Volumetric Data
(UBC Ethics #H12-01672)

Principal Investigators: Dr. Karon MacLean; Associate Professor; Dept of Computer Science;
Student Investigator: Louise Oram; M.Sc. Candidate; Dept of Computer Science;

The purpose of this project is to evaluate the effectiveness of different prototype touch devices for interaction with volumetric data. In this study you will be asked to perform a series of simple tasks that involve interaction with the prototype(s), and to provide feedback on your experiences during or immediately after the study. You may also feel vibrotactile feedback when interacting with the device and be asked about your perception of this feedback. You will also be asked to provide general demographic information such as your age and familiarity with touch screens.

Your hands may be videotaped as you interact with the device. Videotapes will be used for analysis and may also be used for research presentations and videos. If visible, your face will be blurred. Your comments may also be recorded with an audio recorder. You have the option not to be videotaped or audio recorded. If you are not sure about any instructions, do not hesitate to ask.

REIMBURSEMENT:	$15
TIME COMMITMENT:	1 × 1.5 hour session
CONFIDENTIALITY:	*You will not be identified by name in any study reports. Data gathered from this experiment will be stored in a secure Computer Science account accessible only to the experimenters.*

You understand that the experimenter will ANSWER ANY QUESTIONS you have about the instructions or the procedures of this study. After participating, the experimenter will answer any other questions you have about this study.

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without jeopardy. Your signature below indicates that you have received a copy of this consent form for your own records, and consent to participate in this study.

If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Info Line in the UBC Office of Research Services at 604-822-8598.

Version 1.1 / July 11, 2012 / Page 1 of 1

RESEARCHER'S COPY
CONSENT FORM

Department of Computer Science
2366 Main Mall
Vancouver, B.C. Canada V6T 1Z4
tel: (604) 822-3061
fax: (604) 822-4231

Project Title: Active-touch Form Factors to Support Radiologist Interactions with Volumetric Data
(UBC Ethics # H12-01672)

Principal Investigators: Dr. Karon MacLean; Associate Professor; Dept of Computer Science;
Student Investigator: Louise Oram; M.Sc. Candidate; Dept of Computer Science;

The purpose of this project is to evaluate the effectiveness of different prototype touch devices for interaction with volumetric data. In this study you will be asked to perform a series of simple tasks that involve interaction with the prototype(s), and to provide feedback on your experiences during or immediately after the study. You may also feel vibrotactile feedback when interacting with the device and be asked about your perception of this feedback. You will also be asked to provide general demographic information such as your age and familiarity with touch screens.

Your hands may be videotaped as you interact with the device. Videotapes will be used for analysis and may also be used for research presentations and videos. If visible, your face will be blurred. Your comments may also be recorded with an audio recorder. You have the option not to be videotaped or audio recorded. If you are not sure about any instructions, do not hesitate to ask.

REIMBURSEMENT:	$15
TIME COMMITMENT:	1 × 1.5 hour session
CONFIDENTIALITY:	*You will not be identified by name in any study reports. Data gathered from this experiment will be stored in a secure Computer Science account accessible only to the experimenters.*

You understand that the experimenter will ANSWER ANY QUESTIONS you have about the instructions or the procedures of this study. After participating, the experimenter will answer any other questions you have about this study.

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without jeopardy. Your signature below indicates that you have received a copy of this consent form for your own records, and consent to participate in this study.

If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Info Line in the UBC Office of Research Services at 604-822-8598.

You hereby CONSENT to participate and acknowledge RECEIPT of a copy of the consent form:

Printed Name_________________________Date___________Signature______________________
Email _______________________________

Version 1.1 / July 11, 2012 / Page 1 of 1

A.2 Questionnaires

Below you will find the questionnaires used when interviewing the radiologists, the Likert scale questions were repeated for each task example.

You will also find the study questionnaire, which has a background questionnaire that was administered at the beginning of the study, and a ranking (of devices and annotation modality) and open ended feedback section that was administered at the end of the study. The Likert questions on the middle page were repeated for every combination of device and annotation modality (9-12 times depending on if the subject had time to complete the 4th device).

A.2.1 Radiologist Questionnaire

Open-ended questions at beginning:

How many years have you been practicing radiology?

Do you use any touchscreen devices? If so, name:

Have you experienced any issues of fatigue or pain from repetitive use of the workstation? Describe:

Describe the various radiology job settings you work or have worked in, and name PACS station types/brands used:

Over the last several months, in your radiology job settings and comparing to the full range of your image interpretation tasks:

How **important** is this task?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

How **frequently** do you perform this task?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

How **difficult** is this task?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

How well **supported** is this task, given the tools currently available to you?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

[The above is repeated for each of the 6 task examples]

Open-ended questions at end:

Are there any tasks you feel are **important** that are not mentioned here?

Are there any tasks you perform **frequently** that are not mentioned here?

What is one or more of the most **difficult** tasks you do?

Name any tasks you feel are **well supported** by the current system:

Name any tasks you feel are **not well supported** by the current system:

Name any mouse interactions you find particularly tiring or repetitive:

A.2.2 Study Questionnaire

Background Questionnaire Subject #:_______

Gender (circle one): M F Age: ________

Occupation (if student list department):________________________________

Do you use a **mouse**? If so, describe (# of buttons, shape):

How often do you use a **mouse** (within the last month)?

Frequently and/or for long periods every day	A few times per day	A few times a week	Less than once a week
☐	☐	☐	☐

Do you own any devices that employ **touch input** – e.g. touchscreens, touchpads? If so, name:

Do you frequently use any **touch input** devices that you do not own? If so, name:

Do any of these devices use vibrotactile display (e.g. buzzing when typing on a touch keyboard)? If so, please describe the feel and/or purpose:

How often do you use **touch input** devices (within the last month)?

Frequently and/or for long periods every day	A few times per day	A few times a week	Less than once a week
☐	☐	☐	☐

Have you experienced any issues of fatigue or pain from repetitive use of mice or touch input devices? If so please describe:

Interface: _____ Subject #: _____

Scrolling (for this set of trials):

How **accurately** did you feel you could scroll?

Very	Somewhat	Neutral	Not very	Not at all
☐	☐	☐	☐	☐

How **frustrated** were you when scrolling?

Very	Somewhat	Neutral	Not very	Not at all
☐	☐	☐	☐	☐

How **confident** did you feel when scrolling?

Very	Somewhat	Neutral	Not very	Not at all
☐	☐	☐	☐	☐

How much **attention** did you need to devote to scrolling?

A lot	Some	Neutral	A little	Very little
☐	☐	☐	☐	☐

Highlighting (for this set of trials):

How easy was it to **notice** the highlighting?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

How **helpful** did you feel the highlighting was?

Not at all	Not very	Somewhat	Very	Extremely
☐	☐	☐	☐	☐

Please rank which interaction device you preferred:
(Scroll wheel, Click & drag, Touch, Tilt)

Worst ______ ______ ______ ______ Best

Please rank which highlighting you preferred:
(Visual, Vibration, Both)

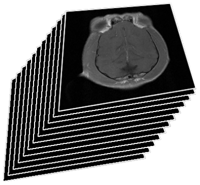
Worst ______ ______ ______ Best

Open-ended feedback (about devices, highlighting, the study in general):

Thanks so much for your time and participation in this study!

A.3 Study Protocol

Task Description:

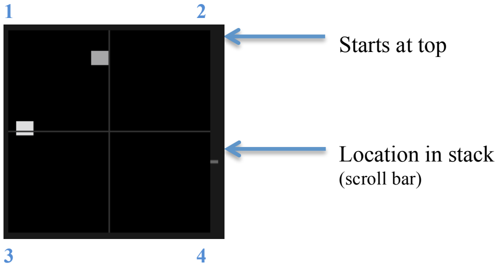


You will be using a program that mimics the task of finding targets (tumors) in a *stack* of radiology images.

- For each *stack* of images there is **always one target** (tumor).
- The target is a **perfect square**, which you will have to find among many rectangular distractors. It is also in the middle of the range of greys used.

The task is broken into the following steps:

1) Scroll through the *stack* looking for the target.
2) Once you have found the target, select the # key (1-4) corresponding to the quandrant the **majority** of the square is in.
3) The system will automatically take you to next *stack* to repeat the task.



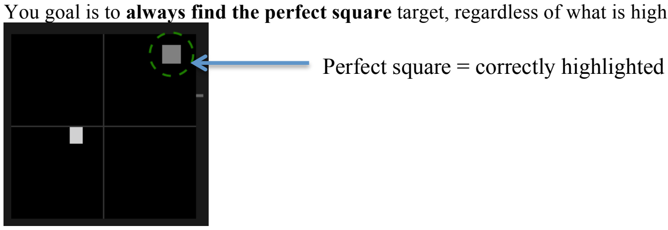
Highlighting:

- Often the target will be **found** and highlighted by the *automatic detection algorithm*
- Some of the time the target will be **missed** by the program. When it is missed the computer will highlight something else it thinks is correct (a false find).

You will do tasks with 3 types of highlights:

- **visual** around the target, as seen below
- **vibration** from the scrolling device as you approach the target (based on your scrolling speed).
- Both **visual** and **vibration**

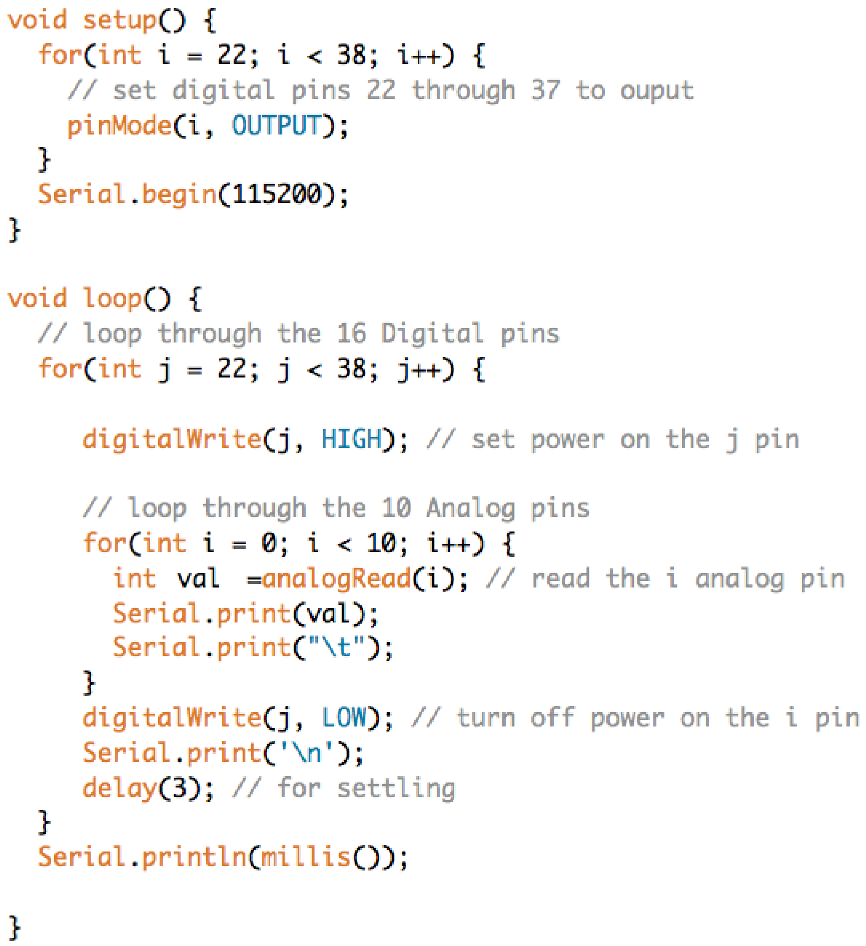
You goal is to **always find the perfect square** target, regardless of what is highlighted.



You will get the highest score by doing the task as ***quickly and accurately as possible.*** You will repeat the task several times for each highlight condition, as well as with different devices. The first trial of each set will be for practice.
If you have any questions do not hesitate to ask the experimenter!

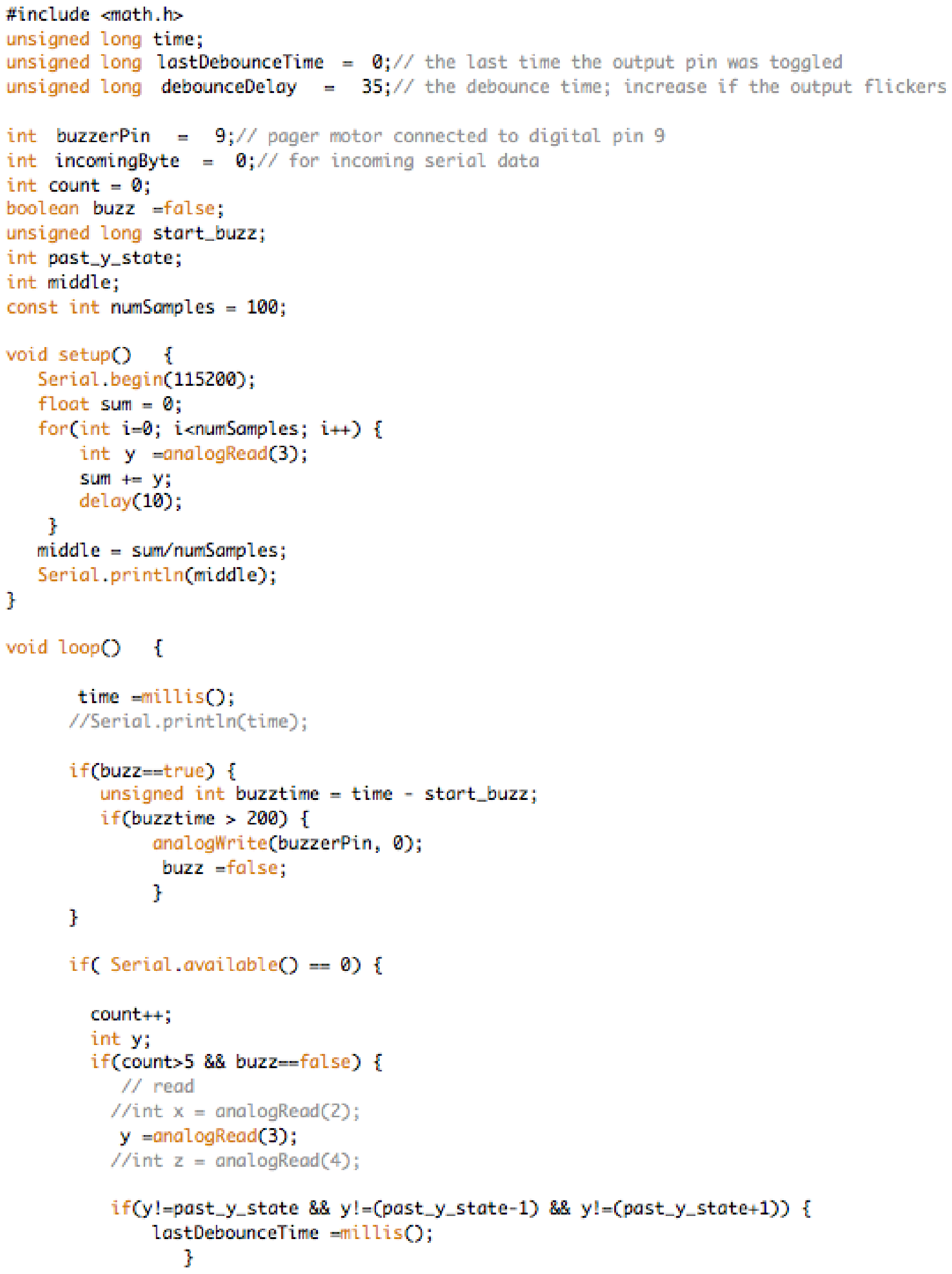
A.4 Arduino Code

A.4.1 Arduino Code for FSR Matrix

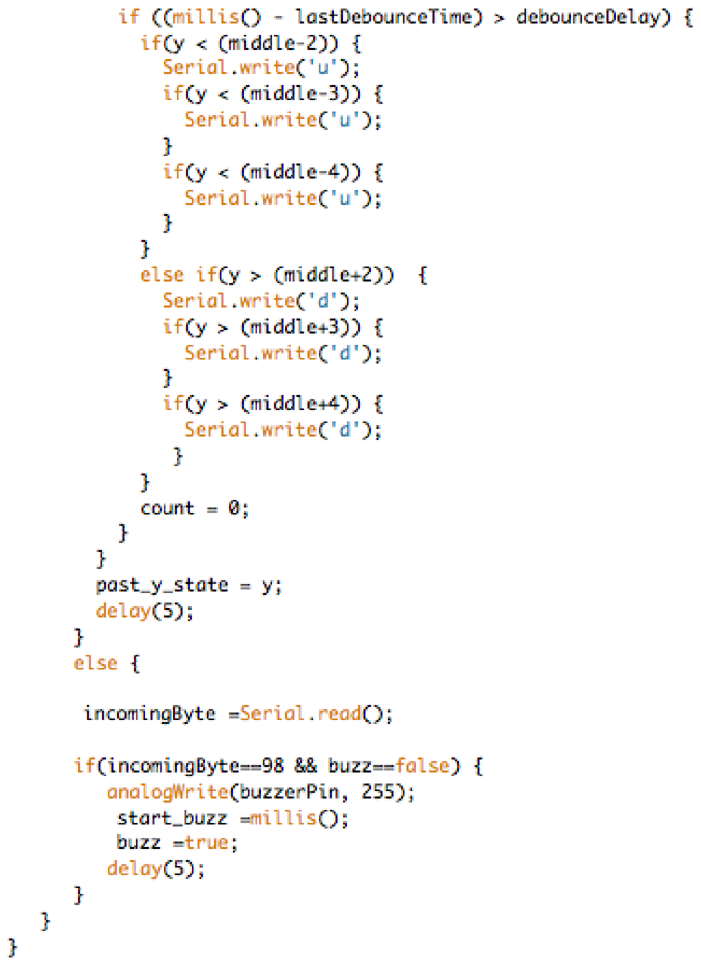


```
void setup() {
  for(int i = 22; i < 38; i++) {
    // set digital pins 22 through 37 to ouput
    pinMode(i, OUTPUT);
  }
  Serial.begin(115200);
}

void loop() {
  // loop through the 16 Digital pins
  for(int j = 22; j < 38; j++) {

     digitalWrite(j, HIGH); // set power on the j pin

     // loop through the 10 Analog pins
     for(int i = 0; i < 10; i++) {
       int val  =analogRead(i); // read the i analog pin
       Serial.print(val);
       Serial.print("\t");
     }
     digitalWrite(j, LOW); // turn off power on the i pin
     Serial.print('\n');
     delay(3); // for settling
  }
  Serial.println(millis());

}
```

A.4.2 Arduino Code for Pager Motor and Accelerometer



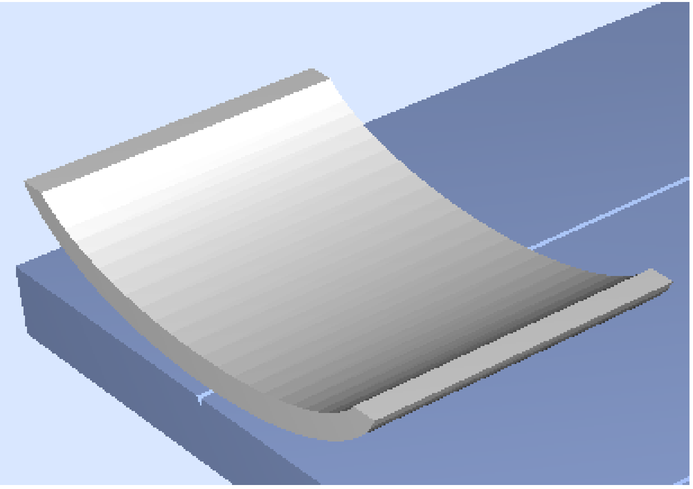
```
#include <math.h>
unsigned long time;
unsigned long lastDebounceTime  =  0;// the last time the output pin was toggled
unsigned long debounceDelay   =   35;// the debounce time; increase if the output flickers

int  buzzerPin   =   9;// pager motor connected to digital pin 9
int  incomingByte  =  0;// for incoming serial data
int count = 0;
boolean buzz  =false;
unsigned long start_buzz;
int past_y_state;
int middle;
const int numSamples = 100;

void setup()   {
   Serial.begin(115200);
   float sum = 0;
   for(int i=0; i<numSamples; i++) {
       int  y  =analogRead(3);
       sum += y;
       delay(10);
    }
   middle = sum/numSamples;
   Serial.println(middle);
}

void loop()   {

       time =millis();
      //Serial.println(time);

      if(buzz==true) {
         unsigned int buzztime = time - start_buzz;
         if(buzztime > 200) {
              analogWrite(buzzerPin, 0);
               buzz =false;
              }
      }

      if( Serial.available() == 0) {

        count++;
        int y;
        if(count>5 && buzz==false) {
           // read
          //int x = analogRead(2);
           y =analogRead(3);
          //int z = analogRead(4);

          if(y!=past_y_state && y!=(past_y_state-1) && y!=(past_y_state+1)) {
              lastDebounceTime =millis();
                 }
```



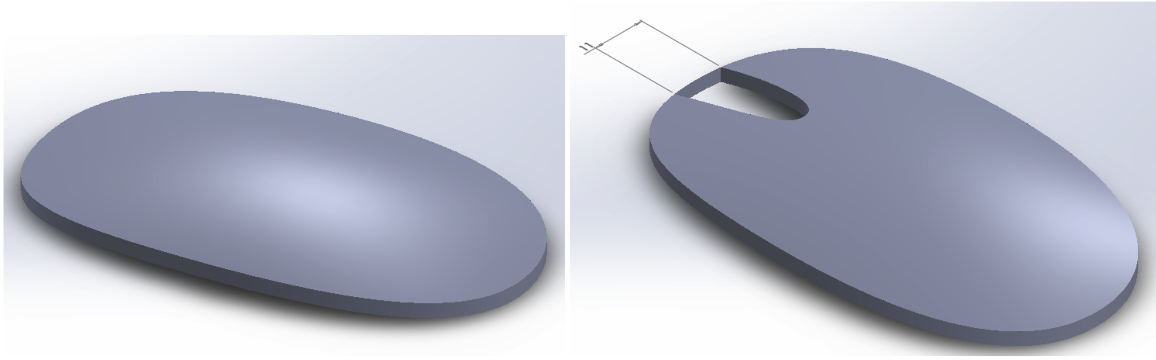
```
          if ((millis() - lastDebounceTime) > debounceDelay) {
            if(y < (middle-2)) {
              Serial.write('u');
              if(y < (middle-3)) {
                Serial.write('u');
              }
              if(y < (middle-4)) {
                Serial.write('u');
              }
            }
            else if(y > (middle+2))  {
              Serial.write('d');
              if(y > (middle+3)) {
                Serial.write('d');
              }
              if(y > (middle+4)) {
                Serial.write('d');
               }
            }
            count = 0;
          }
        }
        past_y_state = y;
        delay(5);
      }
      else {

       incomingByte =Serial.read();

      if(incomingByte==98 && buzz==false) {
         analogWrite(buzzerPin, 255);
          start_buzz =millis();
          buzz =true;
         delay(5);
      }
   }
}
```

A.5 3D printing

Image of the 3D printed rocker for the tilt prototype.



Images of the 3D printed tops for the tilt and click+drag/wheel prototypes, mimicking the the apple magic mouse top surface.



A.6 C++ Code

To open DICOM files I used the GDCM: Grassroots DICOM Library (which can be found at: http://gdcm.sourceforge.net/. The code to read the DICOM files can be found in the file: DCMimage.cpp. The program also relies on the Boost library, as well as several graphics libraries (GTK, GDK, & Cairo).

The serial port communication uses the Boost serial library, and the asynchronous implementation was used from: http://www.webalice.it/fede.tft/serial_port/serial_port.html. This code can be found in: AsyncSerial.cpp.

The implementation of an image stack can be seen in: Stack.cpp. There are 3 types of stacks that can be created: DCMstack (a DICOM stack), JpegStack (a stack of jpeg images), and DrawStack. DrawStack is the stack that was used for the experiment, and it draws the target and distractors on a black background.