The Effects of Integrating Kinesthetic Feedback, Force Feedback and Non-Speech Audio Feedback in Human Computer Interaction

by

Alice Suganthy Thomas

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Applied Science

in

Systems Design Engineering

Waterloo, Ontario, Canada, 1998

0-612-32872-4

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

This research examines combining kinesthetic and force feedback in input devices with non-speech audio feedback (earcons) in interfaces for "everyday" computing tasks. A study involving human participants was designed to test the effects of this type of multi-modal feedback in "point and select" computer tasks. This was a between-subject experiment (3 mouse-feedback groups) with repeated measures (2 x 2). The mouse-feedback groups were no feedback (using a normal mouse), kinesthetic feedback, and kinesthetic and force feedback. Repeat measures were sound condition (on and off) and task condition (single or dual). All participants performed a series of trials on a custom-built Feedback Calculator application that also recorded statistical data on each trial. Objective measures such as speed and accuracy as well as the NASA TLX workload measures along with Subjective Questionnaire responses were used for the analysis.

The results confirmed performance differences due to single and dual task conditions. Trial times increased ($p < 0.001$) and error rates increased ($p < 0.001$) for dual task conditions compared to single task conditions. However, results suggest that performance in the two task conditions were handled differently by each mouse-feedback group based on task difficulty. The results show that a mouse with kinesthetic and force feedback is useful in improving speed and accuracy in dual task conditions which required increased visual information processing when compared to single task conditions ($p < 0.033$). In single task conditions, the multi-feedback mouse was also useful in improving accuracy while maintaining the same speed as a normal mouse.

For single task conditions, auditory feedback did not significantly improve performance for any of the groups. However, in dual task conditions, participants using the normal mouse made use of auditory feedback to a greater level than participants using the kinesthetic mouse or kinesthetic and force feedback mouse. The results also indicate

that using a kinesthetic mouse was not as effective as a kinesthetic and force feedback mouse for both simple tasks and difficult tasks. Nonetheless, participants did not make use of the proactive auditory feedback provided by the earcons.

The workload analysis showed that there were significant differences between the feedback conditions on multiple factors. Participants using the kinesthetic and force feedback mouse had significantly higher physical demand ratings ($p < 0.016$). Further, participants rated the force feedback mouse as having less ease of movement ($p < 0.001$). Subjective measures indicated that participants found the trials with sound less mentally demanding ($p < 0.001$), less temporally demanding ($p < 0.001$), and took less effort ($p < 0.001$) to complete than trials without sound even though there were no significant differences in terms of actual performance (time and accuracy).

This research has examined one combination of multi-modal feedback that has not been studied before. The results were encouraging and suggest that feedback can be used to improve various aspects of human computer interaction. However, this research has found that the effects of feedback can be difficult to interpret because it influences performance, workload measures, and user preference measures differently. Further work in this area could contribute to a better understanding of the appropriate use of multi-modal feedback in a variety of different task situations.

Acknowledgements

I would like to thank my supervisor, Dr. Carolyn MacGregor for her continuous friendship, guidance, mentorship, and financial support which were crucial in bringing this thesis to completion.

I would like to thank Control Advancements Inc. who provided the Virtual Reality Mouse for this study. Additionally, I am grateful to Dr. Eric Kubica and Dan Madill of Control Advancements Inc. for their assistance and support throughout this project.

I would also like to acknowledge my readers, Dr. Tom Carey and Dr. Daniel Stashuk for reviewing my thesis and providing valuable feedback.

I would like to express my appreciation to the University of Waterloo Statistical Consulting Service, for reviewing my experimental design and data analysis.

I would like to thank David Walsh and Kevin Kraul for providing technical assistance and computing equipment when needed.

I would like to thank Apa, Amma, John, Marisa and Fern for their love and support. Finally, I would like to thank my husband, for his help, continuous encouragement, friendship, and love.

Trademarks

Pentium is a registered trademark of Intel Inc.

Microsoft Word, Microsoft PowerPoint, Microsoft Windows NT and Microsoft Windows '95 are registered trademarks of Microsoft Corporation.

Virtual Reality Mouse is a registered trademark of Control Advancements Inc.

Haptic Mouse is a registered trademark of Haptic Technologies Inc.

UltraSound MAX is a registered trademark of Advanced Gravis Computer Technology Ltd.

Phantom Haptic Device is a registered trademark of SensAble Technologies Inc.

Force Mouse is a registered trademark of Force Agent.

Rockin' Mouse is a registered trademark of Alias | Wavefront Inc.

Any other trademarks referred to in this thesis but that have not been acknowledged, are omissions and should be given their proper acknowledgement.

Dedication

To my grandparents: John and Helen Thomas, and Thambiraj and Leelavathy Simpson.

Table of Contents

List of Tables

List of Illustrations

Chapter 1

Introduction

1.1 Overview

This thesis examines the use of multi-sensory feedback to improve human computer interaction. The need for this research has emerged in response to the development of various technologies that make use of multiple modalities including force feedback and audio feedback. Multi-modal systems try to exploit the multi-sensory nature of humans by integrating two or more different interactive modalities (e.g. sound, graphical interaction, force feedback, gesture) into simultaneous use. So far, multi-modal technologies have addressed the needs of various niche markets including: virtual reality applications (Buxton 1996; Gunther 1997), telerobotics, computer accessibility for the blind (Bliss, Katcher, Rogers, and Sheppard 1970), games and simulations, electronic musical instruments, and navigation equipment. Industry and researchers are exploring the use of multi-modal feedback for personal computing and are developing new and innovative devices. Yet, there appears to have been very little scientific research conducted to study the effectiveness, advantages, and disadvantages of using multi-modal technologies for every day computer related tasks from a human performance standpoint.

People experience multi-modal feedback in many aspects of everyday life. For example, a car driver knows when to shift gears just by hearing the different sounds from the engine. Additionally, the driver knows when the intended gear has engaged simply by the movement and feedback from the gear selector stick, and subsequently, the sound of the engine. Visual feedback could have been used to perform the same gear-shifting task (by looking at the tachometer and the final placement of the stick shift). In this case, the visual modality was freed up and allowed the driver to navigate, keep track of stop signs, watch out for pedestrians and other cars, with less disruptions. Thus, the driver has benefited from the multi-modal feedback from the car. Using similar logic, it follows that using multi-modal feedback in computer systems may benefit the user in many ways. Thus, many researchers are exploring different ways of using the multi-sensory capabilities of humans (Gaver, 1989; Akamatsu and MacKenzie 1996). Buxton (1991) writes that the use of multi-sensory, multi-channel, and multi-tasking in interfaces will allow for a more holisitc and user-centered system.

Adding kinesthetic and force feedback characteristics to mice provides an additional sensory feedback directly to the hand. The use of this feedback type in input devices is a recent development but is growing in popularity. Microsoft (1996) recently purchased Exos Inc., a large producer of force feedback devices for its input devices group and has developed a force feedback API (application programming interface) for Windows 95 games. The Virtual Reality Mouse (VRM) developed by Control Advancements Inc. is another example of a mouse that provides kinesthetic and force feedback characteristics. The VRM was originally developed in 1994 by Daniel Madill, a PhD student in Electrical and Computer Engineering, and Kevin Krauel from the University of Waterloo as part of Madill's PhD thesis. The VRM is kinesthetic and exerts force feedback to the user depending on where the cursor is placed on an object on the screen. This technology has been specially targeted for people with visual disabilities. The VRM has been already added (1998) to a list of aids for disabled people approved by the Ontario Assistive Devices Program (ADP). Additionally, there is potential for the VRM to be used as an input/output device for "everyday" computing by the sighted population of users. Most likely, this will only happen if users perceive a direct benefit to using the VRM or another feedback device over a normal mouse. This is one of the reasons a scientific evaluation is necessary and needed. Quantitative and qualitative measures can help determine the usability and applicability of the VRM for the general population of computer users.

Adding sounds to the interface (activated by the mouse cursor) provides additional audio feedback to the user. Auditory feedback does not require the user to focus on something to hear to the extent that visual feedback requires users to focus and pay attention to a display. Researchers and developers have tried to make more use of the audio channel in computing systems. Earcons are one way to integrate sounds in interactive systems. Blattner (1989, p. 13) was the first researcher to propose the concept of earcons. Blattner defines earcons as "non-verbal audio messages that are used in the computer/user interface to provide information to the user about some computer object, operation or interaction". They are typically based on musical sounds. Studies of earcons by Brewster, Wright & Edwards (1992, 1993, 1994) have shown that they are an effective means of communicating information. Even though most computing systems already come with audio capabilities, audio feedback has not been fully integrated with the interfaces of software applications.

A multi-modal system that integrates visual feedback, kinesthetic and force feedback in pointing devices and non-speech audio feedback in user interfaces via the mouse cursor could improve "everyday" computing as we know it today. However, experimental methods should be used to verify that this type of multi-modal system would in fact help users perform tasks faster, more efficiently, and easily. Currently, only certain games and simulators have integrated both sound and force feedback in human machine interaction.

There appears to be no studies that have been carried out that test this combination of multi-modal system using human performance and preference measures. There may be many advantages and disadvantages in using multi-modal feedback computing systems. These systems may even help reduce injuries associated with computing tasks such as repetitive strain injuries. Comparing the effects of force feedback, and audio feedback with visual feedback alone will allow researchers to understand the impact of each on human performance, and help determine where specific feedback mechanisms may be useful, effective, and desired.

1.2 Overall Research Objectives

The main purpose of this research is to evaluate and compare the effects of visual feedback alone to visual feedback augmented by kinesthetic feedback, kinesthetic and force feedback, and non-speech audio feedback using human performance measures such as speed and accuracy, perceived workload, and preference measures. Secondly, to evaluate whether combining the feedback types will improve speed and accuracy. Thirdly, to determine whether using additional feedback modalities reduces the demand on the visual modality and allows users to process more information. The thesis is concerned with studying the effects of the feedback types. It does not explore issues associated with the thresholds or ergonomic factors such as earcon qualities (loudness, frequency, pitch) or force properties (amount, type) or anthropometrics of the devices.

1.3 Thesis Organization

The thesis is organized as follows:

- Chapter 2 reviews background material covering human computer interaction and human information processing. Additionally, it provides a review of studies that are concerned with non-speech audio feedback and force feedback in computing applications.
- Chapter 3 presents the research objectives and discusses the development of a research tool that was specially designed by the author for studies with human participants. The research tool is an online Calculator that has custom built features to display test data, track time to enter data into calculator, record cursor movement and cursor actions, save entries, track errors, and provide audio feedback in the form of earcons. The Calculator application represents tasks that are inherent in all graphical user interfaces, like pointing and selecting. It was built to test the effects of multi-modal feedback on representative "everyday" computing tasks.
- Chapter 4 describes a study that was undertaken as part of the thesis to evaluate feedback types. Participants were asked to perform a set of trials that involved selection and navigation tasks on the Calculator application using different input devices. Objective and subjective measures were collected.
- Chapter 5 presents the results of the study. The results of the section aim to answer specific research questions: does multi-modal feedback improve performance; do kinesthetic and force feedback in input devices improve performance; and, does audio feedback improve performance.
- Chapter 6 contains a detailed discussion of the major results and limitations of this research.
- Chapter 7 presents the conclusions and recommendations for the possibilities of future research.
- Appendix A contains all experimental forms used in the study described in Chapter 3.
- Appendix B contains the statistical results from the study used in the Results section.

Chapter 2

Background

2.1 Human Computer Interaction

Human computer interaction (HCI) is defined as "a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" (SIGCHI 1992, p.6). The field of ergonomics is concerned with the interaction between human and machine. Ergonomics has been defined as "the study of the physical characteristics of the interaction: how the controls are designed, the physical environment in which the interaction takes place, and the layout and physical qualities of the screen" (Dix, Finlay, Abowd, Beale, & Finley 1998, p.110).

Baecker, Grudin, Buxton & Greenberg (1995; Chapter 7) identify the need to focus research on improving HCI by exploring new and appropriate sensory modalities of interaction. Buxton (1991) states that in order for design of computing systems to be user-centred and holistic, it should be multi-sensory, multi-channel, and multi-tasking. He defines each as (1991, p, 6):

- multi-sensory: design that utilizes multiple sensory modalities;
- multi-channel: design that utilizes multiple channels, of the same or different modalities;
- multi-tasking: design that recognizes that (as driving a car demonstrates) humans can perform more than one task at a time.

Figure 2-1 illustrates a simplified model of the relationship between human and machine (MacKenzie 1995).

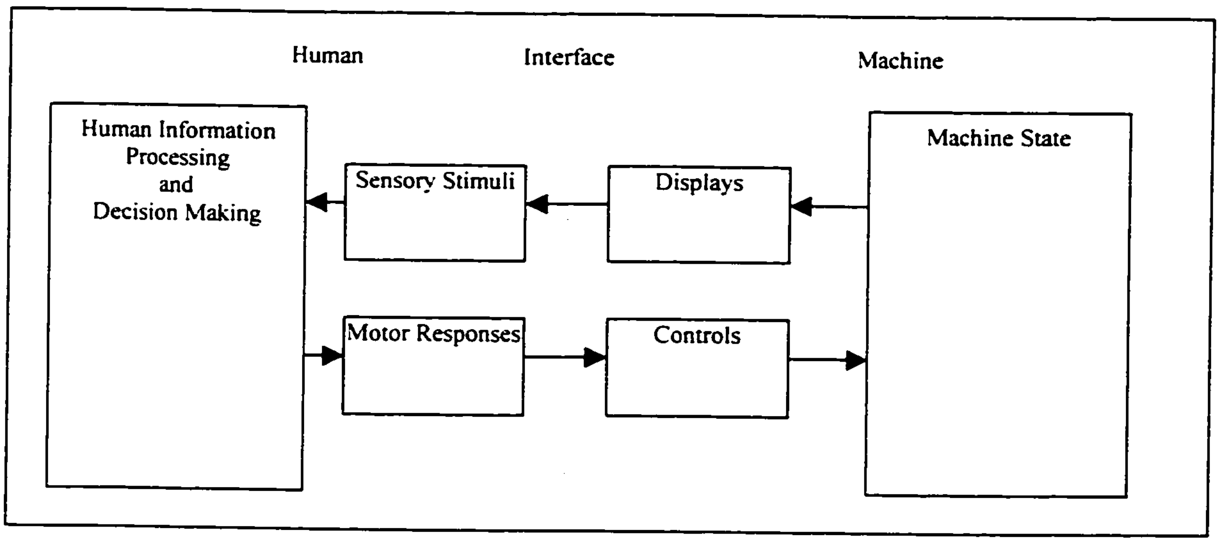


Figure 2-1. A simplified model of the closed-loop human-machine system (MacKenzie 1995, p.437).

When interacting with a computer, the user receives information that is *output* by the computer and responds by providing *input* back to the computer (Dix et al. 1998). Each has unique capabilities and limitations. The computer's *output* becomes the user's *input* and vice versa. The interaction between human and machine forms a closed-loop system. The two main components in the human machine system are "Human Information Processing and Decision Making", and "Machine State". The human information processing component is critical to processing the information that is presented by the machine. The machine uses displays (e.g., monitors, 3-D heads-Up displays, stereoscopic sound, etc.) to present the information to the human. The human information processor receives the information as stimuli and processes it. When a decision is made, the actions are carried out by human motor responses using the controls (e.g., input device, dials, keyboard entries, etc.) provided by the machine. The interaction between human and machine takes place at the interface (dotted line in Figure 2-1).

2.2 Role of Feedback in Human Information Processing & Decision Making

The Human Information Processing System is complex. According to Wickens (1992), information processing is said to occur in stages: sensory processing, perception, decision and response selection, response execution, and attention. Basically, stimuli are first received by various senses via sensory receptors and stored in the short-term sensory system (STSS). It is then processed by the memory systems: working memory and long term memory. Working memory allows humans to calculate, examine, compare and evaluate different mental representations. Working memory and long term memory work together to process the information and make decisions. Once the decision is made, the user makes a response. A model of human information processing can be seen in Figure 2-2. Attention is required for all the stages following the STSS. The model of human information processing presented in Figure 2-2 describes the critical stages of information processing as it pertains to human performance (Wickens 1992). The model provides a framework for understanding human information processing and should not be taken literally.

Each sense (vision, hearing, touch, taste, and smell.) is hypothesized to have its own *sensory memory*. Sensory memory retains information only for a brief period of time. In terms of HCI, visual sensory memory and auditory sensory memory are very important. Visual sensory memory is very brief and auditory sensory memory is longer in duration (Wickens 1992). According to Wickens, "visual sensory memory...is so short as to be unimportant in engineering design" (p.213).

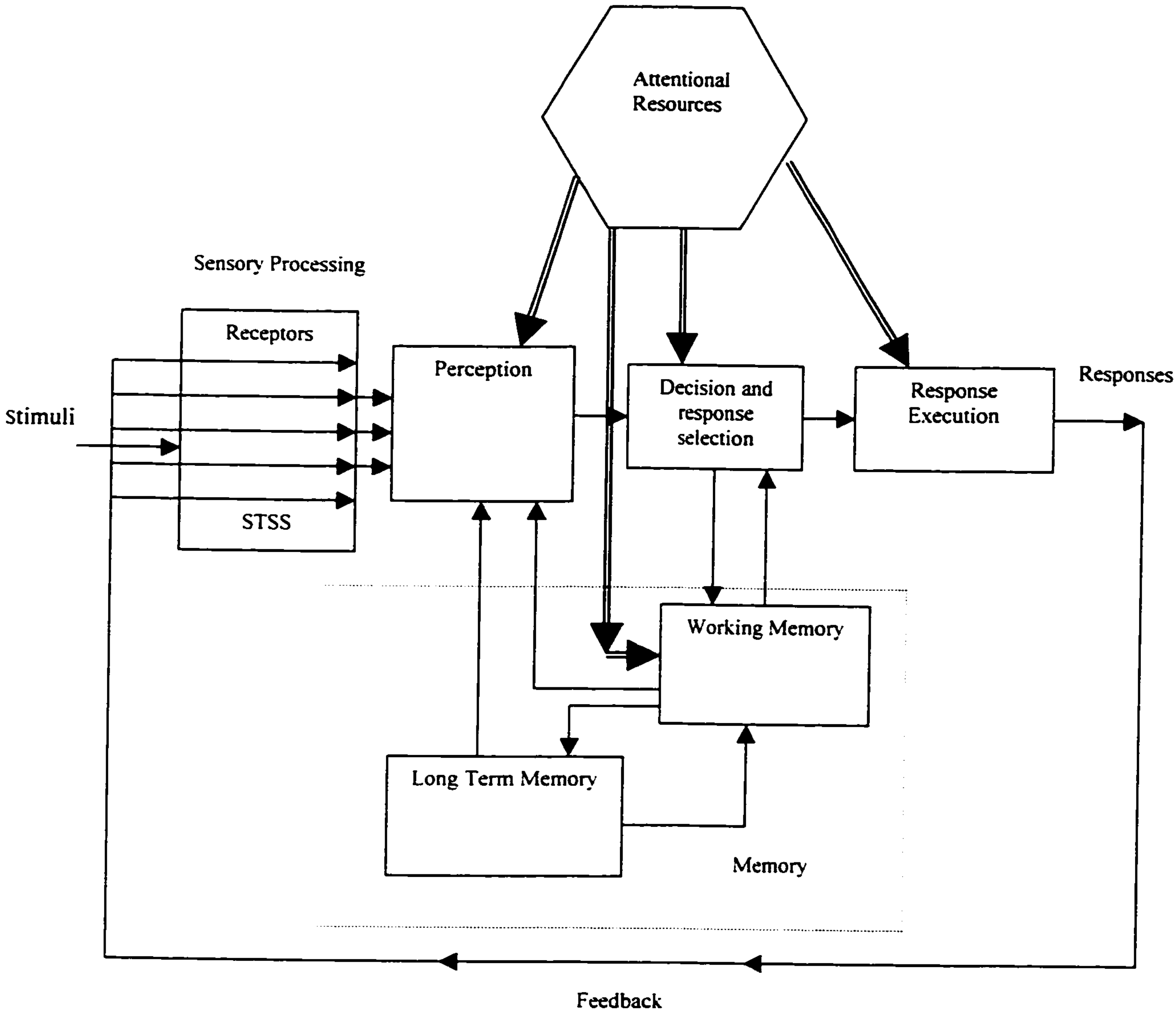


Figure 2-2. A model of human information processing (Wickens 1992, p.17).

Feedback plays an important role in human-computer interaction because it provides information to the user about "what action has actually been done, what result has been accomplished" (Norman 1988, p.27). Feedback is "the use of sensory information to change the course of action" (Douglas & Mithal 1998, p.48). Thus, feedback governs the initiation and completion of movement. Within the traditional framework of human-computer interaction, the visual modality has primarily provided information required for navigation and interaction with the machine. For example, cues of cursor position on screen, current mode, and what the user has entered are generally provided by the visual sense.

According to Sellen, Gordon and Kurtenbach (1992), feedback in HCI is based on presentation, delivery and control. Feedback can be presented in a *proactive* or *reactive* manner (Sellen et al. 1992). Proactive feedback occurs when the feedback is provided before the action takes place. For example, when the cursor passes a button, the button label appears automatically with a description of the function, and is commonly referred to as "touch sensitive help". Reactive feedback occurs after the user selects the button. For example, when the user selects a button, a beep is heard.

Feedback can be delivered in either a transient or sustained fashion. For example, a simple short beep (transient) can indicate that the cursor is over a button as opposed to a continuous beep or sound that can be used to indicate that the cursor is still over a button (sustained).

Control refers to how the system mode is maintained and controlled. In their experiment, Sellen et al. (1992) studied the effects of user maintained mode through the use of foot pedals that provided kinesthetic feedback. Participants were required to actively maintain insert mode by depressing the foot pedal for entry into the *vi* editor.

2.3 Feedback Modalities

As mentioned, humans have five major senses that include: vision, hearing, touch, smell and taste. Of these, vision, hearing, and touch are central and are most exploited in human computer interaction. The following discussion focuses on the three modalities most directly related to HCI and briefly summarizes the most important aspects of the sense as it relates to HCI.

2.3.1 Visual Feedback

Human vision is very complex "with a range of physical and perceptual limitations, yet it is the primary source of information for the average person" (Dix et al. 1998, p. 14). The visual modality provides most of the sensory feedback when interacting with the computer. For example, a normal mouse requires the visual sense to maintain awareness of cursor position on the screen. The mouse software translates the movements of the mouse made by the user, into spatial coordinates on the computer screen as seen by the mouse cursor. The mouse cursor provides no additional cues to the user about its current position. In order to assist the users in maintaining an awareness of cursor position on the screen at all times, additional features can be activated at the user's discretion. For example, enlarging the cursor pointer arrow, adding a mouse trail, or using a blinking cursor. However, in all cases, the user must still visually monitor where the cursor is positioned, and whether the target location has been entered before initiating an action. Further, visual information is usually withheld until the user initiates an action (e.g., by pressing a button). For example, the "Drawing" toolbar in Word for Windows appears only when the corresponding icon is selected. This is common in many applications.

The lack of screen space and the need to reduce information overload for a user, are major reasons for removing the number of options and information on the screen at one time. Less used functions and options are rarely seen on the start-up screen of an application by default. Additionally, the use of dialogs, menu options, and secondary windows are methods used by designers to reduce the clutter on the screen because the visual modality is primarily used for all aspects of search and selection. Users are forced to find less-used functions from a multitude of secondary menus and dialogs by using the Help system, scanning the contents, or using a search feature.

The extensive use of color in user interfaces is common. Color has been used in a variety of ways. It has been used to provide warning, act as a secondary cue, provide mappings, show patterns, and categorize screen elements. Again, the strategy of designers, in order to assist the user to process visual information more quickly has been to use the capabilities of the visual modality rather than to use alternative modalities to provide groupings, clear mappings, show patterns, or categories.

2.3.2 Auditory Feedback

Auditory perception is important for communication in everyday life. We use both speech and non-speech sounds to communicate with other people directly (face to face communication) and indirectly (use of telecommunications equipment like the phone). Integrating non-speech sounds in interactive systems is currently being investigated. In the real world, non-speech sounds convey information about physical events (e.g., pen dropping), invisible structures (e.g., determining where a nail can be inserted), dynamic changes (e.g., as cola fizzles), abnormal structures (e.g., equipment malfunction) and about events in space (e.g., knock at the door). Auditory feedback is an integral part of communication. So far, the auditory modality has been moderately used in HCI. Researchers believe the use of non-speech sounds in interfaces can improve interaction with the computer (Buxton 1989; Mountford & Gaver 1990; Brewster 1997, 1998). Meanings of non-speech sounds are learned by the user rather than known instantly. For example, a person learns that a telephone ring indicates someone is calling on the phone.

O'Malley (1995) has categorized three types of non-speech sounds that are frequently used in human machine systems: alarms and warning systems; status and monitoring indicators; and encoded messages and data. Alarms and warnings indicate a problem or impending problem with the current system state. Status and monitoring indicators are used to convey information about ongoing processes. Finally, messages can be sent to the user in the form of audio files.

The main difference between the auditory sense and other senses is that audio is ubiquitous. Users do not have to focus on objects on the screen or feel items using their tactile feedback. Users hear from all around. This makes the auditory channel very powerful.

2.3.3 Tactile, Kinesthetic, and Force Feedback

All mice provide visual feedback on the computer screen and proprioceptive feedback through tactile and kinesthetic sensations. Proprioceptive information is derived from both the signals from muscles, tendons, and joints and from the signals issued by the brain to the muscles; it provides an awareness of body postures and changes in posture as well as of movement through space (Clark and Horch 1986). The term "tactile" refers to the sense of touch, while "kinesthetic" information refers to a sense of position, motion and force (Brooks 1990). MacKenzie (1995) states:

> *"The act of grasping an object such as a mouse issues forth a tactile sensation in addition to the kinesthetic sensation obtained through muscle and joint receptors. This is important, for example, to confirm touching the object, and in this sense tactile information is a substitute for visual information" (p. 816).*

According to Shimoga (1993), there are four different types of touch receptors (or mechanoreceptors) present in the human hand. The output response of each receptor decreases with time for any stimulus and is called "stimulation adaptation". The 2-point discrimination ability is very important. The index finger pulp is able to sense all points with a distance of more than 2 mm whereas in the centre of the palm two points that are less than 11 mm apart feel like only one. Further, the amplitude and the vibration frequency of the stimulus are also important for discrimination ability. In the case of the mouse, users typically use the whole palm of the right hand to grasp the device. The palm's sensitivity to small forces made by the mouse is generally low because the whole palm is used to move the device.

The term force feedback can be confusing. Force is required to move all objects that have mass. When force is applied, it provides resistance of some sort. There are four types of resistance: elastic (i.e., spring loading), static or sliding friction, viscous damping, and inertia (Douglas & Mithal 1998). When using a pointing device, people typically interpret the sliding friction and inertia from a pointing device as the 'feel' of a device. Since all moveable objects fit this description, a better description of force feedback is required. For this thesis, force feedback is defined as the controlled force exerted by an object.

2.4 Interaction Styles in HCI

There are a variety of ways that a computer system supports dialog with the user. The type of dialog between the computer and the user is known as the interaction style. Some of the most frequently used interaction styles are:

- Command line interface - enter instructions to the computer directly
- Menus - use mouse or alphanumeric keys to select from a set of options to execute selection
- Question/answer and query dialog - user is asked a series of questions in order to lead the interaction
- Form-fills and spreadsheets - primarily used for data entry

- Direct manipulation - use of windows, icons, menus, pointers, pull down menus otherwise known as Graphical User Interfaces (GUIs) via a mouse or keyboard for interaction

The most common interaction style for personal computing today is Direct Manipulation (Schneiderman 1982). Generally, direct manipulation with GUIs involves the interaction with objects on the screen in a rapid, reversible, and incremental fashion. Preece, Rogers, Sharp, Benyon, Holland, S. & Carey (1994) describe a growing trend in output techniques or medium to include multimodal feedback capabilities. According to Preece (Preece et al. 1994), the use of multi-modal feedback types (visual, auditory, kinesthetic, or force feedback) can improve the usability of the GUI.

More and more GUIs are being designed in a process with the user in mind, commonly referred to as "user-centered design". This is done using a variety of usability design techniques. Generally, only visual feedback is considered in the design of GUIs and interactions. Figure 2-3 illustrates the number of functions that are potentially available to the user at one time in a popular application like MS Word. This example illustrates how dependent computing systems are on the use of the visual modality to processing information. A solution to reduce the demand on the visual modality is to make use of multi-modal systems. There are potentially many ways to integrate various feedback types in user interfaces and input devices. For example, one possibility is for designers to consider using sound to categorize icons or present related information. Another example may be to consider using force feedback to indicate the edges of windows so users do not move their cursor past the working window inadvertently.

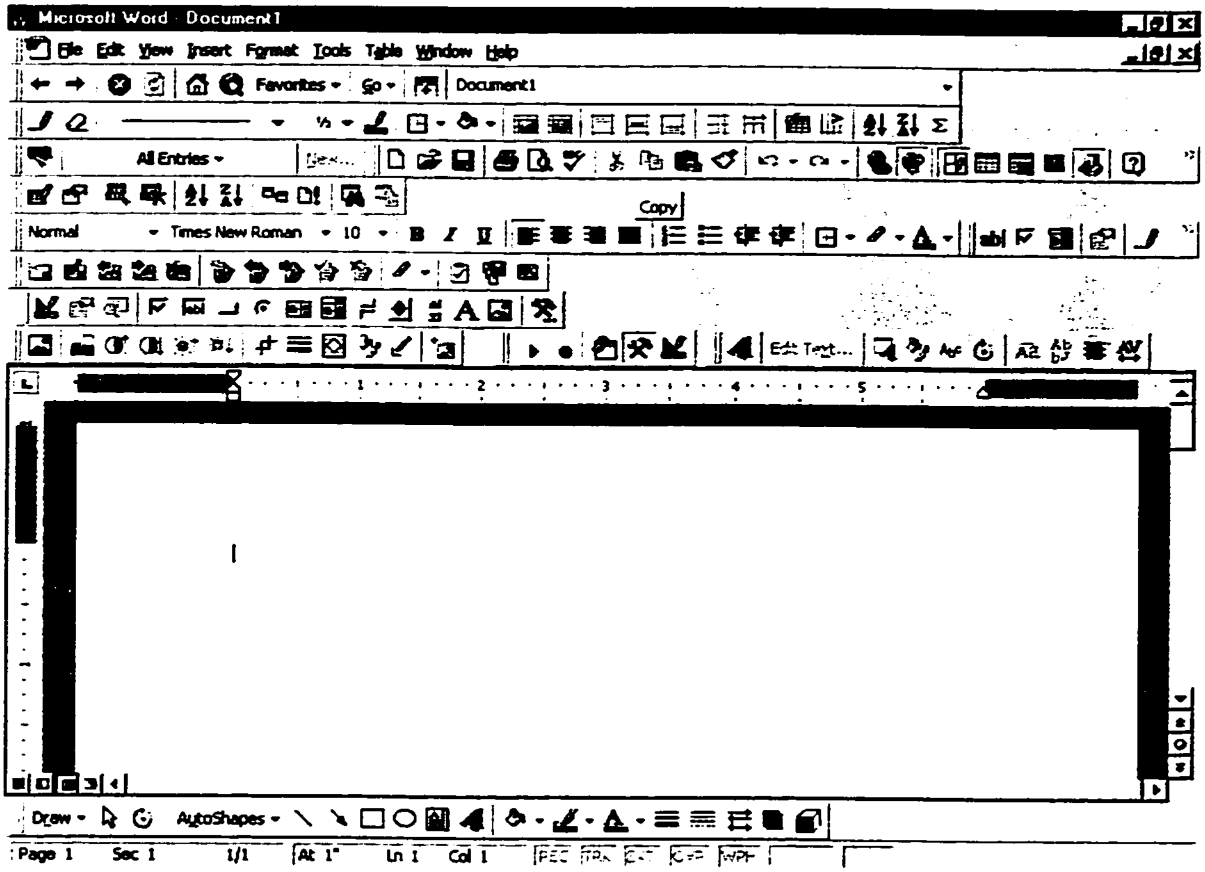


Figure 2-3. Screen shot from Word for Windows interface with all toolbars displayed.

Direct manipulation using GUIs is currently the most common interaction style. The fundamental set of operations that generally make up any task on a GUI, regardless of the application software was identified. This fundamental set of operations forms the task taxonomy for a particular interaction style. Douglas & Mithal (1998) have suggested the taxonomy for GUIs consist of pointing and selection, dragging, drawing, and device switching.

Pointing and selecting was stated as the most common operation performed with GUIs (Douglas & Mithal 1998, p.52). Generally speaking, the cursor is moved from one point on the display to a target object on the screen (e.g., button, icon, graphical object, window) and users click on it using a button on the input device.

Dragging involves moving an object on the screen from one position to another target position. Generally, users hold down the mouse button to select the object and move it. The mouse button is released when the target position is reached in order to drop the object. Drawing is important for graphical tasks such as CAD/CAM. Again, this is accomplished with a pointing device using the path of the cursor movements for tracing.

Drawing is a complex task in comparison and is governed by the software used. Device switching refers to physically moving a body part (i.e., the hand) from one device to another. This occurs frequently when a user switches between the mouse and a keyboard.

Another aspect of the direct manipulation interaction style is the idea of discrete, continuous, and semi-continuous tasks (Douglas & Mithal 1998). Discrete pointing tasks are distinguished by pauses between action. Continuous pointing tasks are performed without pauses. Semi-continuous pointing tasks are considered to be a series of continuous tasks. For example, selecting an item in a menu involves first selecting the main menu, moving down to the item, and clicking on it. Depending on the objective of the task at hand, pointing and selecting tasks can be considered either discrete or continuous. A study that uses a generic test task must then include both aspects of direct manipulation.

2.5 Usability

2.5.1 Human Performance Measures

Usability is an ambiguous term. Researchers and human factors professionals have differing definitions for what it means and how it should be measured. The ISO 9241 standard defines usability as the:

> *"Effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments. In turn, effectiveness is the accuracy and completeness with which specified users can achieve specified goals in particular environments; efficiency is the resources expended in relation to the accuracy and completeness of goals achieved; and satisfaction is the comfort and acceptability of the work system to its users and other people affected by its use"*
>
> *(ISO 9241, as cited in Douglas & Mithal 1998, p. 3).*

According to Douglas & Mithal (1998), the ISO 9241 definition is inadequate because it addresses usability in a broad sense. It fails to provide explicit measures with which researchers can accurately observe and measure human behaviour. With respect to evaluating pointing devices, Douglas & Mithal advocate using more precise human performance (direct) measures like time to learn to use device, error rates during learning, practiced task time, and practiced error rates when evaluating or comparing devices. Additionally, indirect measures such as physical comfort, muscular fatigue and stress, and user acceptability based on subjective and qualitative measures should also be used to evaluate the usability of pointing devices. Ergonomics research in the area of pointing devices has found that there is "a complex interaction of factors which determine human performance directly and attitude indirectly" (ibid, p. 5). Thus, experimental methods that collect both *direct* and *indirect* measures are the most comprehensive way to evaluate pointing devices and interfaces.

As mentioned, the usability of a pointing device should take into account *indirect* measures such as physical comfort, muscle fatigue, stress, and user acceptability. Because of the increased use of pointing devices in the workplace, there is a heightened awareness of the legal liability companies face due to repetitive strain injury, measures of muscle fatigue and stress. Further, because of the awareness of these issues, many users prefer to use one device over another based on the design and perceived ergonomic characteristics. Companies have labelled their products "ergonomic keyboard" or "ergonomic mouse" to give the impression and association that ergonomic factors have been taken into consideration when designing the products.

2.5.2 Workload

Subjective workload has been identified as an important part of test and evaluation. It has been used for years in human factors studies because it is easy to administer, is not intrusive, highly sensitive to changes in workload, and has a high degree of accuracy (Charlton 1996). Many methods have been developed to assess workload. The findings have suggested that "task performance and workload are not monotonically linked" (ibid, p.182). In other words, if workload increases, performance will not necessarily decrease (Hart 1986, as cited in Charlton 1996). Workers are able to adjust effort based on the perceived workload and performance requirements. As Charlton explains:

> *"Over a wide range of moderate workload levels, subjects are able to adjust their level of effort and maintain acceptable levels of performance (Lysaght et al. 1989). For moderately difficult tasks, subjects may not be able to increase their effort enough to meet the task demands and thus increased workload is associated with poorer performance (O'Donnell & Eggemeier 1986). For very difficult tasks, subjects may not continue to expend the extra effort in the face of what are perceived as unreasonable task demands; instead they reduce their effort and allow task performance to deteriorate in order to return to normal levels of workload (Charlton 1991)." (Charlton 1996, p. 182)*

Individual differences play a large part in workload, effort, and performance among workers. There is no single metric or measurement that prevails and is used universally. Researchers have found that workload has multidimensional aspects. Therefore, measurement of workload has also been multidimensional. NASA-Task Load Index (TLX) (Hart 1987; Hart & Staveland 1988) is one of the most frequently used metrics for measuring workload phenomenon. It is based on six sub-scales and uses a weighting scheme to normalize the ratings for each participant. The TLX sub-scales include: mental demand, physical demand, temporal demand, performance, effort, and frustration. Participants rate each linear sub-scale (See Appendix B). Following ratings on the sub-scales, participants make paired comparisons of each sub-scale to select which was more relevant measure of workload for the task they performed. This data is used to provide a weight for each sub-scale. Finally, the sub-scale ratings and weights are combined to yield a score from 1 to 100 for each sub-scale. An overall workload rating can also be found using the NASA TLX procedure (Hart & Staveland 1988).

2.6 Mice

The mouse has been the dominant input device used for navigation, dragging, dropping, and selection in a GUI. Similarly, the QWERTY keyboard has traditionally been the dominant input device used for data entry, editing, and shortcuts, in both graphical user interfaces and command-based interfaces.

Mouse technology has been around for a long time. It was invented in the 1960's at Stanford Research Laboratory (English, Engelbart & Berman 1967) to be a cheap replacement for light-pens which were used since 1954 (Goldberg, and Robson 1979). It became commercial with the Xerox Star but its wide spread acceptance and use was due to the popularity of the Apple™ Macintosh, which was introduced in 1983 (Perry & Voelchker 1989). Currently, mice are the most common forms of input device for GUIs. They come in many different shapes and sizes. Most have two or three buttons which are used for navigation, dragging, dropping, and selection.

Alternative forms of computer input devices have continued to be developed for a number of years by researchers and industry (Buxton 1983; Bleser & Sibert 1990; Jacob 1992). For example, the Rockin' Mouse by Alias|Wavefront is being used to provide an additional degree of freedom to assist users specifically for 3D manipulation of graphics packages (Balakrishnan, Baudel, Kurtenbach, Fitzmaurice 1997). They acknowledge that the popularity of the normal mouse was a reason behind developing the Rockin' Mouse to look and feel like a regular mouse in terms of form, stability, relative frame of reference, and control specifics.

Research and development of new devices have not yielded radically different input devices. The popularity of the present mouse style has remained very high since its inception. Generally, the new input devices are simply alternate forms of keyboard and mouse (e.g., touch pads). Generally, the differing technologies are evaluated using a battery of tests. Researchers commonly use both objective and subjective measures to generate and compare results. The results help determine the *usability* of the different technologies.

The properties of mice can improve or reduce the limits of human performance. These properties are usually fixed and cannot be changed or adjusted. They can include:

frame of reference, system inertia and friction, and clutching. The effects of each are not fully understood because there is very little research concentrated in this area.

In relative frame of reference, the movement of the cursor is relative to mouse position. In an absolute frame of reference, the cursor position on the display maps to specific, absolute spatial co-ordinates on the device. The frame of reference is a property among devices that determines the types of actions that may be faster on one device but slower on another.

Another important property that influences the efficiency of actions is "clutching." MacKenzie defines clutching as "the process of disengaging, adjusting, and re-engaging the input device to extend its field of control" (p.350). The traditional mouse allows users to grab the mouse, pick it up, and move it, for example, when it reaches the end of a mouse pad. Other devices, such as a track ball do not provide this ability. The characteristic may affect performance but is difficult to measure because it is highly dependent on the task at hand (MacKenzie 1995).

A concept that has not been studied is the effect of friction and inertia in pointing devices. To use a normal mouse, users try to minimize surface friction with the use of clean, uniform, and smooth surface. For example, a mouse pad is commonly used to optimize surface friction. The mouse pad usually contains a uniform surface and allows for effective mouse ball movement. A cord from the computer is usually attached to the mouse. The effects of the cord are negligible because it is usually light-weight and does not restrict a user's mouse movements on the mouse pad. Some new devices that are in the market do not make use of mouse balls but rely on servomotors. For example, the Haptic Mouse by Haptic Technologies Inc. or the Virtual Reality Mouse by Control Advancements Inc. The servomotors are affixed to the pointing devices usually with a metal arm, rack and pins, or a combination of multiple arms and joints. This system increases inertia and friction that a user must adapt to (knowingly or unknowingly).

A factor that was briefly mentioned earlier in this thesis was stimulus discrimination and delivery method. The Virtual Reality Mouse and Haptic Mouse are normal mice that have been altered. The stimuli are delivered to the whole hand (palm) as the user traverses the screen objects. Other devices (see Akamatsu, MacKenzie & Hasbrouc 1995) have used a mouse that delivers stimuli via a solenoid-driven pin through a hole in a mouse button. There are a variety of ways that stimuli can be delivered. The delivery mechanism and method could affect the user's sensitivity depending on environmental conditions (e.g., mice used on a desk at home versus used on a moving Aeroplane or Train).

There is lack of research examining the issues surrounding the introduction of inertia or friction into a pointing device. The effects of fixed or relative frame of reference on task performance and issues surrounding clutching have also not been examined. The diversity of environments that the pointing device could be used in should be acknowledged and different devices and type of stimulation to suit the environment could be provided to improve usability. In summary, the properties of devices need to be studied in more detail to test how they help or hinder performance.

2.7 Kinesthetic and Force Feedback Via Input Devices

Force feedback "is feedback in which the control system reflects large inertial or viscous friction forces back to the operator through the control stick" and in this case, the pointing device such as a mouse. (Douglas & Mithal 1998, p. 49). A number of studies that evaluate force feedback and multi-modal technologies from a human performance perspective are summarized in Greenstein & Arnaut (1988) and Milner (1988).

It is important to note that researchers define tactile, kinesthetic, and force feedback with respect to the mouse in different ways. This can be a source of confusion when reading papers in this area. For the purposes of this study, a "normal" mouse provides only tactile feedback. As described in section 2.3.3, tactile feedback refers to a sense of touch. A mouse that is restricted to a fixed frame of reference on the computer screen and is constrained by a planar workspace is said to provide kinesthetic feedback. This is because it gives both a sense of touch and a sense of position on the screen to the user. Finally, a mouse is said to provide force feedback when an active force is delivered to the hand from the mouse. Thus, a mouse with a fixed frame of reference, constrained to a planar workspace and that provides an active force to that hand is referred to as having kinesthetic and force feedback.

Force feedback can be varied according to presentation, delivery, and control just like all other feedback types. In terms of presentation, it can deliver force proactively or reactively. For example, the feedback can be delivered when the cursor passes over a button, and is therefore proactive. The feedback can be delivered after the user presses a button, and is therefore reactive. The delivery can be transient or sustained. When the mouse cursor passes over a button, the user feels a force briefly versus feeling a force (e.g., vibrations) until the cursor leaves the button area. Finally, the feedback can be used to indicate whether the user controls mode or its system maintained. There are many ways that feedback can be varied.

Akamatsu & Sata (1992) performed a target acquisition study to evaluate a modified mouse that provided tactile and force feedback. A solenoid-pin was inserted under the mouse button for tactile feedback and an electromagnet near the base for feedback. When the cursor crossed the edges of screen objects, tactile stimulus was delivered to the finger tip. Force feedback was provided directly on the hand by passing current through the electromagnet to increase friction between the mouse and the mouse pad (made of iron). Friction was varied depending on the greyscale of the screen objects. They found that the addition of tactile and force feedback improved time and accuracy compared to the visual feedback alone condition.

In a study, Akamatsu, MacKenzie & Hasbrouc (1995) compared tactile[1], auditory, and visual feedback in a pointing task using a mouse-type device. Five condition groups (normal, auditory, color, tactile, and combined) were tested. No differences were found in overall response times, error rates, or bandwidths but significant differences were found in final positioning times. The tactile feedback condition was the quickest and the normal feedback condition was the slowest. When the microstructure of movements were examined, tactile feedback allowed the participants to select targets quickly once the cursor entered the target.

Akamatsu & MacKenzie (1996) performed a study with a multi-modal mouse (provided tactile and force feedback) in a target selection task with three target distances and three target sizes. Four condition groups (normal, tactile[2], force, tactile+force) were tested. The multi-modal mouse provided tactile and/or direct feedback by delivering force stimulus directly to the hand and finger tip. Akamatsu and MacKenzie found significant reductions in the overall movement times and in the time to stop the cursor after entering the target. This effect was highest for the tactile condition and for small targets. However, error rates for the tactile and tactile+force conditions were higher compared to normal feedback. From the results of these studies, Card, Mackinlay & Robertson (1991) and Akamatsu & MacKenzie (1996) have predicted that adding force and tactile feedback to mice should also reduce the processing demands of the visual channel, and free up this capacity for other purposes.

---

[1] Akamatsu et al. (1995) defines a normal mouse as not having tactile feedback. Tactile feedback was then achieved via a solenoid-driven pin projecting through a hole in the left mouse button. The pin rises when the cursor enters the target and lowered when the cursor leaves the target. In this thesis, this kind of "tactile feedback" would be referred to as "force feedback".

[2] Akamatsu & MacKenzie (1996) define a normal mouse as not having tactile feedback. Tactile feedback was then achieved via a solenoid-driven pin projecting through a hole in the left mouse button. The pin rises when the cursor enters the target and lowered when the cursor leaves the target. In this study, this "tactile feedback" would be referred to as "force feedback".

Sellen, Gordon & Kurtenbach (1992) conducted two experiments to test whether kinesthetic[3] feedback helped prevent mode errors. In the first experiment, subjects navigated and inserted text into a pre-existing document using the *vi* text editor on a Sun workstation. They were asked to complete the task as quickly as possible and additionally, make any changes to spelling mistakes they saw on the screen. In addition to the primary task, subject also performed a concurrent distractor task: an adjacent computer would beep at random intervals of time at which point the subjects had to note the number (i.e., 1 to 6) that appeared on that screen and enter in on the computer. The effectiveness of kinesthetic versus visual feedback was compared. It showed that kinesthetic feedback through the use of foot pedals was more effective than visual feedback in terms of reducing mode errors and in terms of reducing the cognitive load associated with mode changes. The second experiment by Sellen et al. (1992) verified the hypothesis that using kinesthetic feedback ensured that subjects actively maintained insert mode. This helped show the superiority of kinesthetic feedback.

In summary, kinesthetic and force feedback in input devices is becoming common. There are many aspects to the feedback type that can be explored. Research in this area continues to expand. The use of kinesthetic and force feedback to improve HCI can contribute to reducing the load on the visual modality.

2.8 Auditory Feedback in Interfaces

Sound is a largely unexplored medium for feedback and output, even though it plays an integral role in everyday life (Gaver 1989; Buxton 1989). Auditory feedback has been available for use with computing applications for years. However, most software applications make minimal use of sound. For example, if a user clicks a button to minimize an application window, there may be an audible beep heard to show that the application was minimized.

[3] Sellen et al. (1992) refer to kinesthetic feedback in their study. In this study, this kind of kinesthetic feedback can be thought of as equivalent to force feedback because of differences in terminology.

Generally, sound is considered a redundant feedback mechanism rather than an integral part of an application. An exception may be the extensive use of sound for games.

There is a growing body of evidence to show that sound can improve human computer interaction (Alty, 1995; Cohen, 1992; Brown, 1989). However, the studies have been diverse in nature. They have been used to test various aspects of feedback. As mentioned, feedback in HCI is based on presentation, delivery and control (Sellen et al. 1992). In terms of audio feedback, there is a lot of flexibility with each of these factors. The presentation of feedback can be proactive or reactive. The delivery of the audio feedback can be transient or sustained depending on the need. Audio feedback can also be used to indicate control (e.g., mode). The feedback factors along with the characteristics of the sounds themselves add to the complexity of studying this feedback type. Sound is measured in terms of frequency and intensity (Bailey 1989). Frequency is defined as the number of cycles per unit of time in a periodic vibration and is measured in hertz (Hz). Psychologically, the auditory sensation of frequency is referred to as pitch. The intensity of sounds is the amplitude of the sound wave that strikes the eardrum and is measured by the decibel (dB). Psychologically, the auditory sensation of intensity is referred to as loudness. All these characteristics of sound can vary and thus the specifics of audio feedback can be very flexible. This may be one reason why it is so difficult to develop standards for the use of sound in HCI.

Researchers have performed studies to test various aspects of audio feedback. Some of the major studies will be documented here. The diversity of studies in this area is interesting. Nielsen & Schaefer (1993) undertook a study where non speech audio was used in a computer paint program on users who were between 70 and 75 years old. They found that the older users did not find the program more usable than those tested in the control silent group. It should be noted that non speech audio in this study made use of special effect sounds. For example[4], "bloop-bloop", "crackle-crackle", "bubble-bubble", and "drip-drip".

Gaver (1989) developed an auditory interface called the SonicFinder for Apple Computer, Inc. Information is conveyed using auditory icons combined with the standard visual user interface. Gaver found that the sounds naturally fit into the user interface and "increased feelings of direct manipulation and provided flexibility with the model world of the computer". Generally, the auditory icons were found to increase user satisfaction with the user interface. He predicts that auditory feedback can increase conceptual mapping of the user interface components.

Brown, Newsome, & Glinert (1996) performed an experiment that compared the effectiveness of auditory and visual cues in the performance of a visual task. They found that complex auditory cues could be successfully used to replace visual cues. This study has implications for multi-modal workstations.

Brewster, Wright, Dix, & Edwards (1995) performed an experiment to test whether sonically-enhanced buttons or earcons can overcome the problems of slipping off a graphical button by mistake and not noticing. Timing, error rates and workload measures were used in the experiment. Error recovery was significantly faster and required fewer keystrokes with the sonically-enhanced buttons than with standard ones. The workload analyses showed participants significantly preferred the sonically-enhanced buttons to standard ones.

In summary, auditory feedback is an ideal area for research. There are many aspects of the feedback types that haven't been fully investigated. In 1989, Blattner et al. suggested that the use of earcons and audio feedback in HCI was just beginning. Presently, earcons continue to be studied by researchers and industry in order to find effective ways of using them for various systems and functions.

---

[4] Note that these are onomatopoeic words used to describe the actual dropping, bubbling, etc. sound effects.

Chapter 3

Research Objectives

3.1 Main Research Objectives

As mentioned in the first chapter, the main objective of this research was to investigate whether adding kinesthetic and force feedback in pointing devices, and non-speech audio feedback to computer interfaces will help users perform everyday computer tasks faster, more accurately, and more easily. It was also the goal of this research to determine if multi-modal feedback would reduce the demands on the visual modality used to perform computer tasks. The study needed to include objective performance measures (speed and accuracy) on a generic computer-based task in order to test the effects of the feedback types. To help answer these questions, a study with human participants had to be designed. There were no specific guidelines on the selection of tasks for such a study. Traditionally, researchers have created specific tasks for studying research questions and the tasks selected should be such that they transfer to real world application. Given the relatively new interest in multi-modal feedback and the fact that a battery of tasks and tests were not available, it was necessary to develop an application and test environment specifically for this research.

3.2 Feedback Calculator Application

3.2.1 Justification

In order to investigate the research questions of interest, it was necessary to build a test application that would allow a user to receive a range of feedback while performing a fairly basic computing task. For the study, an application that incorporated features of typical day-to-day computing operations that are mouse-driven was needed. Tasks were desired that would allow participants to perform the task with little or no training. Further, it had to have components that were easy to identify and measure under varying feedback conditions. To this end, an online Calculator type application was developed. The rationale for using a calculator type application was that most operating systems come packaged with an on-line Calculator.

A typical online Calculator can be mouse-driven or keyboard-driven. The Calculator application basically involves precise "point" and "select" operations in a semi-continuous fashion with a mouse. Thus, using a Calculator application as part of the study increases ecological validity in that it is an application requiring little or no training on the part of the user.

Most online Calculator interfaces are very similar in appearance regardless of the operating system. The interface usually contains buttons and windows that are standard screen elements used by most current software applications. The task of navigating with a mouse, pointing and selecting objects on the screen is inherent in all GUI based applications and can be easily represented in a calculator task. Finally, the calculator task is fairly simple, easy to understand and to perform with little training and previous knowledge. The Feedback Calculator application fulfilled much of the criteria and was therefore chosen for use in the study.

Although the Feedback Calculator application represents the point and select aspect of the GUI taxonomy, it does not represent Dragging, Drawing, and Device Switching. It should be noted that because dragging, drawing, and device switching require practice, these aspects were not included in the Feedback Calculator application.

Tasks for the Feedback Calculator were specifically chosen to measure performance such as speed and accuracy over a set of trials. Participants would be asked to enter a series of mathematical expressions and find the answers using the Calculator application. Performance in the tasks would help determine the effects of the feedback types.

3.2.2 Software Specifications

A Feedback Calculator software application was specifically designed and developed for the requirements of this study using Visual C++. The software was written to include the Calculator application combined with a data tracking application. The Feedback Calculator only accepts input from the mouse left click button. It provides control over audio feedback conditions (sound or no sound), and records relevant statistics. All buttons of the Feedback Calculator user interface contain distinguishable musical sounds (earcons) that could be activated when the mouse cursor passed over the buttons. The sounds are saved in individual wave (.wav) files that can be easily modified in future studies. Under the current setup, the numerical buttons used Violin pizzicato notes in C major. The numbers from one to nine and then zero were mapped to the scale (doe to ta) respectively. The other buttons were mapped to various sampled sound effects.

The Feedback Calculator application was specially developed for a research study involving human participants. Since the Feedback Calculator was also designed to be a reusable research tool for use with other studies in this area, its design was kept flexible. The application was designed to fulfil certain criteria: First, the calculator was designed to have the look and feel of a typical online calculator. Further, it needed to work in the same manner as an online Windows Calculator[1], not using reverse polish notation (RPN) for example. It should allow researchers to track participant information and performance.

In keeping with the primary task, the Feedback Calculator application was designed to display expressions on a separate scratch pad. These expressions represent the expected strings of point and select sequences. The numerical expressions are displayed one by one on the non-editable scratchpad. The participant enters the string as accurately as possible into the Feedback Calculator. The calculator application automatically tracks the relevant time statistics. A task time represented how long it took a participant to enter one mathematical expression (in milliseconds). In order to start the timer, a special button was needed. A button called "Save" was used for this purpose. The timer started when the participant selects "Save" to display the expression and it stops when the participant selected the "equals" operand to calculate the answer for the expression.

Figure 3-1 is the start-up screen for the Feedback Calculator application. The start-up screen records information on the trials including: participant id, output file name, and feedback condition. The output file saves the recorded information as well as input by the user. This provides the researcher with the flexibility to examine the data at a later time.

Figure 3-2 is a screen shot of the scratch pad that appears with mathematical expressions. Figure 3-3 displays the Feedback Calculator application interface. The

[1] The calculator was designed to have the same look and feel of a Windows Calculator only because of the popularity of Windows.

interface was designed to have the "look and feel" of a typical Windows calculator. This was primarily done because of the popularity of Windows products and as such should increase familiarity with screen objects.

Figure 3-4 displays a typical output file. As mentioned, the output file automatically records the participant number, condition group, trial number, time to complete each trial, and recorded all mouse "clicks" in an output file. If a participant selects the mouse button when the cursor is not on top of a button, this is registered in the output file as "M" for extra mouse click. This type of error could indicate the participant slipped off a button just before selecting or simply pointed incorrectly. Thus, all mouse button clicks were automatically recorded.

Additionally, the program automatically performs a rudimentary error checking operation between the actual input strings and the expected string. The error checker performs a one-to-one comparison of each character in the string with the expected character. If the two characters are different in any way, a "*** SEQ ERROR" character is inserted in front of the string identification. Once the output file is created, the researcher manually identifies how many errors were made by the user by looking for the "*** SEQ ERROR" and can manually check the entered string with the expected string. This simple method was selected for a number of reasons as opposed to a more complex error checking function. If the user enters an additional character within a full string, an automatic error checker would identify a multitude of errors instead of just one. Performing the checking manually ensures that "slips" are caught and the participant is not penalized for them inadvertently. This type of error checking also reduces the amount of coding required. Additionally, the more error checking that is coded, the more difficult it is for application reusability. Because the intent of the application was to make it accessible for research purposes, it was kept flexible and easy to modify.

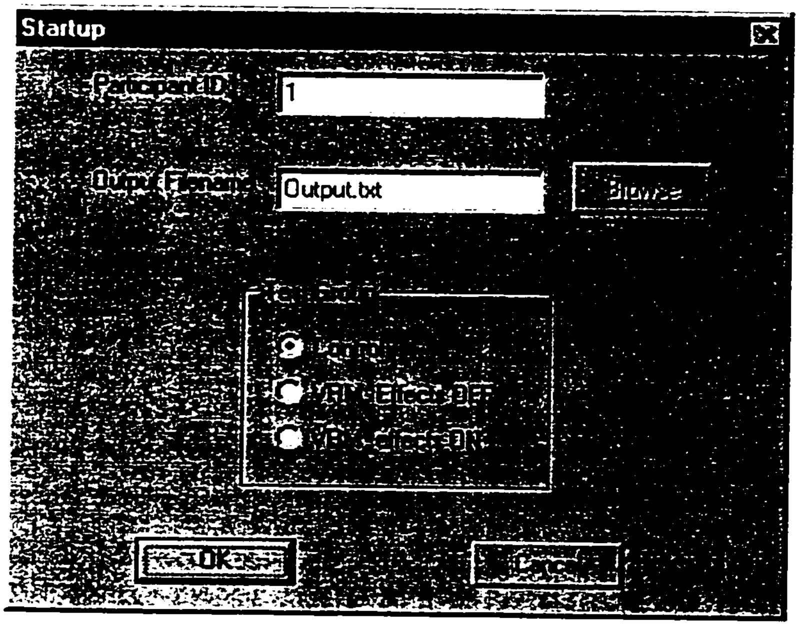


Figure 3-1. The start-up screen for the Calculator application.

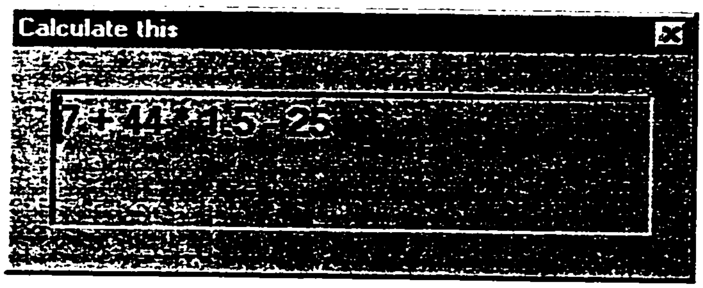


Figure 3-2. Scratch pad that was placed above the calculator interface.

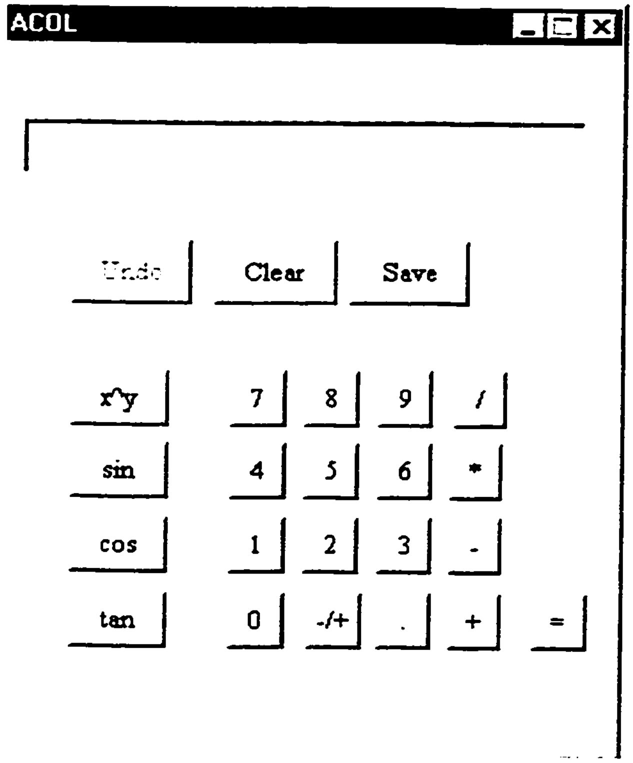
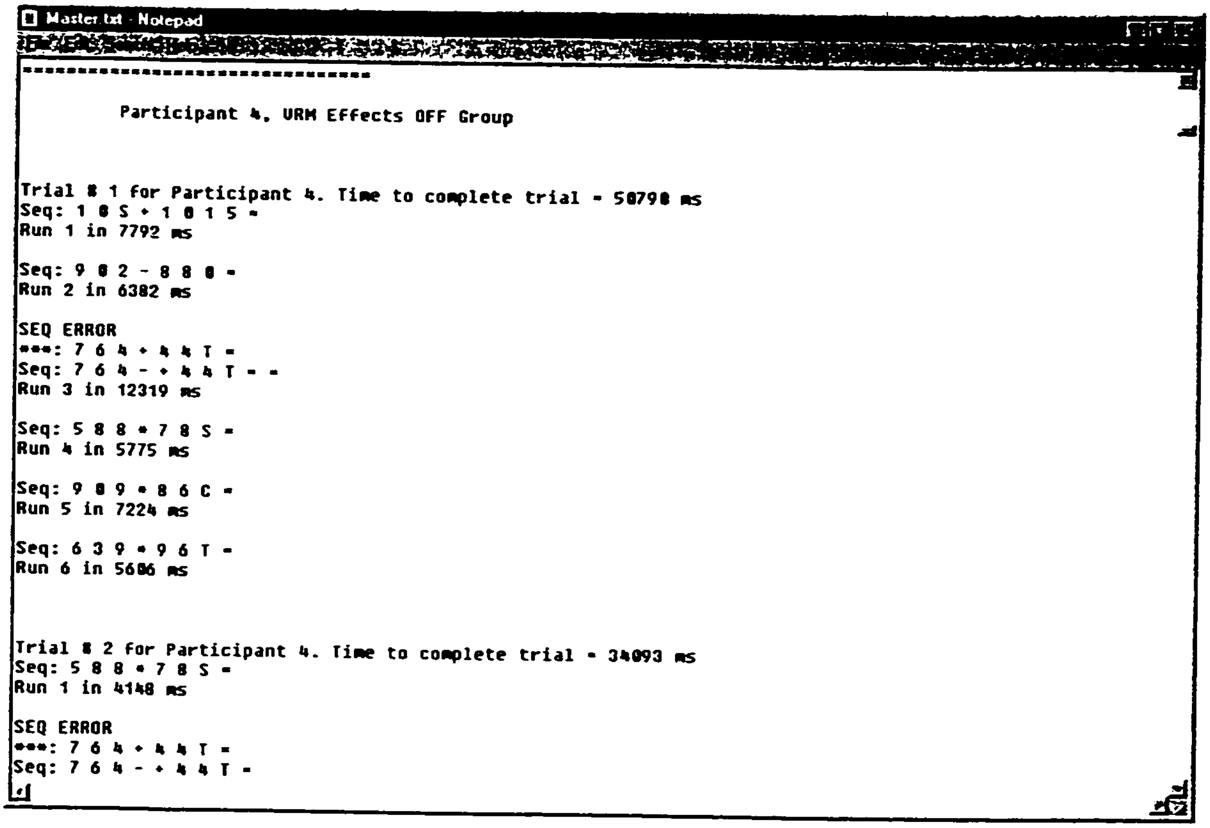


Figure 3-3. Custom built Calculator interface.



Master.txt - Notepad

```
================================

        Participant 4, URM Effects OFF Group


Trial # 1 for Participant 4. Time to complete trial = 50798 ms
Seq: 1 0 S + 1 0 1 5 =
Run 1 in 7792 ms

Seq: 9 0 2 - 8 8 0 =
Run 2 in 6382 ms

SEQ ERROR
***: 7 6 4 + 4 4 T =
Seq: 7 6 4 - + 4 4 T = =
Run 3 in 12319 ms

Seq: 5 8 8 * 7 8 S =
Run 4 in 5775 ms

Seq: 9 8 9 * 8 6 C =
Run 5 in 7224 ms

Seq: 6 3 9 * 9 6 T =
Run 6 in 5606 ms


Trial # 2 for Participant 4. Time to complete trial = 34093 ms
Seq: 5 8 8 * 7 8 S =
Run 1 in 4148 ms

SEQ ERROR
***: 7 6 4 + 4 4 T =
Seq: 7 6 4 - + 4 4 T =
```

Figure 3-4. Screen capture of a typical output file.

Chapter 4

Research Methodology

4.1 Study Overview

As mentioned in the first chapter, the main objective of this thesis was to determine whether the addition of non-speech audio feedback to computer interfaces, and kinesthetic feedback, and kinesthetic and force feedback to pointing devices will help users perform "everyday" computer tasks faster more accurately and more easily than visual feedback alone. It was also the goal of this research to determine if the multi-modal feedback can serve to reduce the information processing demands of the visual modality used to perform computer tasks.

To answer these questions, a study involving human participants was designed and conducted. The Feedback Calculator software application was used in the study (explained in Chapter 3). All buttons on the Feedback Calculator user interface could play earcons when the mouse cursor passed over by the buttons. A specialized piece of mouse hardware by Control Advancements Inc. called the Virtual Reality Mouse (VRM) was used to provide kinesthetic, and kinesthetic and force feedback. Participants were asked to enter a series of mathematical expressions and find the answers using the Calculator application. Participants performed sixteen trials. Speed and accuracy were the objective human performance measures that were recorded. NASA Task Load Index was used to record the perceived workload of the trials.

For half the trials, participants were asked to perform a secondary task. The secondary task was also visual in nature and designed to divert and occupy the participant's visual attention. The dual task was designed so that it would require more information processing to complete both the primary and secondary task. In effect, it was used to test the hypothesis that multi-modal feedback could reduce the load on the visual modality and free up this capacity for other purposes (Akamastu, MacKenzie, & Hasbrouq, 1995).

This particular secondary task was modelled after an experiment that compared the effectiveness of kinesthetic feedback via a foot pedal compared with visual feedback (Sellen, Gordon & Kurtenbach 1992). In the experiment, participants navigated and inserted text into a pre-existing document using the *vi* text editor on a Sun workstation. Participants were asked to complete the task as quickly as possible and additionally, make any changes to spelling mistakes they saw on the screen. In addition to the primary task, participants also performed a concurrent distractor secondary task: an adjacent computer would beep at random intervals of time at which point the subjects had to note the number (i.e., 1 to 6) that appeared on that screen and enter in on the computer. The effectiveness of kinesthetic versus visual feedback was compared. It showed that kinesthetic feedback through the use of foot pedals was more effective than visual feedback in terms of reducing mode errors and in terms of reducing the cognitive load associated with mode changes.

Although the secondary task for this study was modelled after the experiment by Sellen et al. (1992), it was modified for use here. The changes were: the number of choices for the secondary task was limited to three items rather than six; the three items were alphabetic rather than numeric; and beeping was removed. The rationale for the changes were as follows:

- The number of items in the secondary task was reduced because the pilot participants commented that the secondary task was too much of a distraction and subsequently the primary task was too difficult to perform simultaneously.
- The three items that were to be displayed for the secondary task were alphabetic (i.e., a,b,c) rather than numerical because it was felt that the numerical digits would interfere directly with the primary task which was also numeric in nature. When a letter appeared, participants were asked to enter it on a keyboard using the arrow keys "←", "↓", and "→" that were mapped to "a", "b", and "c" respectively.
- The beeping audio prompt for the secondary task was removed because it may have conflicted with the primary task that can also make use of the audio.
- The researcher emphasized the importance of both the primary and secondary task at the start of the corresponding set of trials. Participants were led to believe their scores on both tasks were being recorded.

4.2 Participants

Thirty students were recruited from the University of Waterloo to participate in the study. The criteria for inclusion in the study was self-reported experience with any Windows 3.1 applications or higher, proficiency with a mouse input device, having good vision with or without corrective lenses, and no hearing difficulties. In order to increase generalizability, efforts were made to have an equal number of male and female participants between the ages of eighteen and thirty. Participants were paid $10 for participating in the study.

Participants were recruited primarily from the Engineering and Math faculties at the University of Waterloo to reduce individual differences in spatial ability, computer skills, and experience with Calculator applications. The study took place at the University of Waterloo, in the Systems Design Engineering USE-IT Lab, in E2-3367.

This study received ethics clearance from the Office of Human Research and Animal Care at the University of Waterloo. Before running the actual study, a pilot study was performed to test several items: equipment usability, task appropriateness, experimental set up, visual and auditory feedback, and appropriateness of pencil and paper tasks. After observing participants during the pilot study and recorded verbal comments, some minor changes were made to the experimental setup as noted in section 4.1.

4.3 Apparatus

4.3.1 Hardware

Two Personal Computers (PC) were used in the study. The primary computer had 80586 166 MHz processor with 32 megabytes of RAM and 15 inch monitor. The second computer had 80486 66 MHz processor with 16 megabytes of RAM and 17 inch monitor. Both computers ran Windows 95.

The second computer was used for the secondary task. Both computers had colour monitors that were placed beside each other on a desk approximately 3 feet away from the participant. The primary computer's monitor was placed directly in front of the participant's field on view. The second monitor was placed to the right of the participant's direct visual field.

The primary computer had a SoundBlaster sound card, speakers, and headphones attached. The sound was enabled or disabled depending on the condition. The headphones ensured that participants heard accurate and undisturbed sounds. Participants wore the headphones throughout the study regardless of whether the sound was active. The headphones helped remove any ambient or white noise.

A standard QWERTY keyboard was also placed on the desk in front of the participant. The participant was asked to place the keyboard and input device in an ideal position for ease of use and familiarity on the desk in front of monitor. Participants were allowed to adjust their seating position and mouse for all trials. However, participants were not allowed to adjust the computer monitors.

For the mouse input device, participants either used a standard Logitech mouse or the Virtual Reality Mouse depending on the mouse-feedback group. The mouse was placed to the right or left of the participant depending on the participant's hand preference. With the VRM, the force feedback effects can be enabled or disabled depending on the mouse-feedback group. The study set up is illustrated in Figure 4-1. The participant was seated in front of the primary computer.

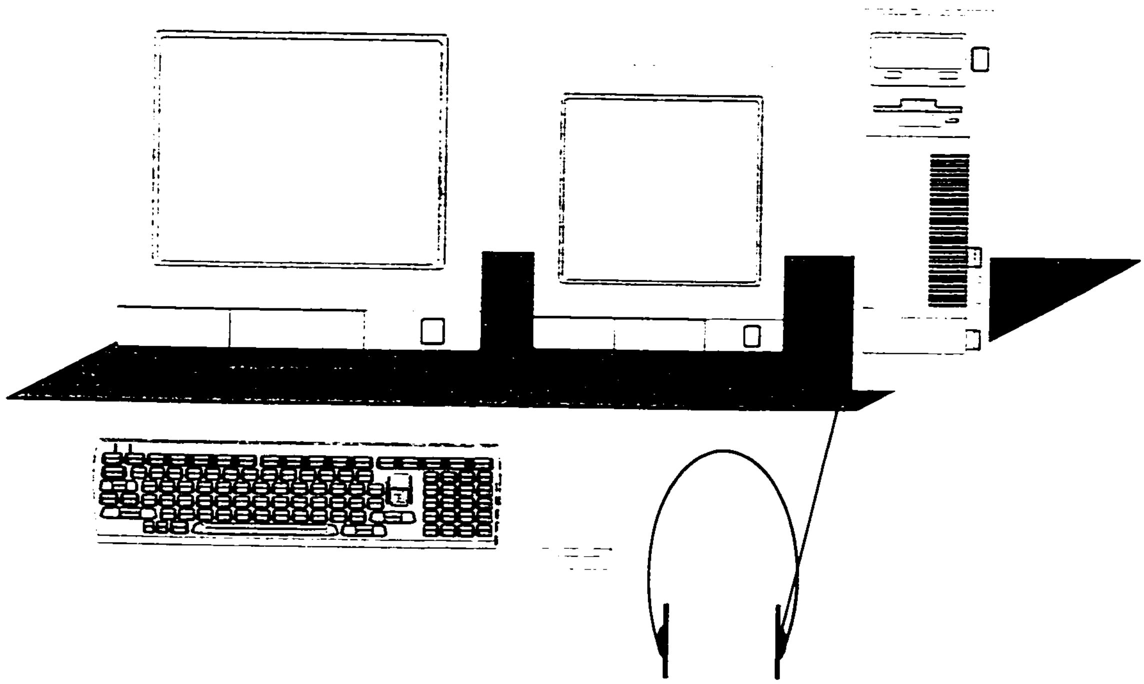


Figure 4-1. Setup for study. The monitor on the left side was used to display the primary task and the monitor on the right hand side was used to display the secondary task.

4.3.2 Setup of the Virtual Reality Mouse

The VRM is a specialized piece of computer hardware developed by Control Advancements Inc. that allows the user to physically interact with a Windows application, navigate the mouse cursor around the screen, and define a custom "feel" over the movement of the mouse. It is a 3-button symmetrical mouse such that both dextrals and sinistrals (i.e., people with dominant left or right handed respectively) can use it. Figure 4-2 provides an illustration of the VRM. It was modified and equipped with servo-motors by Control Advancements. Additionally, the mouse was constrained to a planar workspace of approximately 4" by 3 " (these dimensions match screen resolution e.g. 800x600 pixels) via a two degree-of-freedom mechanism which can be independently driven in each direction with a DC motor coupled with encoder feedback. The VRM maintains a fixed frame of reference with the screen so that all mouse movements made by the user translate exactly to the cursor movements on the screen. For these reasons, the VRM can be referred to as a kinesthetic mouse.

The VRM has a special driver that functions in a Windows environment such that any Windows-based software can exhibit virtual effects via the VRM. The driver was programmed to use closed-loop transfer functions that were equivalent to mechanical models of springs, dampers, mass or combinations of these elements to produce the desired effects. The system uses digital force feedback to model effects such as the feel of a wall, gravity or potential wells, frictional surfaces, and spring and damper effects as the VRM pointer crosses various regions on the screen. All force effects are activated only as the cursor passes over the screen element. As such, the force feedback is delivered in a proactive manner and in a transient fashion.

Pre-defined effects for screen elements were activated as the mouse pointer traverses their boundaries and were reflected to the user via force-feedback of the mouse. Although the VRM has an array of effects (e.g., springs, dampers, walls) that can be activated, only buttons and window borders providing force feedback effects were used in the study. The following effects describe the general feedback mechanisms:

- If the cursor passes over a button border, the user feels a 'bump' and then a light pull to the middle of the button. The aim of the effect was to recreate the feeling a user would have when their finger was passing over buttons of a physical calculator.
- If the cursor comes near the border of a window at a relatively slow speed (e.g., the Calculator window), the user will feel resistance and will not be able to move across the window. It's modelled as a virtual wall which was typically like two compression springs back-to-back so when you first contact the wall you're compressing the first spring and the force was proportional to the distance moved. Once you cross the mid-point in the wall, the other compression spring was effectively compressed (it's the same as if you entered the wall from the opposite direction) so you get pushed out of the wall.

The general objective was to represent the Feedback Calculator application virtually such that a person could imagine it as a physical calculator. The Calculator interface contains only buttons of various sizes. All effects were presented in a proactive manner.

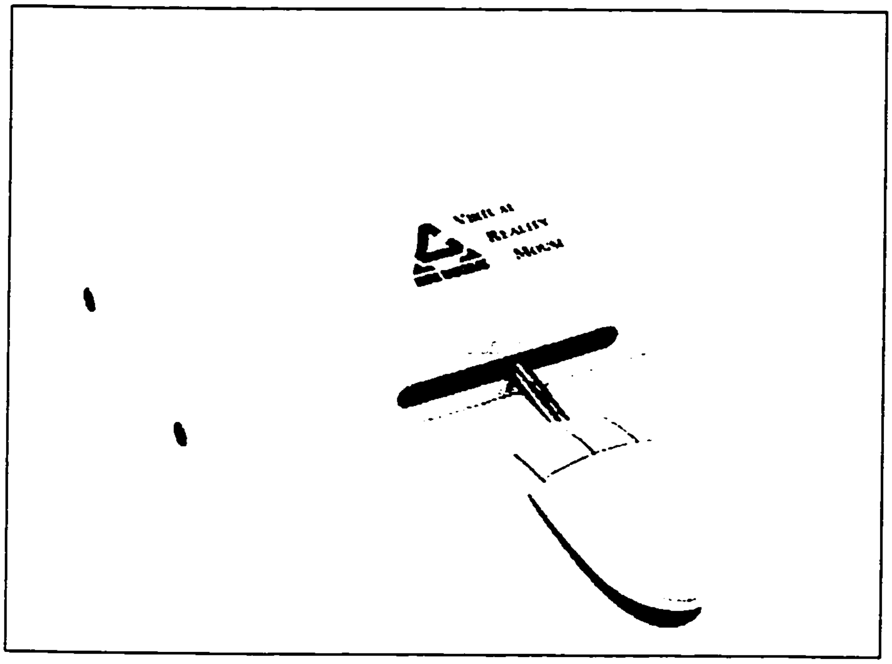


Figure 4-2. Picture of the Virtual Reality Mouse by Control Advancements Inc. (1998).

4.3.3 Software

As described in Chapter 3, the Feedback Calculator application was specially designed for this study. It was written in C++. MS Word and MS PowerPoint were used in combination for the secondary task. The Calculator application was run on the primary computer. The custom-built Calculator only accepted input from the mouse left click button. As explained in the previous chapter, the Calculator application automatically recorded the participant number, mouse-feedback group, trial number, time to complete each trial, identified whether the participant made errors in entering an expression, and recorded all mouse clicks in an output file. The output file was later used to compare results between all mouse-feedback groups. A second computer ran Word and PowerPoint to run the secondary task.

4.4 Experimental Design

This study was a between-subject study with repeated measures. All participants were randomly assigned to one of three equal size mouse-feedback groups:

1. standard mouse (Control condition);
2. VRM with no force feedback (Kin condition); and
3. VRM with force feedback (Kin+Force condition).

While the specifics of feedback varied between groups, the basic approach and methodology remained the same. The study was a mixed-subject design. Participants were designated to one of the three feedback groups and performed 2 x 2 repeated measures. The repeat measures were sound (off and on) and task (single and dual). Figure 4-4 illustrates the experimental design. To avoid possible order effects, the auditory feedback condition was counterbalanced on the sound condition. The sound condition alternated between participants so that an equal number of participants started the trials with sound while the others received the no sound condition first.

All mouse-feedback groups used the same hardware and software apparatus except the Control condition that used a standard Logitech mouse instead of the VRM. For the Control condition, no additional feedback was delivered when the cursor passed over buttons on the Calculator application. The Control condition provides a basis for comparison between the "typical" or "normal" mouse with the VRM. For the Kin condition, the VRM was used but all force effects were turned off. In other words, no additional force feedback was delivered when the cursor passed over buttons on the calculator application. The VRM imposed a grid constraint of movement for the user and which provided kinesthetic feedback. The Kin condition was important to provide a control for the Kin+Force condition. For the Kin+force condition, participants used the VRM mouse that inherently provided kinesthetic feedback with force feedback effects turned on. In other words, the force feedback was activated when the cursor passed over buttons via the mouse.

Participants performed four blocks of trials. Each block consisted of four trials. In total, participants performed sixteen trials. Within each mouse-feedback group, participants had to perform both the primary and dual task with sound off and sound on. The primary and dual tasks used different data sets. Different data sets were used to minimize learning of the data sets. In other words, the same data was not used in all sixteen trials. The single task conditions used one data set while the dual task conditions used a second data set. Both data sets contained the same number of mouse movements. To ensure consistency of learning across groups, the test data for each trial remained the same for all participants and was performed in the same order so that the effects of learning the feedback and calculator interface were balanced. The expressions were constructed to ensure that each trial contained the same number of keystrokes and total travel distance.

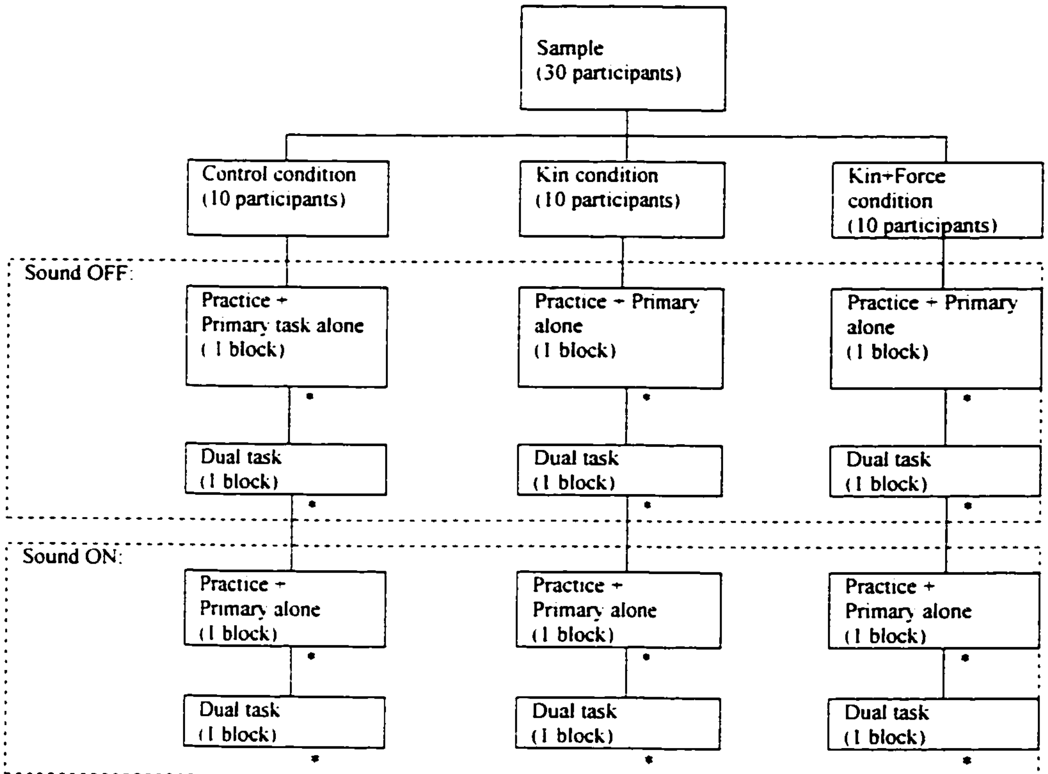


*Figure 4-4. Experimental design for the study. NASA TLX was performed after each block consisting of four trials (shown here by the *). Please note that the sound on or off at the start of the experiment alternated between subjects.*

4.5 Procedure

The study took approximately an hour to complete and each participant was asked to perform several tasks during the course of the experiment. All experimental forms are included in Appendix A.

Before the start of the study, participants were asked to read the Information Letter for a description of the study. Participants were then asked to sign the Consent Form before the study began. All instructions to the participant were read from a Script. This was to ensure that all participants received identical information. Participants were allowed to ask questions if they were confused about the procedures.

The study was divided into three phases. The first phase consisted of the participant completing a Background Questionnaire, and performing practice trials. It took approximately fifteen minutes. The participant performed four blocks of trials (16 trials in total) in the second phase. After each block, the participant performed the NASA TLX workload ratings. This phase took approximately forty minutes. In the final phase, participants completed the Final Participant Questionnaire, were given the Feedback Letter and payment for participation. It took five minutes. Figure 4-1 presents the experimental design and procedure.

4.5.1 Participant Background Questionnaire

The participants were asked to complete the "Participant Background Questionnaire" for demographic data and computer proficiency. The purpose of the participant demographics was to ensure that the four experimental conditions were more or less balanced with respect to any participant characteristics that might influence performance on some of the tasks in the study. None of the personal information collected was used as a screening tool. Participants were asked to state some general information: gender, age, enrolled faculty, and hand preference.

Clearly, any kind of auditory or visual impairment may influence a participant's performance on the computerized tasks in this study. Also, it was possible that participants' computer experience might influence their performance. If the participant wasn't comfortable using a standard mouse, they might hesitate in their actions. Similarly, if the participant had extensive use of the Windows Calculator, it will influence their performance positively. These questions were designed to ensure that groups were similar in background experience prior to the experimental trials.

The participant was then given an overview of the apparatus. Once the participant was seated comfortably in a chair facing the primary monitor, it was turned on. The monitor displayed a *calculator application* and a *scratch pad* that were controlled by the mouse and keyboard. Cursor movement, selection, and data entry for the calculator was controlled by the mouse exclusively. Participants were then given a training session during which time the researcher described how to use the equipment and software applications. They were asked to practice entering a set of expressions with the calculator application, keyboard, and the assigned mouse for about ten minutes. On average, participants entered approximately 20 expressions using the calculator and mouse.

4.6 Experimental Trials

4.6.1 Dependent Objective Measures

Primary Task

For each trial, participants entered six numerical expressions into the Calculator. The numerical expressions were displayed one by one on a non-editable scratchpad above the Calculator interface. After the participant enters the expression, selects the equals button (i.e., "=") to find the answer and then selects "Save" on the calculator, the next expression was displayed. The following is a sample of the numerical expressions that were presented for one trial:

- 20 Cos 40 + 15 =
- – 15 Tan 30 =
- 18 Sin 30 * 23 =
- 55 Sin 15 – 30 =
- 40 / 70 Cos 18 =

- 10 – 93 Tan 50 =

The participants were asked to enter the mathematical expressions as quickly and accurately as possible. After the participant completed each trial, the experimenter asked the participant to take a brief break.

Direct performance measures that were recorded were time and number of errors committed in a trial. Trial time indicated how long it took the participant to enter the set of six expressions. More precisely, the trial time was the summation of the time it took to enter each of the six expressions and find the answer. Secondly, the number of errors were calculated by summing the number of wrong entries using the mouse with the number of mouse clicks that were made on the Calculator window rather than on top of a button.

Dual Task

For the dual tasks, the participant performed the Calculator primary task and a secondary monitoring task concurrently. Participants were asked to visually monitor a second computer screen and enter what was seen. The second screen displayed one of three letters (i.e., A,B,C) on Power Point in a pre-randomized order at varied time intervals (between 0.5 and 4 seconds). The researcher emphasized that both tasks were of equal importance and that both time and accuracy were important parts of the study (See Appendix A: Script for Researcher). MS Power Point and MS Word were both used for the secondary task. The set up of the secondary task applications is shown in Figure 4-5. Three lower case letters (a, b, and c) were displayed in New Times Roman 80-point font at varied time intervals and in a pre-randomized sequence. Participants were asked to view the second monitor intermittently so that they did not miss viewing the letter. Participants simply entered the letter that appeared in the PowerPoint display window into the Word file. The arrow keys "←", "↓", and "→" on the keyboard were mapped to the letters "a", "b", and "c". Again, trial times and number of errors were recorded for the primary task in the dual task condition.

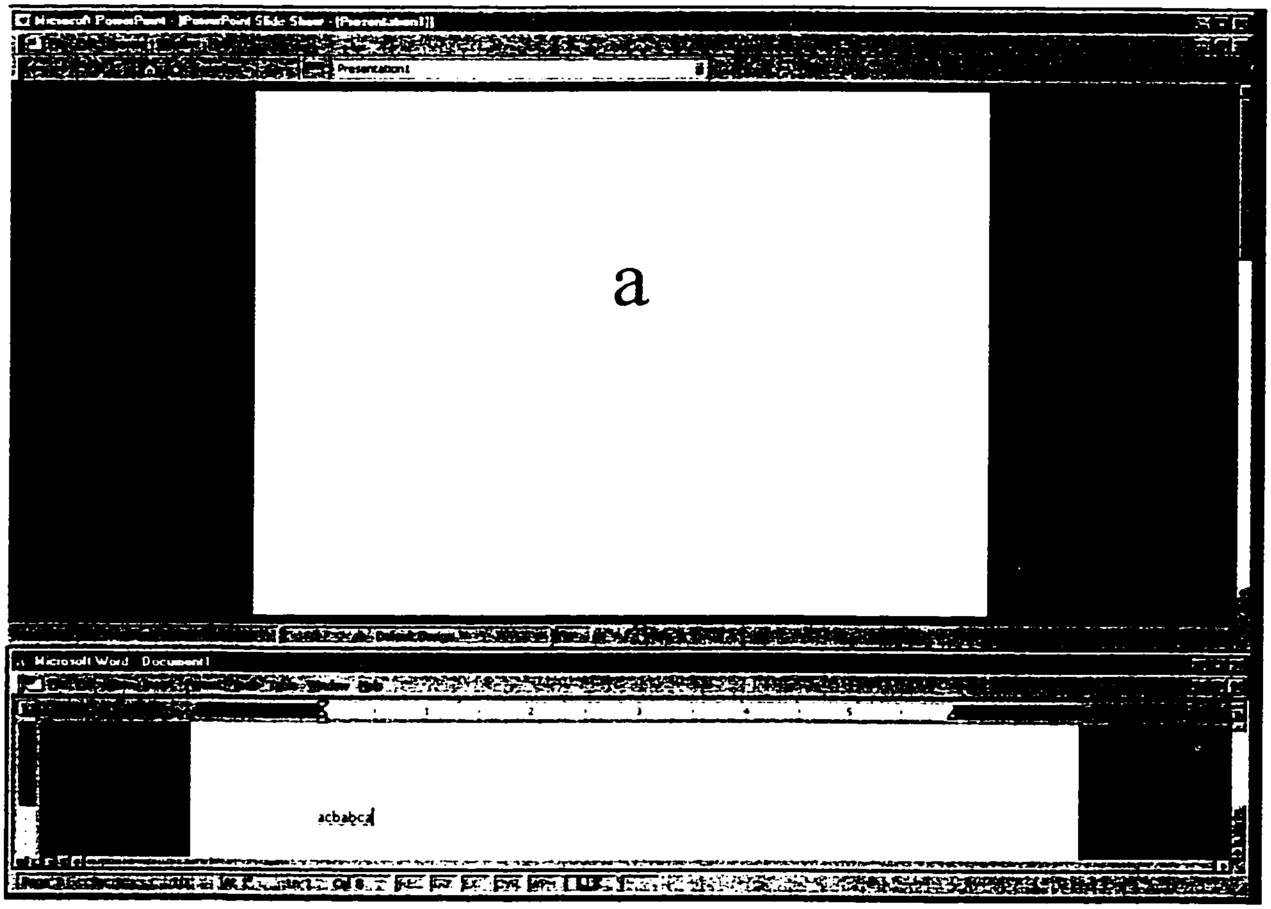


Figure 4-5. The screen set up for the secondary task. The bottom window is MS PowerPoint that displays characters at predetermined random time intervals. The lower window is MS Word that displays the characters the user enters which are supposed to be the same as the one shown in the above screen.

4.6.2 Dependent Subjective Measures

NASA TLX

After each block of trials, participants were asked to complete the NASA TLX. The NASA Task Load Index was a multi-dimensional subjective rating procedure that provides an overall workload score based on a weighted average of ratings on six sub-scales: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort, and Frustration. Three dimensions relate to the demands imposed on the subject (Mental, Physical, and Temporal Demands) and three to the interaction of a subject with the task (Effort, Frustration and Performance). Besides the six scales, an overall weighted measure of task load was also calculated on the basis of the scales. A shareware program called "NASA TLX Version 1.0" was used to collect the subjective workload assessments with the NASA Task Load Index on the study computer. This procedure for collecting workload ratings was developed by the Human Performance Group at NASA Ames Research Centre. These measures provided additional information on the difficulty of tasks performed.

Final Participant Questionnaire

After completing the trials, the participants were asked to complete a Final Questionnaire. The questionnaire was subjective in nature. It was used to get a subjective impression of how useful and effective the equipment and feedback mechanisms were and to what extent participants relied on them. It also contained questions regarding preferences. The following questions were asked of each participant depending on the mouse-feedback group: rate the usability of the VRM compared to a normal mouse, rate the ease of navigation using the mouse, rate the ease of button recognition because of feedback from the mouse, etc. The copy of the Final questionnaire is in Appendix A.

Following this, a feedback letter was read to each participant. The feedback letter thanked the participant for taking part in the study and reiterated the confidential nature of the study. The copy of the feedback letter is in Appendix A. Finally, participants were given time to ask any additional questions of the researcher pertaining to the study before departing.

Chapter 5

Results

5.1 Overview of Statistical Analysis

As described in the previous chapter, this was a between-subject experiment (3 mouse-feedback groups) with repeated measures (2x2). The mouse-feedback groups were called Control, Kin, and Kin+Force. The Control group used a 'normal' mouse. The Kin (kinesthetic) group used a constrained mouse with a fixed frame of reference. The Kin+Force group used a constrained mouse with a fixed frame of reference and active force feedback. The specifics of the mouse-feedback groups were described in Section 4.4. The repeat measures were sound (Off or On) and task type (Single or Dual). The results of the statistical analysis are presented in this chapter. SPSS 7.5.3 for the Windows operating system was used for all statistical data analysis. Criterion level of $\alpha = 0.05$ was used to determine acceptable level of significance. Due to the exploratory nature of this research, alpha levels of 0.10 or less were also reported.

The analysis examined both subjective and objective measures that were collected for each participant in the study. Objective measures were considered to be trial times and error rates. Subjective measures were considered to be results from the NASA TLX and from the Subjective Questionnaires. Both measures were used to analyzed to evaluate the effectiveness of the feedback types.

5.2 Treatment Group Demographics

A total of 30 people participated in the study. Table 4.1 lists the breakdown of participants in each of the mouse-feedback groups. Efforts were made to have an equal number of male and female participants in each mouse-feedback group to remove any possible gender bias.

Table 5-1. Breakdown of male and female participants by mouse-feedback groups.

Gender (sex)	Mouse-Feedback Groups			
	Control	Kin	Kin+Force	Total
Male	6	5	5	16
Female	4	5	5	14
Total	10	10	10	30

As described in Section 4.6.1, at the beginning of the study, all participants completed a background questionnaire. Initially, participants rated their experience using a 6-point interval ratings scale (scale represents: 0 hrs, 1-2 hrs, 3-4 hrs, 5-6 hrs,7-8 hrs, and 9+ hrs respectively). Figure 5-1 provides a graphical illustration of the results. Appendix B contains a summary of the results for each experimental mouse-feedback group.

A one-way analysis of variance (ANOVA) was run to determine if there were any significant differences in participant demographics between the feedback groups. (Refer to Appendix B for a summary of these ANOVAs). The results from the ANOVAs indicate that there were no significant differences in participant demographics among any of the three experimental feedback groups.

Participants identified physical impairments (visual, hearing) as well as physical discomfort usually experienced after computer use on a 2-point nominal scale (1 = No, 2 = Yes). Generally, participants were in good physical health. Only one participant in the Control group experienced minor hearing difficulties. One participant in the Kin group and one participant in the Kin+Force group noted experiencing visual fatigue problems while performing computing tasks. Chi-square tests run for each of the factors between feedback groups showed no significant differences on these variables.

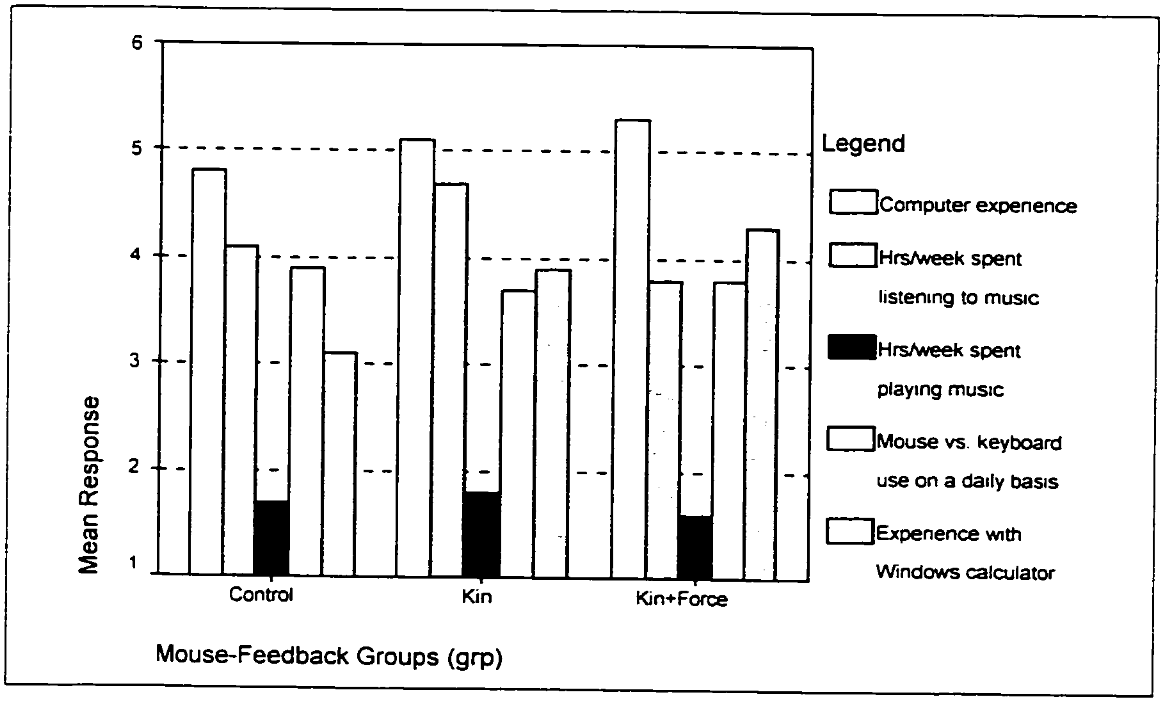


Figure 5-1. Demographics of participant backgrounds compared with condition groups.

Figure 5-1 illustrates how participant backgrounds were similar among condition groups. Participant scores were the highest for number of hours spent using the computer per week. Participants slightly preferred the keyboard to the mouse for navigation.

The time it took for each participant to enter the six mathematical expressions was calculated as the trial time (in milliseconds). Next, the number of errors for each trial was calculated by summing all the errors committed by a participant while entering the set of six expressions. In a two-step process that is described below, abnormally high trial times and error values were identified and removed from the analysis. Additionally, in order to be consistent, if trial times were removed then the corresponding error values for the same trials were also dropped from the analysis and vice versa.

Trial times by individual participants that were greater or less than three standard deviations from the total means were dropped. Additionally, the corresponding error rates for the same participant trial were also removed. Similarly, the error values that were greater or less than 3 standard deviations from the mean were considered abnormal and removed and the corresponding trial times were also removed. Next, an average trial time was found based on the set of four (minus abnormal data) trial times. These four average times were converted to seconds from milliseconds for the analysis. Similarly, an average error rate was found based on the set of four (minus abnormal data) trial error values. This procedure removed twenty-three data points from further analysis from four hundred and eighty data points (i.e., 30 participants performed 16 trials each).

At this point, the second step was to identify and remove deviant block data. Average trial times (i.e., block times) by individual participants that were greater or less than three standard deviations from the respective mouse-feedback group means were dropped. Additionally, the corresponding block error value for the same participant trial was also removed. Similarly, the error values that were greater or less than 3 standard deviations from the respective group means were considered abnormal and were dropped. As before, the corresponding block time was also removed. This procedure eliminated one block time from the Control group and one block time from the Kin group.

5.3 Descriptive Statistics for Time and Errors

Table 5-2 presents the mean trial times for each block and standard deviation based on mouse-feedback group and trial specifications. Table 5-3 presents the mean number of errors and standard deviation for each mouse-feedback group based on trial specifications.

Table 5-2. Mean trial times and Standard Deviations by mouse-feedback groups for each block.

Feedback Groups		Control N=9	Kin N=9	Kin+Force N=10
Single Task, Sound Off	Mean	36.6 (4.01)	38.9 (7.69)	37.9 (4.29)
Single Task, Sound On	Mean	37.7 (3.55)	41.0 (5.21)	38.7 (6.47)
Dual task, Sound Off	Mean	53.8 (11.38)	54.9 (9.18)	48.6 (5.02)
Dual task, Sound On	Mean	50.3 (9.87)	56.8 (9.39)	52.3 (11.13)

Table 5-3. Mean number of errors in trials by mouse-feedback groups for each block.

Feedback Groups		Control N=10	Kin N=10	Kin+Force N=10
Single Task, Sound Off	Mean	1.5 (1.92)	1.7 (1.48)	0.9 (0.98)
Single Task, Sound On	Mean	1.8 (2.65)	1.7 (1.42)	0.8 (1.03)
Dual task, Sound Off	Mean	2.8 (2.82)	2.2 (1.81)	1.9 (1.74)
Dual task, Sound On	Mean	3.2 (4.66)	2.0 (1.30)	1.9 (1.42)

5.3.1 Speed Accuracy Trade-Off

In typical speed-accuracy trade-offs, as a participant tries to go faster the number of errors they make are likely to increase. Thus, one expects speed and error rates to be negatively correlated. Linear correlations were run for block time and corresponding errors to investigate the possibility that the two measures reflected such a trade-off. While all of the relationships were in a negative direction, none reached levels of significance ($p > 0.05$). Therefore, while the two measures are not clearly independent, we can view them as contributing different pieces of information to the understanding on the effects of feedback on performance.

5.3.2 Effects of Ordering of Sound Conditions

The order of sound presentation in the trials was balanced so that an equal number of participants started the trials with sound and the other half started without sound. This was done to balance the order of the sound condition to lessen the potential carry over effects due to practice that might be given to the Sound On condition if it was always presented after the Sound Off trials. To ensure that order of sound presentation did not have a significant effect on performance, preliminary 4-way ANOVAs (Group[1] x Order x Task x Sound) were carried out for both time and error rates.

The results showed unanticipated order effects for both time and error rates. For time, significant interactions were found for sound and mouse-feedback group (Sound x Order; $F(1,23) = 110.64$, $p < 0.001$), as well as a weak interaction for mouse-feedback group and order (Group x Order; $F(2,23) = 3.15$, $p < 0.06$). Figure 5-2 illustrates the effects of sound on mouse-feedback groups on time and Figure 5-3 illustrates the effects of mouse-feedback group with order on time. Figure 5-4 illustrates the effects of sound and order on time. For error rates, a significant 4-way interaction was found (Group x Order x Task x

Sound; $F(2,23) = 6.91$, $p < 0.004$), as well as a significant interaction for sound and order; Sound x Order; $F(1,23) = 9.23$, $p < 0.006$. Figure 5-4 to Figure 5-7 are plots that illustrate the effects of sound and order for each mouse-feedback group on errors.

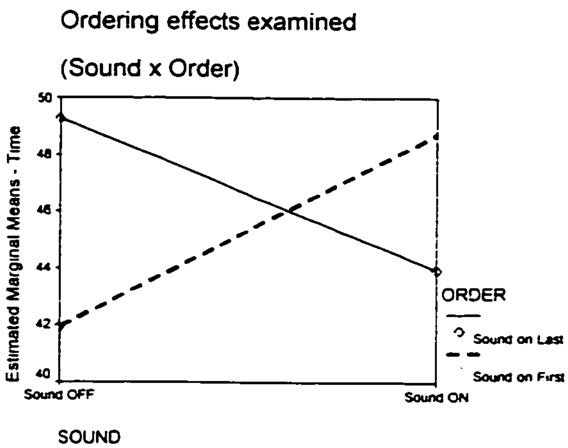


Figure 5-2. Time - Significant interaction for sound depending on order presented.

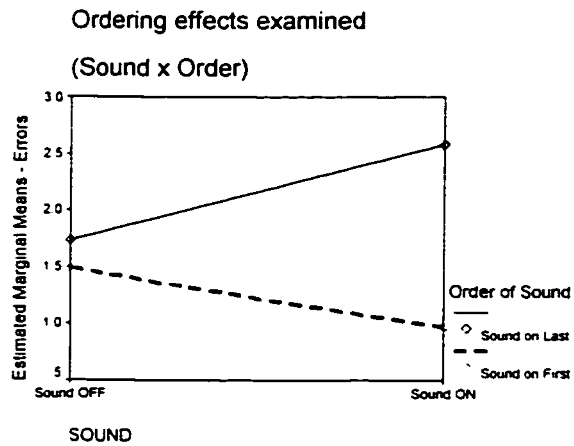


Figure 5-3. Errors – Significant interaction for sound depending on order presented.

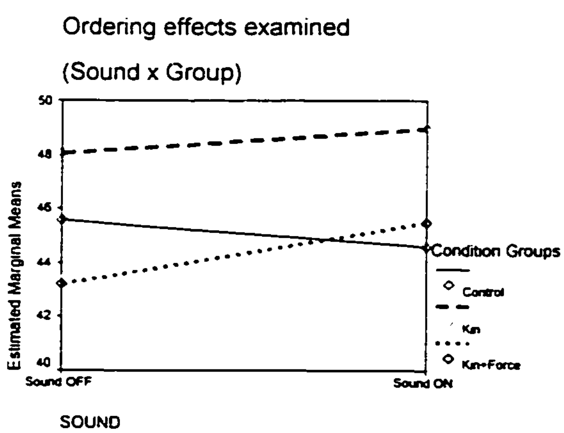


Figure 5-4. Time - Slight interaction for sound depending on mouse-feedback group.

[1] Please note that Group refers to Mouse-Feedback Group for the remainder of this thesis.

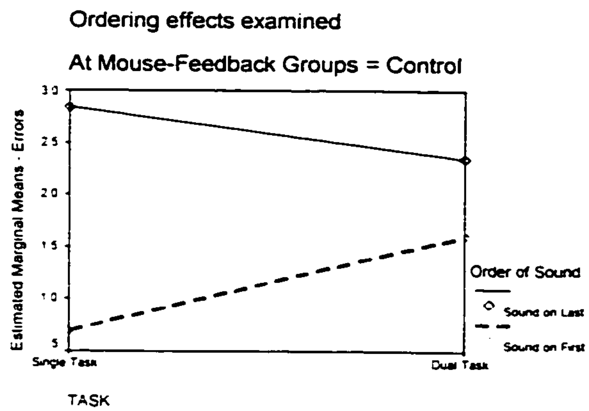


Figure 5-5. Errors examined for Control group according to task type and order.

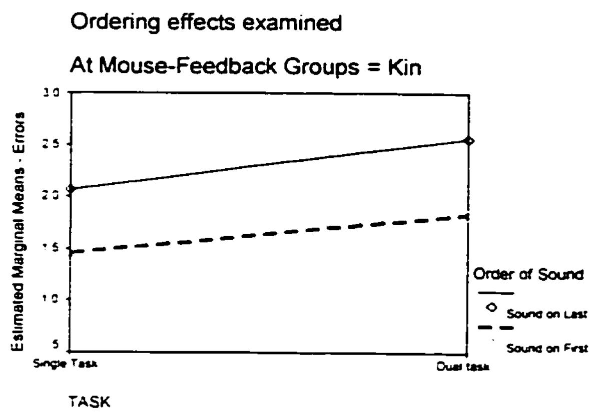


Figure 5-6. Errors examined for Kin group according to task type and order.

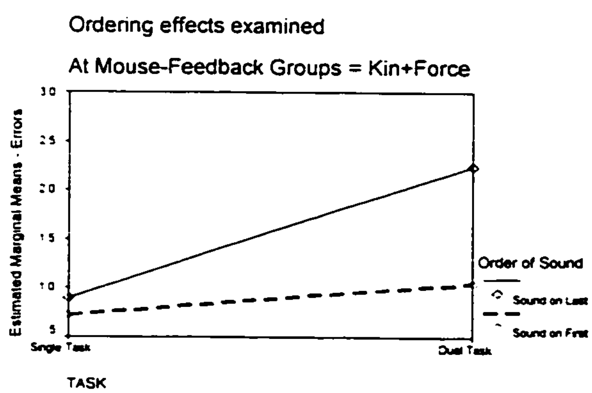


Figure 5-7. Errors examined for Kin+Force group according to task type and order.

From examination of the ordering effects, it appears that the blocks with Sound On trials had an effect on performance. In general, participants performed faster on their first set of trials, regardless of whether sound was on or off. In terms of error rates, in all mouse-feedback groups, participants who received the Sound On condition initially made fewer errors than the participants who received the Sound Off condition. In the Single Task condition (first 2 blocks of trials), both the Control Group and the Kin Group showed more noticeable spread between the error rates according to the order of sound presentation. Given that order effects were not anticipated, further detailed analyses of these order effects cannot be reasonably conducted as parsing the mouse-feedback groups according to order of sound would leave fewer than 5 participants in some of the cells.

5.3.3 Analysis of Variance

The experimental design was a repeated measure design, where a treatment group, (in this case, the use of feedback) was applied repeatedly over sets of participants at different levels (Single task or Dual task) under varying mouse-feedback (with sound or without sound). Thus, the data points produced were not independent, uncorrelated terms. As a result, a standard ANOVA cannot be applied. Repeated measures ANOVA provides uncorrected pair-wise comparisons among estimated marginal means for the main effects, for both between- and within-subjects factors (SPSS 7.5 1997). To further refine any significant results from the repeated measure analysis of variance, interaction effects between two aspects of the experiments (i.e., simple effects) were also examined.

Planned comparisons were used to compare the performance of the feedback groups on a specific block of trials. The following points outline the presentation of the results and describes the main research questions each section examines:

- Examining the effects of multi-modal feedback: Does the combination of kinesthetic and force feedback mouse with audio feedback help users perform tasks faster and more accurately?
- Does the kinesthetic and force feedback mouse without audio feedback help users perform tasks faster and more accurately?
- Does audio feedback alone help users perform tasks faster and more accurately?

All statistical results are presented in Appendix B as reference.

5.4 ANOVAs for Objective Measures

5.4.1 Average Time

An ANOVA was run on average trial time for mouse-feedback group with task type and auditory feedback (Group x Task x Sound). Significant main effects were found for task; $F(1,26) = 146.8$, $p < 0.001$. As it was expected, for all mouse-feedback groups the time for the Dual task was significantly slower than that for the single task, regardless of whether the Sound was On or Off. This main effect indicates that the Dual task condition was more difficult for the users than was the single task condition.

A significant 3-way interaction was found for average trial time; $F(2,26) = 3.91$, $p < 0.033$. Figure 5-8 and Fig 5-9 presents a plot of the means for the mouse-feedback groups according to sound by task type. The plot for the Single task did not reveal a simple interaction for mouse-feedback group by sound.

While main effects were not significant, an interesting pattern of results occurred in the single task. For all three mouse-feedback groups, performance did not appear to be affected by the auditory feedback. For both the Sound Off and Sound On conditions, trial times were faster for the Control condition, followed by the Kin+Force condition, and with the Kin condition having the slowest times. Pairwise comparisons of means did not reveal any significant differences between the actual performance times of the three mouse-feedback groups.

In the dual task, a significant interaction was not present. While main effects of group, sound, and pair-wise comparisons were not found to be significant, the observed patterns of results differed from those of the single task. When auditory feedback was absent, the Kin group and the Control group performed more slowly than the Kin+Force group. However in the sound on condition, the performance of the Control group improved with the addition of auditory feedback, while the Kin and Kin+Force groups show a decrement in performance.

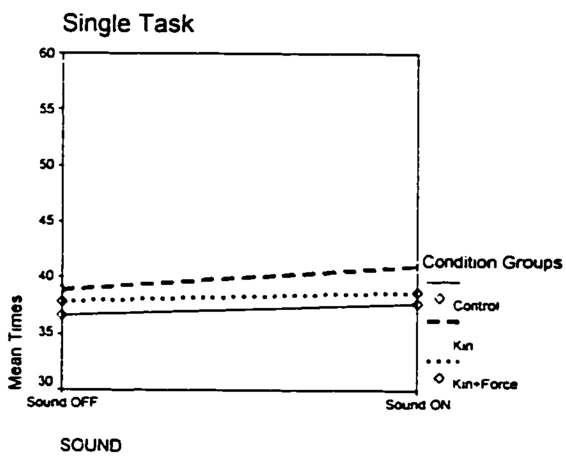


Figure 5-8. Estimated marginal mean time performance by mouse-feedback groups in the Single task when sound was Off or On. The single task appears to be faster for all group compared to dual task: (p <0.001).

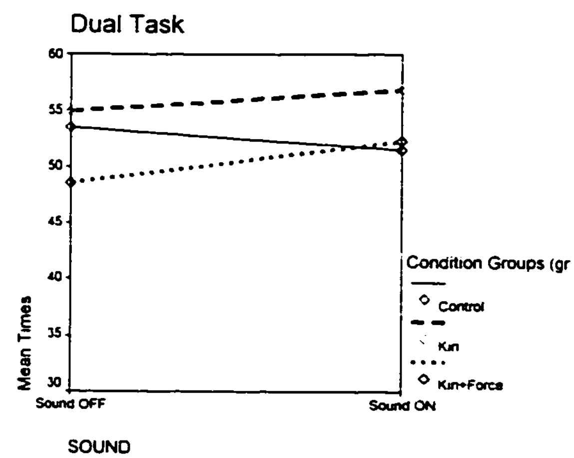


Figure 5-9. Estimated marginal mean time performance by mouse-feedback groups in the Dual task when sound was Off or On. For all groups, the dual task was performed significantly slower than the single task for both sound on and off: (p < 0.001).

5.4.2 Error Rates

Similar to the analyses for Average Time, a three-way ANOVA was carried out for Error rates (Group x Task x Sound). Figures 5-10 and 5-11 present a plot of error rates by task type and sound, as was done for Time. While an interaction appears to be present in the Single task condition, the three-way interaction was not statistically significant ($p > 0.05$). None of the two-way interactions were found to be significant either. The only main effect that was significant was for Task; $F(1,26) = 5.23$, $p < 0.031$ with error rates for the dual task being significantly higher than for the single task. Again, this confirms that the dual task condition was more challenging than the single task condition.

Pairwise comparisons of the mean Error Rates did not reveal any statistically significant differences. An examination of Figures 5-10 and 5-11 show some interesting patterns in the results. Mean Error Rates were lowest for the Kin+Force Group in both the Single and Dual Tasks, and highest for the Kin Group. For the Kin+Force group, there was a slight decrease in errors when the sound was on (as opposed to off) in the Single task conditions, and an increase in errors when the sound was on in the Dual task conditions. For the Kin Group, error rates for both the Single and Dual task showed slight decreases when the auditory feedback was on as compared to when it was off. For the Control Group, a slight increase in Error Rate was displayed in the Single Task when the auditory feedback was on. However, for this group Error Rate was largely unaffected by the presence or absence of auditory feedback in the dual task situation.

Results indicated high variance in the error scores for all mouse-feedback groups in all tasks. Figures 5-12 to 5-15 present individual plots for error rates in each block according to mouse-feedback groups along with standard deviations. This likely affected non-significant results for error analysis.

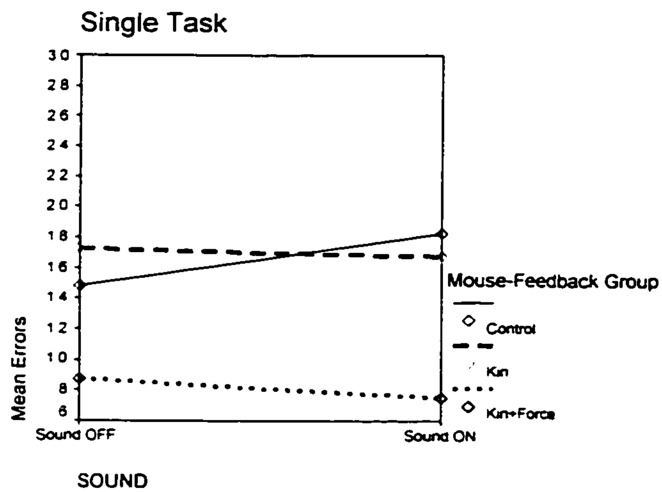


Figure 5-10. Presents plots for Single Task by sound for each group for errors.

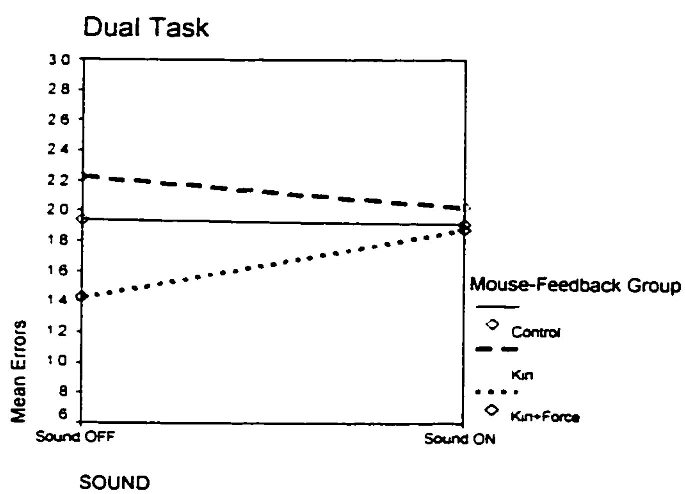


Figure 5-11. Presents plots for Dual Task by sound for each group for errors.

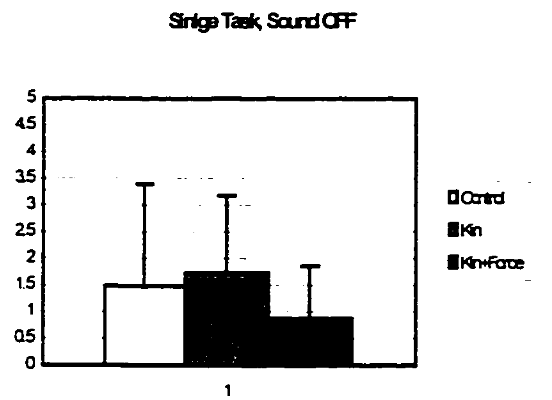


Figure 5-12. Mean time performance in Single task with Sound Off.

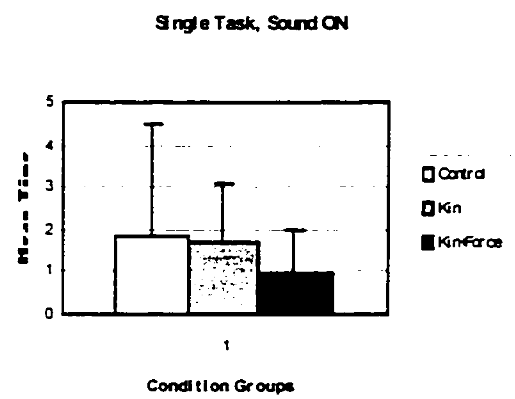


Figure 5-13. Mean time performance in Single task with Sound On.

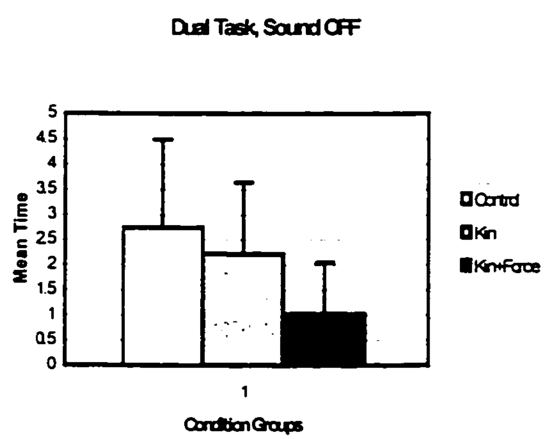


Figure 5-14. Mean time performance for Dual task with Sound Off

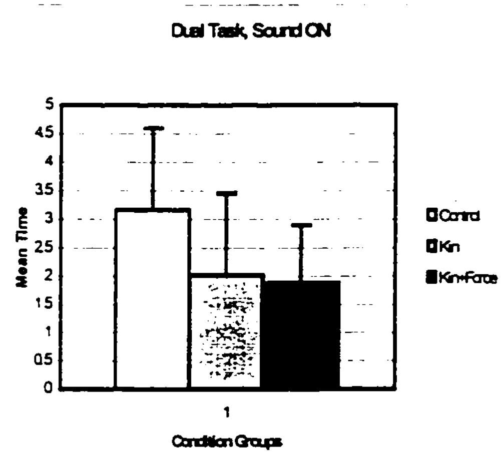


Figure 5-15. Mean time performance for Dual task with Sound On.

5.5 Subjective Workload Measures

NASA TLX consists of six linear rating scales: Mental demands, Physical Demands, Temporal Demands, Effort, Own Performance, and Frustration. The scales display a value from 1 to 100 (or percent value) for low to high respectively (with exception for rating of Own Performance that was from 1 to 100 for good to poor respectively). Additionally, an overall weighted workload rating that was generated by the NASA TLX program (See Chapter 4) was also examined. Each scale was examined separately in the following analysis.

Similar to the objective measures, repeat measures 2x2 ANOVA were run to test whether differences in workload (based on each of the 6 scales) demands existed for the within-subject measures (Single or Dual task type, and sound enabled or sound disabled) with the between-subject measure (3 mouse-feedback groups). As well, paired t-tests between mouse-feedback groups were used to examine any differences between blocks.

Although there were originally 30 participants in the study, the scores for one participant (in the Kin+Force Feedback group) were not analyzed in the NASA TLX because the corresponding data was lost (file became corrupted) after the trial completion from the NASA TLX program. This is not expected to alter the results significantly since only one participant score was not considered.

5.5.1 Mental Demands

A weak three-way interaction was seen for sound, task type, and mouse-feedback group; $F(2, 26) = 3.01$, $p < 0.067$. Figure 5-16 and Figure 5-17 illustrates the effects of mouse-feedback group and auditory feedback on mental demands based on task type. While the two-way interactions were not significant, main effects were found for Sound as well as for Task. Generally, participants rated the tasks without auditory feedback more mentally demanding than with auditory feedback; F (1, 26) = 13.24, p <0.001. Participants in all groups rated the dual task more mentally demanding than the single tasks; F(1, 26) = 94.93, p < 0.001.

Although the three-way interaction was weak, it indicated that participants generally had decreased mental demands ratings for both the single and dual task when auditory feedback was on than off except for the Kin+Force group. The Kin+Force group actually had slightly higher mental demands ratings in the single task when auditory feedback was on compared to when it was off.

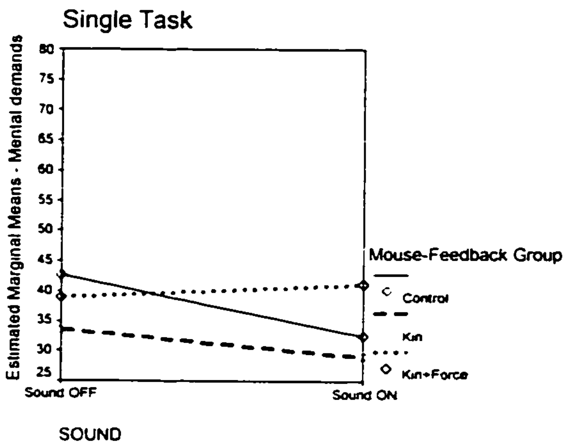


Figure 5-16. Mental demands ratings for single task by mouse-feedback group for sound.

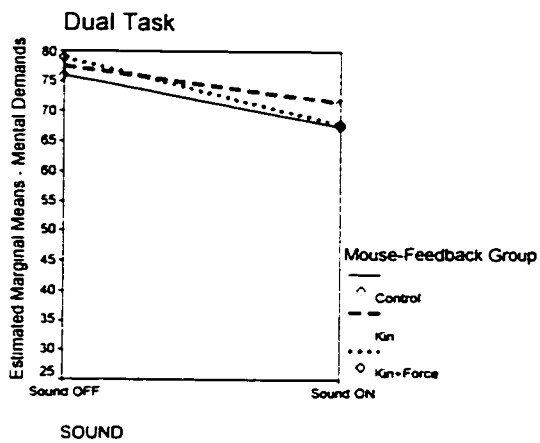


Figure 5-17. Mental demands ratings for single task by mouse-feedback group for sound.

The interaction between sound (On or Off) and task type (Single or Dual) showed a weak trend; F (2, 26) = 2.99, p < 0.095. Participants perceived reduced mental demands while performing the Single task when sound was on compared to the Single task when sound was off. Similarly, participants perceived less mental demands for the Dual task when sound was on compared to the Dual task when sound was off. Further, when sound was on, both the Single and Dual task mental demands ratings were lower than when sound was off.

The effects of sound (on or off) showed significance overall in all the trials; $F(1, 26) = 13.23$, $p < 0.001$. The results show that when the sound was on, participants perceived reduced mental demands to complete tasks than when the sound was off.

5.5.2 Physical Demands

The Control group rated all tasks as being less physically demanding than did the Kin+Force Group ($p < 0.023$). The Kin Group also rated the tasks as being less physically demanding than did the Kin+Force Group; $F(2, 26) = 4.9$, $p < 0.016$. As expected, the effects of task type (Single or Dual) showed significant differences in physical demands; $F(1, 26) = 4.71$, $p < 0.039$. As expected, the Single task had reduced physical demands than the Dual task.

When physical demands were compared based on mouse-feedback groups across all trials, significant differences existed overall; Participants in the Kin+Force group rated physical demands highest, followed by Kin group, and Control group.

T-tests showed that the Control group and Kin+Force group had significantly different ratings for the Single task with Sound Off ($p < 0.034$). The Control group had lower physical demands ratings than the Kin+Force. Similarly, when Kin and Kin+Force group were compared, several significant differences existed. First, in the Single task with Sound Off ($p < 0.010$), the Kin group had much lower phyiscal demands ratings than the Kin+Force group. In the Single task with Sound On ($p < 0.013$), the Kin group was lower again. However, in the Dual task with Sound Off ($p < 0.057$), the Kin group had a lower mean than the Kin +Force group which was interesting.

5.5.3 Temporal Demands

As expected, the effects of task type (Single or Dual) showed significant differences in temporal demands; $F(1, 26) = 37.52$, $p < 0.001$. The Single task required less temporal demands than the Dual task.

The effects of sound (on or off) for temporal demands ratings also showed significance; $F(1, 26) = 23.41$, $p < 0.001$. The Sound Off conditions increased temporal demands compared to Sound On for trials.

5.5.4 Effort

As expected, effort ratings based on task type (Single or Dual) showed significance; $F(1, 26) = 84.06$, $p < 0.001$. Participants found the Single task required less effort than the Dual task.

The effects of Sound (on or off) showed significance; $F(1, 26) = 14.03$, $p < 0.001$. The Sound Off conditions increased effort compared to Sound On conditions.

5.5.5 Frustration

A trend indicated a weak three-way interaction in ratings of frustration according to task, sound, and group; $F(2, 26) = 2.8$, $p < 0.079$. The scores were plotted and Figure 5-18 and Figure 5-19 illustrate the differences in frustration for Sound On and Sound Off conditions according to Single task and Dual task respectively.

When the Control and the Kin groups were compared, significant differences existed in the Dual task with Sound On ($p < 0.039$). The Kin group had lower frustration mean ratings than the Control group.

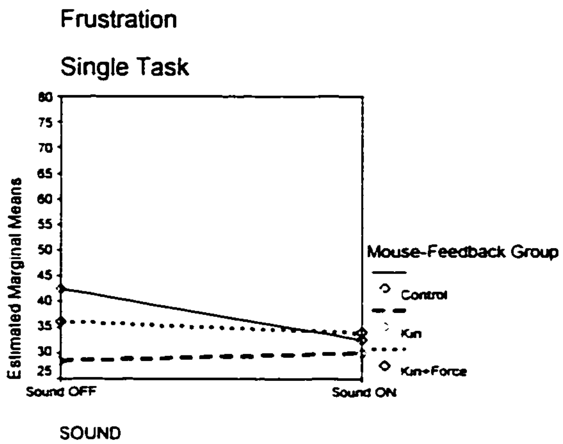


Figure 5-18. Frustration ratings for Single task for each group according to sound.

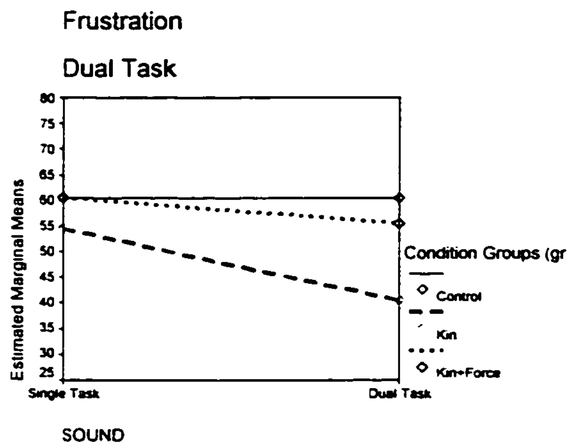


Figure 5-19. Frustration ratings for Dual task for each group according to sound.

5.5.6 Own Peformance

The effects of task type (Single or Dual) showed significance; $F(1, 26) = 41.64$, $p < 0.001$. Participants indicated their performance in the Single task condition was better (scale for "own performance" from 1 to 100 represents good to bad respectively) than their performance in the Dual task condition.

5.5.7 Overall Weighted Workload

The effects of task type (Single or Dual) was found to be significant; $F(1, 26) = 7.00$, $p < 0.014$. Participants indicated the Single task required less overall workload than the Dual task.

The effects of sound (Off or On) showed significance; $F(1, 26) = 123.99$, $p < 0.001$. Participants rated total workload higher for Sound Off ($M = 58.6$) than for Sound On ($M = 53.9$).

When the Control and Kin groups were compared, significant differences existed in the Single task when sound was off ($p < 0.045$). The Control group had higher overall workload ratings than the Kin group. Again, in the Single task with Sound On ($p < 0.081$), the Control group reported higher workload ratings than the Kin group. When the Kin and Kin+Force groups were compared in the overall ratings, significant differences existed for the Single task with Sound Off block ($p < 054$). Here, the Kin+Force group had a higher mean than the Kin group.

5.5.8 Overall Usability Ratings

At the end of the study, all participants were given a final questionnaire in order to evaluate their ability to detect mistakes because of the feedback experienced, as well as to provide subjective and qualitative feedback about the impact of the equipment and the sounds used. Some of the questions did not apply to the participants in the Control group or Kin group. The following sections will document the statistical results that showed significant differences.

Participants were asked to rate the ease of movement of the mouse for the calculator application on a scale from 1 to 6 (difficult to easy respectively). T-tests were used to compare the Control and Kin+Force groups; $t(17) = 5.48$ $p < 0.001$. Participants in the Control group (using a normal mouse) found it easier to move around the calculator application than the Kin+Force group. Similarly, participants in the Kin group had higher ratings than the Kin+Force groups; $t(18) = 4.58$, $p < 0.001$.

Participants were asked to rate the ability to detect mistakes they made because of the feedback from the mouse. T-tests showed significant differences in answers between the Control and Kin group; $t(9) = -2.51$, $p < 0.033$. Participants in the Kin group had higher ratings.

Participants were asked to rate their ability to detect mistakes (e.g., selecting wrong buttons during task) because of the feedback from the sounds. T-tests showed significant differences in the answers between the Control and Kin+Force groups; $t(17) = -2.55$, $p < 0.021$. Participants in the Kin+Force group generally rated the sounds more useful compared to the Kin and Control groups.

Chapter 6

Discussion

6.1 Effect of Multi-modal Feedback on Performance Measures

The results of the study examined the role of multi-modal feedback. For this research, multi-modal feedback referred to combining kinesthetic and force feedback in input devices and auditory feedback in interfaces via a mouse cursor with visual feedback. It was expected that proactive sound with a constrained mouse system that provided force feedback would help users perform computer tasks faster, more accurately, and with more ease than lesser combinations of feedback especially for more complex tasks. Similarly, it was expected that performance and preference would improve with the presence of either auditory feedback or force feedback for the dual task, as they would help reduce the information processing demands of the task.

6.1.1 Effect on Trial Times

The results confirmed that this combination of multi-modal feedback would affect task performance from a time perspective. A significant three-way interaction between mouse-feedback group, task and auditory condition was found ($p < 0.033$). As expected, participants in all mouse-feedback groups performed faster in the single task than the dual task. However, patterns in the results for the single task differed from patterns in the results for the dual task. When examined closely, in the single task condition, performance by the Control group was the fastest compared to the other groups. However, in the single task condition, the control group trial times were not found to be significantly different from the Kin+Force group trial times. Additionally, both these groups were faster than the Kin group in the single task. This suggests that participants using a mouse with kinesthetic and force feedback will not perform at a slower pace than participants using a normal mouse in terms of time.

As mentioned, patterns in the performance of each of the mouse-feedback groups differed in the dual task compared to the single task. In the dual task, participants in the Kin+Force group performed the fastest overall. Followed by the Control group and the Kin group. This suggests that a task that has components that impose heavily on the visual system may benefit from a kinesthetic and force feedback mouse.

Another interesting aspect of performance that arises from the three-way interaction is the effect audio feedback had on time performance in the single task and the dual task. In the single task, all groups performed faster when the sound was off than when the sound was on. Auditory feedback was expected to help improve performance. Similar patterns were present in the dual task results for the Kin and Kin+Force groups but not for the Control group. The Control group performed better when the sound was on than when the sound was off in the dual task. One reason for this may be that the single task condition was simple and required less information processing than the dual task condition. Participants made greatest use of the visual sense to perform the single task conditions. This supports the well-known psychological observation regarding dominance of the visual modality (Edwards & Stevens 1997). However, in the dual task, participants in the Control group had to make use of the audio feedback to support the visual sense because the task difficulty increased. This may not have been necessary in the other two groups because of the inherent kinesthetic (and kinesthetic and force) feedback provided by the VRM.

These findings suggest that performance in a fairly simple task condition may be very different from performance in a dual task condition depending on feedback. Thus, further examination of feedback types and task conditions is needed to fully understand the impact feedback with respect to task complexity.

Results from t-tests were used to compare the effects the mouse-feedback groups had on the sets of trials. As mentioned, the Control group used a normal mouse. The Kin group used a constrained mouse that only provided kinesthetic feedback. The Kin+Force group used the constrained mouse with force feedback. The analysis of trial times showed that the Kin+Force feedback performed better than the Kin group in the dual task when sound was off ($p < 0.070$). Although the difference wasn't significant, the scores indicated that the Kin+Force group also performed better than the Control group in the dual task when sound was off. Data analysis did not show any significant differences in error rates. It may be that the addition of the force feedback served to aid the user in completing the dual task. The fact that significant differences were not found for error rates suggests that the advantage of the Kin+Force feedback was not simply due to a speed-accuracy trade-off.

The Kin group's poorer performance in trials in terms of time could be explained by a couple of factors. The kinesthetic mouse device maintained a stronger surface friction and system inertia than a normal mouse. This was because of the metal arm that was affixed to the mouse with the servomotor system (See description of the VRM in Chapter 3) that controls the two degrees-of-freedom movement of the mouse. This may have contributed to higher trial times. In a "normal" mouse (Control mouse-feedback) the surface friction and inertia are almost non-existent for the system as a whole (at least psychologically, what the user feels). However, participants using the VRM with active force feedback performed with fairly fast speeds compared to the Control group. This was surprising to some extent because participants' have to use enough force to move between buttons and clear the surface friction and active force feedback. Again, one could conclude that the force feedback played a significant role in increasing participants awareness of cursor position and thus, it reduced cursor travel distance. Reducing the travel distance and making discrete mouse movements could have significantly reduced the mean time for trials compared to the same device without force feedback.

6.1.2 Effect on Error Rates

It was expected that multi-modal feedback would improve error rates for both the single and dual task. The three-way ANOVA of error performance between the groups did not reveal any significant differences. However, a significant main effect for task was found ($p < 0.05$). Error rates were higher for the dual task conditions than the single task conditions.

One possible reason a three-way interaction of error performance between the groups did not show significant differences was because of the large variance in the mean error rates between mouse-feedback groups. When the mean error data was examined, the Kin+Force group had the lowest error rates under all conditions.

Performance by the mouse-feedback groups in the single and dual task when sound was on and off yielded plots that were perplexing. In order to understand them, accuracy graphs and corresponding time graphs were examined together. This was a reasonable course of action because of the speed-accuracy trade-off principle. In the single task, participants in the Kin and Kin+Force groups performed more slowly when the sound was on than when the sound was off. This seems to have had an impact on the errors, because as time increased, the errors decreased.

The reciprocity of speed and accuracy were also observed in the dual task performance based on sound state. Participants in the Control and Kin groups performed with less errors but took more time to complete dual tasks when sound was on compared to when sound was off. However, participants in the Kin+Force group performed with more errors while taking more time to complete dual tasks when sound was on than when sound was off. One could suggest that kinesthetic and force feedback requires more resources than visual alone and that compounding it with auditory feedback can cause confusion which results in increased time as well as increased errors.

A couple of possible explanations for a reduction in the number of errors by the Kin+Force group (although not significant) can be attributed to the proactive force feedback from the mouse. Participants in this group could feel when the cursor passed over a button. Thus, at this point, they were able to click the button if it was their desired target location using the visual sense. Visual attention need only be diverted to the calculator interface as confirmation or a secondary cue especially in the dual task conditions. Participants in the Control group could only determine when the cursor reached or left the target position visually. Slip-off errors are common in GUIs. Brewster (1997) states that "a common difficulty is slipping off a button by mistake or without noticing". The kinesthetic and force feedback ensured that participants were aware if the cursor slipped off a button and this could be another reason accuracy was improved in this group.

6.2 Effect of Multi-modal feedback on Subjective Measures

Unlike some of the positive performance measures for the Kin+Force group, the results from the NASA TLX showed that participants in the Kin+Force group reported the highest ratings for physical demands. When the results were compared, the Kin+Force group had a much higher physical demands rating compared to the Kin group or the Control group.

The NASA TLX showed that the use of sound affected various perceived workload factors. The mental demand analysis showed that participants rated sound off trials with higher mental demands than sound on trials ($p < 0.001$). The temporal demand analysis showed that participants rated the sound off trials with higher temporal demands than sound on trials ($p < 0.001$). The effort analysis showed that participants rated the sound off trials with higher effort than the sound on trials. The overall workload analysis showed that sound off trials resulted in higher overall workload than the sound on trails. Final Questionnaire results indicated that participants generally found the sounds useful. In summary, the subjective ratings clearly indicated that participants perceived the sounds to be useful and helpful for the tasks they performed.

While participants interpreted the auditory feedback to be helpful, this advantage was not supported by the performance data. Participants reported higher mental demands for the single and dual task conditions when there was no auditory feedback compared to when there was auditory feedback. However, there was one exception: in single task conditions, participants in the Control group reported that tasks required more mental demands to perform tasks when sound was on than when sound was off. In terms of physical demands, the Kin+Force group had significantly different ratings. Participants using the kinesthetic and force feedback device reported the highest physical demands for all blocks. Frustration levels varied in a three-way interaction based on mouse-feedback group, sound and task type. These results mainly showed that sound on trials were found to be less frustrating for participants in all groups (except Kin+Force in the single task which stayed at the same level) than sound off trials in the single condition and dual conditions.

The results from the Final Questionnaire showed that participants in the Kin+Force group rated ease of movement characteristics with the device as very poor. When compared with the other mouse-feedback groups, the Kin+Force group had a much lower rating compared to the Control and Kin groups. Participants rated the VRM (without feedback) as "average" compared to the normal mouse (rated high) and the VRM with Force Feedback mouse (rated poor).

In terms of devices, these findings indicate that the participants perceived the kinesthetic device very differently than the kinesthetic and force feedback device. The only difference between the two devices is that the later had the addition of force feedback. Yet, participants rated the kinesthetic device with force feedback as requiring more effort than the kinesthetic device. Therefore it seems reasonable to conclude that the force feedback was not a feature participants interpreted as being as useful in their tasks, even though there is some evidence from the performance measures that the addition of force did help improve time to complete the trials. Correspondingly, a question in the Final Questionnaire asked participants from the Kin and Kin+Force groups whether they preferred the feedback they received from the mouse rather than the audio feedback. The results showed that participants in the Kin+Force group preferred the mouse rather than the audio.

Chapter 7

Conclusions & Recommendations

7.1 Contributions and Limitations of this Research

This study is somewhat ground-breading in that few published studies have reported investigating the effects of combining kinesthetic feedback, force feedback, and non-speech auditory feedback in interfaces for "everyday" computing tasks. This research was intended to contribute to a better understanding of how feedback might be used in HCI-related tasks. Constraining a mouse to a fixed grid (as is the VRM mouse) and providing force feedback does not seem to compromise performance times on simple tasks when compared to performance using a "regular" mouse. For single tasks, performance times and error rates were comparable for all three input-device conditions. In the dual task condition, the performance achieved with the kinesthetic-feedback was similar to that achieved with the regular mouse. More noteworthy was the fact that for the dual tasks the combined effect of kinesthetic feedback and force feedback did appear to enhance performance above that of the regular and kinesthetic-only mouse conditions.

This study was also groundbreaking because it used an "everyday" task with two levels of complexity to investigate basic performance issues. Generally, studies evaluating pointing devices use rapid pointing tasks or very simple tasks to measure speed and accuracy. It was felt that the Feedback Calculator tasks included features that were inherently common in present day computing while demanding more information processing than simple pointing tasks.

Including a dual task condition allowed for the simulation of a more complex work environment. Having participants monitor a second display is not unlike the type of task performed by a stock trader. Stock traders generally use a dual monitor computing systems and regularly monitor a secondary screen that provides stock quotes and current news. Thus the generic nature of the single and dual tasks should increase the generalizability of these results as well as increase the ecological validity of the study.

The Feedback Calculator application was designed to be a reusable research tool. All components can be changed using the source code. The Feedback Calculator application interface can be modified or changed completely to suit the requirements of future studies. In terms of measurement tools, the application provides timers to calculate trial time as well as rudimentary error checking features.

The VRM device by Control Advancements Inc. used in this study is a promising device for further research into kinesthetic and force feedback. The VRM device provides a variety of settings for effects (i.e., wall, spring, damper, etc.) that are user-adjustable. This, in turn, provides a lot of flexibility to a researcher. One limitation in terms of feedback research set-up lies in the fact that there are no guidelines as to the most effective feedback settings (i.e. magnitude of strength of force) for particular tasks or for individual participants. For this study, the effects were selected so that the user would feel as if he/she were placing a finger on a physical calculator keyboard thereby feeling the bumps of the interface. The strengths of the effects were set the same for both male and female participants. It is possible that individual performances may have improved further if the participants in the Kin+Force group had been allowed to set the force feedback level for themselves.

The way in which earcons were used in this study was also groundbreaking. Most studies of the effects of earcons make use of reactive feedback. In this study, simple and brief audio messages (tones) were issued to provide proactive and transient feedback. This allowed for the synchronizing of the auditory feedback with the force feedback. In other words, when both audio and force feedback were active, participants heard the earcons and felt the force feedback at the same time. The auditory feedback was relayed for approximately 0.25 seconds, as was the force feedback (depending on the speed at which the user approached the button). The results showed that auditory feedback was not as useful as force feedback in improving performance. This is contrary to the effects typically found with reactive earcons, wherein reactive earcons have been found to improve performance. As yet there are no standards or clear guidelines aimed at improving the design of auditory feedback and earcons.

Studies to clarify the role reactive and proactive auditory feedback can play in everyday tasks are needed if guidelines for designers are to be set out. Another important aspect of auditory feedback that needs to be explored is how the geometric layout of the data entry interface influences the suitability and impact of auditory cues. In the Feedback Calculator application, the numbers from one to nine and zero were laid out like a numeric keypad as opposed to a one line of buttons layout (i.e., showing linearity). With a numeric keypad layout, users may have to move the cursor across three rows of buttons to get to a desired spot. The movement can consist of hitting numerous buttons along the way and without a particular order. For example, if the user had to move the cursor to the button '9' from '1', he/she could move diagonally across the keypad and hit '5' on the way. Thus the numeric keypad layout may have hampered the user making any associations between numerically proximal digits. If the keypad had been laid out in a single line as opposed to three lines, the impact may have been different. For example, in a linear layout the user wishing to move from '1' to '9' would pass over all the buttons in the middle and heard the patterns and associations between the related earcons. The layout of the buttons and their associated sounds may influence user sensitivity to the auditory feedback. This is also an area that needs to be explored further by researchers because of its potential impact on the design and layout of auditory feedback and interfaces.

7.2 Design Issues and Implications

Multi-modal interfaces for computers are relatively recent phenomena. They are advancing rapidly in terms of technology and popularity. It is important to continue to perform scientific studies in this area to ensure that the benefits to the user are "real" and worthwhile. Research in the area of multi-modal interfaces is just starting to address the issues of performance and preference.

The effects of kinesthetic and force feedback in pointing devices and non-speech audio feedback in user interfaces on "everyday" computing tasks have begun to be addressed through this research. In terms of the simple "everyday" task, the mouse with kinesthetic and force feedback was comparable in time performance to that achieved with the regular mouse. This is an important point because designers need to know that introducing such a device is not likely to hamper overall task performance. Furthermore, for more visually demanding tasks the combination of kinesthetic and force feedback may actually enhance performance over that achieved with a regular mouse. There appears to be a complex relationship between time and accuracy performance for multi-modal feedback that doesn't follow the traditional speed-accuracy trade-offs. Further research may help to clarify this relationship so designers can exploit any performance advantages achieved through multi-modal feedback.

Although performance measures indicated that participants performed better with a mouse with kinesthetic and force feedback, their preference ratings indicated higher mental demands, higher physical demands, greater frustration, and higher overall workload ratings compared to a normal mouse. This finding serves as a warning for designers - it is important to include performance measures as well as preference data when evaluating input/output devices. If a designer is aware that users may subjectively interpret an input device as requiring more effort (even though performance may be better), then there is a chance to re-examine the subjective physical and cognitive demands prior to releasing it to the public domain.

Although this research did not specifically examine the effects of feedback presentation order and how it influences both objective and subjective measures, preliminary analyses indicated that ordering effects could play a role in overall performance. The examination of presentation order for auditory feedback suggests that participants that were introduced to sound at the start of the experiment had significantly different results than participants introduced to sound later in the experiment. In fact, participants generally performed better, both in terms of time and errors, when sound was introduced at the start of the experiment. This may have ramifications for the design of training associated with computing tasks and systems. It may be that people are more receptive to using new feedback methods if they are introduced earlier rather than later in the training process.

The objective performance measures and subjective ratings were contradictory for sound. The performance measures showed no significant differences when sound was on or off. Yet, participants overwhelmingly preferred the "sound-on" conditions to the "no-sound" conditions. Thus, sound may effect a participant's perception of personal performance. To clarify, participants may have thought their performance was superior when sound was on than when sound was off. For example, ratings on the NASA TLX indicated that participants felt the "sound-on" conditions decreased mental demands, temporal demands, effort, and overall workload for both single and dual tasks. This suggests that auditory feedback may be highly salient. Therefore, when designing auditory feedback for computer tasks, it may be better to include sound even if the inclusion of sound does not immediately improve performance. The inclusion of sound feedback may serve to increase user acceptance and satisfaction with a new system.

7.3 Recommendations for Further Research

Buxton (1995) has pointed out the need to develop multi-sensory, multi-channel, and multi-tasking user interfaces in order to take advantage of the full potential of human sensory systems. Such holistic systems seem to be the best route to move human computer interaction forward into the next millennium. Further research and technological advancement is needed to reach this goal. There are many aspects of multi-modal systems that need to be examined. Based on the findings from this study, a deeper understanding of the psychological effects and impact of the different feedback modalities should be examined. This could help designers understand where feedback is particularly needed in a task, when it is desired, and how it should be delivered. Further, this may establish guidelines for exploiting the performance advantages that might be realized when combining certain feedback modalities with particular tasks.

The effect of presenting auditory feedback during the initial training phases warrants further exploration. It is not known whether the audio channel is most affected by the presentation order or whether all feedback types (including kinesthetic and force feedback) could be similarly affected. Knowing when to introduce a new user to a particular mode of feedback so as to maximize user performance and acceptance could have important implications for the design of computing interfaces and tasks and the training associated with such devices and tasks.

In order to improve safety and usability in computing, there is a need to be cautious and judicious in providing alternative approaches to computing. Already, many users experience eye fatigue and repeated strain injuries (RSIs) from prolonged use of computers and input/output devices. Researchers need to examine whether multi-modal feedback could add to or lessen physiological health problems. More importantly, this understanding should be used to increase the safety of the next generation of devices. For example, participants using a mouse with kinesthetic and force feedback rated the devices as imposing higher mental and physical demands than were reported by those using the regular mouse. Given the potential for improving task performance, it is important that multi-modal feedback devices be evaluated over longer periods of use to ensure that they do not contribute to use-relate injuries.

As yet, guidelines for the specific use and integration of force feedback and auditory feedback in human computer interaction are lacking. With respect to sound, many operating systems come packaged with software that allows users to add sounds to interface objects, events, and processes. However, this is usually done in an ad-hoc manner. Brewster & Gray (1998) point out that "there is no guarantee that the sounds (added arbitrarily) will be anything more than gimmicks and so are unlikely to improve usability". Further research is needed to establish guidelines and standards that will help users and designers to avoid gimmicks and utilize sounds effectively.

In summary, there may be many opportunities for integrating multi-modal feedback into everyday computing. Auditory feedback appears to have a positive influence on user perception of task demands, even if they do not actually enhance task performance. Force feedback appears to enhance task performance, particularly for more visually demanding tasks. This study has taken a step towards establishing guidelines for building multi-modal feedback interfaces and devices that will serve to improve human-computer interaction.

Bibliography

Akamatsu, M. & Sato, S. (1994). A multi-modal mouse with tactile and force feedback. International Journal of Human-Computer Studies, 40, 443-453.

Akamatsu, M., MacKenzie, I. S. & Hasbroucq, T. (1995). A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. Ergonomics, 38, 816-827.

Akamatsu, M., & MacKenzie, I. S. (1996). Movement characteristics using a mouse with tactile and force feedback. International Journal of Human-Computer Studies,. 45, 483-493.

Akamastu, M., MacKenzie, I. S., & Hasbrouq, T. (1995). A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. Ergonomics, 38, 816-827.

Alistair D. N. Edwards, (1989). Soundtrack: An Auditory Interface for Blind Users. Human-Computer Interaction, 4(1), 45-66.

Alty, J.L (1995). Can we use music in human computer interaction? Proceedings of HCI'95, (pp. 409-423) Huddersfield, UK: Cambridge University Press.

Baddeley, A. (1986). Working Memory. Oxford Psychology Series, 11. Oxford: Clarendon Press.

Baddeley, A. & Hitch, G.J., (1974). Working memory. In Bower, G. (Ed.), Recent Advances in Learning and Motivation Vol. VIII, New York: Academic Press.

Baecker, R. M., Grudin, J., Buxton, W. A. S. & Greenberg, S. (Eds.). (1995). Readings in human-computer interaction: Toward the year 2000, [Chap. 7: Touch, Gesture, and Marking]. San Francisco: Morgan Kaufmann Publishers.

Bailey, R.W.(1989). Human Performance Engineering, (2nd Ed.) PTR Englewood Cliffs, New Jersey: Prentice Hall.

Balakrishnan, R., Baudel, T., Kurtenbach, G., & Fitzmaurice, G. (1997). The Rockin' Mouse: Integral 3D Manipulation on a Plane. Proceedings of CHI'94, [On-line], Available: http://www.acm.org/sigchi/chi97/proceedings/paper/rb1.htm#U3

Baldis, S. (1997) Input Device Interface for 3-D Interaction. Unpublished master's thesis, University of Washington, Washington, U.S.A.

Billinghurst, M., Savage, J. Oppenheimer, P. & Edmond, C. (1995) The Expert Surgical Assistant: An Intelligent Virtual Environment with Multimodal Input. Proceedings of Medicine Meets Virtual Reality IV, (pp. 590-607). San Diego, CA: IOS Press.

Blattner, M., Sumikawa, D. & Greenberg, R. (1989). Earcons and icons: Their structure and common design prinicples. Human Computer Interaction, 4(1), 11-44.

Blesser, T., & Sibert, J. (1990). Toto: a tool for selecting interaction techniques. Proceedings ACM UIST'90 Symposium on User Interface Software and Technology, (pp. 135-142) Utah: ACM.

Bliss, J. C., Katcher, M. H., Rogers, C. H. & Sheppard, R. P. (1970). Optical-to-tactile image conversion for the blind. IEEE Transactions on Man-Machine Systems, MMS-11, 58-65.

Brewster, S.A. (1997). Using non-speech sound to overcome information overload. Displays, Special issue on multimedia displays, 17, 179-189. [On-line], Available: http://www.dcs.gla.ac.uk/~stephen/publications.HTML

Brewster, S.A. (1998, in press). The design of sonically-enhanced widgets. Interacting with Computers, [On-line], Available: http://www.dcs.gla.ac.uk/~stephen/papers/IwC98.pdf

Brewster, S.A. & Gray, P.D. (1998). Guidelines for the use of sound in multimedia human-computer interfaces, [On-line], Available: http://www.dcs.gla.ac.uk/~stephen/Grants/Grants.HTML

Brewster, S.A., Wright, P.C., Dix, A.J. & Edwards, A.D.N. (1995). The sonic enhancement of graphical buttons. In K. Nordby, P. Helmersen, D. Gilmore, & S. Arnesen (Eds.), Proceedings of Interact'95, (pp. 43-48) Lillehammer, Norway: Chapman & Hall. [On-line], Available: http://www.dcs.gla.ac.uk/~stephen/publications.HTML

Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1992). A detailed investigation into the effectiveness of earcons. In G. Kramer (Ed.), Auditory display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display, (pp. 471-498) Santa Fe Institute, Santa Fe, NM: Addison-Wesley. [On-line], Available: http://www.dcs.gla.ac.uk/~stephen/publications.HTML

Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Eds.), INTERCHI'93, Amsterdam: ACM Press, Addison-Wesley, pp. 222-227. [On-line], Available: http://www.dcs.gla.ac.uk/~stephen/publications.HTML

Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1994). The design and evaluation of an auditory-enhanced scrollbar. In B. Adelson, S. Dumais, & J. Olson (Eds.), Proceedings of CHI'94, Boston, Massachusetts: ACM Press, Addison-Wesley, pp.

173-179. [On-line], Available: http://www.dcs.gla.ac.uk/~stephen/publications.HTML

Brooks, L.R. (1968). Spatial and verbal components in the act of recall. Canadian Journal of Psychology, 22, 349-68.

Brooks, J. (1990). Project GROPE-Haptic Displays for Scientific Visualilzation. ACM Computer Graphics, 24(4), 177-185.

Brown, M.L., Newsome, S.L. & Glinert, E.P. (1989). An experiment into the use of auditory cues to reduce visual workload. Proceedings of ACM CHI'89, (pp. 339-346) Austin, Texas: ACM Press, Addison-Wesley.

Buxton, W. (1983). Lexical and pragmatic considerations of input structures, Computer Graphics, 17, 31-7.

Buxton, W. (1990). A Three-State Model of Graphical Input. In D. Diaper et al. (Eds.), Human-Computer Interaction - INTERACT '90, (pp. 449-456.) Amsterdam: Elsevier Science Publishers B.V. North-Holland.

Buxton, W. (1991). The three mirrors of interaction: a holistic approach to user interfaces. Proceedings of Friend21 '91 International Symposum on Next Generation Human Interface, Tokyo, Japan, Nov. 25-27, 1991. [On-line], Available: http://www.dgp.toronto.edu/people/BillBuxton/index.html.

Buxton, B. (1996). A Directory of sources for input technologies. Alias | Wavefront, Inc., University of Toronto, Toronto, Canada. [On-line], Available: http://www.dgp.toronto.edu/people/BillBuxton/InputSources.html

Byers, J.C., Bittner, A.C., Hill, S.G., Zaklad, A.L. & Christ, R.E. (1988). Workload Assessment of a Remotely Piloted Vehicle (RPV) System. Proceedings of the Human Factor Society, 32nd Annual Meeting, (pp. 1145-1149) Santa Monica, CA: Human Factors Society.

Card, S., English & Burr. (1978). Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys and Text Keys for Text Selection on a CRT, Ergonomics, 21(8), 601-613.

Card, S. K., MacKinlay, J. D. & Robertson, G. G. (1991). A morphological analysis of the design space in input devices. ACM Transactions on Information Systems, 9, 99-122.

Cheng, Li-Te, (1996). Evaluation of tactile feedback in delicate virtual reality operations. Unpublished master's thesis. University of Waterloo, Waterloo, Canada.

Clark, F.J. & Horch, K.W. (1986). In Kinesthesia. Boff, KR Kaufman, L, & Thomas, J.P. (Eds.), Handbook of Perception and Human Performance, V1 Sensory Processes and perception, 13, 1-62. New York: Wiley.

Cohen, J. (1992) Monitoring background activities. Proceedings of ICAD92, pp. (pp. 499-532) Santa Fe Institute. Santa Fe: Addison-Wesley.

Dix, A.J., Finlay, J.E., Abowd, G.D., & Beale, R., Finley, J.E. (1998). Human Computer Interaction, (2nd Ed.) Prentice Hall Press.

Edwards, A. D. N. & Stevens, R. D. (1997). Visual dominance and the World-Wide Web. Proceedings of the Sixth International World Wide Web Conference, Santa Clara, California: Stanford University. [On-line] Available: http://www.cs.york.ac.uk/~alistair/papers/Visual_dominance.html

English. W., Engelbart, D. & Berman, M.L. (1967). Display Selection Techniques for Text Manipulation. IEEE Transactions on Human Factors in Electronics, 8(1), 21-31.

Fukui, Y., & Shimojo, M. (1992). Edge Tracing of Virtual Shape Using Input Device with Force Feedback. Systems and Computers in Japan, 23(5), 94-104.

Fitts, P.M. (1954). The information capacity of human motor responses in controlling the amplitude of movement. Journal of Experimental Psychology, 47, 381-391.

Gaver, W.W. (1989) The Sonic Finder: An Interface that Uses Auditory Icons. Human-Computer Interaction. V 4, N 1, p 67-94.

Gaver, W.W. & Smith, R.B. (1995). Auditory icons in large-scale collaborative environments. Chapter 8 in R.M. Baecker, J. Grudin, W. Buxton & S. Greenberg, S. (Eds.) Readings in Human Computer Interaction: Toward the Year 2000, San Francisco: Morgan Kaufmann Publishers.

Gigante, A. (1993). Virtual Reality: Definitions, History and Applications. In R. A. Earnshaw, M. A. Gigante, & H. Jones, (Eds.), Virtual Reality Systems, Chapter 1. Academic Press.

Good, A. (1992). Participatory Design of A Portable Torque-Feedback Device. In P. Bauersfeld, J. Bennett, & G. Lynch, (Eds.) Proceedings of the Conference on Human Factors in Computing Systems, CHI'92, (pp. 439-446) ACM/SIGCHI.

Greenstein, Joel S. & Arnaut, Lynn Y. (1988). Input Devices. In Helander, M. (Ed.). Handbook of HCI, (pp. 495-519) Amsterdam: North-Holland.

Hart, S.G. (1987). Background Description and Application of the NASA Task Load Index (TLX). Proceedings of the Department of Defense Human Engineering Technical Advisory Group Workshop on Workload, (pp. 95-117) (NUSC 6688) [On-line] Newport, RI: Naval Underwater Systems Center.

Hart, S.G., & Staveland, L.E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human Mental Workload, Amsterdam. The Netherlands: Elsevier

Jacob, R. (1991). The use of eye movements in human-computer interaction techniques: What you look at is what you get. ACM Transactions on Information Systems, 9, 3, 152-169.

Lombardi, V. (1995). Audio on the Internet. Unpublished master's thesis, Oct. 1995 - Updated April 1997. NYU School of Education, Music Technology Program. U.S.A. [On-line], Available: http://www.nyu.edu/acf/atg/library

MacKenzie, I. S. (1992a). Movement time prediction in human-computer interfaces. Proceedings of Graphics Interface '92, (pp. 140-150) Toronto: Canadian Information Processing Society.

MacKenzie, I. S. (1992b). Fitts' law as a research and design tool in human-computer interaction. Human-Computer Interaction, 7, 91-139.

MacKenzie, I. S. (1995). Input devices and interaction techniques for advanced computing. In W. Barfield, & T. A. Furness III (Eds.), Virtual environments and advanced interface design, (pp. 437-470) Oxford, UK: Oxford University Press.

Mackinlay, J. D., Card, S. & Robertson, G. G. (1990). A semantic analysis of the design space of input devices. Human-Computer Interaction, in press.

Massie, T. H. & J. K. Salisbury. (1994). The PHANToM Haptic Interface: a Device for Probing Virtual Objects. Proceedings of the ASME Winter Annual Meeting, Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Chicago.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 64, 81-97.

Milner, N. (1988). A review of human performance and preferences with different input devices to computer systems, In D. Jones & R. Winder (Eds.). People and Computers IV, Proceedings of the Fourth Conference of the British Computer Society Human-

Computer Interaction Specialist Group, (pp. 341-362) Cambridge: Cambridge University Press.

Mountford S.J. & Gaver W.W., (1990). Talking and listening to computers. Laurel, B. (Ed.), The art of human computer interface design, (pp. 319-335) New York: Addison Wesley.

Münch, S. (1996). The Force Mouse. ForceAgent™ [On-line], Available: http://wwwipr.ira.uka.de/~miami/ForceAgent/node3.html

Nelson, R. J., Mccandlish, C. A., & Douglas, V. D., (1990). Reaction times for hand movements made in response to visual versus vibratory cues. Somatosensory and Motor Research, 7, 337-352.

Nielsen, J. & Schaefer, L. (1993). Sound Effects as an Interface Element for Older Users. Behaviour & Information Technology, 12(4), 208-215.

Norman, D. A. (1981). Categorization of Action slips. Psychology Review, 88(1), 1-15.

Norman, D. A. (1988). The Design of Everyday Things. New York: Basic Books.

O'Malley, M. (1995). Speech, Language & Audition, Chapter 8. In R.M. Baecker, J. Grudin, W. Buxton & S. Greenberg, S. (Eds.) Readings in Human Computer Interaction: Toward the Year 2000, San Francisco: Morgan Kaufmann Publishers.

Preece, J., Rogers, Y., Sharp, H., Benyon, D. Holland, S. & Carey, T. (1994). Human-Computer Interaction, Massachusetts: Addison-Wesley.

Quionnes, M.A.P, & Sibert, J.L. (1996). A Collaborative Model of Feedback in Human-Computer Interaction. Proceedings of CHI'96, Vancouver, Canada ACM Press. [On-line], Available: http://www.acm.org/sigchi/chi96/proceedings/papers/Perez/map1txt.htm

Rasmusson, J. (1983). Skills, Rules, and Knowledge: Signals, Signs, and Symbols, and Other Distinction in Human Performance Models. IEEE Transactionson System, Man & Cybernetics, 13, 257-266.

Schuck, M.M. (1994). The use of auditory feedback in the design of touch-input devices. Applied Ergonomics, 25 (1), 59-62.

Sellen, A., Kurtenbach, G. & Buxton, W. (1992). The prevention of mode errors through sensory feedback. Human Computer Interaction, 7(2), 141-164.

SensAble Technologies Inc. (1997). "The PHANToM System Overview." [On-line], Available: http://www.sensable.com/products.htm#Overview/.

Shannon, C.E., & Weaver, W. (1949). The mathematical theory of communication. Urbana, IL: University of Illinois Press.

Shimoga, K. B. (1993). A Survey of Perceptual Feedback Issues in Dexterous Telemanipulation: Part II. Finger Touch Feedback. Proceedings of the IEEE Virtual Reality Annual International Symposium, Piscataway, NJ : IEEE Service Center.

Sutherland, I.E. (1963). SketchPad: A Man-Machine Graphical Communication System. AFIPS Spring Joint Computer Conference, 23, 419-490.

Trankle, U. & Deutschmann, D. (1991). Factors Influencing Speed and Precision of Cursor Positioning using a Mouse. Ergonomics, 34(2), 161-174.

Wickens, C. (1992). Engineering Psychology and Human Performance. (2nd Ed.) New York: HarperCollins Publishers Inc.

Zhai, S. (1996). Human Performance in Six Degree of Freedom Input Control. Unpublished doctoral dissertation, University of Toronto, Toronto. [On-line], Available: http://vered.rose.toronto.edu/people/shumin_dir/papers/PhD_Thesis/top_page.html

Appendix A

Experimental Forms

Application for Ethics Approval

OHR 101 **OHR File No: #8363**

Revised November 1995

UNIVERSITY OF WATERLOO
OFFICE OF HUMAN RESEARCH AND ANIMAL CARE

Application for Ethics Review Of Research with Human Participants

A. GENERAL INFORMATION

1. Title of Project:
(This title should be identical to that of any corresponding grant.)
Evaluating kinesthetic and non-speech audio feedback on computer task performance

2. Faculty Investigator(s) **Department** **Ext. No.:** **e-mail:**
Carolyn G. MacGregor, Ph.D. Systems Design Engineering x2897 cgmacgre@sysoffice.watstar.uwaterloo.ca

3. Faculty Supervisor(s) **Department** **Ext.:** **e-mail:**
Carolyn G. MacGregor, Ph.D. Systems Design Engineering x2897 cgmacgre@sysoffice.watstar.uwaterloo.ca

4. Student Investigator(s) **Department** **Ext. No:** **e-mail:**
Alice Thomas Systems Design Engineering x3018 (Office) x5607 (lab) a2thomas@engmail.uwaterloo.ca

5. Level of Project
Faculty Research _____
Thesis Research
Ph.D. _____ M.A. _____ M.Acc. _____ M.A.E.S. _____ M.A.Sc. __X__
M.Math. _____ M.Sc. _____ M.Phil. _____ Honours _____
Administration _____ Other (Specify) ____________

6. Funding Status
Is this project currently funded? Yes _____ No __X__
Details of funding:
Agency:
MRC _____
NSERC _____

SSHRC _____
Other (specify):
NHRDP _____
OMHF _____
Heart and Stroke _____
____
COEFT _____
CDA _____
Alzheimer's Society _____

Period of Funding: ____________________

If no, is funding being sought? Yes _____ No __X__
Period of Funding: ____________________
Agency:
MRC _____
NSERC _____
SSHRC _____
Other (specify):
NHRDP _____
OMHF _____
Heart and Stroke _____
____
COEFT _____
CDA _____
Alzheimer's Society _____

7. Has this application been submitted to any other Institutional Ethics Review Committee? Yes _____ No __X__
If yes, provide name of committee, date, and decision. Attach a copy of the approval.

____

8. For student thesis research, indicate if this proposal has been approved by a

Department Committee. Yes _____ Approval Date: __________ No __X__
Reason: Not submitted because it is not required by the department.

9. Expected Project Commencement Date: May 25, 1997

Expected Project Completion Date: August 28, 1998

B. SUMMARY OF PROPOSED RESEARCH

1. Purpose and/or Rationale for Proposed Research

(Describe the purpose and background rationale for the proposed project as well as the hypothesis(es)/research questions to be examined. Where available, include a copy of a research proposal.)

Purpose

The purpose of this study is to evaluate and compare the effects of kinesthetic feedback, and non-speech audio feedback on computer tasks. A modified mouse incorporating tactile and force feedback characteristics will be used to provide kinesthetic feedback. Earcons will be used to provide non-speech audio feedback. Earcons are non-verbal audio messages that are used in the computer/user interface to provide information to the user about some computer object, operation or interaction.

Feedback plays an important role in human-computer interaction because it provides information to the user about what has happened. Within the traditional framework of human-computer interaction, the visual modality has primarily provided information required for navigation and interaction. For example, cues for cursor position on screen, current mode, and what was previously entered is usually provided by the visual sense. In order to improve the interaction with the interface, many researchers are exploring different ways to use other sensory modalities. It has also been theorized that making use of kinesthetic and auditory modalities will free up the visual capacity for other purposes.

To answer these questions, human participants will be asked to participate in an study. The study consists of the participant performing a sequence of computer data entry tasks over a period of time on a Calculator software application. More specifically, the participant will be asked to enter a set of mathematical expressions into a mouse-driven Calculator software application. Objective measures such as speed, accuracy, types of errors committed, as well as subjective measures will be used to assess the effectiveness of kinesthetic and non-speech audio feedback compared to visual feedback alone. Secondly, participants will be asked to perform the primary Calculator data entry task as well as a secondary monitoring task. The secondary monitoring task will primarily occupy the visual modality. It is expected that participants will compensate for the lack of visual feedback by using the kinesthetic feedback and non-speech audio feedback characteristics. If this is the case, the results should show that the additional feedback mechanisms may be effectively used to help user's perform concurrent tasks. Details of the experimental design can be found in Appendix A (Experimental Design).

Rationale

This thesis is concerned with feedback mechanisms in human computer interaction for every day computer related tasks. The need for this research has emerged because of the development of various advanced multi-modal input technologies. These new technologies address the needs of various niche markets including:: virtual reality applications, telerobotics, computer accessibility for the blind, game playing, electronic musical instruments, and navigation equipment. Many of these technological innovations have been altered for use with home computers in order to capture a wider sales market. Yet, there has been very little research conducted to study the effectiveness, advantages, and disadvantages of using these input technologies for every day computer related tasks and for complex work environments where users perform multiple tasks simultaneously.

2. Methodology/Procedures

(Describe, sequentially and in detail, all procedures in which the research participants will be involved, e.g., paper and pencil tasks, interviews, surveys, questionnaires,

physical assessments, physiological tests, doses and methods of administration of drugs, time requirements, etc.)

This research program will consist of a study involving human participants. While the specifics of task feedback will vary between groups of subjects, the basic approach and methodology will remain uniform. (See Appendix A: Experimental Design)

The following are brief descriptions of the testing procedures:

Participants will be given a verbal overview of the research procedures (see Appendix B: Script For Researcher), as well as a Letter of Information (see Appendix C: Letter of Information) and two copies of the Consent Letter (see Appendix D: Consent Letter) to read and sign. The participant is asked to keep the Letter of Information and one copy of the Consent Letter.

The researcher will read a scripted set of instructions to the participant throughout the study (see Appendix A). This will ensure that the same information is conveyed to all participants.

The study is divided into three phases. During Phase I, participants will be asked to complete a "Participant Background Questionnaire" (see Appendix E: Participant Background Questionnaire) for demographic information and for computer experience, introducing the equipment to the participants, and the training. During the training, participants are asked to become comfortable with the mouse, computer task, and equipment (i.e., specially developed mouse and keyboard controls). It is expected to take twenty minutes.

During Phase II, the participant will perform eight trials. It is expected to take fifty minutes. Each trial consists of the participant entering ten mathematical expressions into the calculator application and finding the answer. Additionally, participants will be asked to perform a secondary monitoring task for the last four trials. The Calculator application will automatically record relevant objective data (i.e., speed and accuracy). Participants will be asked to rest for about two minutes between each trial. Participants will also be asked to perform the NASA Task Load Index. The NASA TLX is a multi-dimensional subjective rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort, and Frustration. Besides the six scales, an overall weighted measure of task load is also calculated on the basis of the scales.

During Phase III, the participant will complete the final participant questionnaire (Appendix F: Participant Final Questionnaire) and will be given the thank you letter (Appendix G: Participant Feedback Letter) and payment for participation. It is expected to take fifteen minutes.

In total, the study requires approximately an hour and half of the participant's time

3. Participants Involved in the Study

UW Students (General) _ ___
UW Students (Psychology Subject Pool) _____
UW Students (Kinesiology Subject Pool) _____
UW Faculty _____
UW Staff _____
Local community agencies _____
Local school boards _____

Children/adolescents _____
Institutionalized elderly _____
Adults from non-UW community _____
Other (specify) UW Undergraduate and Graduate Students

Forty-eight undergraduate and graduate students enrolled at the University of Waterloo will be recruited to participate in the study. The criteria for inclusion in the study are: self-reported work experience with any Windows 3.1 application or higher (e.g., Excel, Lotus 1-2-3), proficiency with a mouse input device, having 20/20 vision with or without corrective lenses, dominant right-handedness, and no degenerative motor function. The participants will comprise of both genders between the ages of eighteen and thirty. Effort will be made to recruit an equal number of male and female participants.

Participants will be recruited primarily from the Engineering and Math faculty to reduce individual differences in spatial ability, computer skills, and experience. As well, engineering and math students should have a comprehensive mathematical background that will ensure they can understand and efficiently enter the mathematical expressions into the calculator.

Participants will be asked to participate in the study for a total of an hour and half. Participants will be paid $10 for full participation at the end of the study.

4. Recruitment Process

(Describe how and from what source the participants will be recruited. Indicate where the study will take place. Attach a copy of any posters(s), advertisement(s) or letter(s) to be used for recruitment.)

Recruitment posters (see Appendix H: Recruitment Poster) will be posted primarily in the Engineering buildings (e.g., Engineering 1 and 2, and Carl Pollack Hall) and Math buildings.

5. Compensation of Participants

Will participants receive compensation for participation? Yes __X__ No _____
Financial __X__
In-kind _____
Other (specify) ____________________
If yes, please provide details:

6. Feedback to Participants

(Whenever possible, and upon completion of their part in the study, participants should receive a feedback letter expressing appreciation for their involvement and providing general information about the objectives/goals of the study. Upon final completion of the study, an executive summary should be offered. Describe below the arrangements for provision of these two types of feedback, and attach a copy of a generic feedback letter to be used in the first instance.)

Participant will be given a copy of the Feedback Letter (See Appendix G: Feedback Letter).

Additionally, participants will be asked whether they wish to receive a summary of the results. If so, the participant will be asked to leave his/her email address at the end of the study. The summary of the results will be forwarded to those participants who were interested in receiving the summary.

C. POTENTIAL BENEFITS FROM THE STUDY

Discuss any potential direct benefits to participants from their involvement in the project. Comment on the (potential) benefits to the scientific community/society that would justify involvement of participants in this study.

The potential benefits for the individual participants are:

1. exposure and use of kinesthetic feedback devices
2. exposure and use of non-speech auditory feedback in interfaces
3. exposure to usability testing methodologies commonly used in research and industry

The potential benefits to the scientific community lie in the contribution that can be made at both theoretical and applied levels in the area of human machine interfaces. The interaction between human and machine has been limited to a great extent because other feedback mechanisms have not been explored or used to generate new possibilities of interaction.

The potential benefits for society can be tremendous. There are many contexts in every day life where multi-modal technologies can benefit users. For example, a trader on a trading floor will usually work with two or three monitors on his/her desk, talk on the phone with clients, watch the monitor overhead that display's ticker symbols and prices, as well as communicate with other traders and trading assistants on the floor simultaneously. Another example of concurrent task performance is when a mother or father is working on the computer while watching the baby play nearby. These environments require users to perform concurrent tasks that are not easily supported without significant performance degradation because the visual modality is the primary modality in use. Utilizing other feedback mechanisms to perform ordinary computer tasks may reduce the demand on the visual modality.

D. POTENTIAL RISKS FROM THE STUDY

1. Discuss the known and anticipated risks of the proposed research, specifying the particular risk(s) associated with each procedure or test. Consider both physical and psychological/emotional risks.

We anticipate that there will be no risks to the participants other than that normally associated with a computer-based task.

2. Describe the procedures or safeguards in place to protect the physical and psychological health of the participants.

Participants will be able to contact the researcher at all times during the study. Drinking water will be provided upon request. The lab is equipped with a sink and running water, and a phone in the event of an emergency. The principal researcher has been trained in First Aid and CPR.

E. INFORMATION AND CONSENT PROCESS

Attach a copy of a Letter of Information describing the procedures and a separate Consent Form. If written consent will not/cannot be obtained or is considered inadvisable, justify this *and outline the process to be used to otherwise fully inform participants.*

See Appendix C for the Letter of Information. See Appendix D for the Consent form.

In the case of minors, describe the process to be used to obtain permission of parent or guardian. Attach a copy of an information-permission letter to be used.

No minors will be used for this study.

F. CONFIDENTIALITY

All participants will be assigned a numeric code that will be used to identify their particular data. Names and identifying characteristics will not be included in any reports, journal articles, or presentations. Video cameras used for data collection will be positioned so that the possibility of identifying a particular participant is minimized.

All collected data (written records, video/audio tapes, and questionnaires) will be secured in a locked cabinet. to these records will be restricted to the student researcher, USE-IT lab members, and members of her masters committee.

All data will be kept for a minimum of one year. When the time comes to dispose the data, written records and questionnaires will be shredded, and video tapes will be erased.

G. DECEPTION

Yes _____ No __X__
(If yes, please describe and justify the need for deception. Explain the debriefing procedures to be used and attach a copy of the written debriefing and post-debriefing consent form.)
Agreement
I have read the Office of Human Research Guidelines for Research with Human Participants and agree to comply with the conditions outlined in the Guidelines. In the case of student research, as Faculty Supervisor, my signature indicates that I have read and approved the application and proposal, deem the project to be valid and worthwhile, and agree to provide necessary supervision of the student.

Date:

____________________ ____________________________
Signature of Faculty Investigator/Supervisor

Date:

____________________ ____________________________
Signature of Student Investigator

FOR OFFICE OF HUMAN RESEARCH AND ANIMAL CARE USE ONLY:
Date:

____________________ ____________________________
S.E. Sykes, Ph.D., C. Psych.
Director
Human Research and Animal Care

Script for Researcher

Introduction

Hello. We are conducting an experiment designed to determine if incorporating kinesthetic feedback and non-speech audio feedback in user interfaces will help users to perform computer tasks faster, more easily and efficiently. I will be reading the following instructions to each participant so that everyone has the same information.

In this experiment, you will be asked to enter a series of mathematical expressions using a "virtual" scientific calculator and specially designed mouse. For some of the trials, you will be asked to monitor a second screen for specific stimuli.

The purpose of the experiment is to test how various types of kinesthetic and non-speech audio feedback affect performance. This experiment is not a test of your abilities to enter the mathematical expressions.

This experiment is broken down into three phases. The first phase is the training phase. The second phase involves participation in a set of trails on the software application. In the third phase you will give us subjective feedback about the tasks and equipment.

Each phase will be explained in detail before you begin and you will be assigned a numeric code. Any data collected during the session will only be identified with this numeric code to ensure that your identity is kept confidential. The only people who have access to the data will be researchers that are directly involved with this project. All data are kept for a minimum of one year. After we are finished with the data, all written records will be destroyed and all stored data erased. You are free to stop the trials and withdraw from this study at any time.

Participants who complete the entire experiment will be paid $10 at the end of the third session. This experiment will take no more than an hour and half to complete. Do you have any questions at this point?

GIVE PARTICIPANTS 2 COPIES OF INFORMATION LETTER AND CONSENT FORM AND WAIT FOR THEM TO READ IT OVER, AND SIGN IT.

Now that I have given you a basic overview of the experiment, I would like you to read over the Information Letter and sign 2 copies of the Consent Form if you are willing to participate. Please keep one copy of the Information Letter and Consent Form for yourself.

PHASE 1: TRAINING

In this first phase, all participants will be asked to fill out a "Background Questionnaire Form", followed by a training phase. The background questionnaire is simply a set of questions designed to give us information on the demographics of the participants involved in the study and their computer experience. You will be identified with a numerical code on the questionnaire. After completing the background questionnaire you will be asked to spend some time familiarizing yourself with the equipment you will use.

HAND BACKGROUND QUESTIONNAIRE TO PARTICIPANT AND WAIT FOR THEM TO COMPLETE IT.

Thank you.

As you see, you will be working with a standard color monitor, keyboard, and mouse (POINT TO ITEMS). All interactions with the computer calculator (POINT TO CALCULATOR) will be performed using the mouse only. You will be required to use the mouse for navigation and data entry. The mouse you will be using is a newly developed mouse. It will interact somewhat differently from what you are used to. Please get acquainted with it in your training session until you feel comfortable.

IDENTIFY MOUSE TO PARTICIPANT AND ASK IF THEY ARE LEFT OR RIGHT HANDED AND ADJUST THE POSITION OF THE MOUSE ACCORDINGLY.

Please take a few minutes to get acquainted with the set up. For the training task, we would like you to enter the mathematical expressions listed on the scratch pad on the calculator and ensure that the answers are correct. GIVE ANSWER SHEET ALONG TO PARTICIPANT FOR TRAINING TASK.

This will take approximately 5 minutes. I will be here if you have any trouble or have questions. When you are ready and feel comfortable with the application and devices, we will move to the second part of the training phase.

PHASE 2: TRIALS

Now we will begin Phase 2. This phase involves eight trials. The first four trials are the same task as in your training phase. Please remember that both accuracy and speed are important aspects of the experiment.

For each trial, you will be asked to enter 10 numerical expressions into the calculator. The ten expressions will be displayed one by one on a scratchpad above the Calculator application. [POINT TO SCRACH PAD] When the "Save" button is pressed on the calculator, the next expression will be displayed.

RESEARCHER LEAVE TEST AREA

Phase ?. AFTER EACH TRIAL FROM 1-3

Thanks, I will just reset the computer and we will be ready to begin the next trial.

Phase 2, AFTER TRIAL 4

We are not only interested in assessing your performance but also the experiences you had during the task. Right now we are going to describe the technique that will be used to examine your experiences. In the most general sense we are examining the "Workload" you experienced. Workload is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt, The workload contributed by different task elements may change as you get more familiar with a task, perform easier or harder versions of it, or move from one task to another. Physical components of workload are relatively easy to conceptualize and evaluate. However, the mental components of workload may be more difficult to measure.

Since workload is something that is experienced individually by each person, there are no effective "rulers" that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. This set of six rating scales was developed for you to use in evaluating your experiences during different tasks. Please read the descriptions of the scales carefully. If you have a question about any of the scales in the table, please ask me about it. It is extremely important that they be clear to you.

Six rating scales will be displayed on the screen. You will evaluate the task by marking each scale at the point which matches your experience. Each line has two endpoint descriptors that describe the scale. Note that "own performance" goes from "good" on the left to "bad" on the right. This order has been confusing for some people.

Move the arrow with the right and left arrow keys until it points at the desired location. Stop it by pressing the up arrow key. Press the down arrow key to enter your selection. Please consider your responses carefully in distinguishing among the task conditions. Consider each scale individually. Your ratings will play an important role in the evaluation being conducted, thus, your active participation is essential to the success of this experiment, and is greatly appreciated.

START NASA TLX AND LEAVE ROOM

ALLOW THEM TO TAKE 1 MINUTE BREAK

PHASE II. DUAL TASK

We have four more trials left. In addition to the calculator task that you have performed, we want you to perform a monitoring task at the same time. This new task requires that you monitor screen B beside you and note what number is displayed. The number displayed changes at random intervals.

Phase 2, AFTER EACH TRIAL FROM 5-7

TURN COPMUTER SCREEN ON AND SHOW AN EXAMPLE RUN FOR 10 SECONDS

The numbers will be displayed are from 1 to 6 in any order. You must be attentive to the changes. The monitoring task is as important as the calculator task. When you see a new number, please tell me what number you see. I will be sitting beside you. If you miss a number, I will remind you to keep your eyes on the second screen. Please try not to miss a number.

Are these instructions clear to you? Do you have any questions about what you are required to do?

That is the end of Phase 2. Now, I want you to take some time and complete the following short questionnaire.

AFTER TRIAL 8

Now, you will be asked to perform the NASA TLX again.

START NASA TLX AND LEAVE ROOM

TAKE 1 MINUTE BREAK

GIVE FINAL PARTICIPANT QUESTIONNAIRE

Thank you for participating in the study. Before leaving, do you have any questions about this experiment?

GIVE FEEDBACK LETTER AND $10

GET PARTICIPANT TO SIGN RECEIPT BOOK.

Letter of Information

Evaluating kinesthetic and non-speech audio feedback on computer task performance

May 1998 **PARTICIPANT ID: ____**

Letter of Information

Student Researcher: **Alice Thomas**
Systems Design Engineering
University of Waterloo
885-1211 x3018
a2thomas@engmail.uwaterloo.ca

Supervisor: **Prof. Carolyn MacGregor**
Systems Design Engineering
University of Waterloo
885-1211 x2897
cgmacgre@engmail.uwaterloo.ca

Title of Research:
Evaluating kinesthetic and non-speech audio feedback on computer task performance

Study Objectives:
The purpose of the study is to test how various types of kinesthetic and non-speech audio feedback affect performance in computer related tasks.

Tasks:
I will be reading scripted set of instructions to each participant so that everyone has the same information. In this study, you will be asked to enter a series of mathematical expressions using a "virtual" scientific calculator and specially designed mouse. For some of the trials, you will be asked to monitor a second screen for specific stimuli. This study is not a test of your abilities to enter the mathematical expressions.

This study is broken down into three phases. The first phase is the training phase. The second phase involves participation in a set of trails on the software application. In the third phase you will give us subjective feedback about the tasks and equipment. Each phase will be explained in detail before you begin.

Risks:
We anticipate that there will be no risks to the participants other than that normally associated with a computer-based task.

Time Commitment and Payment:

The study will take approximately 1.5 hours. All participants are expected to complete all parts of the study. Total payment for participation will be $10 that will be paid at the end of the study.

Right to Withdrawal:
Participants can decide to withdraw from this study at any time, at which point all data for that participant will be destroyed. No payment will be made if participant does not complete study.

Confidentiality:
At the beginning of each phase, you will be assigned a numeric ID number. Any data collected during the phase will only be identified with this numeric code to ensure that your identity is kept confidential. The only people who have access to the data will be researchers that are directly involved with this project.

All collected data (written records, and questionnaires) will be secured in a locked cabinet. Access to these records will be restricted to the student researcher, USE-IT lab members, and members of her masters committee.

All data will be kept for a minimum of one year. When the time comes to dispose the data, written records and questionnaires will be shredded, and video/audio tapes will be erased.

Ethics Review:
This project has been reviewed and received ethics clearance through the Office of Human Research and Animal Care (OHRAC) and participants who have concerns or questions about their involvement in the project (OHR #8363) may contact the Director, OHRAC. The telephone number for the OHRAC is 519-888-4567, Ext. 6005.

Consent Form

Evaluating kinesthetic and non-speech audio feedback on computer task performance

April 1998 **PARTICIPANT ID: ____**

Consent Form

I agree to participate in the study entitled "Evaluating kinesthetic and non-speech audio feedback on computer task performance". I have read over the Information Letter and have had the opportunity to receive additional details about my participation in this study.

I understand that full participation in this study requires approximately 1.5 hours of my time and that compensation for full participation is $10 which will be paid at the end of the sessions.

I understand that I will be asked to complete a background questionnaire, perform a series of computer tasks, and subjective questionnaires. I understand that the results of the experiment are kept confidential and that I will not be individually identified in any reports or presentations pertaining to this research.

I understand that I will not be given any feedback about my individual performance on any of the tasks I perform during this study.

I understand that I have the right to withdraw my consent to participate in this experiment at any time, and upon doing so any data collected relating to myself or my performance will be immediately destroyed.

Participant's Name: **Participant's Signature:**

Name of Witness: **Signature of Witness:**

Date:

Participant Background Questionnaire

Evaluating kinesthetic and non-speech audio feedback on computer related task performance

April 1998 (n:\msc\parback1.doc) **PARTICIPANT ID: ____**

Participant Background Questionnaire

This information will be used for the purposes of creating general descriptions of the participants involved in this study.

1. Gender: ____Male ___Female

2. Age: ____ Years

3. Which faculty are you enrolled in? ______________________

4. Do you have any hearing impairments?
   ___ NO ___YES
   If yes, then please explain.
   ____
   ____
   ____

5. Do you have problems with your vision while working on a computer?
   ___NO ___YES
   If yes, then please explain.
   ____
   ____
   ____

6. Have you taken a math course in the last 2 years? ___NO ___YES

7. Hand preference (writing hand): ___ RIGHT ___LEFT

8. Do you currently suffer from any dexterity problems with your hands, wrist, or arms?

___NO ___YES

If yes, then please explain.

____

____

____

9. Over the past year, how many hours per week did you spend using a computer?
(Please circle the most appropriate number)

0 1-2 hrs 3-4 hrs 5-6 hrs 7-8hrs 9+hours
|-------------|-------------|-------------|-------------|-------------|

10. Over the past year, how many hours per week did you spend listening to music?
(Please circle the most appropriate number)

0 1-2 hrs 3-4 hrs 5-6 hrs 7-8hrs 9+hours
|-------------|-------------|-------------|-------------|-------------|

11. Over the past year, how many hours per week did you spend playing a musical instrument?
(Please circle the most appropriate number)

0 1-2 hrs 3-4 hrs 5-6 hrs 7-8hrs 9+hours
|-------------|-------------|-------------|-------------|-------------|

12. Reflecting on your use of Windows applications on a daily basis, how much of your interaction with the computer was spent using a mouse as opposed to the keyboard for navigation?
(Please circle the most appropriate number)

0% 20% 40% 60% 80% 100%
|-------------|-------------|-------------|-------------|-------------|

13. What is your preferred method of interaction with computer applications (e.g., light pens, mouse, keyboard, joysticks, voice, etc.)? Please list:

14. Over the past year, how many times have you used the Windows calculator application?
(Please circle the most appropriate number)

0	1-2	3-4	5-6	7-8	9 or more

15. Have you ever used a device that provides force feedback? YES _____ NO _____
If Yes, please explain and list which device:

____

____

____

16. Have you ever used software applications that use sound as part of the interface (other than games)?
YES _____ NO _____

17. If Yes, about how frequently do you use sound in your everyday computer applications?
(Please circle the most appropriate number)

0%	20%	40%	60%	80%	100%

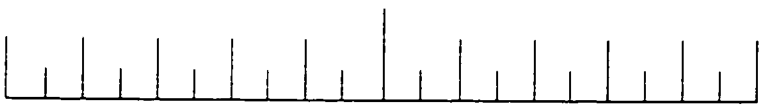
NASA TLX

The NASA Task Load Index is a multi-dimensional subjective rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort, and Frustration. Three dimensions relate to the demands imposed on the subject (Mental, Physical, and Temporal Demands) and three to the interaction of a subject with the task (Effort, Frustration and Performance). Besides the six scales, an overall weighted measure of task load is also calculated on the basis of the scales.

The NASA-TLX is a two-part evaluation procedure consisting of both weightings and ratings. The first part of the procedure involves obtaining numerical rating for each scale that reflects the magnitude of that workload factor in a given task. The second part of the procedure requires from each individual to evaluate the contribution of each workload factor to the total workload of a specific task. The weighting score is an indication of the importance of each factor, relative to the other factors. The weighting accounts for two potential sources of differences between ratters: differences in workload definition within a task, and differences in the sources of workload between tasks.

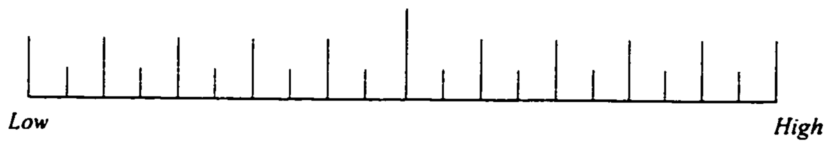
To gauge the difficulty of the tasks, the NASA Task Load Index will be used to measure workload. Workload is a "multi-dimensional psychological construct measuring the subjective experience of work that results from the mental actions performed while perceiving and processing information and executing a response"(Hart and Staveland, 1988). The NASA TLX measures workload on 6 different dimensions including: mental demand, physical demand, temporal demand, performance, effort, and frustration. The dimensions are used to create a picture of the amount and type of mental workload a user experiences during task performance.

<u>Mental Demand</u>: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

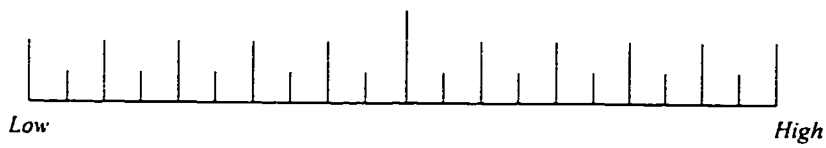


Low *High*

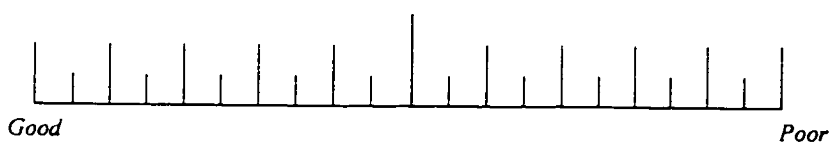
Physical Demand: How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?



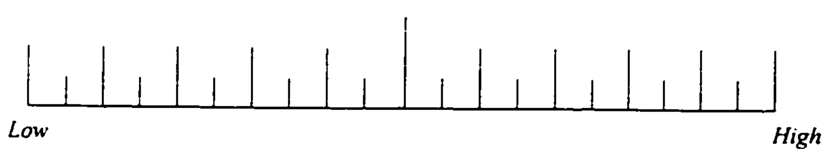
Temporal Demand: How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?



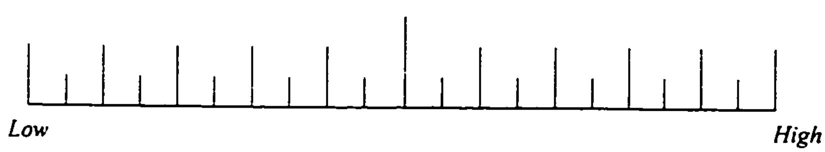
Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?



Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?



Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?



Evaluate the contribution of each workload measure to the total workload of the task you performed (i.e., which one was more important contributer to workload for you):

Mental Demand ☐	☐ Physical Demand
Physical Demand ☐	☐ Temporal Demand
Temporal Demand ☐	☐ Performance
Performance ☐	☐ Effort
Effort ☐	☐ Frustration
Frustration ☐	☐ Mental Demand
Mental Demand ☐	☐ Temporal Demand
Physical Demand ☐	☐ Performance
Temporal Demand ☐	☐ Effort
Performance ☐	☐ Frustration
Effort ☐	☐ Mental Demand
Frustration ☐	☐ Physical Demand
Mental Demand ☐	☐ Performance
Physical Demand ☐	☐ Effort
Temporal Demand ☐	☐ Frustration
Performance ☐	☐ Mental Demand
Effort ☐	☐ Temporal Demand
Frustration ☐	☐ Physical Demand
Mental Demand ☐	☐ Effort
Physical Demand ☐	☐ Frustration
Temporal Demand ☐	☐ Mental Demand
Performance ☐	☐ Physical Demand
Effort ☐	☐ Temporal Demand
Frustration ☐	☐ Performance
Mental Demand ☐	☐ Frustration
Physical Demand ☐	☐ Mental Demand
Temporal Demand ☐	☐ Physical Demand

Performance ☐	☐ Temporal Demand
Effort ☐	☐ Performance
Frustration ☐	☐ Effort

Final Questionnaire

Evaluating kinesthetic and non-speech audio feedback on computer related task performance
Final Participant Questionnaire

n:\msc\Finque.doc Participant ID:________

Thank you for completing the testing process of this experiment. Could you please take some time to complete this questionnaire as clearly as possible. Please circle the item which closest represents your answer.

1. In terms of usability, how would you rate the feedback mouse compared to a regular mouse for the tasks you performed.

Not Applicable ☐

1 2 3 4 5 6

|------------------|------------------|------------------|------------------|------------------|

Terrible Better

2. How easy did you find it to move around the calculator application using the feedback from the mouse?

1 2 3 4 5 6

|------------------|------------------|------------------|------------------|------------------|

Difficult Easy

3. How easily were you able to detect buttons on the calculator because of the feedback from the mouse?

☐ Not Applicable

1 2 3 4 5 6

|------------------|------------------|------------------|------------------|------------------|

Difficult Easy

4. How easily were you able to detect mistakes you made in the calculator task because of the feedback from the mouse?

Not Applicable ☐

1 2 3 4 5 6

|------------------|------------------|------------------|------------------|------------------|

Difficult Easy

5. Did you have any trouble hearing the sounds from the computer?

YES _____ NO _____

6. How easily were you able to detect the buttons on the calculator because of the feedback from the sounds?

1 2 3 4 5 6

|------------------|------------------|------------------|------------------|------------------|

Difficult Easy

7. How easily were you able to detect mistakes you made in the calculator task because of the feedback from the sounds?

1 2 3 4 5 6

|------------------|------------------|------------------|------------------|------------------|

Difficult Easy

8. In general, how would you rate the effectiveness of the feedback from the mouse versus the feedback from the sounds?

Not Applicable ☐

1	2	3	4	5	6
Terrible					Better

9. If you had to choose one feedback mechanism over another, which one would you prefer?

(Check one)

Not Applicable ☐

Feedback from mouse_______ Auditory feedback______ Neither ______

10. If you could change aspects of the feedback from the mouse, what aspects would you change?

Not Applicable ☐

(Please check if you agree)

Magnitude of feedback should be higher:____

Magnitude of feedback should be lower:____

Needed more types of feedback:____

Other:

11. If you could change aspects of the auditory feedback, what aspects would you change?

(Please check if you agree)

Not Applicable ☐

Sounds should be more meaningful:____

Sounds should be less used:____

Sounds should be shorter in duration:____

Sounds should be longer in duration:____

Other:

12. Do you have any general comments about the tasks you performed?

13. Did you experience any physical discomfort during the trials?

YES _____ NO _____

14. If yes, please specify the nature of the discomfort:

____

____

15. If yes, please rate the level of discomfort:

1 2 3 4 5 6

|------------------|------------------|------------------|------------------|------------------|

Painful Minor

Feedback Letter

Evaluating kinesthetic and non-speech audio feedback on computer related task performance

Thank you letter (c:\msc\thank.doc)
Dec. 1997

Dear Participant,

We would like to take the time to thank you for your participation in this study. As a reminder, the focus of this study was on how people use mice to help them perform computer tasks, and not on how well you personally performed on any task.

The data we have collected from the data entry tasks and recall task, as well as the final spatial ability test, will contribute to a better understanding of how people use a mouse and in particular how people use kinesthetic feedback. More specifically, the research findings of this study should help us to determine whether having kinesthetic feedback will help in data entry tasks in a complex work environment.

Please remember that any data pertaining to yourself as an individual participant will be kept confidential. Once all the data is collected and analyzed for this experiment, we plan on sharing this information with the research community through department seminars, conferences, presentations, and journal articles. If you are interested in receiving more information regarding the results of this study, or if you have any questions or concerns about this study, please contact the researcher at either the phone number of email address listed at the bottom of the page.

Thank you for your participation.

Alice Thomas
885-1211 x.3018 or x.5607
Systems Design Engineering
University of Waterloo
a2thomas@novice.uwaterloo.ca

Recruitment Poster

<u>ATTENTION STUDENTS</u>
NEED PARTICIPANTS FOR STUDY ON

Evaluating kinesthetic and non speech audio feedback on computer task performance

We are conducting an experiment designed to determine if incorporating kinesthetic feedback and audio feedback in user interfaces will help with everyday computers tasks in a complex work environment. Previous research in this area shows that people mainly use the visual to perform computer tasks. One potential benefit in adding force and tactile feedback is that the processing demands of the visual channel are lessened, freeing up capacity for other purposes. This experiment will help determine if kinesthetic feedback will improve the usability of software applications. Similarly, the use of non-speech audio feedback in computer interfaces is being explored. Sound is used frequently and in all aspects of our every day lives. Yet sound not been used extensively in interfaces and software applications. By using sounds we may be able to improve human computer interaction.

Who can participate:	**UW Students**
Time commitment:	**1.5 hours**
What does it involve:	**Using VIRTUAL REALITY mouse to perform computer tasks in a simulated complex work environment.**
Payment:	**$10 (for full participation)**

For more information, please contact:
Alice Thomas
Systems Design Engineering,
University of Waterloo
a2thomas@engmail.uwaterloo.ca
<u>885-1211 x. 3018 or x. 5607</u>

This project has been reviewed and received ethics clearance through the Office of Human Research and Animal Care (OHRAC) and participants who have concerns or questions about their involvement in the project (OHR# 8363) may contact the Director, OHRAC. The telephone number for the OHRAC is 519-888-4567, Ext. 6005.

Appendix B

Statistical Results

1.1 Background Questionnaire

Table B-1. Statistical results from the background questionnaire on computer experience.

Condition Groups (grp)		Average hrs/week spent using a computer (b6cx)	Average hrs/week spent listening to music (b7lm)	Average hrs/week spent playing a musical instrument (b8pm)	Mouse use versus keyboard on a daily basis (b9pm)	Number of times a Windows calculator was used (b11wc)
Control	Mean	4.8	4.1	1 7	3.9	3.1
	Std. Deviation	1 23	1 52	1 49	1.29	1.73
Kin	Mean	5 1	4 7	1 8	3 7	3.9
	Std. Deviation	1 52	1.77	1 03	95	1 97
Kin+Force	Mean	5 3	3 8	1 6	3.8	4.3
	Std. Deviation	1 06	1.23	1.17	1 23	1 70

Table Caption: Scores on 6-point interval rating scales (low to high specifications)

Table B-2. Frequency results for question on hearing impairments categorized by mouse-feedback group.

Crosstab

			Condition Groups (grp)			Total
			Control	Kin	Kin+Force	
Hearing Impairments (b15hi)	No	Count	9	10	10	29
		Expected Count	9 7	9 7	9 7	29 0
	Yes	Count	1	0	0	1
		Expected Count	3	3	3	1 0
Total		Count	10	10	10	30
		Expected Count	10 0	10 0	10 0	30 0

Table B-3. Frequency results for question on visual impairments or problems experienced regularly while working on a computer categorized by mouse-feedback group.

Crosstab

			Condition Groups (grp)			Total
			Control	Kin	Kin+Force	
Visual problems (b16vi)	No	Count	10	9	9	28
		Expected Count	9.3	9.3	9.3	28.0
	Yes	Count	0	1	1	2
		Expected Count	7	.7	.7	2.0
Total		Count	10	10	10	30
		Expected Count	10.0	10.0	10.0	30.0

Table B-4. ANOVA of background statistics by groups. As seen, there are no significant differences in the mouse-feedback groups. See Table B-1 for codes.

ANOVA Table

	df	F	Sig.
(b6cx) * Condition Groups (grp)	2 29	.383	.685
(b7lm) * Condition Groups (grp)	2 29	.906	.416
(b8pm) * Condition Groups (grp)	2 29	.064	.938
(b9pm) * Condition Groups (grp)	2 29	.074	.929
(b11wc) * Condition Groups (grp)	2 29	1.147	.333

1.2 Objective Measures

Analysis of Time

Table B-5. Repeated measure results of between-subject factors for mouse-feedback group (grp).

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power [a]
Intercept	241718.8	1	241718.8	1635.140	.000	1635.140	1.000
GRP	296.205	2	148.103	1.002	.381	2.004	.205
Error	3843.519	26	147.828				

a. Computed using alpha = .05

Table B-6. Repeated measures analysis of variance for mean time.

Measure: MEASURE_1
Sphericity Assumed

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
TASK	6045.958	1	6045.958	146.833	.000	146.833	1.000
TASK * GRP	83.223	2	41.612	1.011	.378	2.021	.207
Error(TASK)	1070.571	26	41.176				
SOUND	49.026	1	49.026	.995	.328	.995	.161
SOUND * GRP	39.874	2	19.937	.404	.671	.809	.108
Error(SOUND)	1281.558	26	49.291				
TASK * SOUND	.171	1	.171	.032	.860	.032	.053
TASK * SOUND * GRP	42.205	2	21.102	3.911	.033	7.821	.653
Error(TASK*SOUND)	140.299	26	5.396				

a. Computed using alpha = .05

Table B-7. Statistical results for between-subject pairwise comparisons of mouse-feedback groups for mean time.

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition Groups (grp)	(J) Condition Groups (grp)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Control	Kin	-3.0824	2.793	.280	-8.824	2.659
	Kin+Force	.4832	2.793	.864	-5.258	6.225
Kin	Kin+Force	3.5656	2.719	.201	-2.023	9.154
	Control	3.0824	2.793	.280	-2.659	8.824
Kin+Force	Kin	-3.5656	2.719	.201	-9.154	2.023
	Control	-.4832	2.793	.864	-6.225	5.258

Based on estimated marginal means

T-tests comparisons for mean time

Table B-8. Summary of results for Control and Kin mouse-feedback groups with respect to time for each block.

Group Statistics

	Condition Groups	N	Mean	Std. Deviation	Std. Error Mean
Single Task, Sound OFF	Control	9	36.5752	4.0122	1.3374
	Kin	10	38.9040	7.6870	2.4308
Single Task, Sound ON	Control	9	37.7365	3.5506	1.1835
	Kin	10	41.0417	5.2065	1.6465
Dual Task, Sound OFF	Control	10	53.8279	11.3759	3.5974
	Kin	10	54.9265	9.1820	2.9036
Dual Task, Sound ON	Control	10	50.3368	9.8656	3.1198
	Kin	10	56.8110	9.3855	2.9679

Table B-9. Results of t-test used to compare Control and Kin mouse-feedback groups with respect to time for each block.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Single Task, Sound OFF	Equal variances assumed	3.712	071	-.813	17	427	-2.3289	2.8642
	Equal variances not assumed			-.839	13.846	416	-2.3289	2.7745
Single Task, Sound ON	Equal variances assumed	1.028	325	-1.597	17	129	-3.3052	2.0693
	Equal variances not assumed			-1.630	15.922	123	-3.3052	2.0277
Dual Task, Sound OFF	Equal variances assumed	160	694	-.238	18	815	-1.0985	4.6230
	Equal variances not assumed			-.238	17.233	815	-1.0985	4.6230
Dual Task, Sound ON	Equal variances assumed	090	767	-1.504	18	.150	-6.4742	4.3060
	Equal variances not assumed			-1.504	17.955	.150	-6.4742	4.3060

Table B-10. Summary of results for Control and Kin-Force mouse-feedback groups with respect to time for each block.

Group Statistics

	Condition Groups	N	Mean	Std. Deviation	Std. Error Mean
Single Task, Sound OFF	Control	9	36.5752	4.0122	1.3374
	Kin+Force	10	37.8820	4.2915	1.3571
Single Task, Sound ON	Control	9	37.7365	3.5506	1.1835
	Kin+Force	10	38.7193	6.4722	2.0467
Dual Task, Sound OFF	Control	10	53.8279	11.3759	3.5974
	Kin+Force	10	48.5535	5.0222	1.5882
Dual Task, Sound ON	Control	10	50.3368	9.8656	3.1198
	Kin+Force	10	52.2660	11.1250	3.5180

Table B-11. Results of t-tests used to compare Control and Kin-Force mouse-feedback groups with respect to time for each block.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Single Task, Sound OFF	Equal variances assumed	052	822	-.683	17	504	-1.3068	1.9125
	Equal variances not assumed			-.686	16.967	502	-1.3068	1.9054
Single Task, Sound ON	Equal variances assumed	3.479	079	-.403	17	.692	-.9827	2.4360
	Equal variances not assumed			-.416	14.235	684	-.9827	2.3643
Dual Task, Sound OFF	Equal variances assumed	3.046	098	1.341	18	.197	5.2745	3.9324
	Equal variances not assumed			1.341	12.380	.204	5.2745	3.9324
Dual Task, Sound ON	Equal variances assumed	.550	468	-.410	18	.686	-1.9291	4.7021
	Equal variances not assumed			-.410	17.746	687	-1.9291	4.7021

Table B-12. Summary of results for Kin and Kin+Force mouse-feedback groups with respect to time for each block.

Group Statistics

	Condition Groups	N	Mean	Std. Deviation	Std. Error Mean
Single Task, Sound OFF	Kin	10	38.9040	7.6870	2.4308
	Kin+Force	10	37.8820	4.2915	1.3571
Single Task, Sound ON	Kin	10	41.0417	5.2065	1.6465
	Kin+Force	10	38.7193	6.4722	2.0467
Dual Task, Sound OFF	Kin	10	54.9265	9.1820	2.9036
	Kin+Force	10	48.5535	5.0222	1.5882
Dual Task, Sound ON	Kin	10	56.8110	9.3855	2.9679
	Kin+Force	10	52.2660	11.1250	3.5180

Table B-13. Results of t-tests used to compare Kin and Kin+Force mouse-feedback groups with respect to time for each block.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Single Task, Sound OFF	Equal variances assumed	3.945	062	.367	18	.718	1.0220	2.7840
	Equal variances not assumed			367	14.114	719	1.0220	2.7840
Single Task, Sound ON	Equal variances assumed	701	413	884	18	388	2.3224	2.6267
	Equal variances not assumed			884	17 210	389	2.3224	2.6267
Dual Task, Sound OFF	Equal variances assumed	2.744	115	1.926	18	070	6.3730	3.3096
	Equal variances not assumed			1.926	13.943	075	6.3730	3.3096
Dual Task, Sound ON	Equal variances assumed	1.082	.312	.987	18	.337	4.5450	4.6027
	Equal variances not assumed			.987	17.504	.337	4.5450	4.6027

Error Data

Table B-14. Results for overall between subject comparisons by mouse-feedback groups (grp) with respect to errors.

Measure. MEASURE_1
Transformed Vanable Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig	Noncent. Parameter	Observed Power [a]
Intercept	313 179	1	313 179	46 018	000	46 018	1 000
GRP	10 481	2	5 241	770	473	1 540	167
Error	176 945	26	6 806				

[a] Computed using alpha = 05

Table B-15. Repeated measures analysis of variance for mean number of errors.

Measure: MEASURE_1
Sphericity Assumed

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
TASK	7.626	1	7.626	5.226	.031	5.226	.595
TASK * GRP	1.626	2	.813	.557	.579	1.115	.132
Error(TASK)	37.938	26	1.459				
SOUND	.122	1	.122	.062	.805	.062	.057
SOUND * GRP	.533	2	.266	.136	.874	.272	.069
Error(SOUND)	50.983	26	1.961				
TASK * SOUND	2.398E-03	1	2.398E-03	002	.965	.002	.050
TASK * SOUND * GRP	1.184	2	.592	.488	.619	.976	.121
Error(TASK*SOUND)	31.560	26	1.214				

a. Computed using alpha = .05

Table B-16. Results for overall between-subject pairwise comparisons of mouse-feedback groups for mean errors.

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition Groups (grp)	(J) Condition Groups (grp)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Control	Kin	-.1208	.599	.842	-1.353	1.111
	Kin+Force	.5604	.599	.358	-.672	1.792
Kin	Kin+Force	.6812	.583	.253	-.518	1.880
	Control	.1208	.599	.842	-1.111	1.353
Kin+Force	Kin	-.6812	.583	.253	-1.880	.518
	Control	-.5604	.599	.358	-1.792	.672

Based on estimated marginal means

T-tests for error data

Table B-17. Comparison of statistical data for Control and Kin mouse-feedback groups with respect to errors for each block.

Group Statistics

	Condition Groups	N	Mean	Std. Deviation	Std. Error Mean
Single Task, Sound OFF	Control	9	1.4815	1.9197	.6399
	Kin	10	1.7250	1.4551	.4601
Single Task, Sound ON	Control	9	1.8241	2.6482	.8827
	Kin	10	1.6750	1.4242	.4504
Dual Task, Sund OFF	Control	10	2.7500	2.8186	.8913
	Kin	10	2.2250	1.8084	.5719
Dual Task, Sound ON	Control	10	3.1750	4.6638	1.4748
	Kin	10	2.0250	1.2988	.4107

Table B-18. Results of t-tests for Control and Kin mouse-feedback groups with respect to errors for each block

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Single Task, Sound OFF	Equal variances assumed	002	965	-314	17	758	-2435	7764
	Equal variances not assumed			-309	14 877	762	-2435	7882
Single Task, Sound ON	Equal variances assumed	981	336	155	17	879	1491	9609
	Equal variances not assumed			150	11 985	883	1491	9910
Dual Task, Sund OFF	Equal variances assumed	108	746	496	18	626	5250	1.0590
	Equal variances not assumed			496	15.336	627	.5250	1.0590
Dual Task, Sound ON	Equal variances assumed	7.846	012	.751	18	.462	1.1500	1.5309
	Equal variances not assumed			.751	10.388	469	1.1500	1.5309

Table B-19. Comparison of statistical data for Control and Kin-Force mouse-feedback groups with respect to errors for each block.

Group Statistics

	Condition Groups	N	Mean	Std. Deviation	Std. Error Mean
Single Task, Sound OFF	Control	9	1.4815	1.9197	.6399
	Kin+Force	10	.8750	.9807	.3101
Single Task, Sound ON	Control	9	1.8241	2.6482	.8827
	Kin+Force	10	.7500	1.0274	.3249
Dual Task, Sund OFF	Control	10	2.7500	2.8186	.8913
	Kin+Force	10	1.4250	1.7403	.5503
Dual Task, Sound ON	Control	10	3.1750	4.6638	1.4748
	Kin+Force	10	1.8750	1.4203	.4492

Table B-20. Results of t-tests for Control and Kin-Force mouse-feedback groups with respect to errors for each block.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Single Task, Sound OFF	Equal variances assumed	976	337	881	17	390	.6065	6882
	Equal variances not assumed			853	11 629	411	6065	7111
Single Task, Sound ON	Equal variances assumed	1 959	180	1.190	17	250	1.0741	9026
	Equal variances not assumed			1.142	10.149	.280	1.0741	.9406
Dual Task, Sund OFF	Equal variances assumed	364	554	1.265	18	222	1.3250	1.0475
	Equal variances not assumed			1.265	14.991	.225	1.3250	1.0475
Dual Task, Sound ON	Equal variances assumed	6.791	.018	.843	18	.410	1.3000	1.5417
	Equal variances not assumed			.843	10.655	.418	1.3000	1.5417

Table B-21. Statistical data for Kin and Kin+Force mouse-feedback groups with respect to errors for each block.

Group Statistics

	Condition Groups	N	Mean	Std. Deviation	Std. Error Mean
Single Task, Sound OFF	(grp)	10	1.7250	1.4551	.4601
	Kin+Force	10	.8750	.9807	.3101
Single Task, Sound ON	Kin	10	1.6750	1.4242	.4504
	Kin+Force	10	.7500	1.0274	.3249
Dual Task, Sund OFF	Kin	10	2.2250	1.8084	.5719
	Kin+Force	10	1.4250	1.7403	.5503
Dual Task, Sound ON	Kin	10	2.0250	1.2988	.4107
	Kin+Force	10	1.8750	1.4203	.4492

Table B-22. Results for t-tests comparing Kin and Kin+Force mouse-feedback groups with respect to errors for each block.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Single Task, Sound OFF	Equal variances assumed	2.642	121	1.532	18	143	8500	.5549
	Equal variances not assumed			1.532	15.778	145	8500	.5549
Single Task, Sound ON	Equal variances assumed	361	556	1.666	18	113	9250	.5553
	Equal variances not assumed			1.666	16.371	115	9250	.5553
Dual Task, Sund OFF	Equal variances assumed	293	595	1.008	18	327	8000	7936
	Equal variances not assumed			1.008	17.974	327	.8000	7936
Dual Task, Sound ON	Equal variances assumed	.359	556	.246	18	808	.1500	.6086
	Equal variances not assumed			.246	17.858	.808	.1500	.6086

1.3 Effects of Ordering and Audio Feedback

Mean Time

Table B-23. Repeated measures ANOVA of mouse-feedback group and order on time.

Measure:
Transformed Variable:

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power [a]
Intercept	240314.5	1	240314.5	1849.412	000	1849.412	1.000
ORDER	47.998	1	47.998	369	549	369	090
GRP	380.461	2	190.230	1.464	252	2.928	280
ORDER * GRP	818.831	2	409.415	3.151	.062	6.302	547
Error	2988.643	23	129.941				

a. Computed using alpha = .05

Table B-24. Repeated measures ANOVA of mouse-feedback group and order on time for task and sound conditions.

Measure: MEASURE_1
Sphericity Assumed

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power [a]
TASK	5985.938	1	5985.938	148.850	000	148.850	1.000
TASK * ORDER	2.740E-02	1	2.740E-02	.001	979	.001	050
TASK * GRP	84.367	2	42.183	1.049	366	2.098	211
TASK * ORDER * GRP	145.067	2	72.534	1.804	187	3.607	337
Error(TASK)	924.937	23	40.215				
SOUND	15.085	1	15.085	1.608	217	1.608	229
SOUND * ORDER	1037.890	1	1037.890	110.635	000	110.635	1.000
SOUND * GRP	50.412	2	25.206	2.687	089	5.374	479
SOUND * ORDER * GRP	15.769	2	7.884	840	444	1.681	176
Error(SOUND)	215.769	23	9.381				
TASK * SOUND	161	1	161	029	867	.029	053
TASK * SOUND * ORDER	3.550	1	3.550	628	436	628	118
TASK * SOUND * GRP	43.060	2	21.530	3.806	037	7.613	633
TASK * SOUND * ORDER * GRP	6.509	2	3.254	575	570	1.151	.134
Error(TASK*SOUND)	130.092	23	5.656				

a. Computed using alpha = .05

Error Data

Table B-25. Repeated measures ANOVA for mouse-feedback group and order for errors.

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power [a]
Intercept	326.270	1	326.270	50.221	.000	50.221	1.000
ORDER	24.721	1	24.721	3.805	.063	3.805	.464
GRP	12.800	2	6.400	.985	.389	1.970	.200
ORDER * GRP	3.595	2	1.797	.277	.761	.553	.088
Error	149.424	23	6.497				

a. Computed using alpha = .05

Table B-26. Repeated measures ANOVA for mouse-feedback group and order on task and sound conditions for errors.

Tests of Within-Subjects Effects

Measure: MEASURE_1
Sphericity Assumed

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power [a]
TASK	6.870	1	6.870	5.111	.034	5.111	.582
TASK * ORDER	4.934E-02	1	4.934E-02	.037	.850	.037	.054
TASK * GRP	1.976	2	.988	.735	.490	1.470	.159
TASK * ORDER * GRP	7.011	2	3.505	2.608	.095	5.215	.467
Error(TASK)	30.919	23	1.344				
SOUND	.763	1	.763	.531	.474	.531	.108
SOUND * ORDER	13.271	1	13.271	9.230	.006	9.230	.829
SOUND * GRP	.234	2	.117	.081	.922	.163	.061
SOUND * ORDER * GRP	5.311	2	2.655	1.847	.180	3.694	.345
Error(SOUND)	33.069	23	1.438				
TASK * SOUND	9.992E-03	1	9.992E-03	.012	.914	.012	.051
TASK * SOUND * ORDER	.175	1	.175	.206	.654	.206	.072
TASK * SOUND * GRP	1.066	2	.533	.629	.542	1.257	.142
TASK * SOUND * ORDER * GRP	11.720	2	5.860	6.913	.004	13.826	.886
Error(TASK*SOUND)	19.497	23	.848				

a. Computed using alpha = .05

Workload

Table B-27. Results of overall between subject factors for mental demands according to mouse-feedback group (grp).

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power [a]
Intercept	346508.9	1	346508.9	254 220	000	254.220	1.000
GRP	272.457	2	136.228	100	905	.200	064
Error	35438.750	26	1363.029				

a. Computed using alpha = 05

Table B-28. Summary of repeated-measures experimental results for mental demands analysis.

Measure:

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SOUND	1153.58	1	1153.58	13.24	.001
SOUND * GRP	128.86	2	64.43	74	487
Error(SOUND)	2265 97	26	87 15		
TASK	39484.37	1	39484.37	94.93	000
TASK * GRP	587.11	2	293.56	71	503
Error(TASK)	10813.75	26	415 91		
SOUND * TASK	142.86	1	142.86	3.00	095
SOUND * TASK * GRP	287 11	2	143.56	3 01	067
Error(SOUND*TASK)	1238.75	26	47.64		

a. Computed using alpha = 05

Table B-29. Overall pairwise comparisons for mental demands by mouse-feedback groups.

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition Groups (grp)	(J) Condition Groups (grp)	Mean Difference (I-J)	Std. Error	Sig.
Control	VRM	1.7500	8.255	834
	VRM+Force Feedback	-2.0417	8.482	812
VRM	VRM+Force Feedback	-3.7917	8.482	659
	Control	-1.7500	8.255	834
VRM+Force Feedback	VRM	3.7917	8.482	.659
	Control	2.0417	8.482	.812

Based on estimated marginal means

Table B-30. Results of overall between subject factors for physical demands according to mouse-feedback group (grp).

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Intercept	243178.6	1	243178.6	197.138	.000	197.138	1.000
GRP	12001.485	2	6000.742	4.865	.016	9.729	.753
Error	32072.222	26	1233.547				

a. Computed using alpha = .05

Table B-31. Summary of results for repeated-measures analysis of Physical Demands.

Measure:

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SOUND	24.80	1	24.80	22	641
SOUND * GRP	47 89	2	23.95	22	808
Error(SOUND)	2893.06	26	111.27		
TASK	2560 16	1	2560 16	4.71	039
TASK * GRP	766.00	2	383.00	71	503
Error(TASK)	14118.06	26	543.00		
SOUND * TASK	7 23	1	7.23	11	748
SOUND * TASK * GRP	214.74	2	107.37	1.56	229
Error(SOUND*TASK)	1790 00	26	68.85		

a. Computed using alpha = 05

Table B-32. Summary of results from pair-wise comparison of mouse-feedback group effects on physical demands.

Pairwise Comparisons

Measure:

(I) Condition Groups (grp)	(J) Condition Groups (grp)	Mean Difference (I-J)	Std. Error	Sig.
Control	VRM	4 25	7 85	593
	VRM+Force Feedback	-19.53*	8.07	.023
VRM	VRM+Force Feedback	-23.78*	8.07	007
	Control	-4.25	7.85	593
VRM+Force Feedback	VRM	23.78*	8.07	007
	Control	19.53*	8.07	.023

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

Table B-33. Summary of results for repeated-measures analysis of Temporal Demands.

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power [a]
Intercept	452946.9	1	452946.9	516.572	000	516.572	1.000
GRP	1458.396	2	729.198	832	447	1.663	.177
Error	22797.639	26	876.832				

a. Computed using alpha = 05

Table B-34. Summary of results for temporal demands analysis.

Measure: Temporal Demands

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SOUND	1100.0893	1	1100.09	14.03	001
SOUND * GRP	19.0086	2	9.50	.12	886
Error(SOUND)	2038.7500	26	78.41		
TASK	24422.23	1	24422.23	84.06	000
TASK * GRP	1007.4569	2	503.73	1.73	196
Error(TASK)	7553.7500	26	290.53		
SOUND * TASK	2.5397	1	2.54	02	880
SOUND * TASK * GRP	88.9128	2	44.46	40	671
Error(SOUND*TASK)	2857.6389	26	109.91		

a. Computed using alpha = 05

Table B-35. Pairwise comparisons of temporal demands by mouse-feedback groups.

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition Groups (grp)	(J) Condition Groups (grp)	Mean Difference (I-J)	Std. Error	Sig.
Control	VRM	6.7500	6.621	317
	VRM+Force Feedback	-1.3194	6.803	848
VRM	VRM+Force Feedback	-8.0694	6.803	.246
	Control	-6.7500	6.621	.317
VRM+Force Feedback	VRM	8.0694	6.803	.246
	Control	1.3194	6.803	.848

Based on estimated marginal means

Table B-36. Summary of results for repeated-measures analysis of Effort.

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Intercept	375711.5	1	375711.5	444.171	000	444.171	1.000
GRP	890.120	2	445.060	526	597	1.052	127
Error	21992.639	26	845.871				

a. Computed using alpha = 05

Table B-37. Summary of results for repeated-measures analysis of effort required.

Measure: Effort

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SOUND	1100.09	1	1100.09	14.03	.001
SOUND * GRP	19.01	2	9.50	.12	.886
Error(SOUND)	2038.75	26	78.41		
TASK	24422.23	1	24422.23	84.06	.000
TASK * GRP	1007.46	2	503.73	1.73	.196
Error(TASK)	7553.75	26	290.53		
SOUND * TASK	2.54	1	2.54	.02	.880
SOUND * TASK * GRP	88.91	2	44.46	.40	.671
Error(SOUND*TASK)	2857.64	26	109.91		

a. Computed using alpha = .05

Table B-38. Pairwise comparisons of effort by mouse-feedback group.

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition Groups (grp)	(J) Condition Groups (grp)	Mean Difference (I-J)	Std. Error	Sig.
Control	VRM	5.2500	6.503	427
	VRM+Force Feedback	-1.0694	6.682	874
VRM	VRM+Force Feedback	-6.3194	6.682	353
	Control	-5.2500	6.503	427
VRM+Force Feedback	VRM	6.3194	6.682	353
	Control	1.0694	6.682	874

Based on estimated marginal means

Table B-39. Summary of results for overall between subject measures analysis of "own performance" by group.

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power a
Intercept	216960.8	1	216960.8	274.195	000	274.195	1.000
GRP	1415.946	2	707.973	895	421	1.789	187
Error	20572.847	26	791.263				

a. Computed using alpha = 05

Table B-40. Summary of results for repeated-measures analysis "own performance".

Measure: Own Performance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SOUND	233.75	1	233.75	.84	.369
SOUND * GRP	173.13	2	86.56	.31	.736
Error(SOUND)	7256.18	26	279.08		
TASK	19153.87	1	19153.87	41.65	.000
TASK * GRP	270.60	2	135.30	.29	.748
Error(TASK)	11957.85	26	459.92		
SOUND * TASK	5.95	1	5.95	.04	.841
SOUND * TASK * GRP	72.73	2	36.36	.25	.781
Error(SOUND*TASK)	3794.51	26	145.94		

a. Computed using alpha = .05

Table B-41. Pairwise comparisons of own performance by mouse-feedback groups.

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition Groups (grp)	(J) Condition Groups (grp)	Mean Difference (I-J)	Std. Error	Sig.
Control	VRM	7.8750	6.290	.222
	VRM+Force Feedback	6.5972	6.462	.317
VRM	VRM+Force Feedback	-1.2778	6.462	.845
	Control	-7.8750	6.290	.222
VRM+Force Feedback	VRM	1.2778	6.462	.845
	Control	-6.5972	6.462	.317

Based on estimated marginal means

Table B-42. Summary of results for repeated-measures analysis of frustration.

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power a
Intercept	230528.0	1	230528.0	217.102	.000	217.102	1.000
GRP	2458.101	2	1229.051	1.157	.330	2.315	.232
Error	27607.847	26	1061.840				

a. Computed using alpha = .05

Table B-43. Summary of results for repeated-measures analysis for frustration.

Tests of Within-Subjects Effects

Measure: MEASURE_1
Sphericity Assumed

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power a
TASK	13291.729	1	13291.729	37.029	.000	37.029	1.000
TASK * GRP	149.480	2	74.740	.208	.813	.416	.079
Error(TASK)	9332.847	26	358.956				
SOUND	709.883	1	709.883	1.439	.241	1.439	.211
SOUND * GRP	32.986	2	16.493	.033	.967	.067	.055
Error(SOUND)	12829.514	26	493.443				
TASK * SOUND	55.062	1	55.062	.380	.543	.380	.091
TASK * SOUND * GRP	812.814	2	406.407	2.803	.079	5.606	.503
Error(TASK*SOUND)	3769.514	26	144.981				

a. Computed using alpha = .05

Table B-44. Pairwise comparions for overall ratings by mouse-feedback group for frustration.

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition Groups (grp)	(J) Condition Groups (grp)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Control	Kin	10.6250	7.286	.157	-4.352	25.602
	Kin+Force	2.4722	7.486	.744	-12.916	17.860
Kin	Kin+Force	-8.1528	7.486	.286	-23.541	7.235
	Control	-10.6250	7.286	.157	-25.602	4.352
Kin+Force	Kin	8.1528	7.486	.286	-7.235	23.541
	Control	-2.4722	7.486	.744	-17.860	12.916

Based on estimated marginal means

Table B-45. Results for between-subject analysis for overall performance.

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Intercept	366573.4	1	366573.4	702.152	.000	702.152	1.000
GRP	1717.687	2	858.844	1.645	.212	3.290	.315
Error	13573.847	26	522.071				

a. Computed using alpha = .05

Table B-46. Summary of results for repeated-measures for overall workload statistics.

Measure: Total Workload

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SOUND	631.50	1	631.50	7.00	.014
SOUND * GRP	32.27	2	16.13	.18	.837
Error(SOUND)	2345.03	26	90.19		
TASK	17614.65	1	17614.65	123.99	.000
TASK * GRP	463.13	2	231.56	1.63	.215
Error(TASK)	3693.68	26	142.06		
SOUND * TASK	32.57	1	32.57	.66	.424
SOUND * TASK * GRP	129.97	2	64.99	1.32	.286
Error(SOUND*TASK)	1284.53	26	49.40		

a. Computed using alpha = .05

Table B-47. Pairwise comparisons for Overall workload.

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition Groups (grp)	(J) Condition Groups (grp)	Mean Difference (I-J)	Std. Error	Sig.
Control	VRM	7.8750	6.290	.222
	VRM+Force Feedback	6.5972	6.462	.317
VRM	VRM+Force Feedback	-1.2778	6.462	.845
	Control	-7.8750	6.290	.222
VRM+Force Feedback	VRM	1.2778	6.462	.845
	Control	-6.5972	6.462	.317

Based on estimated marginal means

1.4 Final Questionnaire

Table B-48. Summary of results for the final questionnaire from f1 to f6.

Report

Condition Groups (grp)		Rating for usability of VRM compared to Normal Mouse (f1)	Rating of ease of movement on calculator (f2)	Rating of ability to detect buttons because of mouse (f3)	Rating of ability to detect mistakes because of mouse (f4)	Problems hearing sounds noted (f5)	Rating for ability to detect buttons based on sounds (f6)
Control	Mean	2.0000	4.4444	3.8000	2.4000	1.0000	2.8889
	Std. Deviation	2.8284	1.0138	1.6432	8944	0000	1.2693
	N	2	9	5	5	9	9
Kin	Mean	2.4444	3.9000	3.8333	3.5000	1.0000	2.7778
	Std. Deviation	7265	8756	1.6021	5477	0000	1.6415
	N	9	10	6	6	9	9
Kin+Force	Mean	2.6000	2.3000	3.3000	3.2000	1.0000	3.7000
	Std. Deviation	1.1738	6749	1.4944	1.3166	.0000	1.5670
	N	10	10	10	10	10	10
Total	Mean	2.4762	3.5172	3.5714	3.0952	1.0000	3.1429
	Std. Deviation	1.1233	1.2427	1.5024	1.0911	0000	1.5084
	N	21	29	21	21	28	28

Table B-49. Summary of results for final questionnaire for questions f7-f17.

Report

Condition Groups (grp)		Rating for ability to detect mistakes based on sounds (f7)	Rating of effectiveness of force feedback versus auditory feedback (f8)	Preference for Force feedback versus Auditory feedback (f9)	Experienced physical discomfort during the trials (f13)	Rating of discomfort experienced during trials (f15)	Rating of tiredness after study (f16)	Rating of VRM like versus dislike (f17)
Control	Mean	1.6667	6.0000	1.0000	1.3750	5.0000	3.8750	3.5000
	Std. Deviation	7071			5175	1.4142	1.1260	2.1213
	N	9	1	1	8	4	8	2
Kin	Mean	2.3333	2.6667	2.1667	1.1000	5.0000	4.1111	3.0000
	Std. Deviation	1.1180	.8165	4082	3162	.	1.4530	1.3093
	N	9	6	6	10	1	9	8
Kin+Force	Mean	2.9000	3.6000	2.2000	1.5000	4.0000	4.3000	2.5000
	Std. Deviation	1.2867	1.2649	.7888	5270	1.1547	9487	9718
	N	10	10	10	10	7	10	10
Total	Mean	2.3214	3.4118	2.1176	1.3214	4.4167	4.1111	2.8000
	Std. Deviation	1.1564	1.3257	.6966	4756	1.2401	1.1547	1.1965
	N	28	17	17	28	12	27	20

Table B-50. Summary of t-test results for questions from the final questionnaire f1-f9 comparing Control and Kin groups only.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig	t	df	Sig. (2-tailed)	Mean Difference	Std Error Difference
Rating for usability of VRM compared to Normal Mouse (f1)	Equal variances assumed	27 476	001	- 488	9	637	- 4444	9110
	Equal variances not assumed			- 221	1 030	861	- 4444	2.0146
Rating of ease of movement on calculator (f2)	Equal variances assumed	320	579	1 256	17	226	5444	4334
	Equal variances not assumed			1 246	15 955	231	5444	4369
Rating of ability to detect buttons because of mouse (f3)	Equal variances assumed	031	865	- 034	9	974	-3 33E-02	9812
	Equal variances not assumed			- 034	8 553	974	-3 33E-02	9838
Rating of ability to detect mistakes because of mouse (f4)	Equal variances assumed	1 954	196	-2 514	9	033	-1 1000	4376
	Equal variances not assumed			-2 400	6 391	051	-1 1000	4583
Rating for ability to detect buttons based on sounds (f6)	Equal variances assumed	644	434	161	16	874	1111	6917
	Equal variances not assumed			161	15 047	875	1111	6917
Rating for ability to detect mistakes based on sounds (f7)	Equal variances assumed	1 695	211	-1 512	16	150	- 6667	4410
	Equal variances not assumed			-1 512	13 517	154	- 6667	4410
Rating of effectiveness of force feedback versus auditory feedback (f8)	Equal variances assumed			3 780	5	013	3 3333	8819
	Equal variances not assumed						3 3333	
Preference for Force feedback versus Auditory feedback (f9)	Equal variances assumed			-2.646	5	046	-1 1667	4410
	Equal variances not assumed						-1 1667	

Table B-51. Summary of t-test results for more questions from background questionnaire from f1-f9 comparing Control with Kin-Force group.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Rating for usability of VRM compared to Normal Mouse (f1)	Equal variances assumed	4.939	.050	-.542	10	.599	-.6000	1.1063
	Equal variances not assumed			-.295	1.070	.815	-.6000	2.0342
Rating of ease of movement on calculator (f2)	Equal variances assumed	2.213	.155	5.482	17	.000	2.1444	.3912
	Equal variances not assumed			5.365	13.716	.000	2.1444	.3997
Rating of ability to detect buttons because of mouse (f3)	Equal variances assumed	.044	.836	.592	13	.564	.5000	.8444
	Equal variances not assumed			.572	7.428	.584	.5000	.8737
Rating of ability to detect mistakes because of mouse (f4)	Equal variances assumed	1.236	.286	-1.215	13	.246	-.8000	.6587
	Equal variances not assumed			-1.386	11.410	.192	-.8000	.5774
Rating for ability to detect buttons based on sounds (f6)	Equal variances assumed	.319	.580	-1.231	17	.235	-.8111	.6592
	Equal variances not assumed			-1.245	16.838	.230	-.8111	.6516
Rating for ability to detect mistakes based on sounds (f7)	Equal variances assumed	1.408	.252	-2.546	17	.021	-1.2333	.4845
	Equal variances not assumed			-2.623	14.249	.020	-1.2333	.4702
Rating of effectiveness of force feedback versus auditory feedback (f8)	Equal variances assumed		.	1.809	9	.104	2.4000	1.3266
	Equal variances not assumed			.		.	2.4000	.
Preference for Force feedback versus Auditory feedback (f9)	Equal variances assumed		.	-1.450	9	.181	-1.2000	.8273
	Equal variances not assumed			.	.	.	-1.2000	.

Table B-52. Summary of t-test results from final questionnaire comparing for f1-f9 comparing Kin and Kin+Force groups only.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig	t	df	Sig (2-tailed)	Mean Difference	Std Error Difference
Rating for usability of VRM compared to Normal Mouse (f1)	Equal variances assumed	1 727	206	- 342	17	736	- 1556	4543
	Equal variances not assumed			- 351	15 195	730	- 1556	4432
Rating of ease of movement on calculator (f2)	Equal variances assumed	860	366	4 577	18	000	1 6000	3496
	Equal variances not assumed			4 577	16 905	000	1 6000	3496
Rating of ability to detect buttons because of mouse (f3)	Equal variances assumed	000	996	673	14	512	5333	7920
	Equal variances not assumed			661	10 060	523	5333	8069
Rating of ability to detect mistakes because of mouse (f4)	Equal variances assumed	4 487	053	526	14	607	3000	5707
	Equal variances not assumed			635	12 995	537	3000	4726
Rating for ability to detect buttons based on sounds (f6)	Equal variances assumed	047	830	-1 253	17	227	- 9222	7363
	Equal variances not assumed			-1 249	16 587	229	- 9222	7382
Rating for ability to detect mistakes based on sounds (f7)	Equal variances assumed	023	880	-1 019	17	322	- 5667	5561
	Equal variances not assumed			-1 027	16 986	319	- 5667	5518
Rating of effectiveness of force feedback versus auditory feedback (f8)	Equal variances assumed	1 151	301	-1 606	14	131	- 9333	5812
	Equal variances not assumed			-1 793	13 833	095	- 9333	5207
Preference for Force feedback versus Auditory feedback (f9)	Equal variances assumed	3 675	076	- 095	14	925	-3 33E-02	3501
	Equal variances not assumed			- 111	13 858	913	-3 33E-02	3000

Table B-53. Summary of results from t-tests for final questionnaire (f13-f17) comparing for Control and Kin groups only.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Experienced physical discomfort during the trials (f13)	Equal variances assumed	593	453	-504	16	621	-.1250	2480
	Equal variances not assumed			-505	15.262	621	-1250	2475
Rating of discomfort experienced during trials (f15)	Equal variances assumed	096	763	1.279	9	233	1.0000	7817
	Equal variances not assumed			1.203	5.334	279	1.0000	8309
Rating of tiredness after study (f16)	Equal variances assumed	023	881	-870	16	397	-4250	4886
	Equal variances not assumed			-853	13.757	408	-4250	4985
Rating of VRM like versus dislike (f17)	Equal variances assumed	3.889	077	1.132	10	284	1.0000	8832
	Equal variances not assumed			653	1.086	625	1.0000	1.5312

Table B-54. Summary of t-test results for final questionnaire (f13-f17) comparing Kin and Kin+Force groups only.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Experienced physical discomfort during the trials (f13)	Equal variances assumed	16 000	001	-2 058	18	054	- 4000	1944
	Equal variances not assumed			-2 058	14 737	058	- 4000	1944
Rating of discomfort experienced during trials (f15)	Equal variances assumed			810	6	449	1 0000	1 2344
	Equal variances not assumed						1 0000	
Rating of tiredness after study (f16)	Equal variances assumed	2 315	147	- 339	17	739	- 1889	5571
	Equal variances not assumed			- 332	13 544	745	- 1889	5697
Rating of VRM like versus dislike (f17)	Equal variances assumed	466	504	931	16	366	5000	5369
	Equal variances not assumed			900	12 623	385	5000	5556

Table B-55. Summary of t-test results for final questionnaire (f13-f17) comparing Control and Kin+Force groups only.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig	t	df	Sig (2-tailed)	Mean Difference	Std Error Difference
Experienced physical discomfort during the trials (f13)	Equal variances assumed	593	453	- 504	16	621	- 1250	2480
	Equal variances not assumed			- 505	15 262	621	- 1250	2475
Rating of discomfort experienced during trials (f15)	Equal variances assumed	096	763	1 279	9	233	1 0000	7817
	Equal variances not assumed			1 203	5 334	279	1 0000	8309
Rating of tiredness after study (f16)	Equal variances assumed	023	881	- 870	16	397	- 4250	4886
	Equal variances not assumed			- 853	13 757	408	- 4250	4985
Rating of VRM like versus dislike (f17)	Equal variances assumed	3 889	077	1 132	10	284	1 0000	8832
	Equal variances not assumed			653	1 086	625	1 0000	1 5312

IMAGE EVALUATION TEST TARGET (QA-3)

150mm

6″

APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989