THE EFFECT OF COMMUNICATION STYLE ON TASK PERFORMANCE AND MENTAL WORKLOAD USING WEARABLE COMPUTERS

by

Eric B. Nash

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Industrial and Systems Engineering

Woodrow Barfield, Chairman
Brian M. Kleiner
Maury Nussbaum

10 August 2000
Blacksburg, Virginia

Keywords: Wearable Computers, Mental Workload, Icons, Communication Style

DEDICATION

I would like to dedicate this thesis to my parents for instilling in me the conviction that there can be both a right way and wrong way to do things.

ACKNOWLEDGEMENTS

I would like to thank my chairman, Dr. Woodrow Barfield for the opportunity to work in this field. He has been the most benevolent, generous, and most importantly, accessible professor with whom I have ever had the good fortune of working. While I relied heavily on his professional expertise, I learned the most from casual afternoon conversations, during which I was allowed an occasional peek into his tremendous vision and creativity. These simple and elegant exchanges were seminal to the orchestration of my own thoughts on not only the topics of this thesis but many other ideas and concepts.

I would also like to say a special thank you to the other members of my committee, Drs. Brian M. Kleiner, and Maury Nussbaum for their guidance through the course of this thesis as well as my studies at Virginia Tech.

Finally, I would like to thank the many graduate students within and without the Virtual Environment Lab--Greg Edwards, Mick McGee, Erik Olsen, Jen Thompson, Jimmy Vu, Tolga Durak, and Craig Hamrick--for all the good times and memories.

THE EFFECT OF COMMUNICATION STYLE ON TASK PERFORMANCE AND MENTAL WORKLOAD USING WEARABLE COMPUTERS

Eric B. Nash

Abstract

This thesis measured the mental workload associated with operating a voice activated software application run on wearable computer under five different communication styles (buttons, command line, icon buttons, icon text menus, and text menus). The goal of this thesis was to determine which communication style would be best allow wearable computer users to simultaneously perform other non-computer tasks. Thirty subjects were randomly assigned to using one of five software versions (n = 6), each of which utilized a unique communication style. The mental workload associated with operating each version was assessed by monitoring the performance of secondary tasks. Secondary tasks consisted of completing a block assembly, digit subtraction, and walking along a marked pathway. Each secondary task was performed twice by itself and once while operating one of the software versions, creating a total of nine trials per subject. Block assembly task performance measures included average assembly time, percentage correct blocks, and percentage correct blocks attempted. Digit subtraction measures included percentage of correct digits. And path walking measures included average walking speed. Subjective estimates of mental workload were also collected for those trials in which subjects operated the wearable computer and performed physical tasks using the NASA Task Load Index (TLX). Finally, usability information was collected for each software version via a questionnaire form.

Each of the five versions of the experimental software application was operationally identical to the others, but utilized a separate communication style. The button version displayed available functions via sets of labeled buttons in the control screen. The icon button version replaced the appearance of these buttons with labeled icons. The text menu version displayed available functions textually via a pull down main menu. The icon text version displayed appended icons to the left of each main menu item. Finally, the command line version displayed no labels, buttons, menus, or icons for any functions. The experimental software was designed as a day planner/scheduling application used to set reminder dates on a calendar, edit task lists, and edit phone listings.

Under the multiple resource view of mental workload, it was hypothesized that the different versions and secondary tasks would demand distinct types of mental resource and, consequently, that mental workload would be observed as lowest when the version and secondary task demanded different types of mental resources. In contrast, it was also hypothesized that mental workload would be observed as highest when the version and secondary task demanded the same type of mental resources. Although separate one way ANOVAs performed on all secondary task measures failed to indicate statistically significant differences in mental workload across the versions, secondary task performance was consistently observed as best for subjects using the icon button version. Analysis of NASA TLX subscale data indicated that the block assembly task was rated as requiring less effort and the digit task rated as requiring less mental demand when the icon button version was used. These results generally support using an icon button communication style for wearable computer software applications.

Results of this study are applicable to the design of the user interface of wearable computers. These results not only report subjective and objective measures for assessing the amount of mental effort associated with operating a wearable computer and performing various physical tasks simultaneously, but also provide estimates for determining the amount of physical task performance decrement to expect when wearable computer are also operated. Such data may be used to determine human factors guidelines for matching wearable computer interfaces to physical tasks so that interference between the two is minimal.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Motivation

One of the main goals of wearable computers is to allow users to access computational resources whenever and wherever they are in the physical environment (Barfield and Baird, 1998). Unlike desktop systems, which are characterized by both the user and the computer system remaining stationary in the physical environment, wearable computer systems are designed to be operable under conditions in which both the user and the computer system are subject to a changing environment. Changing environments are likely to impose their own sets of physical and mental demands on the user. Furthermore, these demands may be completely unrelated to operating the wearable computer. Therefore, a dichotomy must be drawn between *computer tasks*, characterized by their pertinence to operating the wearable computer, and *physical tasks*, characterized by their pertinence to interacting with the physical environment.

In order to comply with the previously stated goal of wearable computers, wearable computer software and hardware should be designed such that users can perform both computer tasks and physical tasks simultaneously. Researchers have recommended that wearable computers be designed to interfere with physical tasks as little as possible (Smailagic and Siewiorek, 1996). For example, users should be able to walk across a street corner or manually assemble objects with the same level of performance regardless of whether a wearable computer is also being operated.

Several research studies, however, have reported significant decrements in physical task performance when wearable computers are also operated (Siegel and Bauer, 1997; Ockerman and Pritchett, 1998; Ockerman, Najjar, and Thompson, 1997). These findings suggest that current wearable computer designs significantly interfere with physical tasks. Therefore, wearable computer interface designs are needed which pose minimal interference with physical tasks as they are performed simultaneously. If such interface designs can be determined, such decrements in physical task performance should improve.

1.2. Scope, Goal, and Significance

One strategy for designing wearable computers for improved physical task performance is to rely on existing behavioral theory that provides constructs for explaining and predicting performance of multiple tasks. Such theory may be used to guide the designs of wearable computer systems so that physical tasks are performed as effectively as possible. One theory which addresses the improvement of performing multiple tasks is the *multiple resource view*. The multiple resource view is based on the psychological construct of mental workload, which is defined as the feeling of mental effort or the level of use of the human operator's limited mental resources (Krantz, 1997). Proponents of the multiple resource view assert that several independent mental resources exist and, because of this independence, two tasks demanding different types of mental resources will not require as much mental effort and will be performed better than when two tasks demand the same type of mental resource (Wickens, 1992; Wickens, 1983).

Two types of mental resources purported by researchers to be stored independently are *spatial* and *verbal* resources (Wickens, 1983). Any task which requires a judgment or integration concerning three axes of translation or orientation is defined as spatial. Any task that requires use of language or some arbitrary symbolic coding is defined as verbal. Under the multiple resource view, if one task demands verbal mental resources and one demands spatial mental resources, performing these two tasks together will be better and less mentally demanding than two tasks demanding the same type of resource.

It is possible that the poor physical task performance found in prior wearable computer studies (e.g., Siegel and Bauer, 1997; Ockerman and Pritchett, 1998; Ockerman, Najjar, and Thompson, 1997) was the result of using wearable computer interfaces that competed for the same type of mental resource as the physical task. If wearable computer interfaces can be designed to demand a *different* type of mental resource from that demanded by a physical task, then mental workload and decrements in the physical task performance will be reduced.

The main goal of this thesis was, therefore, to identify specific features of wearable computer software applications that may affect the amount and type of mental resource demanded. Application features were isolated using separate versions of an experimental wearable computer application. It was initially hypothesized that certain versions would require more spatial mental resources to operate and others would require more verbal mental resources to operate. Similarly, it was assumed that certain physical tasks would require more spatial mental resources to perform and others would require more verbal mental resources to perform. For instance, while path walking and block assembly were assumed to measure spatial working memory resources, digit subtraction was assumed to measure verbal mental resources.

Based on these assumptions, it was hypothesized that when a version and physical task demanded the same type of mental resources, greater mental effort would be required and performance of the physical task would decrease. In contrast, it was hypothesized that when a version and physical task demanded different types of mental resources, less mental effort would be required and physical task performance would not decrease. Provided that these results were obtained, this thesis was designed to argue that decrements in physical task performance while operating wearable computer could be prevented if the wearable computer interface was designed such that its operation demanded a type of mental resource different from the physical task.

1.3. Organization

This thesis is organized into six chapters: Introduction, Review of Literature, Methods, Results, Discussion, Conclusions, and Appendices. The first chapter, Introduction, describes the terminology and concepts used throughout the thesis, as well as concepts related to the experimental design and procedure. Chapter 2, Review of Literature, will review findings from previous studies examining the effects of wearable computer operation on physical task performance. The results of these studies will be interpreted in terms of mental workload. Based on this interpretation, *icon usage* and *communication style* will be proposed as two interface features with the potential for demanding specific types of mental resources. Justification for why either feature may demand a specific type of mental resource will be provided via literature reviews on desktop systems. Finally, hypotheses will be made for the effect of communication style on mental workload within the context of wearable computer software applications. Chapter 3, Methods, will cover the experimental protocol followed in this thesis. In chapter 4, graphical and tabular results of all descriptive and inferential statistical will be provided. Chapter 5 will discuss the results in terms of the theoretical terms and constructs provided earlier. References will be provided in Chapter 6. Finally, the raw data analyzed in this experiment will be displayed in the Appendices.

2. REVIEW OF LITERATURE

2.1. Physical Task Performance while Operating Wearable Computers

A wearable computer is defined as "a computer that is subsumed into the personal space of the user, controlled by the user, and has both operational and interactional constancy" (Mann, 1998). The constancy of running a computer that does not have to be opened up and turned on prior to use is a defining characteristic separating wearable computers from other portable devices, such as personal digital assistants (PDAs), laptop computers, and cellular telephones.

The idea of providing users with computing resources *during* performance of jobs at a work site has frequently been described as an application of wearable computers (Mecham, 1997; Starner, Rhodes, and Foner, 1995; Ockerman, Najjar, Thompson, Treanor, and Atkinson, 1996; Thompson, Ockerman, Najjar, and Rogers, 1997). However, in order increase the usefulness of such an application, wearable computers should pose minimal interference while completing jobs (Smailagic and Siewiorek, 1996). The question then becomes how to design wearable computer applications so that they minimally interfere with jobs.

Researchers have indicated that traditional graphical user interfaces optimized for desktop use may not be ideal for wearable computer applications for several reasons (Billinghurst, Bowskill, Dyer, and Morphett, 1998). For instance, users of wearable computers can use non-traditional input hardware to operate the wearable computer in a hands free manner, such as one-handed twiddlers or microphones for voice input. In addition, wearable computer users view output using monoscopic head mounted displays (HMD) with limited resolution and field of view. Finally, in contrast to using traditional computer desktops, wearing the computer on the body means that users will be required to interact with the computer while assuming various positions and locations in the physical environment.

A few studies have compared subjects' performance of various physical tasks while accessing task-relevant information using either a wearable computer or using a paper reference. Many of these studies reported decrements in physical task performance when subjects operated a wearable computer. For example, Siegel and Bauer (1997) showed data indicating that avionics specialists who completed guided aircraft maintenance tasks required

50% more time than those using paper references. Ockerman and Pritchett (1998) reported that pilots performing pre-flight inspection tasks were less thorough when a voice activated wearable computer application was used instead of a paper checklist. In addition, Ockerman, Najjar, and Thompson (1997) showed data indicating that subjects required significantly more time (~60%) to complete an unfamiliar paper-folding task (origami) using a voice activated instructional application on a wearable computer compared to using a paper instruction manual.

2.2. The Role of Mental Workload in Wearable Computing

The concept of *mental workload* may provide some theoretical insight into the above reports of decreased physical task performance when wearable computers were also operated. This thesis will use the terms and definitions provided by Navon and Gopher (1979) for the discussion of mental workload. Mental workload is a condition that arises from the assertion that humans possess a finite amount of processing facilities, referred to in the literature as effort, capacity, attention, and resources. Mental resources are allocated to tasks, which can be characterized across a number of dimensions, including sensory quality, stimulus predictability, availability of memory codes, stimulus-response compatibility, response complexity, and amount of practice. How well a task is performed depends on the amount of mental resources the human allocates to the task and how efficiently the mental resources can contribute to the task. The amount of mental resources required to achieve a desired level of performance on a task is referred to as demand. Humans will supply mental resources to meet the demand for a desired level of task performance to the extent that they are available.

Mental resources can be allocated to the performance of a single, primary task as well as to additional, secondary tasks. However, once the available mental resource limit has been reached for the primary task, performance of additional tasks demanding the same resource will suffer (Wickens, 1984; Wickens, 1992; Meshkati, Hancock, Rahimi, and Dawes, 1995). If operating the wearable computer can be considered the primary task in the wearable computer studies reported above (Siegel and Bauer, 1997; Ockerman and Pritchett, 1998; Ockerman, Najjar, and Thompson, 1997) then it is possible that mental resource demands imposed by operating the wearable computer may have caused fewer resources to be available for the physical tasks they monitored. This decrease in mental resource availability may have resulted in the observed decrease in physical task performance.

If the cause for the decreased performance on the physical task is indeed related to a shortage of available mental resources, one strategy for wearable computer designers is to create interfaces that can be operated while allowing sufficient mental resources to be made available to physical tasks. Researchers have suggested that two tasks can be performed concurrently with minimal mutual interference when the tasks require different stages of processing, different modalities of processing (McLeod, 1977), and/or different hemispheres of processing (Wickens, 1980; Wickens, 1992). Because each of these divisions of processing is suggested to function with separate resource-like properties, two tasks will be performed without conflict and without capacity limitations if they demand separate rather than common resources (Wickens, 1984; Allport, 1980). This notion that humans possess several different types of independent mental resources is known as the *multiple resource view* (Wickens, 1992).

Two hemispheres of processing that have been suggested to function as separate resources are the processes related to spatial and verbal working memory (Wickens, 1984). This separation has been suggested to be related to the large functional distance between localized spatial processing faculties in the right hemisphere of the brain and localized verbal processing faculties in the left hemisphere (Kinsbourne and Hicks, 1978). Wickens (1983) also dichotomizes between *spatial* and *verbal* tasks. Specifically, any task which requires a judgment or integration concerning three axes of translation or orientation is defined as spatial. In contrast, any task that requires use of language or some arbitrary symbolic coding is defined as verbal. According to the multiple resource view, two tasks demanding separate mental resources will be performed more effectively than two tasks demanding the same mental resources. Thus, if a physical task is known to demand one type of mental resource, it may be possible to design wearable computer applications so that they demand another type of mental resource. This way operating the wearable computer will not be affected by resource demands imposed by the physical task.

In order to design wearable computer applications to demand specific types of resources, specific wearable computer features must be isolated and the types of mental resources they demand must be determined. Visual icons are a common feature of user interfaces and are defined as graphic images to represent system objects, options, operations, applications, and messages to the user (Preece, 1998). No research studies have attempted to measure the types of mental resources demanded by visual icons. However, because visual icons can be placed in specific locations within an application, it is possible that the execution of commands may become easier if these spatial locations communicate function information to the user, such as arranging direction arrow icons in a north-south-east-west pattern. Moreover, the icons themselves may communicate information to the user, such as when arrows are used to indicate movement of files, etc. Given these associations with spatial information, it may be hypothesized that operating applications with visual icons may be more demanding in spatial resources than in verbal resources.

Communication style is defined as the way a user interacts with a computer system to exchange information (Preece, 1998). It includes an assemblage of different features for making user inputs into a computer interface. For instance, a push button communication style is characterized by "pushing" screen buttons on the screen to execute specific functions. In contrast, functions may also be executed by selecting items from a pull down menu. Finally, users may execute the functions by entering commands into a text field (command line) using some type of command language (e.g., DOS text commands). Given such variety in the way functions can be performed, it is possible that using these different communication styles may demand different types of mental resources. Since using buttons involves visually searching or recalling the spatial locations of buttons, applications utilizing buttons may demand more spatial resources to operate. Spatial resources may also be demanded for menu applications, but perhaps to a lesser degree since the items on a text menu are typically listed along a single direction. In contrast, because inputting commands into a command line involves recalling a command language syntax, applications utilizing command lines may demand more verbal resources to operate.

Icon usage and different communication styles can be combined within wearable computer software applications to produce a variety of options for executing functions and accessing resources. For instance, icons may be combined with a push button communication style where the buttons appear as icons. Icons may also be combined with text menu items, where icons appear with text labels. If certain features demand distinct types of mental resources, and these features are combined to create an interface, then it can be expected that use of the interface will demand the same types of mental resources. In the following sections, icon usage and communication style usage and the types of mental resource demands specific to each is described in more detail.

2.3. Use of Icons in Wearable Computer Applications

Icons are created by system designers with the goals of being easily remembered and logically "mapped" to what is being represented (Barfield, Rosenburg, and Levasseur, 1991; Hemenway, 1982). Several advantages for using icons over other types of interface displays have been reported using desktop computer systems. Research has shown that icons can communicate information to users more quickly than text. Wandmacher and Müller (1987) provided evidence to show that search and select times for specific commands were faster with a panel of pictures versus a panel of text labels. Arend, Muthig, and Wandmacher (1987) gave subjects computer task descriptions and recorded the time required to find the appropriate option from menus that did or did not display icons. They found that the time to locate the appropriate menu option decreased by approximately one half when icons were used.

Other studies have shown that icons improve recognition and recall of application commands and structures. Barfield *et al* (1991) exposed subjects to hierarchical menus that did or did not also display icons and measured the recognition of menu options using a follow up questionnaire. It was reported that subjects recognized significantly more commands and reported fewer false positives from the list when they had used the menus displaying icons earlier in the experiment. As a follow up procedure, Merwin, Dyre, and Humphrey (1990) measured subjects recall of a previously used database hierarchy by recording the frequency of correct answers written on a partially filled paper diagram of the hierarchy. A significantly greater percentage of correct items were reported when subjects had viewed the hierarchy with icon-text labels than when no icons were present.

In addition, there has been evidence that users perform certain tasks more effectively when icons are used in interfaces. Morse, Lewis, Korfhage, and Olsen (1998) recorded subjects information retrieval time while they used icon lists, text, tables, and other visualization types to locate specific types of information. Results suggested that retrieval time was fastest when icons were used over any other display type.

Though these findings are restricted to desktop systems, they are presently the only basis available for predicting physical task performance with wearable computer systems.

As such, they suggest that interfaces using icons may comprise their own set of advantages and disadvantages for performing physical tasks. For example, based on the recall advantages reported for icons by Merwin, Dyre, and Humphrey (1990), it is possible that users of iconic wearable computer applications will be better able to recall the spatial placement of specific commands within the application. Consequently, users may require less time to visually search for specific commands on the display than if icons were not used. Spending less time searching for commands on the display may enable users to spend more time visually attending to whatever physical task they are also performing. On the other hand, the reliance of iconic wearable computer applications on visual processes may present a potential disadvantage when a physical task also relies on visual processes. From a mental resource perspective, if operating a wearable computer and performing a physical task both require the same types of processes, task performance should decrease. Consequently, iconic interfaces may not be appropriate while performing physical tasks that also demand visual processes.

2.4. Use of Communication Style in Wearable Computer Applications

Communication style is defined as the method by which users exchange information with computers (Preece, 1998). Information exchanges between users and computers can be accomplished by several different means. For example, in a command line communication style, users recall specific commands of a command language from memory and input these (usually textually) into a single field. In contrast, a text menu communication style requires users to recognize and select commands from a list. Finally, users can execute commands by activating screen buttons arranged in meaningful ways within the application (e.g., to indicate directions). For the sake of simplicity, pull down menus will be referred to as text menus and screen buttons will be referred to as buttons from this point forward.

Researchers have attempted to determine how different communication styles between users and computers can affect the performance of various computer tasks. Results on these tasks have often favored the command line communication styles. Durham and Emurian (1998) trained subjects to a criterion performance on command line and text menu communication styles and recorded errors committed during the completion of sets of computer tasks. Although neither communication style was associated with fewer errors during the initial experiment, significantly greater frequencies of errors were reported for the text menu communication style following a four week retention interval. In addition, Whiteside, Jones, Levy, and Wixon (1985) monitored the performance of several file manipulation tasks using command line, text menu, and iconic displays and reported that command lines were generally associated with better performance. Using command line and other menu communication styles to reorder the appearance of a set of data fields in a table, Tullis and Kodimer (1992) reported that when three or four sequencing operations were performed, subjects using the command line communication style completed the task in approximately 60% of the time required when using menu communication styles.

In the case of using wearable computer applications while performing physical tasks, Thompson, Najjar, and Ockerman (1997) have suggested using command line communication styles to the exclusion of more graphical user interfaces. Command line communication styles were recommended on the basis that they may require fewer navigation steps to execute commands than graphical user interfaces. By reducing the amount of navigational steps, command line communication styles would have the advantage of allowing users to concentrate more readily on their physical tasks.

However, there may also be inherent disadvantages associated with command line wearable computer interfaces. Because command line communication styles rely on the user to recall verbal commands from memory, they may present a potential disadvantage when the physical task also requires the processing of verbal information. From a mental resource perspective, if operating a wearable computer and performing a physical task both require the same types of verbal processing, mental workload should be greater and task performance should decrease. Consequently, command line interfaces may not be well suited to performing physical tasks that also demand verbal processes.

Text menu and button interfaces may also have inherent advantages and disadvantages similar to iconic wearable computer interfaces when physical tasks must be performed. While text menu and button communication styles do not require the user to remember verbal commands to execute functions, they do require users to visually search and distinguish commands. The process of searching and distinguishing commands on such interfaces may possibly result in high visual-spatial demands. These demands may be higher for button communication styles since buttons may be sized, spaced, and grouped visually to convey further information (e.g., arranging arrow buttons four directions). In addition, time spent searching for commands on the display may prohibit users from spending time visually attending to whatever physical task they are also performing.

The reliance of text menu and button wearable computer applications on visual-spatial processes may present a potential disadvantage when a physical task also relies on visual-spatial processes. From a mental resource perspective, if operating a wearable computer and performing a physical task both demand the same types of visual-spatial processes, mental workload should increase and task performance should decrease. Consequently, text menu and button wearable computer applications may not be appropriate while performing physical tasks that also demand visual-spatial processes. However, such interfaces may be more appropriate while performing physical tasks that demand resources other than visual-spatial.

2.5. Literature Review Conclusion

The literature reviewed here shows that physical task performance suffers whenever wearable computers must also be operated (Siegel and Bauer, 1997; Ockerman and Pritchett, 1998; Ockerman, Najjar, and Thompson, 1997). In addition, the multiple resource view has been suggested as a theory for interpreting not only why physical performance suffered during use of the wearable computer, but also how future wearable computer interfaces may be designed so that physical task performance will not suffer. Lastly, two types of software features have been reviewed that potentially demand different types of mental resources. Based on these reviews, specific hypotheses will be made concerning how the reviewed software features will affect the performance of different physical tasks. These hypotheses are detailed in the following section.

2.6. Research Hypotheses

The versions were organized *a priori* into a simple continuum showing the versions postulated to require more verbal type resources to the left and versions postulated to require more spatial resources to the right (Figure 1).

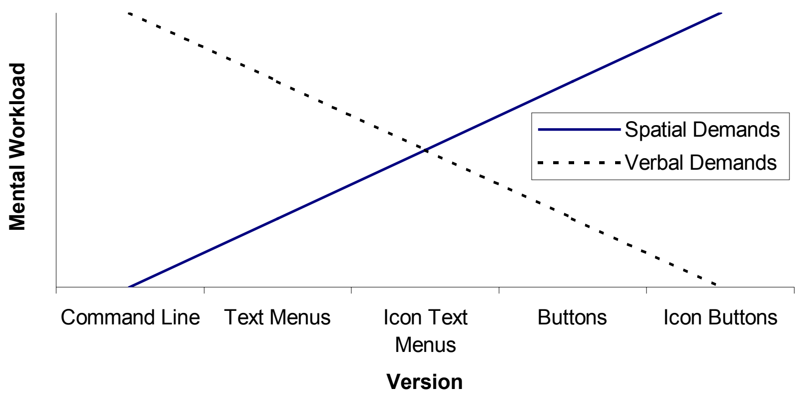

Figure 1. Proposed Continuum of Mental Resource Demands Across Version

Using the multiple resource view, it was hypothesized that subjects using the icon button version would show the greatest increase in mental workload and greatest decrease in secondary task performance when the secondary task also demanded spatial mental resources (e.g., the assembly and path walking tasks). In turn, it was hypothesized that subjects using the icon button version would show the greatest decrease in mental workload and the greatest increase in secondary task performance when the secondary tasks demanded verbal mental resources (e.g., the digit subtraction task). It was further hypothesized that the button, icon text menu, and text menu versions would simply show a less exaggerated version of the same results hypothesized for the icon button version.

Additionally, it was hypothesized that subjects using the command line version would show the greatest increase in mental workload and the greatest decrease in secondary task performance when the secondary task also demand verbal mental resources (e.g., digit subtraction task). Furthermore, it was hypothesized that subjects using the command line version would show the greatest decrease in mental workload and the greatest increase in secondary task performance when the secondary tasks demanded spatial mental resources (e.g., the assembly and path walking tasks).

3. METHODS

3.1. Participants

Thirty participants were recruited from undergraduate and graduate students at Virginia Tech. Eighteen subjects were male and 12 were female and ranged in age from 20-34 years with a mean age of 26.8. None of the subjects had prior experience operating a wearable computer or using the experimental software. Recruitment was through personal contacts and public flyers posted around campus. Each subject attended one experimental session which lasted approximately 90 minutes. Subjects were compensated $10 for their time in the experiment.

3.2. Equipment

The wearable computer used was a Xybernaut 133P with 16 MB RAM running Windows 95™. Test participants viewed the wearable computer display using a monochrome VGA HMD with 640 x 480 resolution, 20 degrees horizontal field of view, and 15 degrees vertical field of view. Verbal inputs were captured using a noise-canceling microphone and Verbex Listen for Windows speech recognition/synthesis software.

All five versions of the experimental software application were developed using MS Visual C++. Pictorial examples of the command line, text menus, icon text menus, buttons, and icon buttons versions are included in Figure 2. This application served as an electronic organizer, allowing users to record events on a monthly calendar, edit items on a task list, and edit phone lists.

All versions included a header panel at the top of the screen that displayed a text field labeled, “Task,” where the current benchmark task could be displayed and a button labeled, “Begin.” Subjects could begin the first task by speaking the command, “Begin” into the microphone, which would display the first task in the text field. Once any task was completed, subjects could display succeeding computer benchmark tasks in the text field by issuing the verbal command, “Next Task.” The wearable computer automatically recorded the time elapsed between activating the “Begin” button and the completion of the final benchmark task.

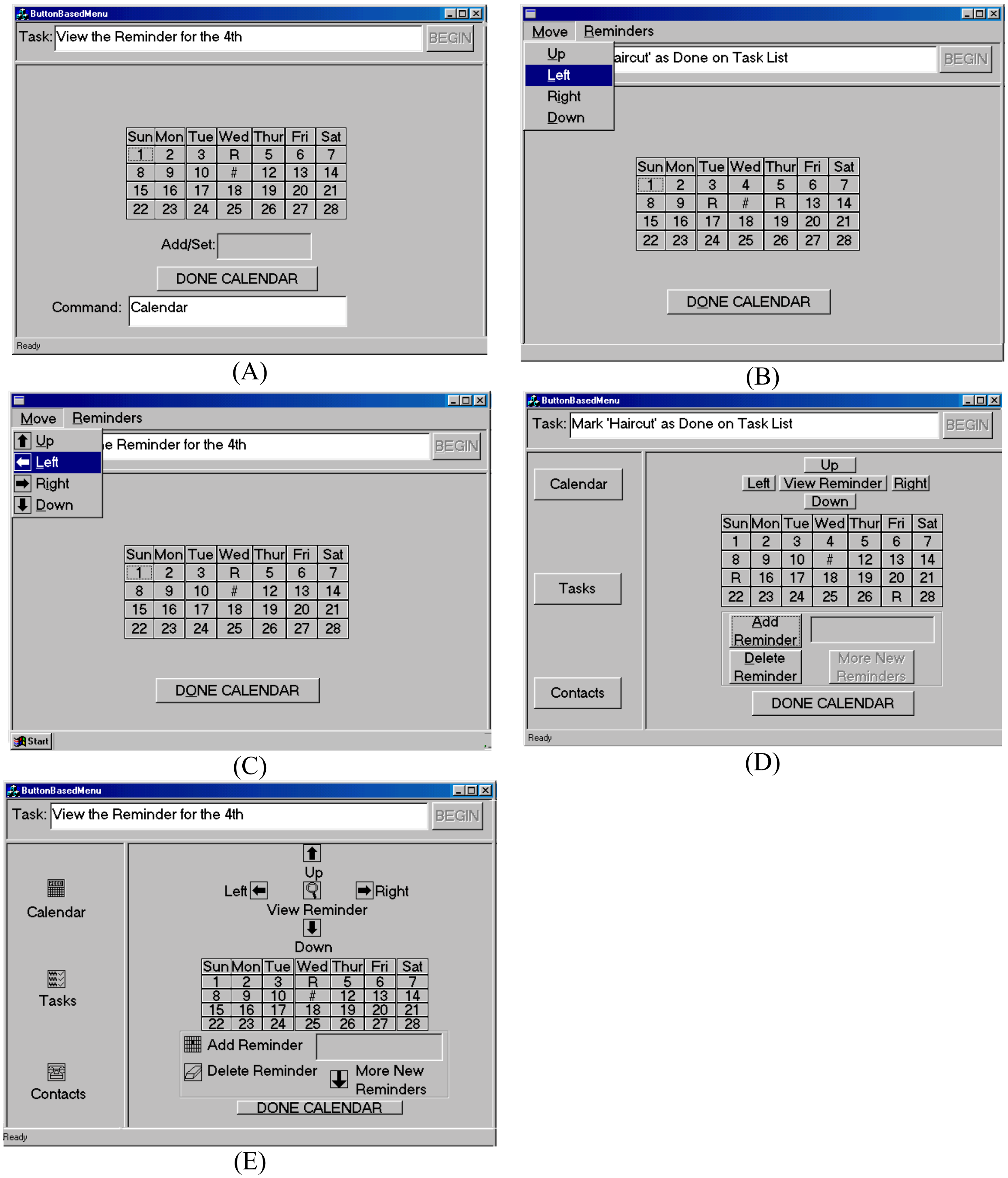

Figure 2. Versions of the experimental software: command line (A), text menus (B), icon text menus (C), buttons (D), and icon buttons (E).

The versions utilizing buttons displayed available functions via sets of labeled buttons in the control screen. A button became activated when the subject spoke the associated label into the microphone. Button labels appeared centered on the buttons themselves (Figure 2-D). For the icon button version, the buttons were replaced by icons with the label appearing in close proximity (Figure 2-E).

The menu versions displayed lists of available functions textually via pull down menus. Subjects were required to issue separate commands to activate functions from the pull down lists: one command to display all the items on the list and one additional command to execute a single item from the list. When icons were used, they were appended to the left of each menu item (Figure 2-C). Otherwise, the labels appeared alone on the menu (Figure 2-B).

The command line version (Figure 2-A) displayed no function labels for any available commands. Rather, commands had to be recalled from memory and spoken into the microphone as appropriate. Once the command was spoken into the microphone, the associated function was executed and the command appeared within the COMMAND field at the bottom of the screen.

The icons used for the icon button and icon text menu versions were selected based on a survey of 8 subjects prior to the experiment. Three icons were created to represent each function and were displayed in series on a paper survey. The survey was administered to rank order the icons representing a given function. Only the icons with the highest mean ranking across subjects were used in the experiment (Table 1).

Table 1. Mean rank values for Icon candidates across Function Description
(1 = Most Representative, 3 = Least Representative)

Function Description	**Icon 1**	**Icon 2**	**Icon 3**
Switching to *calendar* mode. This mode will display a calendar onto which you may set reminders for appointments.	2.75	2.13	**1.13**
Switching to *task scheduling* mode. This mode will display a task list onto which you may edit task items.	3.00	**1.63**	2.00
Switching to *contact* mode. This mode will display a list of contacts onto which you may update phone information.	**1.13**	2.13	2.75
Setting a new appointment reminder on the calendar.	2.50	2.00	**1.50**
Removing an old appointment reminder from the calendar.	2.38	**1.25**	2.25
View more specific information on a particular appointment reminder.	**1.88**	2.00	2.88
Advance through a list of existing appointment reminders.	2.63	2.13	**1.25**
Creating a new task item.	2.38	**1.75**	1.88
Removing an old task item.	**1.75**	2.25	2.00
To place an "X" next to a task on the task list, signifying that the task is "done".	**1.50**	2.00	2.50
Advance through a list of existing tasks.	2.38	2.38	**1.25**
Update a phone listing for a contact.	**1.88**	2.13	2.00
Removing a phone listing for a contact.	**1.50**	2.25	2.25
Advance through a list of existing phone listings.	2.50	2.25	**1.25**
Backspace through the digits in a phone listing	**1.00**	2.00	3.00
Sort through a list of contact names by first name	**1.00**	2.00	3.00
Sort through a list of contact names by last name	**1.00**	2.00	3.00
*Bold ratings indicate highest mean ranking for each set of icon candidates			

An attempt was made to produce an application that conformed to specific design recommendations published by previous wearable computer researchers (Table 2). When appropriate, a control/content screen format was employed in which users issued commands via the control screen and observed changes via the content screen. All text appeared in no less than 14-point font (Siegel, Kraut, John, and Carley, 1995; Siewiorek, Smailagic, Bass, Siegel, Martin, and Bennington, 1998). And active buttons were explicitly highlighted during use with button version (Siegel and Bauer, 1997).

Table 2. Summary of Wearable Computer Design Recommendations

Recommendations	**Researchers**
• Interface's active buttons should be explicitly highlighted. • Related procedures and illustrations should be linked such that navigation is minimized.	Siegel and Bauer (1997)
• Design interfaces with two screens: a control and a content screen. The control screen should occupy one-third of the screen and be used for controlling the display of the content screen by selecting topics or commands. The content screen should occupy the remaining two-thirds of the screen. • Use font sizes of 14-point and above to ensure visibility.	Siewiorik, *et al* (1998); Siegel, *et al* (1995)
• Use acoustically distinct phrases to maximize voice recognition accuracy.	Thompson, Najjar, and Ockerman (1997)

Subjects operated the software by completing sets of 5 benchmark tasks. A benchmark task is defined as a set of performance criteria which a product is expected to meet (McDaniel, 1993). Time constraints prevented the task sets to include more than five benchmark tasks. All task sets were identical in terms of the types of tasks included and the number of inputs, but differed in terms of specific information required to complete the tasks. This was done to ensure that subjects would not memorize the inputs from one condition to another. For instance, all task sets included a benchmark task requiring subjects to update a phone listing. However, the specific digits used to update the phone listing were different across task sets. Examples of some of the benchmark tasks used are provided below:

- Mark 'Haircut' as Done on Task List
- View the Reminder for the 27th
- Update Earl Albert's Phone Number to 956-421-9247

A computer version of the NASA Task Load Index (TLX) questionnaire (Hart and Staveland, 1988) was also used in order to provide a subjective measure of mental workload for each of the different versions. The computer version was developed by the Human Performance Group at NASA Ames Research Center to be operated with a mouse. The TLX was run on a Compaq™ Presario laptop using 32 MB RAM and running Windows 98™. The TLX consisted of 6 different subscales (mental demand, physical demand, temporal demand, performance, effort, and frustration level) on which subjects assigned weights and ratings. A measure of mental workload was then obtained by calculating the mean of the weighted ratings across subscales.

Finally, a short usability questionnaire was constructed to evaluate subjects overall impression of the wearable computer application as well as usability (Appendix 1). The items used in the questionnaire were constructed based on fundamental usability concepts (Preece, 1998) covering ease of learning, quickness of use, error proneness, ease of use, distinguishableness and comprehensiveness of interface items, and display clutter. Subjects answered the questionnaire items by providing ratings on a scale of 1 to 7.

3.3. Experimental Design

Each physical task (assembly task, digit task, and walking task) was treated as a separate 5 x 2 mixed factorial design (Table 3). The version factor consisted of five levels: command line, text menus, icon text menus, buttons, and icon buttons. The wearable computer status factor consisted of two levels: wearable computer off and wearable computer on. Levels of physical task consisted of the Walking, Assembly, and Digits tasks. Physical tasks were counterbalanced across subjects using a balanced Latin square (Table 4).

Table 3. Experimental Design

			Version				
			Command Line	Text Menus	Icon Text Menus	Buttons	Icon Buttons
Physical Task	Assembly	Wearable Computer Off	S_1-S_6	S_7-S_{12}	S_{13}-S_{18}	S_{19}-S_{24}	S_{25}-S_{30}
		Wearable Computer On	S_1-S_6	S_7-S_{12}	S_{13}-S_{18}	S_{19}-S_{24}	S_{25}-S_{30}
	Digit	Wearable Computer Off	S_1-S_6	S_7-S_{12}	S_{13}-S_{18}	S_{19}-S_{24}	S_{25}-S_{30}
		Wearable Computer On	S_1-S_6	S_7-S_{12}	S_{13}-S_{18}	S_{19}-S_{24}	S_{25}-S_{30}
	Path Walking	Wearable Computer Off	S_1-S_6	S_7-S_{12}	S_{13}-S_{18}	S_{19}-S_{24}	S_{25}-S_{30}
		Wearable Computer On	S_1-S_6	S_7-S_{12}	S_{13}-S_{18}	S_{19}-S_{24}	S_{25}-S_{30}

Table 4. Physical Task Counterbalanced Across Subjects

		Task Order		
		1	2	3
Subjects (n=6)	S_1	Assembly	Digits	Path Walking
	S_2	Assembly	Path Walking	Digits
	S_3	Digits	Assembly	Path Walking
	S_4	Digits	Path Walking	Assembly
	S_5	Path Walking	Digits	Assembly
	S_6	Path Walking	Assembly	Digits

Several time and accuracy measures were employed for each of the physical tasks (Table 5). For the assembly task, *average assembly rate* was used to measure the speed of assembly and was defined as the number of assembled blocks (correct or not) divided by the number of seconds lapsed during the trial. Accuracy was assessed using percentage correct blocks and percentage correct blocks attempted. Percentage correct blocks was defined as the number of correct blocks correctly assembled during the trial divided by the total number of blocks required to complete the assembly task (17). Percentage correct blocks attempted was defined as the number of correct blocks correctly assembled during the trial divided by the number of blocks (correct or not) assembled during the trial.

Because the number of digits correctly subtracted during a trial was confounded by the individual length of subjects' trial time, accuracy on the digit subtraction task was assessed using p*ercentage correct digits.* This measure was defined as the number of digits correctly subtracted during the trial divided by the total number of digits subtracted during the trial. Since all subjects received digits at the same rate while performing this task, no time measure was employed.

Time on the walking task was assessed using a*verage walking speed,* which was defined as the number of feet traversed along the path divided by the number of seconds lapsed during the trial. Due to difficulties in defining and measuring deviations from the marked path, no accuracy measures were employed.

Table 5: Dependent Variables and Their Components

Dependent Variable	Definition	Formula
Task time	Number of seconds lapsed during the trial	T
Assembly Task		
Total Blocks Assembled	Number of blocks assembled during the trial	B
*Average Assembly Rate	Total blocks assembled divided by task time.	B / T
Correct Blocks	Number of correct blocks correctly assembled during the trial	C_B
*Percentage Correct Blocks	Correct blocks divided by the number of blocks required to complete the assembly task (17)	C_B / 17
*Percentage Correct Blocks Attempted	Correct blocks divided by total blocks assembled.	C_B / B
Digit Task		
Total Digits	Number of digits subtracted during the trial	D
Correct Digits	Number of correct digits subtracted during the trial	C_D
*Percentage Correct Digits	Correct digits divided by total digits	C_D / D
Walking Task		
Distance Walked	Number of feet traversed along the path during the trial	F
*Average Walking Speed	Distance walked divided by task time	F / T

* Analyzed using inferential statistical procedures

3.4. Procedure

After signing the consent form (Appendix 2), each subject was randomly assigned to one of the versions of the experimental software. Random assignment to the versions was restricted such that equal numbers of subjects (n = 6) would be assigned to each version. Subjects began by donning the wearable computer apparatus, which they wore for the entirety of the experiment. After performing a set of verbal exercises into the wearable computer microphone to acclimate the speech software to their voice, subjects were trained to operate the application using a set of practice benchmark tasks. Subjects continued practicing the same set of benchmark tasks until they were able to complete the same set three times without error. This criterion was chosen because it had been used successfully in previous experiments designed to examine the learning effects associated with different communication styles (Durham and Emurian, 1998). It was assumed that by demonstrating mastery of the task set three times over, differences found within the experimental trials would not be as likely to be confounded by practice effects.

Each physical task was performed over three trials to provide a total of nine trials per subject. In the first trial, subjects performed the physical task with the wearable computer turned off. During the second trial, subjects repeated the physical task while also performing a set of computer benchmark tasks on the wearable computer. Because any differences found between the first and second trial could be confounded by the additional practice subjects had gained during the first trial, subjects repeated the physical task with the wearable computer once again turned off in a third trial. Performance data from the first and third trials were averaged to provide a practice corrected "wearable computer off" trial to compare with the "wearable computer on" trial (trial 2).

Before beginning the second trial for each physical task, the experimenter verbally issued the following instructions, "In this trial, you will be asked to operate the wearable computer while performing the [Assembly/Digit/Walking] task at the same time. Because you are performing two tasks, you should concentrate primarily on completing the computer benchmark tasks as quickly and accurately as possible. Thus, if your performance on either of the two tasks must suffer, it should be the [physical] task."

For all assembly task trials, subjects sat at a table with a dish containing an assortment of Lego™ blocks separated by size and color. (Figure 3). The dish contained far more blocks than the assembly task actually required. Prior to beginning, the subject was given a block assembly to study for sixty seconds. The block assembly was comprised of seventeen random blocks arranged in semi-random order on a flat Lego™ base. The one restriction on the randomization of the block arrangement was that there could be no space on the base unoccupied by a block. Pilot work was used to determine the appropriate size of the base and the length of exposure time such that ceiling and floor effects could be avoided. Following the sixty second exposure period, the experimenter retrieved the block assembly and informed the subject assemble a replica of the block assembly using the blocks in the dish.

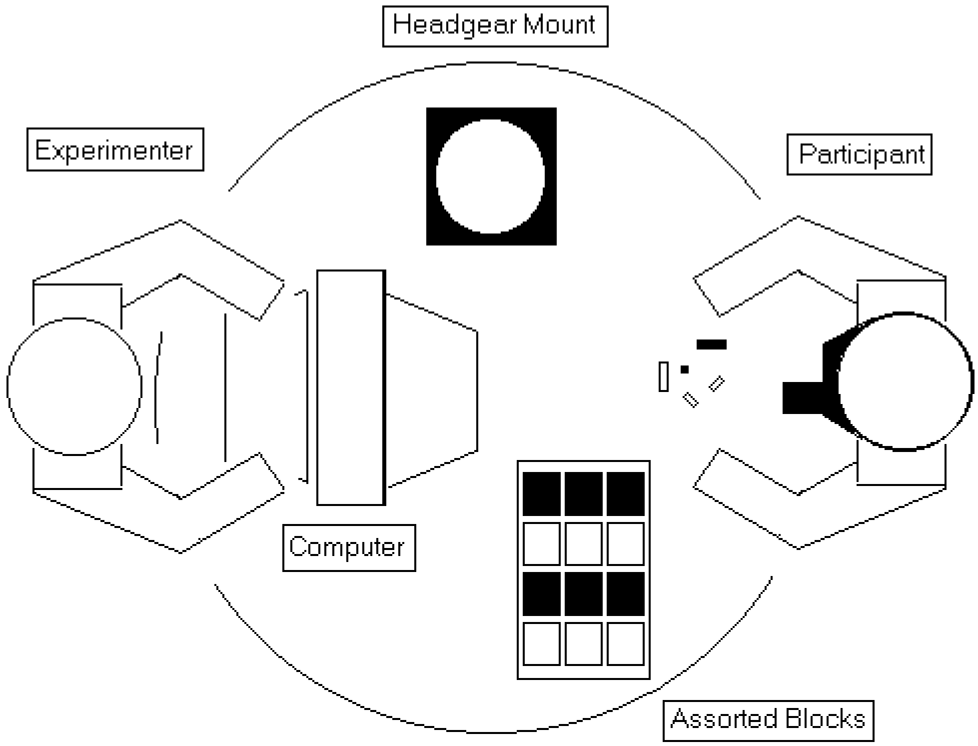

Figure 3. Experimental Setup During Block Assembly Trials

For the first and third trials of the assembly task, the experimenter started the trial by starting a timer and signaling the subject to begin. The subject was instructed to continue assembling using the blocks until they “felt like they were guessing.” These trials ended either when the subject completed the assembly or when subjects informed the experimenter they could not continue the assembly without guessing.

For the second trial of the assembly task, the test participant began the trial by speaking the command, “Begin” into the microphone, which displayed the first computer benchmark task on the HMD, subjects began completing both the set of computer benchmark tasks and the assembly task. Once subjects finished the first computer benchmark task, they displayed succeeding benchmark tasks using the “Next Task” command. The second trial ended after the subject spoke the final command to complete the fifth computer benchmark task. Subjects were not allowed to make any changes or additions to their assembly once the trial ended.

For all trials of the digit task, subjects sat at a table with a pencil and a blank answer sheet. The experimenter began the trial by playing an auditory tape recording of random single digits presented at a rate of 1 every 4 seconds. Although good stimulus response compatibility and consistency with previous researchers (Aretz, 1996; Kantowitz and Knight, 1976) would have been maintained had subjects provided their answers vocally, this practice did not match well with the speech recognition software running on the wearable computer. Therefore, as the recording played, subjects subtracted the current digit from the previous digit and wrote the difference on a paper answer sheet. In order to ensure that subjects did not memorize digits from one trial to the next, a different set of random single digits was played at every trial. After each trial ended, the experimenter collected the answer sheet from the subject. To conserve time, the first and third trials were terminated after the subject subtracted 30 digits (120 seconds).

For the second trial of the digit task, subjects began the trial by speaking the command, "Begin" into the microphone. Subjects then proceeded to perform the computer benchmark tasks while writing the subtracted digits on the answer sheet. Just as with the assembly task, the trial ended after the subject spoke the final command to complete the fifth computer benchmark task.

All trials of the walking task took place within a laboratory hallway. A 150 foot long pathway with 6 different 90 degree turns was marked with masking tape on the hallway floor. Subjects were positioned at one end of the pathway while the experimenter stood off to the side.

For the first and third trials, the experimenter began the trial by starting a timer, which signaled subjects to begin walking along the pathway. The trial ended once subjects successfully walked to the opposite end of the pathway and returned to the starting point. For the second trial, subjects began the trial by speaking the command, "Begin" into the microphone. Subjects then proceeded to perform the computer benchmark tasks while continuously walking back and forth along the pathway. Just as with the assembly task, the application terminated after the subject spoke the final command to complete the fifth computer benchmark task, thus signaling the end of the trial. Subjects were asked to remain standing at the point when the application terminated until the total length traversed by the subject could be recorded.

If subjects using the command line version happened to forget any commands during any condition, the experimenter noted the error and provided the command. In addition, if at any point the speech recognition software failed to produce the correct command issued by the subject, the experimenter could manually provide the intended command using a keyboard. This was done so that the attention of subjects would remain fixed on the tasks at hand and not on the operating status of the speech recognition software.

Once subjects completed all nine trials, they were asked to complete the NASA Task Load Index (TLX) three separate times, once for each trial in which they had to perform a physical task and a computer task. Subjects were instructed to answer the NASA TLX items in the context of how much mental workload was required to perform *both* the computer task and the physical task simultaneously. Following completion of the three NASA TLX questionnaires, subjects completed the usability questionnaire, which completed their participation in the experiment.

3.5. Statistical Procedures

In order to detect any differences in physical task performance across the different versions, separate one-way between subjects ANOVAs were performed across the five levels of the version factor (command line, text menus, icon text menus, buttons, icon buttons) for each of the physical task measures. For each of the three administrations of the NASA TLX, a separate one-way ANOVA was performed on the overall mental workload score as well as the six subscales across the levels of the version factor. Post hoc comparisons were made using the least significant difference (LSD) method. Differences on each of the physical task measures between the two levels of the wearable computer status factor (wearable computer on, wearable computer off) were analyzed separately for each version using two-tailed t-tests for related samples. Finally, for each usability questionnaire item, separate Kruskal Wallace One-Way ANOVAs were used to detect any differences in ratings across the five versions. The alpha level for all statistical tests was set to 0.05.

4. RESULTS

4.1. Assembly Task Measures

Means for each of the physical task measures are provided across the experimental conditions in the figures below.

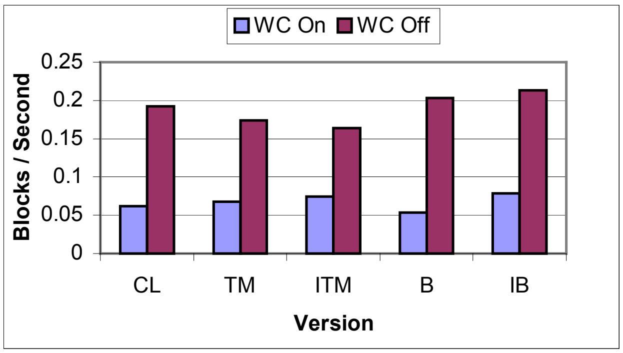

Figure 4. Mean Assembly Rate Across Version and Wearable Computer Status

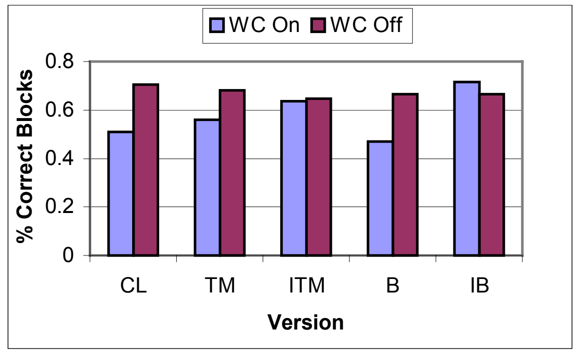

Figure 5. Mean Percentage Correct Blocks Across Version and Wearable Computer Status

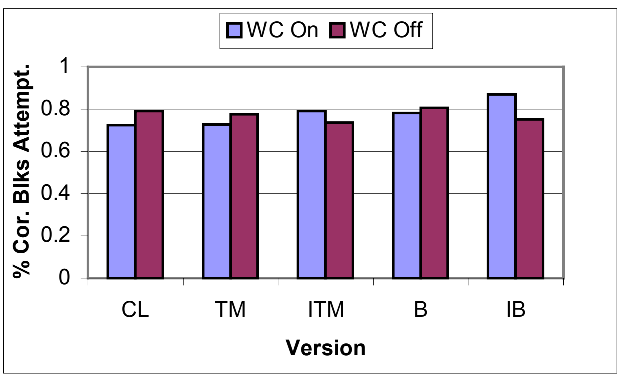

Figure 6. Mean Percentage Correct Blocks Attempted Across Version and Wearable Computer Status

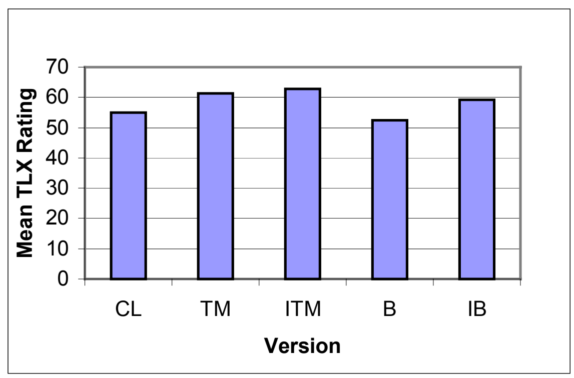

Figure 7. Mean Weighted Workload Rating (NASA TLX) Across Version for the Assembly Task

The results generally showed trends toward increased task performance when the icon button version was used, indicating the highest average assembly rate (0.079 blocks per second), the highest percentage correct blocks (71.6 %), and the highest percentage correct blocks attempted (86.9%). The button version indicated the lowest average assembly rate (0.053 blocks per second) and lowest percentage correct blocks (47%) while the command line version indicated the lowest percentage correct blocks attempted (72.4%). However, results from one way ANOVAs performed respectively on average assembly rate ($F_{4,\ 25}$ =

0.94, $p = 0.46$), percentage correct blocks ($F_{4, 25} = 0.71$, $p = 0.59$) and percentage correct blocks attempted ($F_{4, 25} = 0.27$, $p = 0.89$) across the 5 levels of the Version factor failed to indicate any significant main effect.

Results also provided evidence of better physical task performance when the wearable computer was turned off compared to when it was on, regardless of which version was operated. This trend was most evident for the average assembly rate measure, which when averaged across all versions indicated 2.8 times faster assembly rates when the wearable computer was turned off compared to when it was turned on. Separate two-tailed t-tests for related samples performed for each version indicated significantly faster average assembly rates when the wearable computer was turned off compared to when it was turned on for the buttons ($t_5 = -4.51$, $p = 0.006$), command line ($t_5 = -8.42$, $p = 0.0004$), icon buttons ($t_5 = -5.47$, $p = 0.003$), icon text ($t_5 = -4.27$, $p = 0.008$), and text menu version ($t_5 = -17.56$ $p < 0.0001$). Since all versions showed significant differences, an additional one way between subjects ANOVA was performed across versions to determine whether any particular version showed a more drastic difference in average assembly rate between wearable computer on and off conditions. The result was not significant ($F_{4, 25} = 1.17$, $p = 0.35$).

For the percentage correct blocks measure, the trend toward better performance with the wearable computer off was not as evident. For each of the versions, two tailed t-tests for related samples performed on percentage correct blocks between the two wearable computer status levels indicated no statistically significant difference for buttons ($t_5 = -1.64$, $p = 0.16$), command line ($t_5 = -1.63$, $p = 0.16$), icon button ($t_5 = 0.71$, $p = 0.51$), icon text ($t_5 = -0.09$, $p = 0.93$), or text menu ($t_5 = -1.33$ $p = 0.24$). A one way between subjects ANOVA performed on the difference between the wearable computer on and off measures across all versions failed to indicate any significant differences ($F_{4, 25} = 1.17$, $p = 0.35$).

The same trends found for the percentage correct blocks measure were also found for the percentage correct blocks attempted measure. Two tailed t-tests for related samples performed on percentage correct blocks attempted between the two wearable computer status levels indicated no statistically significant difference for buttons ($t_5 = -0.32$, $p = 0.75$), command line ($t_5 = -0.38$, $p = 0.72$), icon button ($t_5 = 1.7$, $p = 0.15$), icon text ($t_5 = 0.75$, $p = 0.49$), or text menu ($t_5 = -0.67$ $p = 0.53$). A one way between subjects ANOVA performed on the difference between the wearable computer on and off measures across all versions failed to indicate any significant differences ($F_{4, 25} = 0.57$, $p = 0.69$).

The overall weighted workload ratings provided by the NASA TLX were analyzed using a one-way between subjects ANOVA across the five levels of the version factor. No significant main effect was found ($F_{4, 25} = 0.66$, $p = 0.62$). Separate ANOVAs performed on each of the NASA TLX subscales across the levels of the version factor indicated a significant main effect for the physical demand subscale ($F_{4, 25} = 4.52$, $p = 0.007$). Post hoc comparisons performed using the least significant difference method indicated that subjects rated the command line version as requiring a significantly greater physical demand than each of the other versions (least significant difference = 20.99). A significant main effect was also found for the effort subscale ($F_{4, 25} = 2.91$, $p = 0.04$). Post hoc comparisons indicated that subjects rated the icon button version as requiring significantly less effort than the text menu and icon text menu versions (least significant difference = 20.41). No main effects were indicated for the mental demand ($F_{4, 25} = 1.12$, $p = 0.37$), temporal demand ($F_{4, 25} = 0.07$, $p = 0.99$), performance ($F_{4, 25} = 0.66$, $p = 0.62$), and frustration ($F_{4, 25} = 1.15$ $p = 0.36$) subscales.

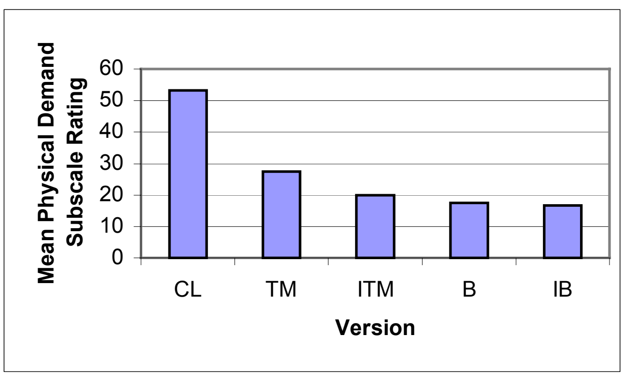

Figure 8. Mean Physical Demand Subscale Rating Across Version for the Assembly Task

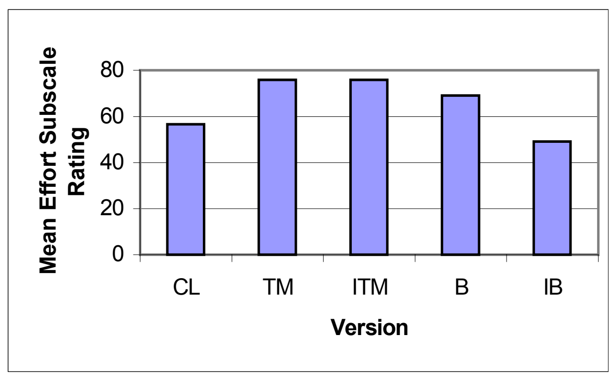

Figure 9. Mean Effort Subscale Rating Across Version for the Assembly Task

Table 6 lists the correlations between each of the task performance metrics recorded during the second trial with subjective measures of mental workload (NASA TLX ratings). Of the three task performance measures, the percentage correct blocks attempted measure provided the highest correlation with the overall workload ratings ($r = -0.33$), indicating a trend towards higher ratings of mental workload when fewer blocks were assembled ($t_{28,\ two\text{-}tailed} = -1.83$, $p = 0.078$). The correlation between percentage correct blocks attempted measure and the frustration subscale (r = -0.32) also indicating a trend towards higher ratings of frustration when fewer blocks were assembled ($t_{28, \text{two-tailed}} = -1.79$, p = 0.085).

Table 6. Correlations Among Assembly Task Performance Measures and Mental Workload Ratings (NASA-TLX)

		Assembly Task Performance Measure		
		Average Assembly Rate	Percentage Correct Blocks	Percentage Correct Blocks Attempted
Mental Workload Rating	Overall Workload	0.08	-0.11	-0.33
	Mental Demand	-0.12	-0.20	-0.21
	Physical Demand	0.01	-0.02	0.10
	Temporal Demand	-0.08	-0.26	-0.25
	Effort	-0.07	-0.08	-0.12
	Performance	0.00	0.03	-0.11
	Frustration	-0.02	-0.17	-0.32
* Significant at the 0.05 level				

4.2. Digit Subtraction Measures

A summary of the mean values for the percentage correct digits measures across version and wearable computer status is provided in Figure 10. Like the assembly task performance measures, subjects using the icon button version recorded a higher average percentage of digits (62%) than subjects using any of the other versions. However, a one-way between subjects ANOVA performed on the percentage correct digits measure across the 5 levels of the Version factor indicated no significant main effect ($F_{4, 25} = 0.85$, p = 0.51).

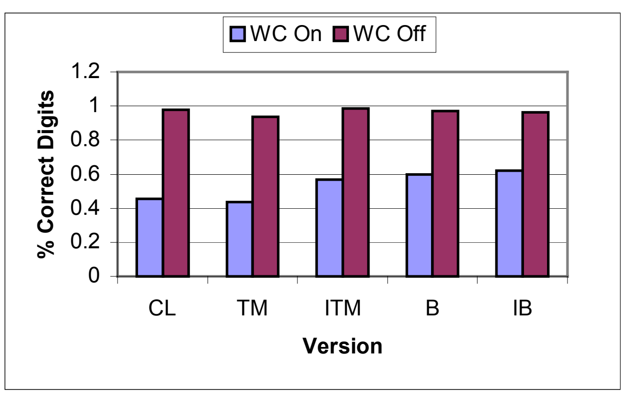

Figure 10. Mean Percentage Correct Digits Across Version and Wearable Computer Status

Separate two-tailed t-tests for related samples performed on percentage correct digits between the two wearable computer status levels indicated statistically significant differences for buttons (t_5 = -3.92, p = 0.01), command line (t_5 = -5.85, p = 0.002), icon buttons (t_5 = -5.97, p = 0.002), icon text (t_5 = -4.19, p = 0.009), and text menu (t_5 = -9.87 p = 0.0002). However, a one way between subjects ANOVA performed on the difference between the wearable computer on and off measures across all versions failed to indicate any significant differences ($F_{4, 25}$ = 0.94, p = 0.46).

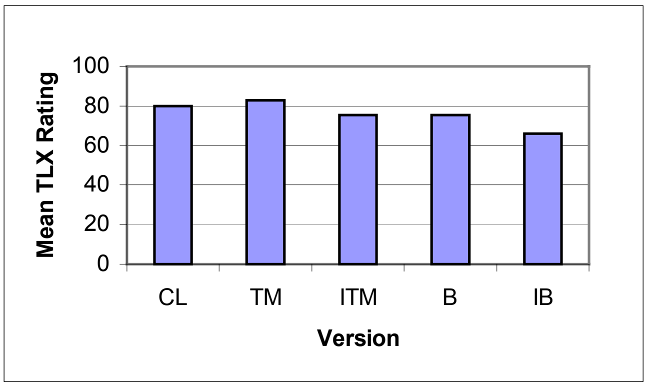

Figure 11. Mean Weighted Workload (NASA TLX) Rating Across Version for the Digit Task

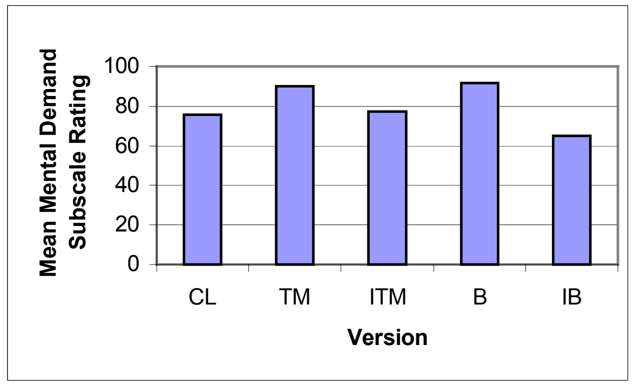

Figure 12. Mean Mental Demand Rating Across Version for the Digit Task

The NASA-TLX ratings were analyzed using a one-way between subjects ANOVA across the levels of the version factor. No significant main effect was found for the overall weighted workload rating ($F_{4, 25}$ = 1.53, p > 0.05). Separate one way between subjects ANOVAs performed on each of the NASA TLX subscale ratings across the levels of the version factor a indicated significant main effect for the mental demand subscale ($F_{4, 25}$ = 2.91, p = 0.04). Post hoc comparisons (Least significant difference = 18.78) indicated that subjects mean mental demand rating for the icon button version (Mean rating = 65) was significantly less than rating for the text menu version (Mean rating = 90) and the button version (Mean rating = 91.7). No main effect was found for the physical demand ($F_{4, 25}$ = 1.28, p = 0.31), temporal demand ($F_{4, 25}$ = 1.18, p = 0.34), effort ($F_{4, 25}$ = 2.21, p = 0.10), performance ($F_{4, 25}$ = 2.11, p = 0.11), or frustration ($F_{4, 25}$ = 2.07 p = 0.12) subscales.

Table 7. Correlations Among Digit Task Performance Measures and Subjective Mental Workload Ratings (NASA-TLX)

		Digit Task Performance Measure
		Percentage Correct Digits
Mental Workload Rating	Overall Workload	-0.22
	Mental Demand	0.12
	Physical Demand	-0.16
	Temporal Demand	-0.12
	Effort	0.03
	Performance	*-0.46
	Frustration	-0.35
* Significant at the 0.01 level		

Table 7 lists the correlations between the percentage correct digits recorded during the second trial and the subjective measures of mental workload (NASA TLX ratings). While a correlation of –0.22 between the percentage correct digits and the overall workload ratings indicated a trend towards higher ratings of mental workload when a fewer percentage of digits was recorded, linear regression failed to detect a significant trend ($t_{28} = -1.19$, $p = 0.24$). However, the percentage correct digits variable was significantly correlated ($r = -0.46$) with the performance subscale ($t_{28} = -2.74$, $p = 0.01$) and showed a trend towards being significantly correlated ($r = -0.35$) with the frustration subscale ($t_{28} = -1.98$, $p = 0.057$).

Path Walking Measures

A summary of the mean average walking speed across version and wearable computer status is provided in Figure 13. Though by a slight margin, subjects using the icon button version recorded a higher average walking speed (1.84 feet per second) than subjects using any of the other versions (note that the icon text menu version was slower by a margin of only 0.06 less feet per second). A one way between subjects ANOVA was performed on average walking speed across the 5 levels of the version factor indicated no significant main effect ($F_{4, 25} = 0.71$, $p = 0.59$).

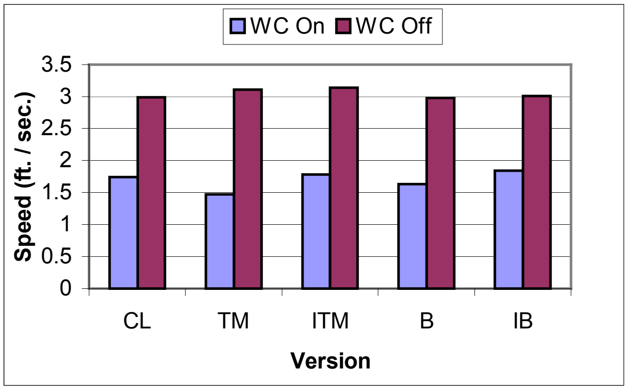

Figure 13. Mean Average Walking Speed Across Version and Wearable Computer Status

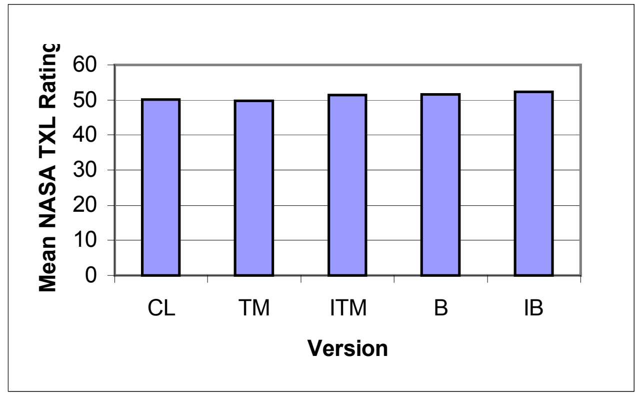

Figure 14. Mean Weighted Workload (NASA TLX) Rating Across Version for the Path Walking Task

Two tailed t-tests for related samples performed on average walking speed between the two wearable computer status levels indicated significantly faster average walking speed when the wearable computer was turned off compared to when it was turned on for buttons (t_5 = -3.22, p = 0.02), command line (t_5 = -10.8, p = 0.00006), icon button (t_5 = -3.22, p = 0.02), icon text (t_5 = -14.5, p = 0.00001), and text menu (t_5 = -7.84 p = 0.0005). Since all versions showed significant differences, an additional one way between subjects ANOVA was performed across versions to determine whether any particular version showed a more drastic difference in average assembly rate between wearable computer on and off conditions. The result was not significant ($F_{4, 25}$ = 0.71, p = 0.59).

The NASA-TLX ratings were analyzed using a one way between subjects ANOVA across the 5 levels of the Version factor. No significant main effect was found for the overall weighted workload rating ($F_{4, 25}$ = 0.02, p = 0.99). Separate ANOVAs performed on each of the NASA TLX subscale ratings across the levels of the version factor indicated no significant main effect for the mental demand ($F_{4, 25}$ = 0.77, p = 0.55), physical demand ($F_{4, 25}$ = 0.66, p = 0.62), temporal demand ($F_{4, 25}$ = 0.46, p = 0.76), effort ($F_{4, 25}$ = 0.59, p = 0.67), performance ($F_{4, 25}$ = 0.31, p = 0.87), or frustration ($F_{4, 25}$ = 0.28 p = 0.89) subscales.

Table 8 lists the correlations between average walking speed recorded during the second trial and the subjective measures of mental workload (NASA TLX ratings). Although the correlation of –0.15 between average walking speed and the overall workload rating indicated a slight trend towards slower walking speeds when mental workload was rated higher, linear regression failed to detect any significant trend ($t_{28} = -0.81$, $p = 0.42$). However, average walking speed did show trends towards being positively correlated ($r = 0.33$) with the physical demand ($t_{28} = 1.85$, $p = 0.075$) and negatively correlated ($r = -0.32$) effort subscales ($t_{28} = -1.81$, $p = 0.082$).

Table 8. Correlations Among Path Walking Task Performance Measures and Subjective Mental Workload Ratings (NASA-TLX)

		Path Walking Task Performance Measure
		Average Walking Speed
Mental Workload Rating	Overall Workload Rating	-0.15
	Mental Demand	-0.06
	Physical Demand	0.33
	Temporal Demand	-0.11
	Effort	-0.32
	Performance	-0.18
	Frustration	-0.28
* Significant at the 0.05 level		

4.3. Relative Task Performance

In order to provide an additional indicator of the relative decrease in physical task performance when operating each version, task performance data collected during the wearable computer on condition was divided by task performance data during the wearable computer off condition and multiplied by 100 (Table 9). For most cases, percentage ratios were under 100%, indicating decreased task performance when the wearable computer was operated. However, in a minority of cases the ratios were greater than 100%, indicating that physical task performance was actually better when the wearable computer was operated relative to when it was turned off. For four of the five physical task performance metrics, the icon button version showed the highest ratio, indicating that subjects using this version performed the task best relative to when no wearable computer was operated.

Table 9. Task Performance Ratios Across Version

Task Performance Measure		Version				
		Buttons	Command Line	Icon Buttons	Icon Text Menus	Text Menus
	Average Assembly Rate	**26 %**	32 %	37 %	**45 %**	39 %
	Perc. Correct Blocks	**71 %**	72 %	**107 %**	98%	82%
	Perc. Correct Blocks Attempted	97%	**91%**	**116 %**	107 %	94 %
	Perc. Correct Digits	62 %	**46 %**	**79 %**	73 %	47 %
	Average Walking Speed	55 %	58 %	**61 %**	57 %	**47 %**

Note: Highest and lowest values for each row appear in bold

4.4. Usability Measures

Differences in questionnaire item ratings were analyzed across the five different versions using Kruskal Wallace one-way ANOVAs. A summary of the results is included in Table 10. Only the item, "How quickly do you feel you were able to perform operations on the computer software application?" was answered significantly differently across the different versions. Specifically, subjects indicated that they could perform operations more quickly using the text menu version than the command line version ($H_{4df} = 10.47$, $p < 0.05$).

Table 10. Kruskal Wallace Values and Mean Usability Ratings Across Questionnaire Items

	Version					Kruskal Wallace Value
	Buttons	Command Line	Icon Buttons	Icon Text Menus	Text Menus	
How quickly do you feel that you learned the computer operations during training? 1= Not at All Quickly 7= Extremely Quickly	5	5	5.5	5.7	5.8	2.47
How much effort do you feel you expended learning the computer interface during training? 1= No Effort 7= Extreme Effort	3.2	3.3	3.2	3.7	3.2	1
How quickly do you feel you were able to perform operations on the computer software application? 1= Not at All Quickly 7= Extremely Quickly	4.5	3	4	4.8	5.2	*10.47
Generally speaking, how error prone was the wearable computer system? 1= Not Error Prone 7= Extremely Error Prone	3.5	5.3	4.3	4.2	3.7	4.51
How would you rate the software applications overall ease of use? 1= Not at All Easy to Use 7= Extremely Easy to Use	5.7	4.7	5.7	4.3	5	4.61
How easily do you feel that you could remember the meaning of specific items? 1= Not at All Easily 7= Extremely Easily	5.7	5	6	5.5	5.3	2.75
How much did you enjoy using the software application? 1= Not at All 7= Extreme Enjoyment	4.7	5.2	4.3	4	5.3	2.65

*Significant at the 0.05 level

5. DISCUSSION

The objective of this thesis was to test hypotheses concerning the type and amount of mental demands required by several versions of an experimental software for use with a wearable computer. Mental demand was assessed using not only a standard subjective measure (NASA-TLX), but also the performance of nonstandard secondary tasks. It was hoped that by employing these measures on several different versions of a single wearable computer interface, a specific version less demanding of mental resources could be determined. Versions found to require fewer mental demands could then be recommended for design based on their propensity for allowing users to effectively perform physical tasks simultaneously. In addition, it was hoped that by using three different physical tasks as secondary task measures, either specific pairings between versions and tasks showing effective performance could be determined, or a single version showing consistent effective performance across all the physical tasks could be determined.

This discussion will begin by summarizing the more general findings concerning subjects' physical task performance when using each of the experimental versions relative to performing the physical task by itself. Next, the assembly task, digit subtraction task, and path walking task will each be discussed in terms of their aptitude as a measure of mental workload based on the predictions made by mental resource theory. The discussion will follow with design recommendations for wearable computer interfaces based on the experimental results. The discussion will conclude with recommendations for future research.

There was strong statistical evidence to suggest that using wearable computers decreases the performance of physical tasks regardless of which version was used. On the assembly task, subjects assembled blocks at slower rates and used fewer correct blocks in their assemblies when they also operated the wearable computer. On the digit subtraction task, subjects wrote down lower percentages of correct digits when the wearable computer was operated. Finally, subjects walked at slower speeds on the path walking task when they also operated the wearable computer. These results were expected given previous findings of decreased performance on physical tasks when wearable computers are used. For example, Siegel and Bauer (1997) found that subjects generally performed a manual aircraft engine assembly task 1.5 times more slowly when a wearable computer was used for instruction rather than a paper-based manual. Ockerman and Pritchett (1998) observed that subjects performed an aircraft inspection task more slowly when a wearable computer was used for instruction rather than a paper based manual.

While performance on all physical tasks was worse when subjects operated the wearable computer, there was evidence to suggest that accuracy was less affected by the use of wearable computers than the time dependent measures. This was most apparent for the assembly task analysis, which detected no difference in the percentage correct blocks or percentage correct blocks attempted variables when the wearable computer turned on or off. Moreover, there were isolated cases where assembly task performance accuracy was slightly better when the wearable computer was turned on (such as when task performance ratios exceeded 100% in Table 9). Other researchers have reported similar anecdotal findings of high accuracy on physical tasks when wearable computers are used. For instance, Ockerman and Pritchett (1998) reported that while subjects performed aircraft inspection tasks more slowly when wearable computers were used, the inspections were more thorough. And Siegel and Bauer (1997) reported that subjects performing an aircraft maintenance tasks were more thorough when they used a wearable computer system when compared to using a paper based system.

While it was expected that subjects would generally perform worse on all physical tasks when the wearable computer was operated, one goal of the analysis was to determine whether a particular interface afforded superior physical task performance. It was expected that if a single version consistently afforded better physical task performance than the others, then there would be a larger difference in performance between the wearable computer on and wearable computer off conditions for this version. However, when these differences were analyzed across versions, no version showed any smaller or larger differences than the other versions.

While statistical tests generally failed to detect any differences in physical task performance among the versions, there did appear to be trends suggesting that the icon button version afforded better physical task performance than other versions. Inspection of the dependent measures for the assembly task revealed that subjects assembled blocks at the fastest average rate (Figure 4), used the greatest percentage of correct blocks (Figure 5), and used the greatest number of correct blocks attempted (Figure 6) when they used the icon button version. On the digit subtraction task, subjects recorded the highest average percentages of correct digits when the icon button version was used (Figure 10). And on the path walking task, subjects using the icon button version walked at the fastest average speeds (Figure 13). In addition, for all measures except for the assembly rate measure, subjects using the icon button version showed the highest proportions of task performance relative to the wearable computer off trials (Table 9).

While the superiority of the icon button version over the other versions was occasionally slight, the consistency in finding superior performance across the three physical tasks supports the notion that using icon button interfaces may be most beneficial when several different physical tasks must also be performed. This finding was not expected given the context of the multiple resource theory, which led to hypotheses that physical task performance would be contingent on which version was used. The following sections discusses the physical tasks and their use as measures of mental workload in greater detail.

5.1. Assembly Task

Prior to performing the experiment, the icon button version was hypothesized to be more demanding in spatial mental resources while the command line version was hypothesized to be more demanding in verbal mental resources. Based on the assumption that the assembly task would be high in spatial mental resources demands, it was predicted from the multiple resource view that mental workload would be highest for this task when the icon button version was used and lowest when the command line version was used (Figure 1

). These predictions were not supported by the data. In terms of the overall mental workload ratings provided by the NASA TLX scores, subjects did not rate any of the versions differently from one another. And differences in TLX subscale ratings among the different versions occurred opposite to what was predicted by the multiple resource view. For instance, subjects rated the task higher on the physical demand subscale when using the command line version than when any of the other versions were used (Figure 8). And subjects rated the task lower on the effort subscale when the icon button version was used than when several of the other versions were used (Figure 9). Like the subjective ratings, analyses performed on the secondary task performance measures not only failed to detect any differences in mental workload across the versions, but any observable trends appeared opposite to what was predicted by the multiple resource view. Subjects using the icon button version tended to be slightly faster (Figure 4) and more accurate (Figure 5 & Figure 6) at the assembly task than when any of the other versions were used.

The failure of both the subjective and task performance data in supporting the predictions made from the multiple resource view may have occurred for several reasons. First, the versions may not have been as differentially demanding in spatial mental resources as originally expected (Figure 1 provides a proposed graph showing the different versions in relation to their predicted demands in spatial resources). Other than the use of visual icons and the dispersion of the controls about the user's screen, the basis for determining that using the icon button version would demand greater spatial resources than the other versions was not particularly strong. At the other extreme of the graph, although subjects using the command line version did not have to visually attend to the screen to look for available commands, they frequently attended to the screen after issuing commands to confirm that the commands had been executed successfully. This frequent visual attendance may have resulted in the command line version demanding greater spatial resources than originally hypothesized. If these more extreme versions within the proposed graph were not different enough in spatial mental resources, then not only should no differences have been expected between them, but no differences should have been expected among any of the less extreme versions on the graph.

Second, it was originally assumed that the assembly task was a spatially demanding task. However, it is possible that the task was not as demanding in spatial mental resources as previously thought. The only initial basis for determining that the assembly task was highly demanding of spatial resources was that successful completion of the task involved recalling the spatial positions of individually colored and shaped blocks and replicating these positions. No prior research compared the assembly task with other more well known spatially demanding tasks for validation. If the assembly task was in fact not as spatially demanding as hypothesized, then any differences in spatial demands across the versions would not be detectable using the assembly task.

Finally, failure to lend support to the predictions made by the multiple resource view may have been related to the methods used in the experiment. The particular measures employed (assembly rate, percentage correct blocks, percentage correct blocks attempted) may not have been sensitive enough measures to detect differences in spatial resource demands among the version types. In addition, the small sample size (n=6) used might have contributed to larger error terms than what might be expected for larger sample sizes, leading to possible Type II errors.

It is important to note that the subjective ratings of mental workload were not generally well correlated with the performance data (Table 6), suggesting that any differences found in task performance among the different versions were not likely attributable to differences in mental workload. Alternative explanations may then be needed to interpret such differences. One possible reason why subjects performed slightly better and rated the task lower on the TLX effort subscale with the icon button version may be related to the advantages for icons with desktop systems. As researchers have shown that icons can communicate information more quickly than text (Barfield, Rosenburg, and Levasseur, 1991; Arend, Muthig, and Wandmacher, 1987), it is possible that subjects using the icon button version may have recognized needed commands more quickly than subjects using the other versions. Faster communication of information on the interface might then have allowed subjects to attend more readily to the assembly task, resulting in better assembly task performance with less reported effort. In addition, though computer expertise was not formally tracked in this experiment, subjects may have been more familiar with applications similar in appearance and function to the icon button version than with the communication styles of the other versions. Higher familiarity with the icon button version may then have allowed subjects to concentrate less on the interface and more on the nuances of the assembly task.

The higher ratings recorded for the physical demand subscale (Figure 8) when the command line version was used are difficult to interpret, especially given the lack of any complimentary difference in other subscale ratings, average assembly rate (Figure 4), percentage correct blocks (Figure 5), or percentage correct blocks attempted (Figure 6) when the command line version was used. However, based on these data, it would appear that using the command line version caused subjects to perceive that the act of assembling the blocks and operating the wearable computer as more physically laborious than subjects using the other versions. Future research is needed to determine the validity and causality of this finding.

5.2. Digit Subtraction Task

The digit subtraction task was included in this experiment based on its prior use by previous researchers as a secondary task more demanding in verbal resources (Aretz, Johannsen, and Ober, 1996; Kantowitz and Knight, 1976). Because operating the icon button version was hypothesized to be more demanding in spatial mental resources demands while operating the command line version was hypothesized to be more demanding in verbal mental resources, it was predicted that mental workload would be highest for this task when the command line version was used and lowest when the icon button version was used (Figure 1). These results were partially supported by both the subjective and performance data. In terms of subjective ratings, subjects rated the task as lower on the TLX mental demand subscale when the icon button version was used. Although subjects did not rate the task as higher on the mental demand subscale when the command line version was used (Figure 12), there was a statistically non-significant trend suggesting higher overall mental workload ratings when the command line version was used (Figure 11). In terms of the physical task performance measure, the results showed trends indicating larger percentages of correctly subtracted digits when the icon button version was used and smaller percentages when the command line version was used (Figure 10).

Although the correlation between the percentage correct digits and the overall mental workload ratings was not statistically significant ($r = -0.22$), both measures generally supported the hypothesis that the icon button version was less demanding of verbal mental resources while the command line version was more demanding of verbal mental resources. One possible reason for why the icon button version may have been less demanding of verbal resources was the spatial placement of icons on the interface. By visually communicating commands to subjects using pictures and spatial positioning, searching for commands may not have been as demanding of verbal resources. Consequently, subjects using the icon button version would have more verbal resources available for performing the digit subtraction task, resulting in the better performance. In contrast, subjects using the command line version had to recall from memory each of the verbal commands required to operate the interface. This exercise may have possibly resulted in a higher overhead in verbal mental resources, resulting in a shortage of verbal resources to allocate to the digit task. As further anecdotal evidence to this trend, subjects appeared to have the most difficulty with the digit task when they also had to issue number commands on the command line interface (such as when updating a phone listing). Consequently, the cause of the suffered performance may have been related to both tasks exceeding subjects' available verbal resources.

5.3. Path Walking Measures

Like the assembly task, the path walking task was also assumed to be high in spatial mental demands. Under the hypothesis that the icon button version was the most demanding in spatial mental resources, the highest mental workload was predicted for this version. Lower mental workload was predicted for the command line version under the hypothesis that this version was more demanding in verbal mental resources. Similarly to the results found with the assembly task, these predictions were not supported by either the subjective or task performance data. No differences in mental workload were found for either the overall workload (Figure 14) or subscale workload measures. In terms of task performance, no differences were detected in walking speed across the versions (Figure 13).

While the lack of difference in ratings and task performance among the versions was not predicted, the consistency of these findings with those from the assembly task supports the case that the versions were not different in terms of spatial resource demand. Like the assembly task performance measures, it is possible that the walking speed measure may not have been a particularly sensitive to differences in spatial demand. However, any differences in mental workload resulting from increased spatial demand should still have been detected using the NASA TLX. Since no differences were found using the NASA TLX measure for either the assembly task condition or path walking conditions, it may be more likely that the lack of differences found across the versions was attributable to a lack of difference in spatial demand among them.

5.4. Usability Measures

Although not statistically significant, subjects rated the button and icon button versions as highest in "overall ease of use" (Table 10). This finding may have occurred for a number of reasons. It is possible that subjects found the button and icon button versions easier to use over the text menu and icon text menu versions because the former only required a single input to execute a function. For instance, in order to move the cursor right one space on the calendar screen, subjects using the button and icon button versions only issued the verbal command, “Right.” Subjects making the same move using the text menu and icon text menu versions issued one command to highlight the main menu item, “Move”, plus an additional command to select within the main menu item list, “Left.”

Subjects may have also found the icon button and button versions easier to use because all commands required to operate the versions appeared directly on top or below the button icon, allowing subjects to issue commands by recognition. Subjects using the command line version, in contrast, had to recall all commands from memory in order to execute them. Within the context of performing a physical task while operating the wearable computer, this difference may have been even more influential in their ease of use ratings.

Finally, subjects may have had the more familiarity with other software using buttons and icon buttons than text menus and command lines. While subjects’ experience with each type of communication style was not verified, the popularity of commercial end user software products using toolbars with icon buttons to perform frequent and/or routine tasks (e.g., choosing the ‘print’ icon button on the toolbar to print a document rather than choosing “File” and “Print) is unquestionable. Such experience with icon buttons may have contributed to the higher ease of use ratings.

5.5. Conclusions

For each of the physical tasks used in this design, task performance was best for those subjects who used the icon buttons version, though the differences were not always statistically significant. These results are contrary suggestions that graphical user interfaces optimized for desktop use may not be applicable to the wearable platform (Billinghurst, Bowskill, Dyer, and Morphett, 1998), results from the present experiment suggest that using the more graphical user interfaces lead to improved task performance and slightly better ratings of usability. And while Thompson, Najjar, and Ockerman (1997) suggest that the usability of voice input on wearable computers may be improved by using simple applications which are command based, the present findings indicate that voice input could be used effectively with a more graphical user interface provided that the interface does not require excessive navigation to perform an operation.

There was a general failure to reject null hypotheses that there were no differences in spatial mental demands across the versions. This failure was supported by the lack of differences in task performance and NASA TLX ratings found across versions for both the assembly and path walking task measures. It should be noted that there had been no previous research to confirm the assembly and path walking tasks as valid, reliable measures of spatial mental demands. However, there was general evidence of differences in verbal demands across the versions. Using a more classic secondary task measure (digit subtraction performance) the trends indicated that the icon button version demand fewer verbal resources than the other versions. Furthermore, overall mental workload ratings as well as mental demand subscale ratings provided by the NASA TLX suggested that the icon button version was less mentally demanding to operate.

In summary, this thesis research provided a means for determining which communication styles to implement on the wearable computer platform so that physical tasks can be effectively shared with wearable computer operation. In addition, this research used multiple measures to determine the different types of mental resources demanded by specific wearable computer application designs. It is hoped that as humans begin using wearable computer technology with the goal of improving the efficiency and effectiveness of all types of tasks, the findings of this research will allow wearable computer designers to determine the interface most appropriate for attaining this goal.

5.6. Future Research

The present study generally showed improved physical task performance when using icon button interfaces on a wearable computer platform. However, further research could be performed on the appearance of the interface to isolate the qualities most attributable to performance improvement. Such research should more systematically compare different spatial locations of icon buttons to determine the types of spatial arrangements which best communicate with the user. Further research could also determine the optimal number of hierarchical levels an icon button interface can contain before it loses its task performance advantages.

The present study used only speech input and visual output to determine which type of communication style maximized physical task performance and minimized mental workload. However, this particular pairing may not be best suited for multiple task performance with wearable computers. Future research should compare multiple types of input/output pairings to determine the most compatible for enhancing task performance and minimizing mental workload.

While this study attempted to measure mental workload using more practical, real world tasks as secondary tasks, there was no empirical basis for assuming their validity as measures of mental workload. While the present study attempted to use subjective mental workload ratings to verify their use as mental workload measures, future research should attempt to further validate the findings here using more classic secondary tasks such as rhythmic tapping, random number generation, probe reaction time, or search tasks. In addition, these more classic secondary tasks could be compared to various types of real world tasks to find which real world tasks were most closely correlated with the established measures. Using such real world tasks in future studies would have the dual advantage of not only being applicable to the outside world but also having been validated with established measures of mental workload.

6. REFERENCES

Allport, D. A. (1980). Attention and performance. In G.L. Claxton (ed.), New directions in cognitive psychology. London: Routledge & Kegan Paul, p. 112-153.

Arend, U., Muthig, K. P., & Wandmacher (1987). Evidence for global feature superiority in menu selection by icons. Behaviour and Information Technology, v. 6, p. 411-426.

Aretz, A. J., Johannsen, C., & Ober, K. (1996). An empirical validation of subjective workload ratings. Proceedings of the Human Factors and Ergonomics Society, p. 91-95.

Barfield, W., and Baird, K. (1998). Issues in the design and use of wearable computers. Virtual Reality, v. 3, p. 157-166.

Barfield, W., Rosenburg, C., and Levassuer, G. (1991). The use of icons, earcons, and commands in the design of an online hierarchical menu. IEEE Transactions on Professional Communication, v. 34, p. 101-108.

Billinghurst, M., Bowskill, J., Dyer, N., Morphett, J. (1998). An evaluation of wearable information spaces. IEEE Transactions, p. 20-26.

Durham, A. G., and Emurian, H. H. (1998). Learning and retention with a menu and a command line interface. Computers in Human Behavior, v. 14, p. 597-620.

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (eds.), Human Mental Workload. Amsterdam: North-Holland, p. 139-183.

Hemenway, K. (1982). Psychological issues in the use of icons in command menus. CHI Proceedings, Conference on Human Factors and Computing Systems, ACM. New York, NY, p. 21-24.

Kantowitz B. H., and Knight, J. L. (1976). Testing tapping time sharing: I. Auditory secondary task. Acta Psychologica, v. 40, p. 343-362.

Kinsbourne, M., and Hicks, R. (1978). Functional cerebral space. In J. Requin (ed.), Attention and performance VII. Hillsdale, NJ: Erlbaum.

Krantz, J. H. (1997). http://psych.hanover.edu/classes/hfnotes3/sld022.html. Last Updated: February 09, 2000.

Mann, S. (1998). Definition of "wearable computer." http://wearcomp.org/wearcompdef.html. Last Updated: May 12, 1998.

McDaniel, George (1993). IBM Dictionary of Computing, Tenth Edition, McGraw-Hill.

McLoed, P. (1977). A dual task response modality effect: Support for multiprocessor models of attention. Quarterly Journal of Experimental Psychology, v. 29, p. 651-667.

Mecham, M. (1997). Wearable computer. Aviation Week & Space Technology, April 7, 1997.

Merwin, D. H., Dyre, B. P., Humphrey, D. G. (1990). The impact of icons & visual effects on learning computer databases. Proceedings of the Human Factors and Ergonomics Society, p. 424-428.

Meshkati, N., Hancock, P. A., Rahimi, M., Dawes, S. M. (1995). Techniques in Mental Workload Assessment. In Evaluation of Human Work, 2nd Ed. Taylor & Francis Ltd: Bristol, PA.

Morse, E. M., Lewis, R. R., Korfhage, & Olsen (1998). Evaluation of text, numeric, and graphical presentations for information retrieval interfaces. User preferences and task performance measures. Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, October 11-14, 1998.

Ockerman, J. J., Najjar, L. J., Thompson, J. C., Treanor, C. J., & Atkinson, F. D. (1996). FAST: A research paradigm for educational performance support systems. In P. Carson & F. Makedon (Eds.), "Educational Multimedia and Hypermedia 1996." Charlottesville, VA: Association for the Advancement of Computing in Education, p. 545-550.

Ockerman, J. J., Najjar, L. J., & Thompson, J. C. (1997). Wearable computers for performance support: Initial feasibility study. In “Digest of papers: First International Symposium on Wearable Computers,” Los Alamitos, CA: IEEE Computer Society, p. 10-17.

Ockerman, J. J., and Pritchett, A. R. (1998). Preliminary investigation of wearable computers for task guidance in aircraft inspection. In: "Digest of papers: Second International Symposium on Wearable Computers." Pittsburgh, PA: IEEE Computer Society, p. 33-40.

Preece, J. (1998). System and Interface Features. In: A Guide to Usability. Addison-Wesley: New York, NY, p. 57-89.

Siegel J., and Bauer, M. (1997). A field usability evaluation of a wearable system. In "Digest of papers: First International Symposium on Wearable Computers," Los Alamitos, CA: IEEE Computer Society, p. 18-20.

Siegel, J., Kraut, R. E., John, B. E., and Carley, K. M. (1995). Empirical study of collaborative wearable computer systems. Human Factors in Computing Systems (CHI) Conference-Proceedings. v 2 1995, ACM, New York, NY, p. 312-313.

Siewiorek, D., Smailagic, A., Bass, L., Siegel, J., Martin, R., Bennington, B. (1998). Adtranz: a mobile computing system for maintenance and collaboration. In "Digest of papers: Second International Symposium on Wearable Computers," Pittsburgh, PA: IEEE Computer Society, p. 25-32.

Smailagic, A., and Siewiorek, D. (1996). Matching interface design with user tasks, modalities of interaction with CMU wearable computers. IEEE Transactions on Professional Communication, February Issue, p. 14-25.

Starner, T, Rhodes, B, and Foner, L. (1995). Standard applications. WWW document. URL: http://www.media.mit.edu/wearables/normal-apps.htm

Thompson, C., Najjar, L. J., Ockerman, J. J. (1997). Wearable computer based training and performance support systems. In: S. Aronberg (Ed.), 19th Interservice/Industry Training, Simulation, and Educational Conference Proceedings. Arlington, VA: National Training Systems Association, p. 746-752.

Thompson, C., Ockerman, J., Najjar, L. J., and Rogers, E. (1997). Factory automation support technology (FAST): a new paradigm of continuous learning and support using a wearable. In "Digest of papers: First International Symposium on Wearable Computers," Los Alamitos, CA: IEEE Computer Society, p. 31-38.

Tullis, .T S., and Kodimer, M. L. (1992). A comparison of direct manipuation, selection, and data entry techniques for reordering fields in a table. Proceedings of the Human Factors Society, 36th Annual Meeting, p. 298-302.

Wandmacher, J., & Müller, U. (1987). On the usability of verbal and iconic command representations. Zeitschrift für Psychologie (in press).

Whiteside, J., Jones, S., Levy, P. S., and Wixon, D. (1985). User performance with command, menu, and iconic interfaces. CHI Proceedings, Human Computer Interaction, New York, NY, p. 185-191.

Wickens, C. D. (1980). The structure of attentional resources. In: R. Nickerson (ed.), Attention and performance VIII. Hillsdale, NJ: Erlbaum, p. 239-257.

Wickens, C. D. (1984). Processing resources in attention. In: R. Parasuraman and R. Davies (eds.), Varieties of attention. New York: Academic Press, p. 63-101.

Wickens, C. D., Sandry, D. L., and Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. Human Factors, v 25, n 2, p. 227-248.

Wickens, C. D. (1992). Attention, Time-Sharing, and Workload. In Engineering Psychology and Human Performance, 2nd Ed. HarperCollins Publishers Inc: New York, NY.

Wickens, C. D. (1992). Memory, Learning, and Training. In Engineering Psychology and Human Performance, 2nd Ed. HarperCollins Publishers Inc: New York, NY.

7. APPENDIX 1: USABILITY QUESTIONNAIRE

Usability Questionnaire

This follow-up questionnaire is designed to assess the usability associated with the software application you used during the experiment. Usability is a broad metric of computer system performance that concerns the safety, effectiveness, efficiency, and enjoyableness of the system. Your answers will be used to evaluate and improve the existing software application so that future applications will be more usable. There are no right or wrong answers or time constraints. Completing the questionnaire will take 5-10 minutes.

For this set of questions, think about the operations you performed on the computer. Please read the following questions and circle the answer that best represents your response. Some items may require you to write a short answer.

1. How quickly do you feel that you learned the computer operations during training?
   1---------------2---------------3---------------4---------------5---------------6------------7-

   Not at All Quickly — Somewhat Quickly — Extremely Quickly

2. effort do you feel you … arning the computer … g training?
   1---------------2---------------3---------------4---------------5---------------6------------7-

   No Effort — Some Effort — Extreme Effort

3. How quickly do you feel you were able to perform operations on the computer software application?
   1---------------2---------------3---------------4---------------5---------------6------------7-

   Not at All Quickly — Somewhat Quickly — Extremely Quickly

4. Generally speaking, how error prone was the wearable computer system?
   1---------------2---------------3---------------4---------------5---------------6------------7-

   Not Error Prone — Somewhat Error Prone — Extremely Error Prone

5. Please indicate which items of the wearable computer system you felt were most error prone during the course of the experiment.

   ____

   ____

6. How would you rate the software applications overall *ease of use*?
   1---------------2---------------3---------------4---------------5---------------6------------7-

Not at All Easy to Use | Somewhat Easy to Use | Extremely Easy to Use

7. How would you rate the *legibility* of the items (how easy was it to read or view items at a glance)?
1---------------2---------------3---------------4---------------5---------------6------------7-

Not Legible | Somewhat Legible | Extremely Legible

8. How would you rate the *distinguishableness* of the items (i.e., how clearly could you tell two items apart)?
1---------------2---------------3---------------4---------------5---------------6------------7-

Not Distinguishable | Somewhat Distinguishable | Extremely Distinguishable

9. How would you rate the *comprehensiveness* of the items (i.e., how easy was it to tell what function a specific item represented)?
1---------------2---------------3---------------4---------------5---------------6------------7-

Not Comprehensible | Somewhat Comprehensible | Extremely Comprehensible

10. How *cluttered* did the items appear?
1---------------2---------------3---------------4---------------5---------------6------------7-

Not Cluttered | Somewhat Cluttered | Extremely Cluttered

11. How easily do you feel you could remember *the locations* of specific items?
1---------------2---------------3---------------4---------------5---------------6------------7-

Not at All Easily | Somewhat Easily | Extremely Easily

12. How easily do you feel you could remember *the meaning* of specific items?
1---------------2---------------3---------------4---------------5---------------6------------7-

Not at All Easily | Somewhat Easily | Extremely Easily

13. How much did you enjoy using the software application?
1---------------2---------------3---------------4---------------5---------------6------------7-

Not at All | Some Enjoyment | Extreme Enjoyment

8. APPENDIX 2: CONSENT FORM

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Informed Consent for Participants of Investigative Projects

Project Title: The Effect of Icon Usage and Communication Style on Mental Workload Using Wearable Computers

Principal Investigator: Eric Nash

1. The Purpose of the Research Project

This study is designed to measure the mental workload associated with using different types of wearable computer software applications. Researchers generally define *mental workload* as the amount of mental effort required to perform a task. This study will measure your mental workload as you operate the wearable computer by monitoring your performance on various tasks. Using the performance information you provide, we can estimate the amount of mental workload the wearable computer software application demands. This information will ultimately help designers of wearable computer applications create future products that demand less mental workload.

2. Procedures

You will be one of 30 subjects participating in this experiment. Following an eye exam, you will don the wearable computer and be briefed on its components. Included in this will be a short verbal exercise, which the computer will need to acclimate itself to the nuances of your speech. Next, you will be trained to perform a set of 9 benchmark tasks. These will include standard scheduling/planning activities, creating task lists, and updating phone listings.

After training, you will be asked to perform three different kinds of tasks a repeated number of times while wearing the computer. These will include walking a path, assembling a Lego™ kit, and a verbal recall task. Sometimes the wearable computer will be turned off while you perform these tasks. Other times you will be operating the computer while you perform the tasks. You will be videotaped during the walking trials so that deviations from the path can be determined.

Following completion of all tasks, you will be asked to complete a follow-up questionnaire. The questionnaire will be used to evaluate your overall impression of the wearable computer application as well as the amount of mental workload you felt it required.

This will conclude your participation in the experiment. Total testing time is estimated to require 1.0-1.5 hours and you will only be required to participate once. All research will take place in the Virtual Environment Laboratory (560 Whittemore Hall) at Virginia Tech.

3. Risks

There are no foreseeable physical or mental risks associated with participation in this experiment.

4. Benefits of this Project

This research will ultimately contribute towards designing future wearable computer software applications that easier to operate.

As a matter of university policy, a statement must be made to the effect that no promise or guarantee of benefits have been made to encourage you to participate in this experiment.

5. Extent of Anonymity and Confidentiality

During the periods of this experiment when you are to perform walking tasks, video will be recorded to determine your performance. These recordings are vital to the analysis of the mental workload. After you've completed the experiment the data you provide on collection forms and video tapes will be labeled arbitrarily with a number for identification purposes. No names will be used on the data collection forms or the video recordings.

During analysis of these video tapes, they will be locked in a filing cabinet in the Virtual Environment Laboratory, where only the principle investigator will have access. The principle investigator will perform all analyses of the tapes. Immediately following analysis, these tapes will be destroyed. At no time will the researchers release the results of the study to anyone other than individuals working on the project without your written consent.

6. Compensation

You will receive $10 compensation for your participation in this experiment.

7. Freedom to Withdraw

As a matter of policy, you should be made aware that all subjects are free to withdraw from this study at any time without penalty. Subjects are free not to answer any questions or respond to experimental situations that they choose without penalty.

8. Approval of Research

This research has been approved, as required, by the Institutional Review Board for Research Involving Human Subjects at Virginia Polytechnic and State University, by the Department of Industrial and Systems Engineering.

9. Subject's Responsibilities

There are no prior responsibilities you must fulfill before participating in this experiment.

10.Subject's Permission

I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project.

If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

_______________________________ _______________________________

Signature Date

Should I have any questions about this research or its conduct, I may contact:

Investigator	Eric Nash	540-231-1791
Faculty Advisor	Dr. Woodrow Barfield	540-231-7520
Chair, IRB Research Division	H.T. Hurd	540-231-1234

9. APPENDIX 3: RAW DATA

Table 11. Task Performance Raw Data for the Wearable Computer Off Conditions

Subject	Version	Percentage Correct Blocks	Percentage Correct Blocks Attempted	Average Assembly Rate	Percentage Correct Digits	Average Walking Speed
2	B	0.79	0.82	0.12	0.95	2.44
4	B	1.00	0.74	0.25	1.00	2.86
12	B	0.79	1.00	0.29	1.00	3.45
13	B	0.85	0.76	0.20	0.99	3.18
23	B	0.82	1.00	0.26	1.00	2.94
24	B	0.65	0.73	0.11	0.88	3.00
8	CL	0.76	0.90	0.17	0.97	2.92
11	CL	1.00	0.82	0.25	1.00	3.05
16	CL	0.82	0.85	0.15	0.97	2.82
22	CL	0.91	0.63	0.14	0.95	3.03
28	CL	0.85	0.61	0.21	0.97	3.13
30	CL	0.97	0.94	0.23	1.00	2.99
3	IB	1.00	0.91	0.28	0.97	3.33
15	IB	0.56	0.58	0.16	0.99	3.40
20	IB	0.76	0.57	0.20	1.00	2.51
21	IB	0.88	1.00	0.17	0.93	2.71
25	IB	1.00	0.88	0.35	0.97	3.70
26	IB	0.71	0.56	0.12	0.94	2.44
5	ITM	0.82	0.54	0.15	1.00	3.29
6	ITM	0.97	0.88	0.20	1.00	3.22
10	ITM	0.41	0.55	0.10	0.97	3.08
14	ITM	0.91	0.52	0.13	1.00	2.74
17	ITM	0.97	1.00	0.17	0.99	3.51
27	ITM	0.91	0.93	0.24	0.97	2.99
1	TM	0.82	0.95	0.21	0.90	3.79
7	TM	0.97	0.37	0.15	0.90	2.56
9	TM	0.82	0.91	0.19	0.99	3.78
18	TM	0.91	1.00	0.17	0.84	2.33
19	TM	0.94	0.78	0.16	1.00	3.45
29	TM	0.79	0.63	0.17	1.00	2.75

Table 12. Task Performance Raw Data for the Wearable Computer On Condition

Subject	Version	Percentage Correct Blocks	Percentage Correct Blocks Attempted	Average Assembly Rate	Percentage Correct Digits	Average Walking Speed
2	B	0.53	0.56	0.05	0.41	1.69
4	B	0.29	0.60	0.03	0.96	1.56
12	B	0.59	0.80	0.04	0.72	2.08
13	B	0.71	0.92	0.06	0.44	1.87
23	B	0.59	0.90	0.06	0.80	1.42
24	B	0.76	0.92	0.07	0.26	1.16
8	CL	0.12	0.00	0.01	0.31	2.05
11	CL	0.53	1.00	0.06	0.49	1.33
16	CL	0.65	0.82	0.06	0.21	1.68
22	CL	0.41	1.00	0.04	0.38	1.66
28	CL	1.12	0.53	0.10	0.45	1.91
30	CL	1.00	1.00	0.10	0.88	1.82
3	IB	1.18	0.90	0.12	0.79	2.65
15	IB	0.29	0.60	0.04	0.50	1.45
20	IB	0.82	0.71	0.07	0.84	1.83
21	IB	0.82	1.00	0.09	0.49	1.65
25	IB	1.12	1.00	0.11	0.59	1.16
26	IB	0.53	1.00	0.05	0.51	2.28
5	ITM	0.94	0.69	0.09	0.91	2.24
6	ITM	0.71	0.83	0.07	0.48	1.96
10	ITM	0.41	0.57	0.04	0.15	1.57
14	ITM	1.00	0.88	0.10	0.64	1.42
17	ITM	0.88	1.00	0.09	0.71	1.79
27	ITM	0.76	0.77	0.07	0.51	1.68
1	TM	0.59	1.00	0.08	0.36	1.63
7	TM	0.71	0.00	0.06	0.32	0.81
9	TM	0.76	0.85	0.08	0.63	2.37
18	TM	0.65	1.00	0.06	0.15	0.96
19	TM	0.82	0.79	0.06	0.60	1.20
29	TM	1.12	0.74	0.07	0.56	1.83

Table 13. Usability Raw Data (Part 1)

Subject	Version	QuickLearn	EffortLearn	QuickPerf	ErrorProne	EaseUse	Legible
2	B	6	3	6	2	6	6
4	B	6	2	5	3	7	6
12	B	4	4	4	3	5	6
13	B	4	6	3	6	4	3
23	B	6	2	5	3	6	5
24	B	4	2	4	4	6	5
8	CL	5	5	4	6	6	6
11	CL	5	4	3	3	4	4
16	CL	4	3	3	7	4	
22	CL	6	1	3	3	6	6
28	CL	4	5	3	7	4	
30	CL	6	2	2	6	4	
3	IB	6	3	6	4	5	5
15	IB	6	3	4	4	7	6
20	IB	5	3	4	5	6	5
21	IB	6	5	4	4	7	6
25	IB	6	3	2	6	4	4
26	IB	4	2	4	3	5	2
5	ITM	6	2	6	7	3	3
6	ITM	5	3	4	6	4	6
10	ITM	5	5	5	4	5	4
14	ITM	5	5	5	3	5	6
17	ITM	7	4	4	2	2	4
27	ITM	6	3	5	3	7	5
1	TM	6	3	7	4	4	5
7	TM	5	3	5	3	4	4
9	TM	7	3	5	4	7	7
18	TM	5	4	5	5	4	5
19	TM	6	3	6	2	6	4
29	TM	6	3	4	3	4	5

Table 14. Usability Raw Data (Part 2)

Subject	Version	Distinguish	Comprehend	Cluttered	RecallLoc	RemMeaning	Enjoy
2	B	5	7	1	5	6	6
4	B	6	5	3	5	6	6
12	B	6	6	2	6	6	4
13	B	3	4	5	6	6	3
23	B	6	5	3	6	6	4
24	B	5	6	3	5	4	5
8	CL	5	4	2	5	5	5
11	CL	5	6	2	5	5	5
16	CL	4	4			4	4
22	CL	6	5	3	4	4	6
28	CL	6	7			5	7
30	CL	6	7			7	4
3	IB	5	5	2	5	6	6
15	IB	7	7	4	6	7	5
20	IB	6	6	4	6	6	3
21	IB	7	7	2	7	7	5
25	IB	4	5	6	4	6	2
26	IB	5	5	3	4	4	5
5	ITM	3	4	2	3	6	1
6	ITM	6	4	2	5	4	4
10	ITM	6	5	2	4	5	3
14	ITM	7	6	1	6	6	5
17	ITM	5	4	1	4	6	4
27	ITM	7	6	2	4	6	7
1	TM	6	5	2	4	4	5
7	TM	5	5	2	4	5	4
9	TM	7	7	2	7	7	6
18	TM	6	6	6	4	6	5
19	TM	6	6	2	6	7	6
29	TM	5	5	1	6	3	6

Table 15. NASA TLX Rating Raw Data for the Assembly Task

Subject	Version	Weighted Workload	Mental Demand	Physical Demand	Temporal Demand	Effort	Performance	Frustration
2	B	38	35	5	20	50	35	45
4	B	37	70	20	25	70	10	35
12	B	58	80	15	10	45	80	25
13	B	55	25	5	45	75	75	70
23	B	65	90	5	80	85	10	10
24	B	62	65	55	90	90	35	45
8	CL	64	75	50	45	60	85	20
11	CL	60	30	70	30	65	85	20
16	CL	45	75	20	5	30	50	15
22	CL	52	80	70	50	65	40	45
28	CL	59	45	55	80	65	35	15
30	CL	50	40	55	25	55	65	15
3	IB	60	65	30	40	15	75	55
15	IB	52	55	10	45	50	50	60
20	IB	77	90	10	75	70	45	85
21	IB	55	55	35	30	25	85	5
25	IB	45	50	5	25	75	30	40
26	IB	66	70	10	65	60	85	30
5	ITM	72	75	25	20	65	80	75
6	ITM	51	75	15	25	70	40	15
10	ITM	48	60	50	50	60	35	30
14	ITM	67	80	10	30	90	50	30
17	ITM	50	85	15	15	80	20	30
27	ITM	89	100	5	100	90	85	70
1	TM	51	75	50	30	70	25	15
7	TM	93	95	5	95	95	70	95
9	TM	51	55	55	15	55	55	30
18	TM	54	50	25	65	70	55	40
19	TM	61	80	10	35	80	30	60
29	TM	58	50	20	10	85	80	40

Table 16. NASA TLX Rating Raw Data for the Digit Task

Subject	Version	Weighted Workload	Mental Demand	Physical Demand	Temporal Demand	Effort	Performance	Frustration
2	B	81	80	10	90	80	75	70
4	B	83	100	10	85	90	45	60
12	B	81	85	5	85	80	90	30
13	B	64	90	5	10	80	20	70
23	B	70	100	10	45	90	25	35
24	B	74	95	5	90	85	40	55
8	CL	72	75	20	75	70	70	55
11	CL	71	70	15	90	65	60	40
16	CL	97	50	10	100	100	100	100
22	CL	85	90	65	90	80	80	75
28	CL	87	90	5	100	80	95	70
30	CL	67	80	10	75	90	45	25
3	IB	63	70	15	65	20	75	60
15	IB	41	35	10	5	55	45	25
20	IB	67	80	10	65	65	35	70
21	IB	73	45	5	90	65	85	20
25	IB	69	70	5	50	85	70	70
26	IB	83	90	20	90	75	70	70
5	ITM	54	65	30	60	70	25	45
6	ITM	86	90	40	85	90	65	90
10	ITM	68	70	40	50	70	65	80
14	ITM	65	50	10	90	90	25	75
17	ITM	92	100	5	90	100	45	90
27	ITM	87	90	10	85	85	70	95
1	TM	89	100	5	90	90	70	70
7	TM	99	100	5	100	100	85	100
9	TM	75	80	10	80	100	55	60
18	TM	63	70	15	65	40	70	65
19	TM	93	100	5	50	100	70	100
29	TM	78	90	40	85	75	80	60

Table 17. NASA TLX Rating Raw Data for the Path Walking Task

Subject	Version	Weighted Workload	Mental Demand	Physical Demand	Temporal Demand	Effort	Performance	Frustration
2	B	48	35	70	40	50	30	25
4	B	35	65	20	25	50	10	25
12	B	34	65	15	15	15	30	10
13	B	42	60	25	45	60	20	30
23	B	66	80	55	10	65	80	5
24	B	85	80	25	75	90	90	90
8	CL	48	50	45	35	55	40	10
11	CL	74	35	85	15	70	60	85
16	CL	30	30	50	10	30	25	10
22	CL	50	90	30	40	50	15	35
28	CL	66	15	15	90	60	75	80
30	CL	33	30	10	50	25	45	5
3	IB	68	75	65	65	15	75	30
15	IB	40	40	50	35	35	40	40
20	IB	44	20	25	60	40	50	25
21	IB	47	35	55	45	25	70	5
25	IB	52	50	10	70	40	60	50
26	IB	63	80	60	30	75	40	55
5	ITM	57	35	60	25	60	45	75
6	ITM	69	50	75	50	75	60	80
10	ITM	55	60	45	45	50	65	40
14	ITM	25	15	30	5	30	25	5
17	ITM	54	65	80	15	65	30	55
27	ITM	48	70	25	55	50	45	30
1	TM	17	50	10	5	15	5	5
7	TM	90	95	10	100	100	65	80
9	TM	61	50	80	55	55	55	15
18	TM	62	45	50	60	60	70	65
19	TM	47	35	25	10	50	55	55
29	TM	21	15	15	30	30	15	20

10. APPENDIX 4: VITA

E-mail soulenoid@cs.com

Eric B. Nash

Education

1998 - 2001 Virginia Tech Blacksburg, VA

Master's of Science in Industrial and Systems Engineering

- United Parcel Service Fellow, Fall 1998-Summer 1999.

1994 - 1997 University of Cincinnati (UC) Cincinnati, OH

Bachelor's of Arts in Psychology

- Summa Cum Laude with High Honors, *GPA* 3.93 / 4.0
- Golden Key National Honor Society, 1996

Relevant Professional experience

2000 – present Oracle Corporation Redwood Shores, CA

Usability Engineer Level II

- Create interactive desktop and mobile application prototypes using HTML, Javascript, XML, WML, WML Script
- Develop VC++ applications for collecting usability measures
- Develop, conduct, and assist in user group requirements exercises, usability walkthroughs, and formalized usability tests
- Report usability findings in oral and written form to product development teams and interaction designers
- Participate in product redesigns with development teams and designers
- Maintain product team's internal website

1997 – 1998 University of Cincinnati Cincinnati, OH

Level 1 Research Assistant (Data Manager)

- Provide data quality protocol, quality control programs, and descriptive statistics for epidemiological studies using SAS v6.12
- Develop user interfaces for data entry using SAS FSEDIT / SCL

Other work experience

1999 – 2000 Virginia Tech Computer Science Dept. Blacksburg, VA

Graduate Research Assistant

1997 – 1998 Postural Stability Laboratory (UC) Cincinnati, OH

Student Research Assistant

1996 – 1997 Introduction to Statistics (UC) Cincinnati, OH

Student Teaching Assistant

Professional memberships

- SIGCHI, 2001
- Human Factors and Ergonomics Society (Student Member), 1999
- Psi Chi, National Honor Society for Psychology Students, 1997