THE EFFECTS OF THREE MEDIA PRESENTATION SYSTEMS ON MAINTENANCE TASK PERFORMANCE

THESIS

John W. Chastain, Captain USAF

AFIT/GAL/LAL/98S-2

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GAL/LAL/98S-2

THE EFFECTS OF THREE MEDIA PRESENTATION SYSTEMS ON MAINTENANCE TASK PERFORMANCE

THESIS

John W. Chastain, Captain USAF

AFIT/GAL/LAL/98S-2

AFIT/GAL/LAL/98S-2

THE EFFECTS OF THREE MEDIA PRESENTATION SYSTEMS ON MAINTENANCE TASK PERFORMANCE

THESIS

Presented to the Faculty of the Graduate School of Logistics

and Acquisition Management of the Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Acquisition Logistics Management

John W. Chastain, B.S.

Captain USAF

September 1998

Acknowledgements

There are many people I would like to thank for assisting me and enduring me throughout this thesis effort. First and foremost, I would like to thank my family for their unwavering patience and support throughout the past several months. It goes without saying that I must thank Marcie for her patience and support throughout this long process. On a more serious note, I would like to give special thanks to Lt Col Stephen G. Atkins for having faith in me, and allowing me to work at my own pace. His professional guidance and support were crucial to this effort. Equally important were the advice and encouragement given me by Dr. Guy S. Shane. Without the dedicated support of these two gentlemen, this thesis may not have earned approval. I would also like to thank Seymour Johnson AFB, NC, especially MSgt. Robert Schaffer of the Repair and Reclamation Flight and SMSgt Pride Johnson of the 334 Fighter Training Unit. A special thanks goes out to the greater community of Goldsboro, NC for supplying us with a wide selection of fine dining, music from 96 Rocks, and fun for all at Mimo's. Finally, I'd like to thank Dave "Why am I here?" Kancler, Alan "Catch 22" Revels, Laurie "No Games" Quill, and Barbara "Mrs. Persistance" Masquelier. For without the help and assistance from this team there is no way this thesis would have been completed.

John W. Chastain

Table of Contents

Page

List of Tables

Table Page

Abstract

Wearable computers and Portable Maintenance Aids (PMAs) may soon be the normal way of doing aircraft maintenance in the Air Force. Currently, the Air Force uses the paper form of Technical Orders (TO's) while doing aircraft maintenance. The purpose of this thesis was to compare the effects of three different media presentations of the information used during aircraft maintenance. The three different presentations compared are the current paper form, a Head Mounted Display (HMD), and an auditory mode. No research has compared an auditory display to any other system in the flight line environment. An experiment was conducted to determine if there was a significant difference between the systems (in terms of task completion times and user preference). Nine F-15E maintenance technicians from the Repair and Reclamation Flight of the 4th Equipment Maintenance Squadron of the 4th Fighter Wing, Seymour Johnson AFB, NC were chosen to participate in this experiment. Each individual accomplished the same task using the three different systems. While some treatment effects were in the predicted direction, differences were not statistically significant. The results weakly suggested that these technicians preferred the newer technology to the current paper form. The primary conclusion is that there may be a use for technology-augmented checklist presentation in the aircraft maintenance arena. This research establishes a foundation for future research efforts.

THE EFFECTS OF THREE MEDIA PRESENTATION SYSTEMS ON MAINTENANCE TASK PERFORMANCE

I. Introduction

Chapter Overview

People are always looking for more efficient and effective ways of doing business. For example, the Air Force has been reducing the number of personnel and facilities while maintaining the same level of operation tempo, putting pressure on each individual to work harder and faster while maintaining the same quality work. Therefore, there is a need for tools to assist aircraft maintenance personnel in their day to day operation on the aircraft. This chapter discusses the general issue of the need for more efficient tools used in aircraft maintenance operation, followed by a specific problem statement that is addressed in this thesis. An outline of the research objective and hypotheses derived from the problem statement are illustrated. This chapter then concludes with a brief discussion of the scope and limitations of the study.

General Issues

In the aircraft maintenance arena it is imperative that maintenance personnel work as efficiently and effectively as possible. Every task accomplished while maintaining an aircraft must be obtained from a technical order (TO) which maintenance personnel must have on-site while performing these tasks. These technical orders are constantly being revised and updated with the changes in the current aircraft inventories of the different military services. It is an accepted fact that the current paper form of the technical orders used in support of aircraft maintenance are poorly organized, cumbersome to use, and sometimes outdated or even incomplete. Due to these inefficiencies of the current technical orders, it is sometimes tempting for the maintenance technician to just set the TO aside and go by personal experience for certain tasks. Technology currently available allows the technician to work completely hands free. Therefore, a logical solution to this problem would be to create a fully automated TO system in which the technician could obtain the needed information but work in a more efficient manner.

Specific Problem

Technology has allowed Armstrong Laboratory to fabricate a lightweight vest that consists of a small drive for data storage, a power supply, and all hardware for a headset including a microphone and a Head Mounted Display (HMD) allowing the technician to accomplish the task via through the HMD. The procedure for a certain maintenance task is electronically stored on the small computer drive allowing the technician to navigate through the procedure as seen on the HMD via various inputs. There has been research done with this configuration comparing the use of voice input versus keypad input to manipulate the needed information.

The keypad requires that the technician remove his hand from the job to input the command needed to navigate through the TO in order to complete the task. Both configurations (the keypad input and the voice input) require that the technician devote his concentration to reading the information presented on the HMD. There is technology available which will give the information completely in the auditory mode, allowing the technician to never remove his/her eyes from the task at hand.

Verbex is a computer software program that is trained to recognize a specific individual's voice and relate it to certain computer input commands while allowing the user to be hands-free. Integrating the Verbex program and removing the HMD with the current configuration will allow the technician to devote his/her hands completely to the maintenance task. Specifically, the problem of inefficiently using the cumbersome paper technical orders may be resolved by retrieving task procedures using electronically stored technical orders.

Research Objective

The objective of this thesis research is to compare the results of three different sources of media presentation in three areas; difference in technician performance, number errors with each source, and user satisfaction with each source.

Experimental Hypothesis

The overall research hypothesis is that technician performance will be enhanced using the voice input/auditory presentation versus the voice input/HMD presentation or the current paperback TO. Detailed hypotheses include a decreased task completion time, less errors, and a greater satisfaction while using the voice input/auditory presentation.

Scope and Limitations

The hardware and software being used and tested by Armstrong Laboratory at the time this research was being conducted were utilized. The vest and HMD configuration is already fabricated and available. The voice input/auditory presentation configuration must be programmed into the Verbex software package. Verbex is currently being used as the voice recognition system in other areas at Armstrong Laboratory and is available for this thesis research.

The tasks for this research will be conducted at Seymour Johnson AFB, NC (SJAFB) by Aircraft Repair (AR) technicians in the 4th Equipment Maintenance Squadron (EMS).

The task selected, Nose Wheel Shimmy Analysis Checkout, was limited due to several reasons. The task had to be a step-by-step procedure having as little diagram involvement as possible. For security and availability reasons the technical orders had to be available here at Wright Patterson AFB, OH.

II. Literature Review

Chapter Overview

The Air Force has been using the paper form of technical manuals to accomplish maintenance tasks on the flight line and in the backshops. From personal experience and feedback from flight line technicians, this is a tedious and cumbersome procedure. There are possibly more efficient and user-friendly methods to accomplish the same tasks with today's technology. Many research projects have addressed this problem by using electronic technical data systems and media displays in place of the current paper form. Air Force Research Laboratory Human Readiness Division has been researching this issue for over a decade, and conducting many tests and evaluations specific to wearable electronic computing systems. These systems consist of their own CPU, power source, and media presentation. Media presentations for electronic systems range from simple portable laptop computers to more complex head mounted displays (HMD).

This chapter discusses the research concerning media presentation and system development leading to the state-of-the-art. The chapter then discusses human-computer interaction and the knowledge necessary to properly compare visual and auditory media presentations.

System Development

Wearable computing systems have been in development for many years and may have a use in today's Air Force maintenance arena. The wearable computer system that has been tested and developed range in a variety of configurations. For example, Masquelier (1991) compared technician use of a portable laptop monitor and a head mounted visual display while performing maintenance tasks (Masquelier, 1991). The study consisted of two groups of maintenance technicians sitting at a workbench performing inspections and fault isolation maintenance on circuit boards. The two forms of media presentations; HMD and auditory modes; were paralleled as closely as possible to the current form of paper technical manuals. The study resulted in non-significant differences between the two display types concerning performance times or number of errors. However, Masquelier's study led to follow-on research dealing with similar tasks in an environment requiring mobility, possibly more appropriate for the wearable computer display (Friend and Grinstead, 1992). This resulted in a faster completion time with the HMD along with detecting more faults.

Another follow-on to Masquelier was Chapman and Simmons (Chapman, 1995) utilizing a wearable computer system in conjunction with the human voice as a source of input. Specifically, they compared technician performance using two different input devices coupled with a HMD: 1) voice recognition and 2) wrist-worn keypad. Because this study was among the first concerning the use of voice recognition as input on the flightline, a careful design had to be constructed. The design had to measure the effects of hardware interaction with the user rather than the software program. As with Masquelier, the study resulted in finding no significant differences in performance times between the input devices. This resulted in a study comparing visual displays and auditory displays. (See Table 1)

Table 1. Chart of Previous Studies

Study	Configuration	Results	Recommendations
Masquelier, 1991	HMD connected to desktop computer compared to regular monitor display	1. No statistically significant performance differences between displays.	1. Evaluate HMD on flight line maintenance tasks
Friend and Grinstead, 1992	Fully portable HMD compared to hand-held computer	1. Task completion times faster with HMD 2. More faults detected with HMD	1. Test more complex tasks 2. Test on more complex weapon system
Carney and Quinto, 1993	Personal laptop computer with programmable soft-keys, dedicated hardware keys, pushbutton keys, cursor keys, and number keys	1. Dedicated hardware keys and number provided greatest user satisfaction 2. Pushbuttons and programmable soft-keys provided lowest user satisfaction	1. Test different types of input devices, such as mouse or touchscreen 2. Evaluate same interface in different environment

When deciding which form of presentation to use, there are certain guidelines that must be met that are shown in Table 2.

Table 2. Use of Auditory presentation versus Visual presentation

Use auditory presentation if:	Use visual presentation if:
1. The message is simple.	1. The message is complex.
2. The message is short.	2. The message is long.
3. The message will not be referred to later.	3. The message will be referred to later.
4. The message deals with events in time.	4. The message deals with location in space.
5. The message calls for immediate action.	5. The message does not call for immediate action.
6. The visual system of the person is overburdened.	6. The auditory system of the person is overburdened.
7. The receiving location is too bright or dark-adaptation integrity is necessary.	7. The receiving location is too noisy.
8. The person's job requires him to move about continually.	8. The person's job allows him to remain in one position.

Source: Deatherage, 1972.

To fully understand human-computer interaction concerning auditory arena, both human and computer-based components of speech must be understood.

Speech Intelligibility and Human-Computer Interaction

Human-Computer Interaction (HCI) has been an area of research since the design of the first computer. Researchers have been looking for easier, quicker ways for humans to accomplish tasks in the work environment by the use of computers. As the use of voice recognition and voice synthesis (the link between human and computers) becomes viable, speech and its constraints must be studied. The areas of concern when evaluating speech are intelligibility and quality.

Intelligibility is the ability to correctly recognize a spoken message. For example, presenting a subject with a list of words and asking them to repeat it out loud is a way of assessing speech intelligibility (Sanders and McCormick, 1993). Intelligibility is also very dependent on context and how they are used. In everyday conversational speech a large portion of words are unintelligible when taken out of context. Therefore, individual words may be affected by the surrounding context and the expectations of the hearer (Sanders and McCormick, 1993).

Quality is important when dealing with speaker identification or user satisfaction with a communication system (Sanders and McCormick, 1993). Quality is a method that many telephone companies use when comparing themselves to an industry standard for user satisfaction. Intelligibility and quality are both affected by several specific components of speech communication. "A speech communication system is thought to consist of the speaker, the message, the transmission system, and the hearer" (Sanders and McCormick, 1993).

Research has made it possible to not only gain an understanding of intelligibility, but to trace certain features that actually affect the overall communication channel. The speaker has a great influence on intelligibility through syllable duration, intensity, pitch, the amount of speech time and pauses, and fundamental vocal frequencies (Bigler, 1955). These factors themselves cannot be modified to any great measure, but appropriate speech training can result in moderate improvements in the intelligibility of the speaker.

The intelligibility of the message is affected by several characteristics, such as the phonemes used, the types of words, and the context. Phonemes are the smallest unit of speech; for example, the *b* in ball. When set alone many letters can be very confusing in sound, such as DVPBGCET and FXSH (Hull, 1976). Therefore, it is best to avoid using letters as codes in most environments. The use of familiar words versus unfamiliar words has a great impact on the intelligibility of the message. The speech communication system must use a vocabulary that the receiver is familiar with and uses in everyday operation. Also, longer words are more intelligible than shorter words. Therefore, if certain parts of a long word are intelligible, then the listener can figure out the remaining, unintelligible parts of the word. If the word, itself, is familiar to the listener, then the context of the words used is another characteristic that affects the message. The intelligibility of the message increases when the words are arranged in meaningful sentences (Sanders and McCormick, 1993). To improve intelligibility, especially under noisy conditions, several guidelines should be observed: 1) use a small vocabulary; 2) use meaningful sentence structure; 3) avoid short words; and 4) familiarize the user with the vocabulary and sentence structure used (Sanders and McCormick, 1993).

The quality of the hardware used for the transmission system is another way to improve the intelligibility of the message. Transmission systems can be altered by use of distortion and filtering the message display. Filtering is simply blocking out certain frequencies allowing only the desired frequencies to pass through to the user. Typically, filters block frequencies above some level or below some level; a low-pass filter or a high-pass filter, respectively (Sanders and McCormick, 1993). High-pass filters and low-pass filters increase intelligibility more effectively only at certain frequencies, specifically between 1000 Hz and 3000 Hz (Sanders and McCormick, 1993).

A major influence on message intelligibility that must be dealt with is the noise environment. Irrelevant noise in the environment can be either internal or external to the communication system. Not only must the location of the noise be detected, but also it must be controlled. The best approach to control for this is to calculate and maintain a desired signal-to-noise (S/N) ratio (Sanders and McCormick, 1993). This is the algebraic difference between the signal (the actual communication) and noise (other than irrelevant noise) in decibels. Due to the constantly changing noise environment of maintenance technicians, it is beyond the scope of this research to maintain a desired S/N ratio. However, S/N rates can be addressed by controlling the distance between the speaker and the listener. The closer the speaker is to the listener, the more intelligible the message will be when using a normal voice.

Definitely important, the listener is the last link in the communication chain. Age, the ability to reasonably withstand situational stresses, and training (on the types of communications to be received) are some factors that affect the listener's ability to correctly understand the message (Sanders and McCormick, 1993). The intelligibility of a message is increased considerably when the listener is already trained and familiar with the communication. When the listener is comfortable with the surrounding environment and stressors, they will be more relaxed resulting in a higher intelligibility. Even the psychological awareness of using a speech recognition system could cause the speaker to produce a noticeable difference in the pronunciation and/or articulation in their speech (Juang, 1991). These characteristic changes in articulation due to environmental influence are known as the Lombard effect (Juang, 1991). With age comes hearing degradation. Consequently, intelligibility decreases considerably after the age of 60 (Bergman, 1976).

Speech Recognition

Human-computer interaction not only deals with the human being trained and familiar with the type of communication, but the computer itself must trained. This is referred to as speech recognition or automatic speech recognition. Limited forms of speech recognition systems are already available on personal workstations for different applications. Currently, speech recognition has proven useful in many areas, such as telephone voice-response systems, digit recognition for cellular phones, and data entry using portable databases (Peacocke, 1995). Although the speech recognition industry has improved dramatically in the recent years, communication in a high noise or fluctuating noise environment pose a problem for the typical automatic speech recognition system. However, there are five factors that are used to control and simplify the speech recognition task; 1) isolated words, 2) single speaker 3) vocabulary size, 4) grammar, and 5) the environment (Peacocke, 1995).

Automatic speech recognition systems can correctly recognize isolated words more easily than continuous speech due to the difficulties in distinguishing word boundaries in continuous speech. Also, "coarticulation effects in continuous speech cause the pronunciation of a word to change depending on its position relative to the other words in a sentence" (Peacocke, 1995). For example, "will you?" is not pronounced the same as "will" + short silence + "you?" Therefore, having the speaker pause between words dramatically reduces the error rates in speech recognition systems.

Each particular speaker has unique characteristics, which affect the parametric representations of speech. Therefore, a single speaker will result in fewer errors using an automatic speech recognition system. These systems are referred to as being speaker dependent--trained for use with a specific individual. The majority of speech recognition systems are speaker-dependent due to the fact that error rates are three to five times smaller than speaker-independent systems (Peacocke, 1995).

The size of the vocabulary being used for the task also influences the recognition accuracy. As with humans; a larger vocabulary increases the chances of ambiguous words. Ambiguous words are those whose pattern-matching templates appear similar to the classification of other words used by the recognizer (Peacocke, 1995). Error rates will be kept to a minimum if the vocabulary is small and confined to the required words. The use of the vocabulary also incorporates the rules of grammar. The amount of constraint on word choice is referred to as the "perplexity of the grammar."

> Systems with low perplexity are potentially more accurate than those that give the user more freedom because the system can limit the effective vocabulary (and search space) to those words that can occur in the current input context. (Peacocke, 1995)

Environment is the final factor that can be used to control and simplify speech recognition. Many recognition systems can maintain low error rates as long as the environmental conditions remain quiet and controlled. Background noise, changes in microphone characteristics, and loudness can all affect recognition accuracy (Peacocke, 1995). The performance of the system always degrades when ambient or random noises are introduced or the external conditions differ from the training session. It has been demonstrated that an isolated word recognizer trained in clean conditions and capable of achieving a recognition accuracy of 95% had an order of magnitude increase in error rate when tested in a noisy environment (Dautrich, 1983). To compensate, the user must always wear a head-mounted, noise-limiting microphone with the same characteristics used during the training session.

Speech Synthesis

In order for the human-computer interaction to be complete, the computer must communicate with the human auditorily. This is accomplished by one of two ways; recorded natural voice or synthesized voice. Advances in recent research have made it possible to synthesize human speech at a relatively low cost (Sanders and McCormick, 1993). This has generated a large number of studies aimed at determining when it is appropriate to use synthesized speech and the influence it has on human performance. There are different types of speech synthesis systems available. Natural voice systems use a recording of the speech made for use during the required task. The major problem with this method is the moving parts eventually wear out and break down (Sanders and McCormick, 1993). The more advanced method is to completely digitize speech and store it in a computer. However, the amount of space required to store the resulting digitized information is impractical for most applications. There are two methods currently used to solve these problems; synthesis by analysis and synthesis by rule.

Synthesis by analysis uses various techniques to reduce the required storage space needed for the more compressed digitized human speech. However, this method is restricted to only those words or phrases that were previously stored in the system (Sanders and McCormick, 1993). Due to the lack of coarticulation(method of joining words or sounds) when linking words, it is said that synthesis by analysis is really not synthesized voice but the same as digitized speech.

Synthesis by rule is considered by many to be true synthesized speech. This method uses the basic rules for generating speech sounds, combining sounds into words, and stressing particular sounds and words to attain the correct prosody of speech (Sanders and McCormick, 1993). Prosody is the correct structure of verse in creating the rhythm quality of natural speech. By not using digitized recordings or stored words, synthesis by rule is capable of a very large vocabulary with relatively small amounts of computer memory.

There are many different uses for synthesized speech. For example, the military uses synthesized speech for cockpit warnings and interactive training systems (Werkowitz, 1980). The telephone industry has been using synthesized speech to present telephone numbers to callers as well as banks having automated phone teller service. In the examples above there is still the comparison to natural speech in both intelligibility and how well people can remember the synthesized messages.

Researchers have used a modified rhyme test (MRT) to test the segmental intelligibility, individual phonemes that make up words, of both natural human speech and synthesized speech. The testing of ten synthesis-by-rule systems were compared to natural speech using the MRT. The synthesized systems had an error rate ranging from 3 percent to 35 percent, while natural speech produced an error rate of less than 1 percent (Logan, 1989). Other tests were conducted comparing correct word identification within meaningful sentences. Natural speech was 99.2 percent intelligible. Out of four synthesis-by-rule systems, the best was 95.3 percent intelligible and the worst was 83.7 percent intelligible (Nusbaum, 1985). Although the initial intelligibility of speech synthesis is not as high as natural speech, intelligibility increases significantly with proper training to a system's voice (Sanders and McCormick, 1993).

Memory recall of synthesized speech is another problem area of application. It is suggested that prosodic differences between synthetic and natural speech present the major difficulty to comprehension of synthetic speech (Logan, 1989). It is generally accepted that listening to synthesized speech requires more processing capacity in maintaining short-term memory than does natural speech. It is evident that the lack of intelligibility of words and meaningful sentences while using synthesized speech systems results in the degradation in short-term memory recall. However, once synthesized speech is encoded, it is stored just as efficiently as natural speech (Sanders and McCormick, 1993).

Conclusion

Past research has shown that wearable computer systems can, in certain applications, improve the user's performance. Voice recognition is also an implementation that has shown favorable results in HCI. The question remains, will combining these current technologies with the task steps presented in the auditory mode improve the technician's performance? Based on previous studies in the field of human performance, we expect to find improvements in the technician's task performance and satisfaction while using the auditory mode. The next chapter will explain the methodology used for this experiment.

III. Methodology

Chapter Overview

There are many different Portable Maintenance Aids (PMAs) that are available and even more being tested for those that need their hands free in their specific job environment. The difficulty is selecting the right PMA for the right job while considering certain requirements such as durability or even the PMA's capabilities. This is especially evident in the aircraft maintenance arena. PMAs can be an excellent tool for the aircraft technician. In order to find the right PMA for the aircraft technician, the advantages and disadvantages must be considered and weighed. The goal of this study was to determine the effect of using different sources for manipulating technical orders in order to complete a maintenance task. This chapter explains the experimental methodology that was used in order to evaluate the differences in technician performance. First, the discussion of the experimental design and the different hypotheses were tested. Then there is a description of the equipment that was used in the experiment followed by a discussion of the subjects and the task chosen for the experiment. Finally, there is an explanation of the data collected and analyses used necessary to support or negate the hypotheses.

Experimental Design

There were a total of 9 maintenance technicians selected. They performed a maintenance task using three different methods of presenting checklist steps from an Air Force maintenance technical order. The 9 technicians were divided into three different groups of three according to the shift that they were working. All three groups performed the task using all three methods in a counter-balanced sequence ensuring that the three different groups used the media presentations in three different orders (i.e. no medium had a sequential disadvantage). Then there is a discussion of the experimental variables and controls as well as a discussion of the repeated measures design used (Montgomery, 118).

Variables

This experiment looked at a single variable. The only independent variable of this study was the type of media presentation being used. The three different methods of presentation in this study were paper, visual (HMD), and auditory. The dependent variables that were used to determine the effect of the different systems were completion times of the task, the magnitude of errors, and customer satisfaction. The completion time was measured from the beginning to the end of the task. Errors (for reasons explained later) were dropped from the study. Customer satisfaction was measured by the ratings and rank order in preference of the different systems obtained by the technician following the study. The technicians were constantly reminded to critique the different systems and its usability not the size, comfort, or appearances due to the fact that the HMD and auditory systems were still prototypes.

Controls

In order for the experimenters to keep even measurement throughout the study, an experimental plan, Appendix A, was developed. This experimental plan was used to ensure the standardization of presentation of instructions with each technician for each test session. All data collection was accomplished at the same location. Not only were the same experimenters conducting the tests and data collection, but also videotape was used to analyze any discrepancies. Each subject received the same training on how the system works and how to use the different methods of TO presentation. The same input commands were used in both the HMD presentation and the auditory presentation. Alternating the order of the presentation method among the three groups of technicians controlled for learning effect. The same subject began the task using the different methods at approximately the same time each day (so that between subject effects might be more consistent).

Experimental Hypotheses

The following hypotheses are used in analyzing the technician performance using the three different presentation methods.

I. The time to complete the task will be faster with the auditory method than either paper or the HMD.

II. The number of input errors with the auditory method will be less than either the HMD or paper.

III. Customer satisfaction with the auditory method will be greater than either paper or the HMD.

Hardware

At the time of this study the paper form was the current method that was used Air Force wide. The technician hand carried the TO to the job site and referred to it for the steps taken to complete the task. There was no actual hardware associated with the paper method. The HMD method consisted of a vest-mounted CPU, a head-mounted display (HMD), and a voice recognition input device. The auditory method consisted of a vest-mounted CPU, an auditory speech synthesis system, and a voice recognition input device.

Software

The voice recognition software used for the HMD and the auditory methods was Verbex 4.0 program and the voice recognition input device was simply a voice recognition microphone. Both the HMD and the auditory methods were paralleled as closely as possible to the paper form in order to control for deviation within the presentation of the steps required to accomplish the maintenance task. Each user was required to train the software to recognize his/her voice before each task was performed while using the HMD and the auditory methods.

Tasks

There was one aircraft maintenance task, Nosewheel Shimmy Analysis Checkout, required to evaluate the technician performance using the three different methods. There were many considerations taken into account while selecting this task. First, there was a time consideration; the overall task must not be over an hour from start to finish to ensure it did not exceed the battery life. Second, the task had to be one that could be paralleled as evenly as possible among the three different methods. The task had to be one that was accomplished on a regular basis and required the technician to work with hands free and the ability to move about the aircraft. These considerations were taken to effectively evaluate the performance differences among the different presentation methods.

Subjects

The nosewheel shimmy checkout procedure is the responsibility of the Aircraft Repair (AR) Flight. There are approximately 25 maintenance technicians in the 4th Maintenance Repair Squadron (MRS) of the 4th Fighter Wing (FW), Seymour Johnson AFB, NC. There were three technicians selected from each shift based solely on availability (due to the operational tempo of the 4th FW). At the time of this study the 4th FW was an operational wing ensuring total compliance and adherence to current TO's and regulations.

Background information was collected on each technician ensuring that they were as similarly paralleled as possible. The technicians had received the required training on the selected task prior to our experiment.

Data Collection

Both quantitative task performance and self-reported user satisfaction data were collected during this experiment. The self-reported user satisfaction data was collected to evaluate hypothesis III while quantitative data was collected to evaluate hypotheses I and II (unsuccessfully for hypothesis II - see below).

Self-Reported User Satisfaction

Self-reported user satisfaction data was collected to evaluate the preference and satisfaction level of the technicians while using the three different presentation methods. They gave us a sense of how well the system worked for them and how user friendly the systems were going to be in the field. This was collected by using several different scaled questions pertaining to the level of the user's satisfaction while interacting with the three different systems. These questions were selected based on the relevance of how well the user was able to interact with the system including user's reactions to the system, learning difficulty, and system evaluations. Similar type of questions were used in the research of Chapman and Simmons (1991) which compared voice recognition input versus keypad input. A sample of the questionnaires can be found in Appendices C, D, E, and F.

Quantitative Performance

Quantitative performance data was collected to evaluate hypotheses I and II. The task completion times were recorded by a built-in computer timer while using the HMD and the auditory presentation methods. The command input errors were recorded by the computer software while using the HMD and the auditory presentation method. The experimenter and video recordings of the technicians accomplishing the task verified all of the collected data.

Data Analysis

There was both self-reported user satisfaction and quantitative data collected during this experiment. Our efforts were intended to allow analysis of the following hypotheses.

Hypothesis I

The time completion data was imported into the SPSS program to analyze the distribution. The differences between the means were then compared to find any significant difference between the three presentation methods.

Hypothesis II

The number of command input errors were defined as both the number of times that the computer did not properly understand or execute the command and the user could not understand or read the technical data. This hypothesis only included the HMD and the auditory methods. Again, the collected data was to be put into the SPSS program to analyze the distribution and means were then to be compared to find any significant differences (however, we had to abandon this effort - see below).

Hypothesis III

This self-reported user satisfaction data was collected through questionnaires, numbered responses, and preferences among the three presentation methods. This data was used to evaluate the overall satisfaction of the technicians concerning the three methods. The questions were used to gain insight to the users' preference and gain additional information concerning the different methods. The numbered responses were used to evaluate the customer satisfaction with each presentation method. The numbered ranking was used to compare the order of preference concerning the different presentation media.

Summary

This chapter explained the methodology used to evaluate the differences between the technician performances while using the different presentation media. The experimental design and hypotheses were discussed. Next a description of the hardware and software used was described. Then a discussion of the task and subjects chosen was given. Next the intended data collection and analyses process was described. The results and analyses of the collected data follow in Chapter IV.

IV. Results and Analysis

Chapter Overview

This experiment was designed to collect data using the different systems in three different categories: task completion times, number of user input errors, and user satisfaction. Task times were recorded at intervals for the auditory and the head mounted systems using the CPU with the wearable vest. The task times for the paper system were recorded manually at the corresponding intervals. Due to certain constraints that were out of the experimenters' control, the number of user input errors could not be recorded (see discussion below). The 9 subjects performed qualitative evaluations of the three different systems (i.e. media) after each time the maintenance task was completed. There were several questions for each system that evaluated the user satisfaction based on a five-point scale along with open-ended questions discussing likes and dislikes about each system (Appendices C, D, and E). There was also an overall questionnaire that was conducted by the experimenter orally comparing the overall impression of the different systems (Appendix F).

Quantitative Performance Results

The task times were collected at four different intervals throughout the maintenance task: the first, second, and third measurements including the task completion time. These times were collected for all nine subjects and were compared using a repeated measures design with SPSS checking for significant differences between the performance time mean values of the different systems. The different user satisfaction questions were also compared across media.

Hypothesis I - Task Completion Times

This hypothesis predicted that the completion times would be faster using the auditory system than with either the HMD system or the paper. All nine subjects were able to complete the task. Task completion time is defined as once the user says "Start Checklist" until the user says "Checklist Complete". Subject 8 experienced the only severe technical difficulties encountered during the data collection period. There was an aircraft at idle in which the intakes were facing directly at us. This caused the system not to correctly recognize the users commands. However, when it began to significantly influence the task time, the experimenters decided to stop the task, wait until the aircraft shutdown the engines, and start the task again. Data collected for subjects one through ten are shown below in Table 3(the tenth subject was not from the R/R flight but volunteered).

Table 3. Task Completion Times

	Auditory	HMD	Paper
Subject 1	24.15	13.53	12.15
Subject 2	23.85	16.6	12.15
Subject 3	13.87	20.25	10.88
Subject 4	16.58	17.77	9.13
Subject 5	19.9	18.73	18.83
Subject 6	30.3	15.18	32.33
Subject 7	19.48	16.53	17.67
Subject 8	22.13	26.15	14.17
Subject 9	24.77	13.5	17.85
Subject 10	17.08	13.3	16.65

Using SPSS the means were compared for significant difference for both within subjects and between subjects effects. The means for audio (1), HMD(2), and paper(3) are 21.211, 17.154, and 16.181 minutes respectively. Table 4 shows the means and standard deviations.

Table 4. Means and Std Dev.

	Mean	Std. Dev.
Audio	21.21	1.523
HMD	17.15	1.244
Paper	16.18	2.075

Upon visual inspection, there appeared to be a substantial difference between the auditory system and the paper system. After looking at the data set and the resulting times, we went back and looked at the task at hand. Noting that it was not recommended to use an auditory system when there are long sentences or instructions, it was observed that in the Technical Order there were several long "Notes" and "Warnings" in the beginning and end of the task. In order to evaluate the systems on an equal level, it was decided to evaluate this corrected task time without the long "Notes" and "Warnings" only using the time collected between the first and third measurement. This decreased the times significantly. The corrected completion times for each subject using the different systems are shown in Table 5.

Table 5. Corrected Completion Times

	Auditory	HMD	Paper
Subject 1	3.43	6.13	5.55
Subject 2	7.54	7.4	5.86
Subject 3	4.11	7.45	6.01
Subject 4	5.87	9.16	4.09
Subject 5	7.43	8.5	7.87
Subject 6	10.74	5.87	10.65
Subject 7	5.34	7.28	10.52
Subject 8	7.43	14.12	7.8
Subject 9	8.66	5.81	6.93

The new corrected task time means for the auditory, HMD, and paper systems are 6.69, 7.74, and 7.16 minutes respectively. The means and standard deviations are shown below in Table 6.

Table 6. Corrected Means and Std Dev.

	Mean	Std. Dev.
Audio	6.73	.763
HMD	7.97	.858
Paper	7.25	.739

Analyzing these corrected completion times via repeated measures analysis of variance, between subjects differences were significant ($F = 212.3$, sig. level $< .01$), but differences between the media presentation systems were not significant ($Chi^2 = 1.56$, sig. level = .46 via Friedman's non-parametric ANOVA for matched samples at k-levels). A non-parametric repeated measures approach (Devore, 1982) was used to test within subjects effects because the appropriate parametric test assumes sphericity (i.e. once between subjects effects are removed, within subjects effects yield negligible correlation and similar variances in the matched samples). This assumption is traditionally tested via the Mauchly Spherecity Test. Our sample size was too small to test this assumption (i.e. for the three matched samples, the Mauchly Test is trying to assess 6 parameters simultaneously - in our case, with N=9). Friedman's matched sample ANOVA was one approach recommended by Cooper and Emory (1995). By just comparing the mean values, the auditory system was faster than the others were but the difference is not statistically significant.

Summary of Results for Hypothesis I

Using the first data set of task completion times there appeared to be substantial differences between the auditory and the paper system. The auditory system was actually significantly slower than the paper system. However, after correcting the task completion times, it was observed that there were not any statistically significant differences between the different systems. Therefore, Hypothesis I was not supported.

Hypothesis II - User Input Errors

This hypothesis predicted that there would be less user-input errors with the auditory system than with either of the other systems. The plan for this hypothesis was to use a false measuring tool and collect the number of times that a subject incorrectly measured the nose-strut. However, due to multiple constraints this hypothesis was not successfully tested. First, the available aircraft that was used did not have a Nose-Wheel Steering Unit; therefore, the free-play could not be measured and was simulated. Secondly, the false measuring tool was not scaled to the proper increments and therefore was invalid.

Self-Reported User Satisfaction Results

Hypothesis III - User Satisfaction

This hypothesis predicted that the user satisfaction would be greater using the auditory system than with either the paper system or the HMD system. Each subject answered a questionnaire directly after using the different systems. Each of these questionnaires (see Appendices C, D, and E) dealt directly with that immediate system used. These different questionnaires used a 5-point scale to rate satisfaction and had open-ended questions (also pertaining to user satisfaction). After the subject had used all three systems and answered the different questionnaires, an oral interview was conducted asking several open-ended questions comparing and contrasting the three different systems. We also asked the subjects to rank the systems in order of preference.

Internal Consistency Analysis of User Satisfaction Surveys

Crombach's alpha analysis suggested that the internal consistency of our satisfaction scales (taken together) would be improved by deleting the system speed appraisal item from each sub-scale. Responses to this item did not seem to vary systematically with user satisfaction. This left us with two six-item surveys (HMD and Auditory) and one four-item survey (paper). First, we tested the internal consistency of the HMD and Auditory scales. We then calculated the internal consistency of the four items, which were common among all three scales. Given our intended comparisons, we calculated the following five internal consistency alphas:

Six- item auditory scale alpha = .6319

Six-item HMD scale alpha = .6811

Four-item Paper scale alpha = .6887

Four- item auditory scale alpha = .5885

Four-item HMD scale alpha = .5079

These calculated alphas are moderately high enough to suggest that our different survey scales are each measuring a somewhat uni-dimensional attitude (regarding the given checklist media), which we call user satisfaction.

Analysis of User Satisfaction Data

Since the three surveys had only 4 items in common, we tested the user response means in two steps. First, we averaged each subject's attitude towards the two technology-augmented systems and compared these to their attitudes towards the conventional paper form (using the four items that the three surveys had in common). We used the following non-parametric paired sample means comparison (at Type I error = .05):

Table 7. Sample Means: Audio/HMD vs. Paper

	Avg of Audio and HMD	Avg of Paper
Subject 1	4.5	4
Subject 2	4.92	4.75
Subject 3	3.83	4.25
Subject 4	4.42	4.75
Subject 5	4.5	4
Subject 6	4.75	4.25
Subject 7	4.33	4.75
Subject 8	4.08	3.75
Subject 9	4.17	3.25
Avg Mean	4.389	4.194

Using a Wilcoxon Signed Ranks Test to analyze the above means resulted in a significance level of .19 (not statistically significant). Given that the subjects slightly (but not significantly) preferred the technology-augmented systems to the conventional paper form, we then compared the full 6 item surveys assessing the attitudes about the audio condition versus the HMD condition (since these surveys shared like items). The numbers are outlined in Table 8 below.

Table 8. Sample Means: Audio vs. HMD

	Avg of Audio	Avg of HMD
Subject 1	4.67	4.33
Subject 2	4.83	5.00
Subject 3	3.83	3.83
Subject 4	4.17	4.67
Subject 5	4.33	4.67
Subject 6	4.67	4.83
Subject 7	4.33	4.33
Subject 8	3.67	4.50
Subject 9	3.83	4.50
Avg Mean	4.25	4.51

Again using a Wilcoxon Signed Ranks Test to analyze the above means resulted in a significance level of .075 (not statistically significant). Apparently, the subjects slightly (but not significantly) preferred the HMD system to the audio system.

Since the item assessing the user's satisfaction of the media system's speed failed to behave consistently as a facet of user satisfaction (see internal consistency section - page 28), we were forced to analyze it separately. The correlation between subject speed (i.e. task completion time) and each subject's impression of the system speed was non-significant for Auditory (sig. level = .17, N = 9) and also for HMD (sig. level = .41, N = 9). In other words this "perceived system speed" item doesn't seem to vary systematically with our logically related variables. Since this item was not used for the paper condition, we were left with only a direct comparison between HMD and auditory (and of course, we were skeptical of any analysis of the item given its lack of clear relations with anything we measured). The mean "perceived system speed" response for the HMD was 4.44 (N = 9). The mean response for the auditory was 4.11 (N = 9). This difference was tested via a Wilcoxon and was non-significant (p = .18, N = 9).

We also looked at the results of the individual areas in which the surveys collected data. These different areas are analyzed below for the three different systems.

Auditory System Questions (see Appendix E)

The questions pertaining to the auditory system were broken into three areas. The areas covered overall reaction to the system, learning, and system evaluation. Two questions pertained to overall reaction to the system, three pertained to learning the system, and two pertained to system evaluation. A summary of the means for each question is shown below in Table 9.

Table 9. Auditory System Question Means

Question	Mean
1	4.22
2	4.00
3	4.22
4	4.11
5	1.11
6	4.11
7	4.11

Discussion of Auditory System Questions

Question 1. Evaluated the user's overall reaction to the system with the answers ranging from "Bad" to "Good." The mean response was 4.22 that indicates a slightly positive reaction to the system.

Question 2. Evaluated the user's reaction to the use of the system with answers ranging from "Difficult" to "Easy." The mean response was 4.00 that indicates a slightly positive reaction to the system use.

Question 3. Evaluated the user's ability to learn the operation of the system with answers ranging from "Difficult" to "Easy." The mean response was 4.22 that indicates a slightly positive reaction to learning the system operation.

Question 4. Evaluated the user's ability to learning and remembering the input commands with answers ranging from "Difficult" to "Easy." The mean response was 4.11 that indicates a slightly positive reaction to learning and remembering the input commands.

Question 5. Evaluated the use of a "Reference Sheet" with answers ranging from "Infrequent" to "Frequent". The mean response was 1.11 that indicates infrequent use of the "Reference Sheet" for using the input commands.

Question 6. Evaluated user's impression of the systems speed with answers ranging from "Slow" to "Fast." The mean response was 4.11 that indicates a slightly positive reaction to the system speed.

Question 7. Evaluated the systems ability to navigate through the Technical Order with answers from "Difficult" to "Easy." The mean response was 4.11 that indicates a slightly positive reaction to the system's ability to navigate through the Technical Order.

Open-Ended Auditory Questions

There were two questions pertaining to the users likes and dislikes of the auditory system. Question one focused on the users likes of the auditory system. The majority of the subjects enjoyed the hands-free aspect of the auditory system. The second question pertained specifically to the dislikes of the users while using the auditory system. Most subjects complained about the computer generated voice being hard to understand and that the system was bulky.

The self-reported data collected pertaining to auditory system show somewhat of a positive acceptance. The subjects seem to be satisfied with the system's ease of use, the speed, and the hands-free capability. However, they seem to dislike the bulkiness and computerized voice. The open-ended questions again support these findings.

HMD System Questions (see Appendix D)

The questions pertaining to the HMD system were broken into three areas. The areas covered overall reaction to the system, learning, and system evaluation. Two questions pertained to overall reaction to the system, three pertained to learning the system, and two pertained to system evaluation. A summary of the means for each question is shown below in Table 10.

Table 10. HMD System Question Means

Question	Mean
1	4.22
2	4.00
3	4.88
4	4.66
5	1.22
6	4.44
7	4.55

Discussion of HMD System Questions

Question 1. Evaluated the user's overall reaction to the system with the answers ranging from "Bad" to "Good." The mean response was 4.22 that indicates a slightly positive reaction to the system.

Question 2. Evaluated the user's reaction to the use of the system with answers ranging from "Difficult" to "Easy." The mean response was 4.00 that indicates a slightly positive reaction to the system use.

Question 3. Evaluated the user's ability to learn the operation of the system with answers ranging from "Difficult" to "Easy." The mean response was 4.88 that indicates a slightly positive reaction to learning the system operation.

Question 4. Evaluated the user's ability to learning and remembering the input commands with answers ranging from "Difficult" to "Easy." The mean response was 4.66 that indicates a slightly positive reaction to learning and remembering the input commands.

Question 5. Evaluated the use of a "Reference Sheet" with answers ranging from "Infrequent" to "Frequent". The mean response was 1.22 that indicates infrequent use of the "Reference Sheet" for using the input commands.

Question 6. Evaluated the user's impression of systems speed with answers ranging from "Slow" to "Fast." The mean response was 4.44 that indicates a slightly positive reaction to the system speed.

Question 7. Evaluated the systems ability to navigate through the Technical Order with answers from "Difficult" to "Easy." The mean response was 4.55 that indicates a slightly positive reaction to the system's ability to navigate through the Technical Order.

Open-Ended HMD System Questions

There were two questions pertaining to the users likes and dislikes of the HMD system. Question one focused on the users likes of the HMD system. The majority of the subjects enjoyed the hands-free aspect of the HMD system and seemed to find it easy to read. The second question pertained specifically to the dislikes of the users while using the auditory system. Most subjects complained about the HMD display blocking their vision and that the rest of the system was bulky.

The self-reported data collected pertaining to HMD system show somewhat of a positive acceptance. The subjects seem to be satisfied with the system's ease of use, the speed, and the hands-free capability. Several subjects commented on the convenience of have the step right there at all times. However, they seem to dislike the bulkiness and had a fear of damaging the display due to it being in the way. The open-ended questions again support these findings.

Paper System Questions (see Appendix C)

The questions pertaining to the paper system were broken into three areas. The areas covered overall reaction to the system, learning, and system evaluation. Two questions pertained to overall reaction to the system, three pertained to learning the system, and two pertained to system evaluation. A summary of the means for each question is shown below in Table 11.

Table 11. Paper System Question Means

Question	Mean
1	4.44
2	3.89
3	4.22
4	4.22

Discussion of Paper System Questions

Question 1. Evaluated the user's overall reaction to the system with the answers ranging from "Bad" to "Good." The mean response was 4.44 that indicates a slightly positive reaction to the system.

Question 2. Evaluated the user's reaction to the use of the system with answers ranging from "Difficult" to "Easy." The mean response was 3.89 that indicates a slightly positive reaction to the system use.

Question 3. Evaluated the user's ability to learn the operation of the system with answers ranging from "Difficult" to "Easy." The mean response was 4.22 that indicates a slightly positive reaction to learning the system operation.

Question 4. Evaluated the systems ability to navigate through the Technical Order with answers from "Difficult" to "Easy." The mean response was 4.22 that indicates a slightly positive reaction to the system's ability to navigate through the Technical Order.

Open-Ended Paper System Questions

There were two questions pertaining to the users likes and dislikes of the paper system. Question one focused on the users likes of the paper system. The majority of the subjects enjoyed the different diagrams that are accompanied with the paper system and felt they were biased since they were already use to this form. The second question pertained specifically to the dislikes of the users while using the paper system. Most subjects complained about having to constantly refer back to the TO in order to complete the task. They also complained of having to carry the TO with them during the task.

The self-reported data collected pertaining to paper system show somewhat of a positive acceptance. The subjects seem to be satisfied with the system's ease of use. However, there were several negative comments concerning the paper system. Not only do the pages in the TO get dirty, but they have been blown away, gotten torn out, and are hard to maintain. The open-ended questions again support these findings.

Summary of Results for Hypothesis III

Results show that the subjects were pleased and satisfied with all three systems. However, there definitely were disadvantages to all three. When asked to rate the systems in order of preference, technicians slightly preferred the HMD to the auditory form. This is summarized in the Table 12.

Table 12. System Rank Order

	Auditory	Video	Paper
Subject 1	1	3	2
Subject 2	3	1	2
Subject 3	3	1	2
Subject 4	3	2	1
Subject 5	2	1	3
Subject 6	1	3	2
Subject 7	1	3	2
Subject 8	2	1	3
Subject 9	2	1	2
Subject 10	2	1	3
Avg.	**2**	**1.7**	**2.2**

(lowest average being the preferred)

Therefore, Hypothesis III was rejected due to the fact that the subjects apparently preferred the HMD system to either the auditory or the paper system. The information in Table 12 should be viewed with extreme caution. It is not treated statistically since we forced within-subjects variance artificially (i.e. we failed to offer our subjects the option of recording ties in these ranks). Given this artificially forced variance, these differences are probably trivial in size.

Summary

Three types of data were planned on being collected. However, only two types of data were actually collected for this experiment: task completion times and user self-reported satisfaction with each system. For the auditory system and the HMD system the CPU collected the task times while the task times with the paper system were collected manually by the experimenter. Self-reported evaluations were taken from each subject after the use of each system. There were seven questions evaluating the auditory and the HMD systems, while there were only four questions used to evaluate the paper system.

The difference between surveys was due the fact that the paper form did not require input commands or system quickness evaluations. Hypothesis I was not supported by the data collected due to the lack of significant difference. Hypothesis II was not analyzed due to constraints out of the experimenters' control. Hypothesis III was not statistically supported; however, it did suggest that the subjects preferred the HMD system to the auditory system.

V. Findings and Conclusions

Chapter Overview

This chapter discusses the results found during this experiment. There is a discussion of all three experimental hypotheses and whether they were supported or not. There is a discussion of both the technicians' and our own recommendations for system improvements that were discovered during this experiment. Recommendations for further research are presented. We then present our final conclusions.

Discussion of Quantitative Performance

Quantitative performance data was collected for the first experimental hypothesis. Task completion time data was collected for Hypothesis I. There was supposed to be data collected for the second hypothesis, but due to constraints discussed earlier this was not possible.

Hypothesis I

The data collected during this experiment does not support experimental Hypothesis I. Hypothesis I stated that the task completion times would be faster using the auditory system than either the HMD system or the paper form. There was not a statistically significant difference across the media systems. However, comparing the mean corrected task completion times, the auditory system was faster than the others (though insignificantly). We believe there were several factors contributing to the lack of statistical significance. These factors include type of task used, the type of user evaluated, and outside constraints.

The task that was used during this experiment was one that the technicians were familiar with. However, some of the technicians had performed the task more than others. The task really had no set guidelines on how to actually perform certain steps within the task. Therefore, there were several different techniques of accomplishing the task. Some of the steps that were presented by the HMD and the auditory systems did not follow some of the recommended guidelines in Table 2. For example, some of the steps were probably too long for auditory use. The task selected was limited to only the technicians from the Aircraft Repair Flight; therefore, limiting the number of available participants and constraining statistical power.

Hypothesis II

Data was not collected in support of experimental Hypothesis II. Due to constraints outside our control, the technicians could not make measurements on the nosewheel landing gear with sufficient precision.

Discussion of Self-Reported User Satisfaction

Self-reported data was collected via questionnaires using a 5-point scale to rate satisfaction in support of Hypothesis II. Each of these questionnaires (see Appendices C, D, and E) dealt directly with that immediate system used. After the subject had used all three systems and answered the different questionnaires, an oral interview was conducted asking several open-ended questions comparing and contrasting the three different systems. We also asked the subjects to rank the systems in order of preference.

Hypothesis III

Hypothesis III predicted that the user satisfaction would be greater using the auditory system than with either the paper system or the HMD system. Hypothesis III was not supported due to the lack of significant difference between the data collected for each system. However, the subjects slightly (but not significantly) preferred the technology-augmented systems to the conventional paper form. Even further analysis revealed that the majority slightly (but not significantly) preferred the HMD to the auditory system.

Improvements

There are several areas that both the technicians and the experimenters feel could be improved (in creating a better system). These areas include improvements with the software and hardware.

Software

The software that was used in the HMD and the auditory system could be improved to create a more user-friendly system. First, the voice used in the auditory system was a computer-generated voice and very monotone. Both of these characteristics proved negative according to the technicians. The technicians stated that it would be easier to understand the voice if it were more human-like and had normal fluctuations and pauses typical of everyday conversation. The eyepiece on the HMD proved to work well; however, the technicians commented that it would be nice if there were diagrams displayed in the HMD. Other than the tedious computer-training portion of the voice recognition system, it also proved to work very well. The noise-canceling microphone worked great even in the noisy environment of the hangar.

Hardware

The largest complaint from the technicians concerning the hardware was the bulkiness. The vest itself was bothersome and hot while working in the heat of the flight line. The vest could be made of a breathable mesh-like material and have pockets designed specifically for the equipment needed to run the system. The HMD was described as "being in the way" by the technicians. Using better technology, a smaller display could be implemented in the HMD system. It was suggested that combining two media (auditory and HMD) might provide a more effective system.

Recommendations for Further Research

During the accomplishment of this experiment it was obvious that more research would have to be done in order to get a final working product. I still believe that voice recognition should be pursued as a source of input. An additional study could reveal more significant results and prove to be beneficial in the technology arena. Future research should consider task selection and subject size used.

There are several factors that must be considered when choosing the task to be used in the experiment. It must be a step-by-step process without any long areas of text. For example, in the task used for this experiment we had long "Notes" and "Warnings". The complexity of the task plays a large role in how well the subjects understand the steps needed to be accomplished. Unless the HMD and auditory systems are combined, the task must not contain any diagrams in order to accomplish the task.

When selecting the subjects to be used you must consider their expertise with the task selected. Either they must be very familiar with the task or have never accomplished the task before. This puts the subjects on a more comparable level as far as ability to complete the task (i.e. with less variance between subjects). Finally, one should make sure there is larger pool of subjects for selection.

Conclusion

The two technology-augmented systems may someday be beneficial in the aircraft maintenance arena. Even though there was not a statistically significant difference across the three systems, the technicians did slightly favor the technology-augmented systems to the paper form. The fact that the technicians favored the technology-augmented systems is possibly of some importance. The future may hold a place in aircraft maintenance for these types of systems.

Appendix A. Experimental Plan

Description of Evaluation

Purpose

The purpose of this study is to evaluate performance differences between technicians performing tasks using three different methods of presenting the required steps from Air Force maintenance technical orders (TO). The three methods being used to present the information are paper, HMD, and auditory forms.

Hardware

There are different types of hardware for the different presentations. The first media presentation had no particular hardware due to the fact that it was in paper form. The technician merely reads the maintenance steps directly from the paper form of the TO. The other media presentations had similar pieces of hardware. The visual method consisted of a vest-mounted computer, a HMD used to display the information, and a voice recognition device. The auditory method used the same type vest and voice recognition device but the computer provided the information using auditory means. Therefore, it consisted of an auditory speech synthesis system.

Software

The voice recognition software used for the HMD and auditory methods for this experiment consists of Verbex. For the auditory method Verbex developed the speech synthesis.

Subjects

There were a total of 9 maintenance technicians from the 4th Equipment Maintenance Squadron used in this experiment. The technicians were divided into three groups of three. All three groups performed the task using the three different methods in the various different orders.

Tasks

Each maintenance technician performed the task with the three different methods: paper (task 1), HMD (task 2), and auditory (task 3). Using a Latin Square design, the first group performed task 1 first followed by task 2 and 3. The second group performed task 2 first then 3 and 1. The third group performed task 3 first then 1 and 2.

Conditions

The subjects were assigned according to availability and familiarity with the required maintenance task. The subjects had to have at least a skill level 3 and at least 1 year on the weapon system (F-15E). The order of the different methods used was alternated to control for maturation effects.

Hypothesis

1. Time to complete the task will be faster than either paper or HMD using the auditory method.

2. The number of errors (input) will be less than either paper or HMD using the auditory method.

3. The customer satisfaction will be greater when using the auditory method than when using either paper or HMD to complete the task.

Data Collection

The computer system collected the start and stop times of the task. During the experiment, notes, observations, and input errors were documented by the experimenter on note paper and by using videotapes. Videotapes were used if the data collected for an individual seemed to be an outlier. A questionnaire was administered following the experiment to measure the satisfaction of the subjects using each method. The following information was collected during the experiment:

- Task completion time
- Command input errors
- User satisfaction data

Controls

The following actions were performed to control for experimental variation:

1. All data collection was performed at the same location.

2. All subjects were assigned to one of the three groups.

3. Both computer methods, HMD and auditory, were as closely paralleled as possible to control for variation in the presented material when using the different methods.

4. Maturation effect was controlled by alternating the order which the subject uses the different method.

5. The same subject began the task using the different methods at approximately the same time each day.

6. The subjects received identical training on how to use the different systems.

7. All questionnaires were performed at a private location.

Conducting the Experiment

Sequence of Events (for group one)

Day One

- First, the technicians of Group 1 (day shift) received the shift inbrief.
- A technician of Group 1 will then be given the individual briefing and introduced to the system that they will use that day (HMD for example).
- The technician will then train the system to recognize his/her particular voice.
- The technician will then be trained on the different voice input commands that will be used with the system.
- The same technician will then perform an introductory task to ensure they are comfortable with different commands and when to use them.
- The technician will then perform the maintenance task using the HMD.
- The technician will complete HMD questionnaire.
- The same steps above are performed with the other technicians of Group 1 (if available).
- Same process above using the paper method with Group 2 (swing shift).
- Same process above using the auditory method with Group 3 (mid shift).

Day Two

- Same process as above with Group 1 using the paper method.
- Same process as above with Group 2 using the auditory method.
- Same process as above with Group 3 using the HMD method.

Day Three

- Same process as above with Group 1 using the auditory method.
- Same process as above with Group 2 using the HMD method.
- Same process as above with Group 3 using the paper method.

Note: Due to technicians' work schedule, task completion time and time to train each person, This process will more than likely take more than three days.

Introduction

The technicians received a briefing of the purpose of the experiment and the instructions required to complete the experiment. The briefing covered the background of the research, responsibilities of the technicians, what data was being collected, how the data will be used, and the privacy of their performance and responses.

Training

Technicians received an introduction on the different systems and what they consist of and how they are used. They received more thorough training using the different systems immediately prior performing the task. They were given a "reference sheet" of voice commands used for input during system operation. Experimenters were available during the task to answer any questions the technicians may have.

Debriefing

Questionnaires were given to each subject after completion of each task. These were a measure of the satisfaction of the system operation. After each task completion, the subjects were debriefed. The debrief included thanking the technician for his/her participation, any other feedback, instructions not to talk with other subjects about the experiment until all the proper data had been collected, and explaining how their data was to be used in the experiment.

Appendix B. Personal Background Form

Personal Background

Name: __________________________ Subject #:_________

1. Circle one: Military/Civilian
2. Time in Service:____________(yrs/mos)
3. Rank:___________
4. Job Title (AFSC):________________
5. Number of years on current Weapon System:_____________(yrs/mos)
6. Skill level:__________

Appendix C. Paper Questionnaire

Subject #:________

Overall reactions to paper form

Bad Good
1 2 3 4 5

Difficult to use Easy to use
1 2 3 4 5

Learning

Learning how to use paper form
Difficult Easy
1 2 3 4 5

Paper evaluation

Navigation through technical order
Difficult Easy
1 2 3 4 5

Open Ended

What did you like about the paper form?

What did you not like about the paper form?

Appendix D. HMD Questionnaire

Subject #:________

Overall reactions to system

Bad Good
1 2 3 4 5

Difficult to use Easy to use
1 2 3 4 5

Learning

Learning operation of system
Difficult Easy
1 2 3 4 5

Remembering input commands
Difficult Easy
1 2 3 4 5

Use of "Cheat Sheet"
Infrequent Frequent
1 2 3 4 5

System evaluation

System speed
Slow Fast
1 2 3 4 5

Navigation through technical order
Difficult Easy
1 2 3 4 5

Open Ended

What did you like about the HMD system?

What did you not like about the HMD system?

Appendix E. Auditory Questionnaire

Subject #:________

Overall reactions to system

Bad Good
1 2 3 4 5

Difficult to use Easy to use
1 2 3 4 5

Learning

Learning operation of system
Difficult Easy
1 2 3 4 5

Remembering input commands
Difficult Easy
1 2 3 4 5

Use of "Reference Sheet"
Infrequent Frequent
1 2 3 4 5

System evaluation

System speed
Slow Fast
1 2 3 4 5

Navigation through technical order
Difficult Easy
1 2 3 4 5

Open Ended

What did you like about the Auditory system?

What did you not like about the Auditory system?

Appendix F. Overall System Questionnaire

Subject #:________

Open Ended

What did you like best about the Paper form?

What did you like best about the Auditory System?

What did you like best about the HMD (visual) System?

What did you not like about the Paper form?

What did you not like about the Auditory System?

What did you not like about the HMD (visual) System?

Which system would you prefer to use on a daily basis?
Rank order:

____Paper
____HMD
____Auditory

Any suggestions to improve the hardware or software (vest, HMD, voice, or input commands)?

Any suggestions for future study?

Appendix G. Shift Inbrief

I am Capt. Chastain/Dave Kancler and you are about to participate in a field experiment that is on the leading edge of technology and may have a future impact on the way the Air Force operates.

You will be accomplishing a common maintenance task using 3 different media presentations: 1) current paper form 2) HMD visual display 3) auditory presentation. The HMD and auditory presentation uses voice as a source of input; meaning you get from step to step by using your voice for the commands. Over the next three days you will perform a maintenance task three times using a different presentation media each time. The visual display and the auditory presentations allow for the technician to be hands free. We are testing the different systems we are not testing you as the user.

All information collected from you as the technician is held completely confidential. The only request we have for you is to not discuss any part of the experiment and/or equipment with any of your fellow workers for the duration of the experiment (approximately 10 days).

If you would now fill out the consent form and Background Info sheet that is in front of you. Again, this information is going to be held completely confidential. Bring your sheet up front when you are done.

Any questions?

Appendix H. Individual Briefing

You are going to be asked to accomplish a maintenance task using HMD/auditory presentation. You will be communicating with the computer using voice commands. Therefore, we have to train the computer to recognize your voice and train you on the different commands that will be used. You will be using our measurement tools and go with the measurement you read.

System Brief

This is the vest and headgear that you will be wearing during the task. While using these systems, keep in mind that they are prototypes. The fabrications are somewhat rough, but the software/hardware used is on the cutting edge of technology. Do you have any questions before we get started?

Appendix I. Human Release Form

INFORMATION PROTECTED BY THE PRIVACY ACT OF 1974

CONSENT FORM

NEW TECHNOLOGIES FOR MAINTENANCE AND LOGISTICS INFORMATION SYSTEM STUDIES

1. I have been invited to participate in studies to evaluate new technology applications to the maintenance and/or logistics planning environments. The purpose of these studies is to evaluate such factors as data recall techniques, formats, and demonstration systems for presenting technical information prior to their incorporation into test systems. Field tests will be used to evaluate demonstration systems developed using Laboratory developed techniques, software, and hardware. The studies will be designed to evaluate the various techniques and demonstration systems in terms of their ability to effectively provide the user with the required information, acceptability to the user, and ability to support the mission of the maintenance and logistics organizations.

2. My participation in this study may require me to wear a vest-mounted computer with a head-mounted display device, a monocular or bi-ocular display device and/or a head-mounted microphone activated voice recognition package, a wrist keypad, and/or mouse or similar control device. These tools will be the aids used to complete various limited maintenance and troubleshooting tasks. The performance of these tasks using any of this equipment will be monitored by one of the investigators and/or videotaped. Data will be collected on task performance (speed and accuracy), and any discomfort or limitations found with the equipment during the test

3. My participation will not involve risks greater than I encounter performing my normal duties. I understand depending on the equipment being tested, some discomfort may be encountered with headbands, goggles or glasses.

4. My participation in this study will help to ensure that the application and further development of these technologies are designed to meet my needs. The ultimate benefit of this project will be to make maintenance and logistics personnel more effective and make their jobs easier.

5. The only other way to obtain the required information would be to conduct studies in a laboratory setting using non-maintenance personnel. These people would not be representative of maintenance personnel, and the information gathered would not reflect the true needs of maintenance personnel.

6. Entitlements and Confidentiality:

a) Records of my participation in this study may only be disclosed according to federal law, including the Federal Privacy Act, 5 U.S.C. 552a, and its implementing regulations.

b) I understand my entitlements to medical and dental care and/or compensation in the event of injury are governed by federal laws and regulations, and that if I desire further information, I may contact the base legal office.

c) If an unanticipated event (medical misadventure) occurs during my participation in this study, I will be informed. If I am not competent at the time to understand the nature of the event, such information will be brought to the attention of my next of kin.

d) The decision to participate in this research is completely voluntary on my part. No one has coerced or intimidated me into participating in this program. I am participating because I want to. Ms Masquelier, AL/HRGO, DSN 785-2606, has adequately answered any and all questions I have about this study, my participation, and the procedures involved. I understand that Ms Masquelier will be available to answer any questions concerning procedures throughout this study. In understand that if significant new findings develop during the course of this research, which may relate to my decision to continue participation, I will be informed. I further understand that I may withdraw this consent at any time and discontinue further participation in this study without prejudice to my entitlements. I also understand that the medical monitor of this study may terminate my participation in this study if she or he feels this to be in my best interest.

____________________________ ______

VOLUNTEER SIGNATURE AND SSAN DATE

____________________________ ______

PRINCIPAL INVESTIGATOR SIGNATURE DATE

____________________________ ______

WITNESS SIGNATURE DATE

Bibliography

Beacker, Ronald M. and others. Human-Computer Interaction: Toward the Year 2000 (Second Edition). San Francisco: Morgan Kaufmann Publishers, 1995.

Bergman, M. and others. "Age-related Decrement in Hearing for Speech," Journal of Gerontology, 31:533-538 (1955).

Bigler, R. and others. A Further Investigation of the Relationships Between Voice Variables and Speech Intelligibility in High Level Noise. Contract N6ori-104. Lafayette IN: Purdue University, 1955.

Card, Stuart K. and Thomas P. Moran. "User Technology: From Pointing to Pondering," in Human-Computer Interaction: Toward the Year 2000 (Second Edition). Ed. Ronald M. Beacker. San Francisco: Morgan Kaufmann Publishers, 1995.

Chapman, David and J. Simmons. A Comparative Evaluation of Voice versus Keypad Input for Manipulating Electronic Technical Data for Flight Line Maintenance Technicians. MS thesis, AFIT/GAL/LAL/95S-2. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1991 (AD-B234 261).

Cooper, Donald R. and C. William Emory. Business Research Methods (Fifth Edition). Chicago IL: Irwin, 1995.

Dautrich, B. A. and others. "On the Effects of Varying Filter Bank Parameters on Isolated Word Recognition," IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-31: 793-806 (1983).

Deatherage, B. "Auditory and Other Sensory Forms of Information Presentation," in Human Engineering Guide to Equipment Design. Ed. H. Van Cott and R. Kinkade. Washington DC: Government Printing Office, 1972.

Devore, Jay L. Probability and Statistics for Engineering and the Sciences Belmont CA: Brooks/Cole, 1982.

Friend, Jeffrey A. and Randy S. Grinstead. Comparative Evaluation of a Monocular Head Mounted Display Device Versus a Flat Screen Display Device in Presenting Aircraft Maintenance Technical Data. MS thesis, AFIT/GSM/LSM/92S-8. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1992 (AD-A258 684).

Gaver, William W. and Randall B. Smith. "Auditory Icons in Large-Scale Collaborative Environments," in Human-Computer Interaction: Toward the Year 2000 (Second Edition). Ed. Ronald M. Beacker. San Francisco: Morgan Kaufmann Publishers, 1995.

Hull, A. "Reducing Sources of Human Error in Transmission of Alphanumeric Codes," Applied Ergonomics, 7(2): 75-78 (1976).

Juang, B. H. "Speech Recognition in Adverse Conditions," Computer Speech and Language, 5: 275-294 (1991).

Logan, J. and others. "Segmental Intelligibility of Synthetic Speech Produced by Rule," Journal of the Acoustical Society of America, 86(2): 566-581 (1989).

Luce, P. A., Timothy C. Feustel, and David B. Pisoni. "Capacity Demands in Short-Term Memory," Human Factors, 25(1): 17-32 (February 1983).

Masquelier, Barbara L. Comparative Evaluation of Monocular Display Devices Relative to Portable Computer for Display of Aircraft Maintenance Technical Data. MS thesis, AFIT/GSM/LSM/91S-18. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1991 (AD-B161 858).

McGrath, Joseph E. "Methodology Matters: Doing Research in the Behavioral and Social Sciences," in Human-Computer Interaction: Toward the Year 2000 (Second Edition). Ed. Ronald M. Beacker. San Francisco: Morgan Kaufmann Publishers, 1995.

Nusbaum, H. and D. Pisoni. "Constraints on the Perception of Synthetic Speech Generated by Rule," Behavior Research Methods, Instruments, and Computers, 17(2): 235-242 (1985).

O'Malley, Michael H. "Text-To-Speech Conversation Technology," in Human-Computer Interaction: Toward the Year 2000 (Second Edition). Ed. Ronald M. Beacker. San Francisco: Morgan Kaufmann Publishers, 1995.

Peacocke, Richard D. and Daryl H. Graf. "An Introduction to Speech and Speaker Recognition," in Human-Computer Interaction: Toward the Year 2000 (Second Edition). Ed. Ronald M. Beacker. San Francisco: Morgan Kaufmann Publishers, 1995.

Salvendy, Gavriel. Handbook of Human Factors and Ergonomics (Second Edition). New York: John Wiley and Sons, 1997.

Sanders, Mark S. and Ernest J. McCormick. Human Factors in Engineering and Design (Seventh Edition). New York: McGraw-Hill Inc., 1993.

Shneiderman, Ben. "A Taxonomy and Rule Base for the Selection of Interaction Styles," in Human-Computer Interaction: Toward the Year 2000 (Second Edition). Ed. Ronald M. Beacker. San Francisco: Morgan Kaufmann Publishers, 1995.

Werkowitz, J. and C. Thomas. "Why is synthetic speech harder to remember than natural speech?" in CHI-85 Proceedings. New York: Association for Computing Machinery, 1985.

White, George M. "Natural Language Understanding and Speech Recognition," in Human-Computer Interaction: Toward the Year 2000 (Second Edition). Ed. Ronald M. Beacker. San Francisco: Morgan Kaufmann Publishers, 1995.

Wiecha, Charles and others. "ITS: A Tool for Rapidly Developing Interactive Applications," in Human-Computer Interaction: Toward the Year 2000 (Second Edition). Ed. Ronald M. Beacker. San Francisco: Morgan Kaufmann Publishers, 1995.

Vita

Captain John Wesley Chastain III is from Tyler, Texas and was born on 22 April 1971 in Lufkin, Texas. He graduated from the United States Air Force Academy in 1994 with a Bachelor of Science Degree in Engineering Mechanics. John received his commission on 1 June 1994. After completing the Aircraft Maintenance and Munitions Officer Course (AMMOC) at Sheppard AFB, Texas in January 1995, he was assigned to the 4th Fighter Wing (FW) at Seymour Johnson AFB, NC.

During his tour at Seymour Johnson AFB, Captain Chastain held several maintenance positions in support of the F-15E Strike Eagle. These positions include Sortie Support Flight Officer in Charge (OIC) in the 334th Fighter Training Unit (FTU), Sortie Generation Flight OIC in the 334th FTU, Accessories Flight Commander in the 4th Component Repair Squadron (CRS). In January 1997, Captain Chastain entered the Air Force Institute of Technology at Wright-Patterson AFB, Ohio, and graduated in 1998 with a Masters degree in Acquisition Logistics Management. He then was assigned to the Avionics Integration System Project Office at Wright-Patterson AFB, Ohio as a Project Officer.

Permanent Address: 10939 Knollwood
Tyler, TX. 75703

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1998	Master's Thesis

4. TITLE AND SUBTITLE
THE EFFECTS OF THREE MEDIA PRESENTATION SYSTEMS ON MAINTENANCE TASK PERFORMANCE

5. FUNDING NUMBERS

6. AUTHOR(S)
John W. Chastain, Captain, USAF

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
2750 P Street
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GAL/LAL/98S-2

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFRL/HESR Bldg. 190
2698 G Street
WPAFB, OH 4533-7604

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

Wearable computers and Portable Maintenance Aids (PMAs) may soon be the normal way of doing aircraft maintenance in the Air Force. Currently, the Air Force uses the paper form of Technical Orders (TO's) while doing aircraft maintenance. The purpose of this thesis was to compare the effects of three different media presentations of the information used during aircraft maintenance. The three different presentations compared are the current paper form, a Head Mounted Display (HMD), and an auditory mode. An experiment was conducted to determine if there was a significant difference between the systems (in terms of task completion times and user preference). Nine F-15E maintenance technicians from the 4th Fighter Wing, Seymour Johnson AFB, NC were chosen to participate in this experiment. Each individual accomplished the same task using the three different systems. While some treatment effects were in the predicted direction, differences were not statistically significant. The results weakly suggested that these technicians preferred the newer technology to the current paper form. The primary conclusion is that there may be a use for technology-augmented checklist presentation in the aircraft maintenance arena. This research establishes a foundation for future research efforts.

14. SUBJECT TERMS
Presentation Display, Auditory Display, HMD, Media Presentation, Wearable Systems, IMIS

15. NUMBER OF PAGES
70

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

AFIT Control Number AFIT/GAL/LAL/98S-2

AFIT RESEARCH ASSESSMENT

The purpose of this questionnaire is to determine the potential for current and future applications of AFIT thesis research. **Please return completed questionnaire** to: AIR FORCE INSTITUTE OF TECHNOLOGY/LAC, 2950 P STREET, WRIGHT-PATTERSON AFB OH 45433-7765. Your response is **important.** Thank you.

1. Did this research contribute to a current research project? a. Yes b. No

2. Do you believe this research topic is significant enough that it would have been researched (or contracted) by your organization or another agency if AFIT had not researched it?
a. Yes b. No

3. **Please estimate** what this research would have cost in terms of manpower and dollars if it had been accomplished under contract or if it had been done in-house.

Man Years ____________ $ ____________

4. Whether or not you were able to establish an equivalent value for this research (in Question 3), what is your estimate of its significance?

a. Highly Significant b. Significant c. Slightly Significant d. Of No Significance

5. Comments (Please feel free to use a separate sheet for more detailed answers and include it with this form):

____
Name and Grade

____
Organization

____
Position or Title

____
Address