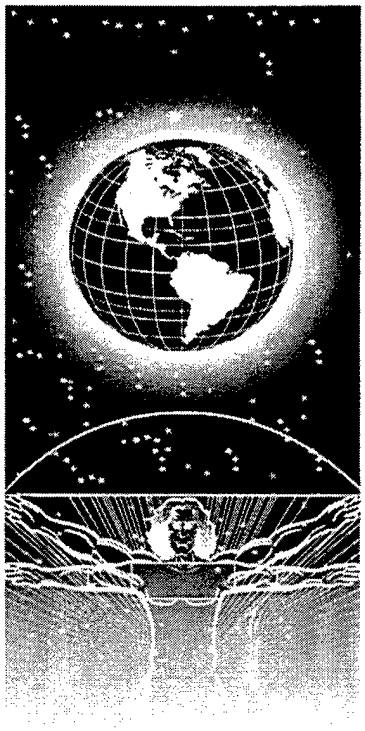
AFRL-HE-WP-TR-2002-0008



UNITED STATES AIR FORCE RESEARCH LABORATORY

Point of Maintenance Spiral One Usability Study

Carlton Donahoo
Megan Gorman
David Kancler
Laurie Quill
Allen R. Revels

University of Dayton Research Institute
Human Factors Group
300 College Park
Dayton, OH 45469-0158

August 2001

Interim Report for the Period September 2000 to August 2001

Human Effectiveness Directorate
Deployment and Sustainment Division
Logistics Readiness Branch
2698 G Street
Wright-Patterson AFB OH 45433-7604

NOTICES

When US Government drawings, specifications or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Air Force Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161

Federal Government agencies registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
8725 John J. Kingman Rd., Ste 0944
Ft. Belvoir, VA 22060-6218

DISCLAIMER

This Technical Report is published as received and has not been edited by the Air Force Research Laboratory, Human Effectiveness Directorate.

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2002-0008

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Mark M. Hoffman

MARK M. HOFFMAN
Deputy Chief
Deployment and Sustainment Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2001	Interim - September 2000 to August 2001

4. TITLE AND SUBTITLE
Point of Maintenance Spiral One Usability Study

5. FUNDING NUMBERS
C: F33615-99-D-6001
DO: 14
PE: 62202F
PR: 1710
TA: D0
WU: 09

6. AUTHOR(S)
Carlton Donahoo, Megan Gorman, David Kancler, Laurie Quill, Allen R. Revels

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Dayton Research Institute
Human Factors Group
300 College Park
Dayton, OH 45469-0158

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory, Human Effectiveness Directorate
Deployment and Sustainment Division
Air Force Materiel Command
Logistics Readiness Branch
Wright-Patterson AFB, OH 45433-7604

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-HE-WP-TR-2002-0008

11. SUPPLEMENTARY NOTES

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

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The goal of the Point of Maintenance (P-O-Mx) program is to demonstrate timely, accurate, and effective data collection, and logistics operations processing to meet the maintainer's needs at the point of origin. The Spiral One Usability Test consisted of human factors and user-centered evaluations of the P-O-Mx activity. User level evaluations were conducted to determine the maintenance effectiveness of P-O-Mx devices in presenting, modifying, and transmitting maintenance and service data. The evaluations were administered to technicians in the real-world maintenance environment. The purpose of the evaluation was to compare five hardware platforms (Libretto, Itronix, FM-Net, Xybernaut-Wrist, Xybernaut-Head) for opening work orders from the flightline aircraft location. Previous empirical studies conducted by the Air Force Research Laboratory Logistics Readiness Branch (AFRL/HESR) have evaluated both portable maintenance aids and wearable computers for flightline maintenance. Both platforms have been shown to improve maintenance performance over use of paper-based platforms. While these empirical studies have compared wearable computer technologies with portable computers, usability inspection methods (such as usability testing) have not compared these devices. This study demonstrates the usability comparisons of the five hardware platforms using Usability Test Methods.

14. SUBJECT TERMS			15. NUMBER OF PAGES
Point of Maintenance (P-O-Mx)	Portable Maintenance Aids	Wearable Computers	32
Usability Inspection Methods	Usability Test Methods	Flightline Maintenance	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

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

Introduction

The Point-of-Maintenance (P-O-Mx) Spiral 1 Usability Test consisted of human factors and user-centered evaluations of the P-O-Mx activity. The experiment emphasized operational effectiveness of P-O-Mx applications in U.S. Air Force aircraft maintenance environments.

User level evaluations were conducted to determine the maintenance effectiveness of P-O-Mx devices in presenting, modifying, and transmitting maintenance and service data. These evaluations were administered to technicians in a real-world maintenance environment. The purpose of the evaluation was to compare five hardware platforms for opening work orders from the flightline aircraft location. Usability test methods used for the current evaluation are based on principles outlined by Dumas and Redish (1993). For a full description of the method used refer to the Usability Test Plan (Quill, Kancler, Revels & Donahoo, 2001).

Given that the overall goal of this line of research is to improve flightline maintenance activities, two criteria must be considered when selecting devices to assist in achieving the goal. First, actual maintenance performance must be improved—that is, time to perform tasks must be reduced and errors must be reduced. Second and of equal importance, usability of the devices in the actual maintenance environment must be considered. No matter how much a device improves performance, if it is not usable, overall effectiveness of the device is reduced (often it will not be used at all).

Empirical studies conducted by the Air Force Research Laboratory Logistics Readiness Branch (AFRL/HESR) have evaluated both portable maintenance aids and wearable computers for flightline maintenance. Both platforms have been shown to improve maintenance performance over use of paper-based platforms (Carlson, Smith, Smith, Thomas, & Smillie, 1992; Thomas, 1995; Friend & Grinstead, 1992). In addition, empirical testing of performance has shown that wearable computing systems result in better performance than portable maintenance aids under certain maintenance conditions (Friend & Grinstead, 1992). These studies have shown that technicians perform more quickly and accurately when using a wearable computer than when using a portable maintenance aid. While these empirical studies have compared wearable computer technologies with portable computers, usability inspection methods (such as usability testing) have not compared these devices. In other words, while the technician's performance may be improved, it has not been determined how usable wearable computers are in comparison with portable aids. The purpose of the current study is to make these usability comparisons.

Background

Since 1990, the Air Force Research Laboratory (AFRL) has been researching wearable computing systems and their peripheral devices. Research methods including demonstrations, empirical tests, and usability inspection methods have provided abundant data pertaining to the usability and usefulness of wearable computing systems. Karat (1994) notes that as a system evolves, it is necessary that research, design, and development efforts include *all* of these types of research methods. The following

paragraphs review empirical studies conducted at AFRL/HESR, and emphasize the need for usability testing of these devices as well.

Over the past decade, AFRL has demonstrated wearable computer technologies in numerous settings. The purpose of these demonstrations was to allow potential users to perform an informal walkthrough using the wearable system. Potential users provided informal feedback pertaining to the potential usability of the hardware and software components. In addition to these demonstrations, many field tests have been conducted which have provided both empirical and usability inspection data.

In addition to demonstration, AFRL has conducted numerous empirical studies and usability tests concerning wearable computing systems. Results of selected studies are summarized below.

In 1990, Masquelier compared two hardware visual displays. The study compared use of a light emitting diode (LED), monocular, head mounted display (the Private Eye™) with use of a standard Grid™ laptop computer monitor (set on a table top). In addition to this hardware comparison, Masquelier evaluated the interaction of the hardware with maintainers. Two groups of maintenance technicians (experienced and inexperienced) participated in the study—sixteen subjects in total. Each subject performed two avionics communications tasks in an intermediate level maintenance task environment. Subjects sat at a workbench while performing inspection and fault isolation maintenance on computer circuit boards—mobility around the flightline was not an issue in this test. Software was standardized for both displays and the presentation interface was consistent with paper technical manual presentations. Significant differences between the two display types, maintainers, and interactions were not found for measures of either task performance times or number of errors made. Recommendations for follow-on experiments stated that similar tasks in environments requiring movement in and around the aircraft might be more appropriate for the wearable computer display (Masquelier, 1990).

Acting on Masquelier's recommendation, in 1992, Friend and Grinstead conducted a study that compared two similar types of display devices in a flightline maintenance environment. In the flightline task, technicians were required to move around the airplane conducting fault isolation and inspection tasks. To provide mobility, a portable computer was used for the "standard" display device while a wearable computer was used with the head-mounted display (HMD). Both displays were liquid crystal displays. The HMD was built by Intervision™. In keeping with Masquelier's design, two levels of experience were selected—experienced and inexperienced.

In contrast with Masquelier's study, the results of this study yielded several significant effects and interactions. For the fault isolation task, inexperienced technicians took longer to isolate the defect when using the portable computer; whereas, experienced technicians performed equally well using either display device (portable or HMD). Additionally, in the inspection task, experienced technicians found more faults (i.e., fewer diagnostic errors) when using the HMD than when using the portable computer; whereas, inexperienced technicians performed equally with either device. The main effect of display type showed technicians using the HMD found five more faults, on average, than their portable-equipped counterparts.

The overall results showed that when comparing a wearable computer with a portable one, performance was better in some circumstances with the wearable computer.

AFRL/HESR has conducted several usability studies. One such usability study evaluated a monocular, occluding head-mounted display. The usability study included evaluation of the HMD in a variety of flightline conditions, including: sunlight and nighttime conditions, heavy and light aircraft, a variety of body positions (sitting in the cockpit, under the wing, in a wheel well, in a crew rest bunk, etc.), varying weather conditions from extreme summertime heat to near freezing temperatures in springtime. While the HMD was "usable" in all these environmental conditions, these conditions identified areas for design improvement.

Usability studies, such as the one just described, have provided valuable feedback about one wearable computer or components of a given wearable computer. However, they have not directly compared the usability of a wearable computer with a portable computer.

The current study compares 5 computing systems for usability in an Air Force flightline environment.

Method

Five hardware configurations were compared in the current study. The FM-Net consisted of a Motorola handheld radio on which work orders were opened via voice commands. A notebook configuration was provided through the Itronix ruggedized computer. The Libretto computer was a miniature, very lightweight computer. Two configurations of the Xybernaut wearable computer were tested: one included a head-mounted display and the other included a wrist-worn display (see Figure 1). All computers were full computers running MS Windows.

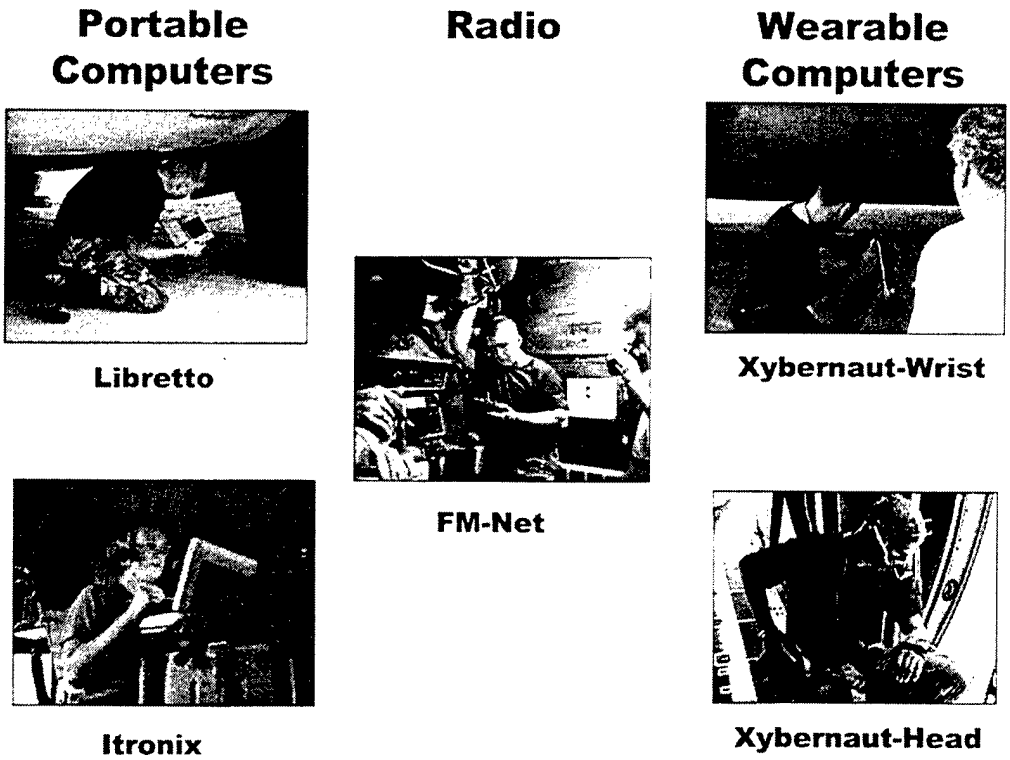


Figure 1. Hardware Configurations

For scheduling purposes, a total of 4 hours were needed to test a participant. Inbrief and training required ½ hour; testing required no more than 2½ hours; the completion of Post-Test Questionnaires required ½ hour. Experimenters needed an additional ½ hour to reconfigure hardware and prepare for the next participant.

Hardware Requirements

1. FM Net radio with converter box
2. FM net workstation, located in the MOCC for receipt of radio transactions
3. Itronix ruggedized notebook computer for use on the flightline

4. Libretto miniature notebook computer for use on the flightline
5. Xybernaut wearable computer
6. Head-mounted display for Xybernaut
7. Wrist-worn keyboard for Xybernaut
8. Wrist-mounted display for Xybernaut
9. Light Meter
10. Sound Meter

Users

Users were C-130 Talon II maintainers qualified to open work orders. These maintainers included Crew Chiefs and Communication/Navigation Specialists. Ten users participated in the test, 5 Crew Chiefs and 5 Comm/Nav Specialists. Experience level was not controlled in the study and ranged from inexperienced (new to the career field) to very experienced (dual AFSC qualified and many years in the career field).

Facilities

The 15th Aircraft Maintenance Unit (AMU) located at Hurlburt Field served as the test facility. The test aircraft was the MC-130H Combat Talon II.

Data Collection Team

The data collection team consisted of four individuals: 1) a subject matter expert, 2) a videographer, 3) a subject coordinator, and 4) a flightline coordinator. The subject matter expert provided the majority of the interaction with each crew chief and specialist. The videographer was responsible for video and audio documentation, via camcorder, of each experimental session. The subject coordinator was responsible for scheduling participant testing times via the Hurlburt POC, and arranging and carrying out the inbrief and outbriefs sessions of each participant. The flightline coordinator was responsible for ensuring that all necessary hardware (e.g., computers) was available and ready.

Procedure

Users were required to open/create 3 work orders using the FM Net radio, Itronix computer, Libretto notebook computer, Xybernaut head-mounted display configuration, and the Xybernaut wrist-mounted display configuration (Department of Defense, 2000). Only the third work order for a given equipment configuration was observed for data collection. Before testing the subject coordinator conducted an inbrief session with the participant. This inbrief included familiarization with the software (see Appendix B for description of the software used for the study). After completion of the inbriefing the flightline coordinator assisted the participant in becoming familiar with the specific equipment configuration. Once the participant was familiar with the equipment the subject matter expert directed the participant through the scenarios. When the participant had finished the scenarios for a given equipment configuration the participant completed a Post-Condition Questionnaire. Upon completion of the Post-Condition Questionnaire,

the flightline coordinator helped the participant become familiar with the next equipment configuration. This process continued until all five devices had been tested. When each participant was finished with the hardware testing they completed a Post-Test Questionnaire and received an outbrief session with the subject coordinator.

Results

Both objective and subjective data were collected during the Spiral 1 testing. Results are presented for each type of data collected.

Objective Data

Completion times and errors were collected as a part of this usability test. However, as stated in the Introduction, the purpose of this test was not to compare the devices from an objective, performance-based perspective. The purpose is to compare devices from a subjective, usability-based perspective. The Objective Data presented in this section are included for completeness. The Subjective Data section provides results most pertinent to the Usability Test.

Completion Time

For each device the participants were timed from the initiation to the submission of opening/creating a work order. Figure 2 shows the average completion time for each device. Clearly the FM-Net condition resulted in the shortest performance times. From an empirical standpoint this finding should not be overlooked. While statistical measures of comparison were not made, the indication is that the voice recognition technology offered through the FM-Net condition may substantially reduce the time required to open work orders on the flightline.

In contrast, the Xybernaut HMD condition resulted in the longest performance times. The completion times for the remaining three devices were within approximately 10 seconds of one another.

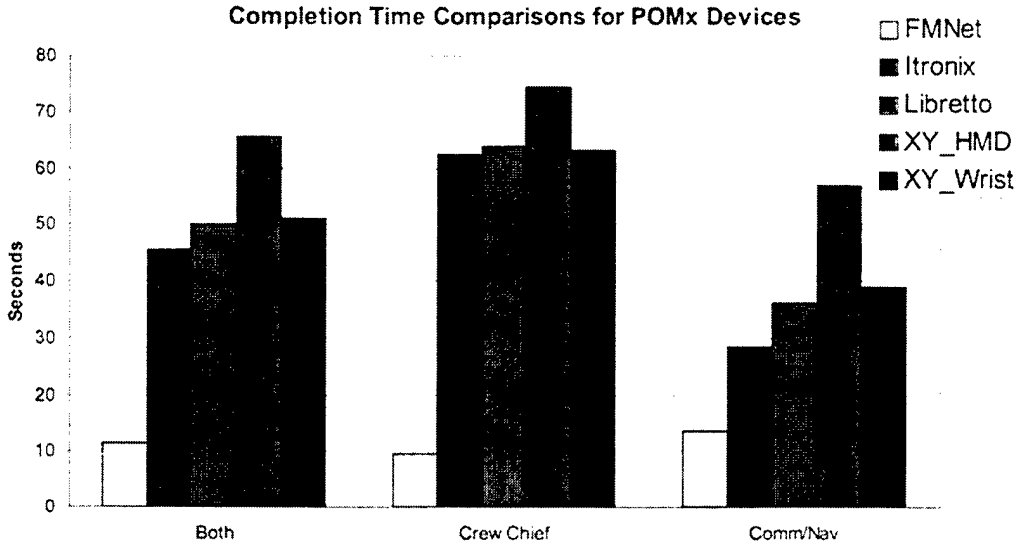


Figure 2. Completion Times

Errors

Due to slight differences in the testing scenarios, the results for the FM Net trials are discussed separately from the trials for the other devices.

For Itronix, Libretto, Xybernaut HMD and Xybernaut Wrist, eleven data entry errors were found in the test trials. Nine trials were in error because WD (When Discovered Code) was not populated. The two remaining errors were contained in the WCE Narrative field. Of these eleven errors, eight were associated with the Xybernaut devices (4 each with Xybernaut HMD and Xybernaut Wrist). Two errors were associated with the Libretto, and one with Itronix (Figure 3). Not included in this total are the two errors discussed below for FM Net.

Comm/Nav Specialists were more likely to commit errors in their data entry than were Crew Chiefs. Of the eleven errors discussed above, seven were recorded by Comm/Nav Specialists, while only four were recorded by Crew Chiefs.

Two errors were recorded for the FM Net trials, both due to inaccurate field population (one each in Aircraft # and System). Both of these errors were committed by Comm/Nav Specialists. Crew Chiefs did not commit any errors during FM Net trials.

In six of the ten FM Net trials, subjects did not populate the Discrepancy field. While this was not considered an error for this particular test (as long as the System field was populated), in the future, there should be some visual feedback to users about the fact

that the System information they are verbalizing is being used to fill in the Discrepancy field.

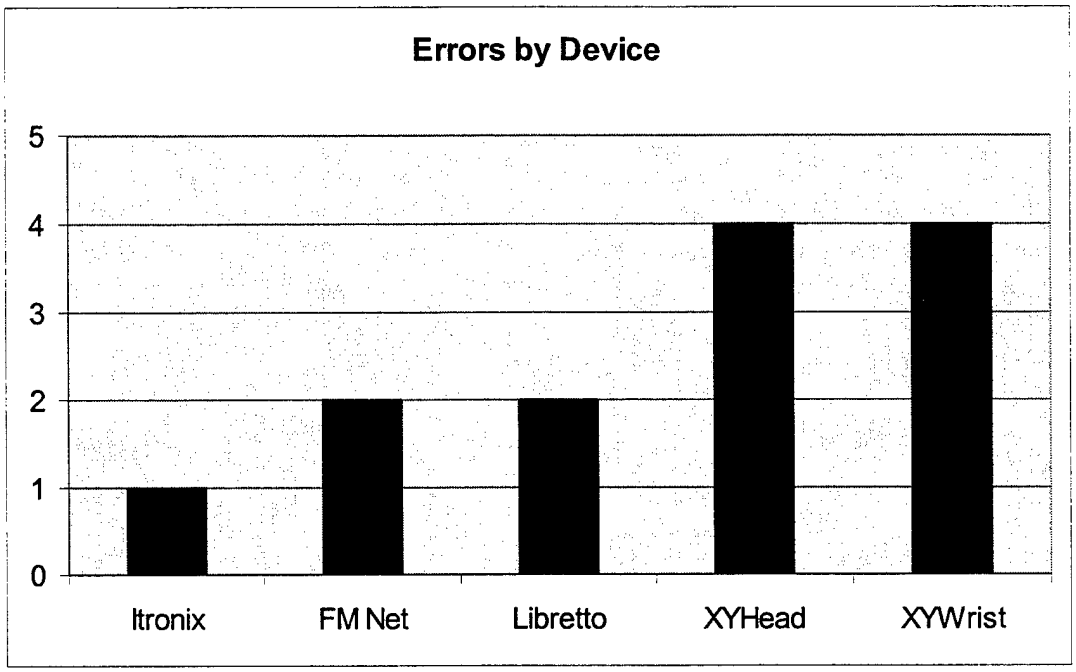


Figure 3. Errors by Device

Subjective Data

Post-Condition Questionnaires

The primary question for the Post-Condition Questionnaires dealt with the ease in opening work orders. These questions were on a 5-point Likert scale (1= Easy, 5= Not at all easy). Figure 4 shows a graph of the responses to the question of ease of use for each device. For the FM-Net, Itronix, and Libretto conditions, most subjects rated the devices Easy-to-Use. In the two Xybernaut wearable computer conditions, however, ratings were slightly lower and for the Xybernaut HMD condition, two subjects rated the device in an unacceptable range (4 and 5 respectively).

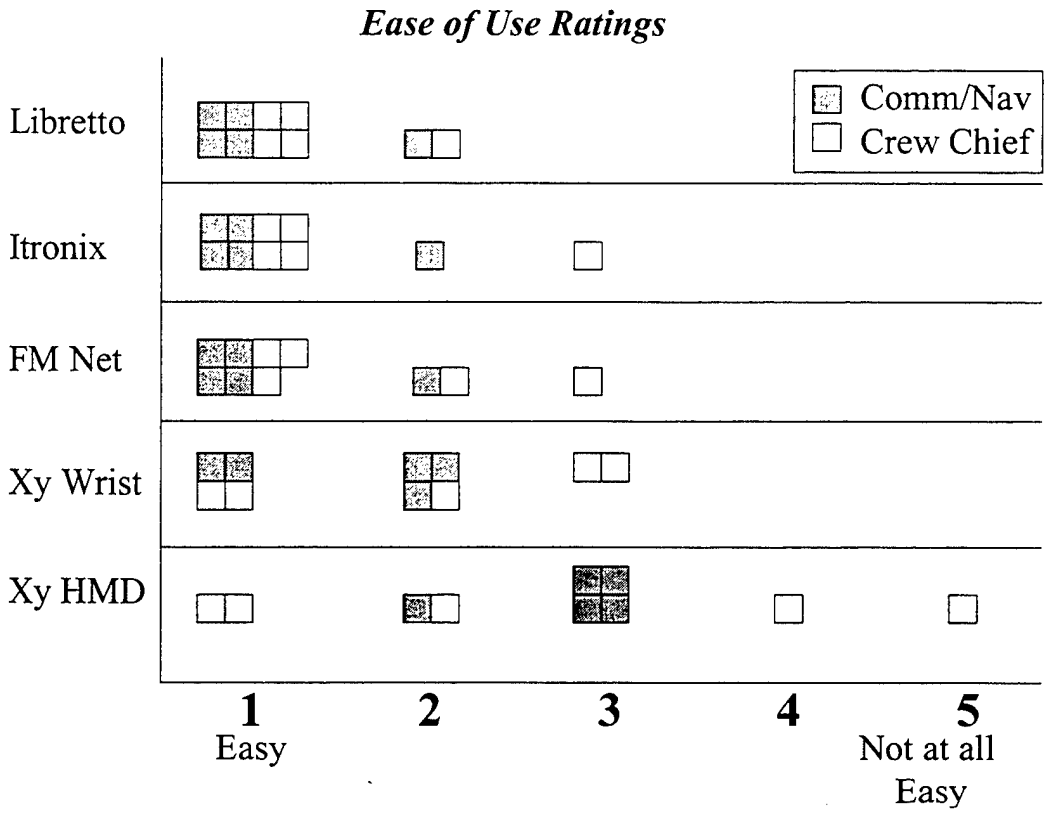


Figure 4. Post-Condition Ease of Use Rating for all 5 equipment configurations

In usability evaluations, the trends or patterns of responses must be carefully considered. In this case, for the FM-Net and Itronix conditions only one person rated the device a 3 (middle of the scale); however, for the Xybernaut Wrist condition two rated it a 3 and there were an even number that rated it as a 1 or 2. This indicates a small negative shift in usability in comparison with the other three devices. The Xybernaut HMD condition indicates even a further negative shift in usability, with two people rating the device in the unacceptable range.

Comments from users further substantiate these results. Users indicated that for the Xybernaut HMD condition the equipment was too bulky and got in the way of moving around in tight places. They also indicated the equipment needed to be cordless, and that the screen was not visible in sunlight. These are all issues that substantially reduce the usability of a device. With respect to the Xybernaut Wrist condition, users gave similar comments. Users did not like the bulkiness or cords; however, they did like the screen size for the wrist-mounted display and they liked the touch screen offered by this display.

Issues identified by users with respect to the other three devices were milder. They found the Itronix device too heavy and the FM-Net vocabulary somewhat limited; however, these somewhat negative comments were overshadowed by the positive comments given about these devices. For the Libretto and FM-Net conditions, users specifically stated that they were easy-to-use. They liked the fact that these two devices

were lightweight and had a small footprint. They also indicated that they liked the screen readability and the touch screen capability of the Itronix device (this is consistent with their positive comments about the Xybernaut wrist-mounted display).

Post-Test Questionnaires

The Post-Test Questionnaires addressed the participants' preference of equipment configuration (see Appendix A for a sample of questionnaires administered). The participants rank ordered the devices on 4 criteria: 1) display, 2) text entry, 3) field selection, 4) and overall technology. The following data are averages of the participants' rank orders (see Figures 5-8).

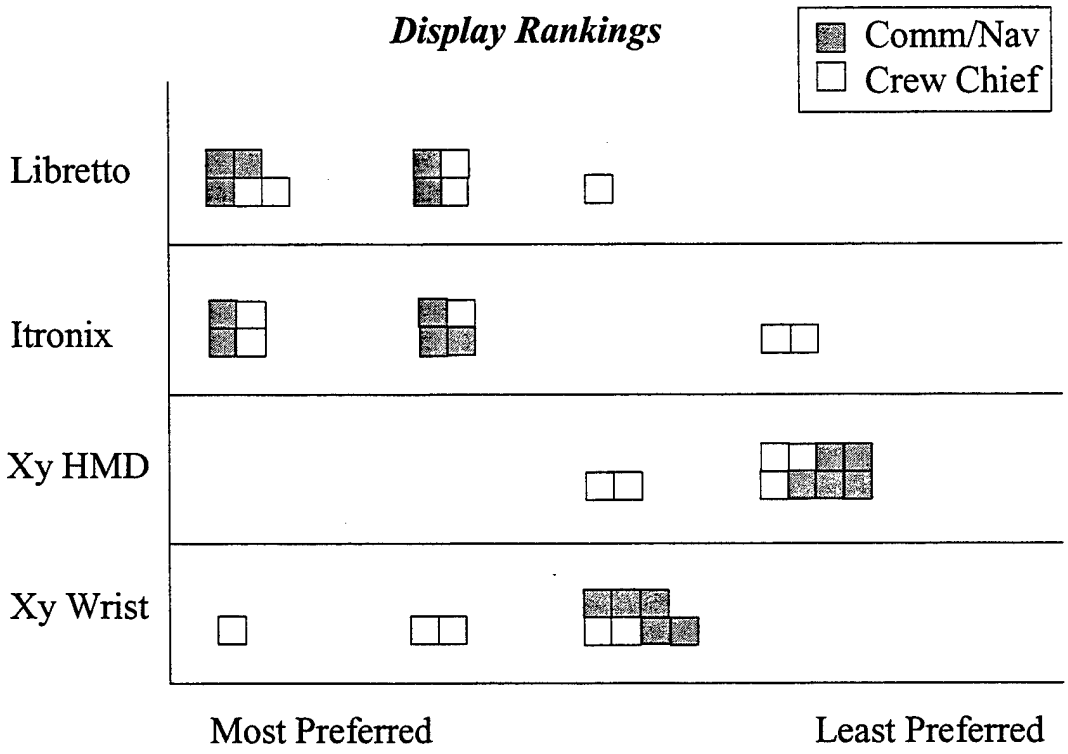


Figure 5. Ranking for Display

Note that the FM-Net condition was a voice only system (without any display device); therefore, it was not included in this rank order comparison.

Ranking of the devices for display clearly shows that the Xybernaut HMD was the least preferred condition. Technicians noted that this monocular head-mounted display device was difficult to see in bright sunlight. Many technicians had to readjust the position of the eyepiece while wearing it. It should also be noted that the study was conducted in Florida, in July. Reported heat index (combined heat and humidity) was between 90 and 110 each day. Flightline conditions were probably worse than this. While technicians did not indicate that the heat had an effect on their rankings, experimenters were required to clean the head-mounted display after each test to remove

perspiration from the device. Other laboratory head-mounted display experiments have also shown this to be problematic with users. With respect to the other devices technicians commented that they liked the size of the Xybernaut-wrist mounted display, although it was ranked second to last for preference.

Users indicated near equal preference for the Libretto and Itronix displays.

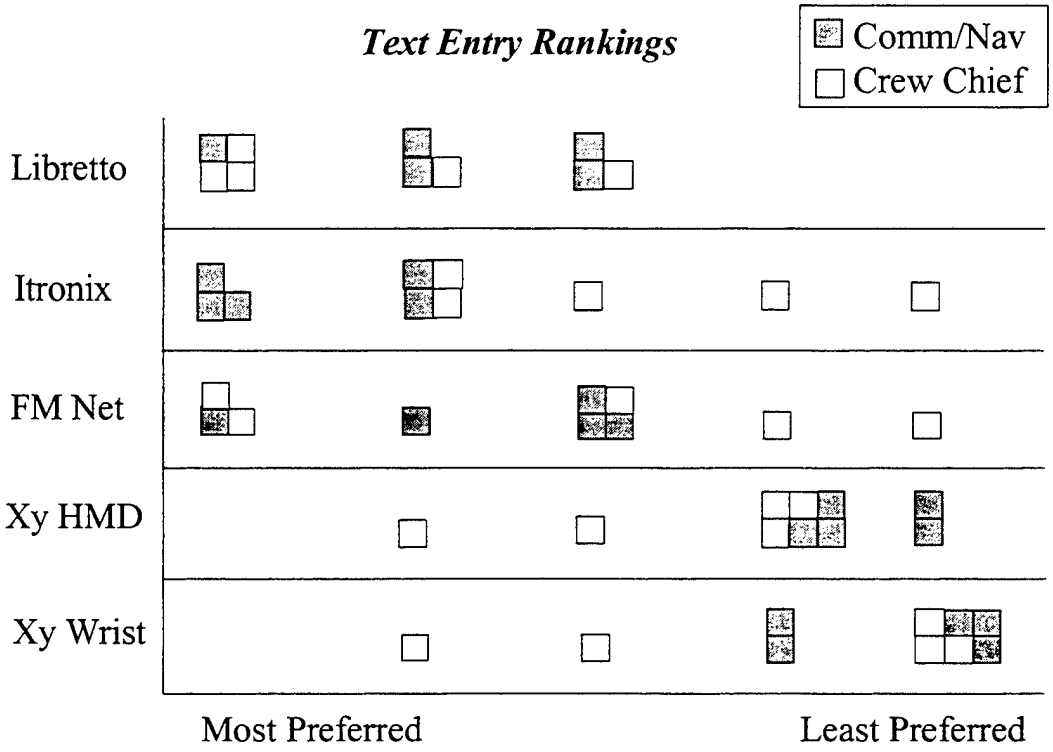


Figure 6. Ranking for Text Entry

For text entry the device least preferred was the Xybernaut with the wrist-mounted display. The Xybernaut Wrist condition had an on-screen keyboard and technicians were required to move the keyboard around the screen to view the fields being filled in. As shown in the figure, they did not care for the on-screen keyboard as presented. Several participants indicated that if an on-screen keyboard is to be used, it should be fixed at the bottom of the screen and not cover any of the material presented on the work order.

The Xybernaut HMD provided a wrist-worn keyboard. User's ranked the wrist keyboard (on the Xybernaut HMD condition) slightly better; however text entry and field selection frequently intermixed. For the Xybernaut HMD condition, technicians were required to select a field with the hip-mounted pointing device and then fill in the text with the wrist-worn keyboard. Experimenters observed that this right-hand movement from the right hip to the left arm keyboard was, at times quite awkward for the technicians. This may have added to the relatively poor ranking given to this device for text entry.

For text entry, the Itronix device and the Libretto offered a full keyboard. It is interesting that users did not identify a preference for verbal text entry (offered by the FM-Net condition) over the keyboard entry. Entering text verbally is very intuitive, and several technicians commented that it was very easy to use. However, other technicians expressed difficulty with entering text in the current application. Two concerns were raised: technicians wanted more flexibility in the vocabulary accepted by the system (i.e., an expanded vocabulary), and they wanted a visual reference to what had been entered.

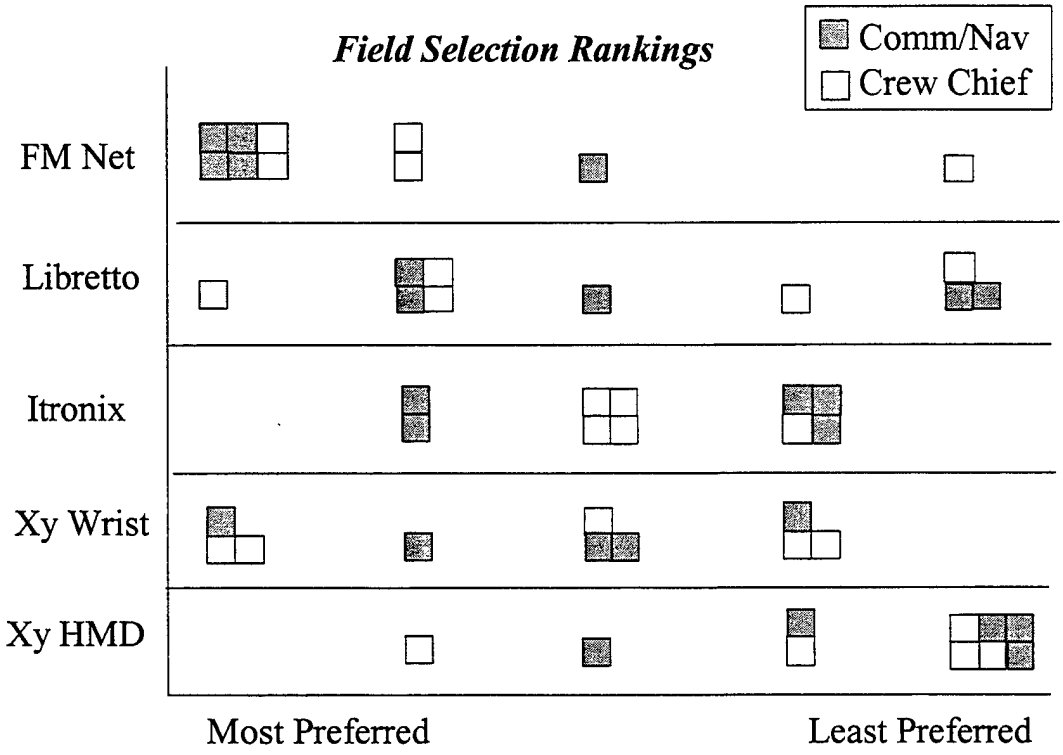


Figure 7. Ranking for Field Selection

Technicians preferred the FM-Net voice recognition system to the other devices. For the FM-Net condition, a crib sheet was attached to each radio. This provided users with a visual reference as to what fields needed to be verbally selected. This visual reference along with the ease of verbal input made this selection easy to users. Interestingly, the Xybernaut Wrist condition was ranked second to the FM-Net condition. This device provided a touch screen for field selection and was mounted on the left wrist (all subjects were right handed). This configuration appears to have been somewhat conducive to field selection.

The Xybernaut HMD condition with the hip-mounted mouse was ranked last in terms of field selection. Again, this appeared to be due to the interaction of the hip-mounted mouse and the wrist-mounted keyboard.

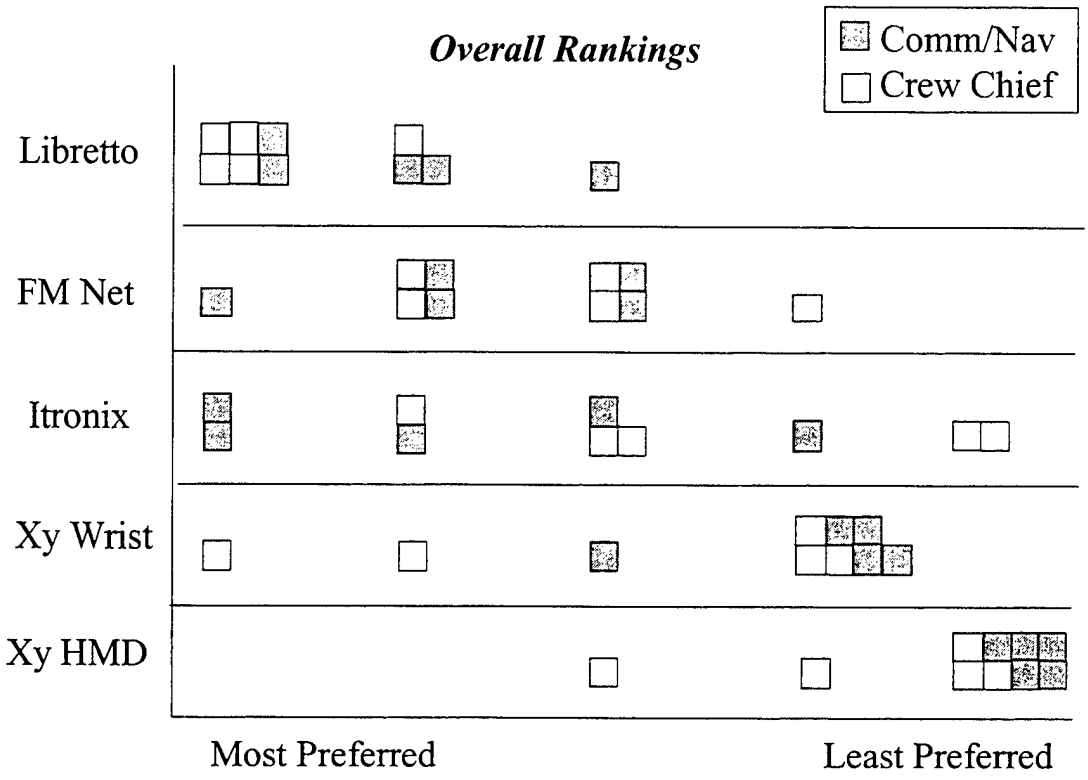


Figure 8. Ranking for Overall Technology

Technicians preferred the Libretto computer to the other devices. This is true for both Crew Chiefs and Comm/Nav Specialists. Technicians indicated that this lightweight, small device was very easy to use. Clearly this ranking indicates their preference for the device. It is noteworthy that the Comm/Nav Specialists' rankings indicate that they also seemed to like the FM-Net and Itronix devices. The least preferred device was the wearable Xybernaut computer with the head-mounted display. The Xybernaut computer with the wrist-mounted display was ranked second to last.

Discussion

FM-Net

The Completion Time data showed that the users took the least amount of time to complete the work orders with the FM Net radio. According to the Post-Condition Questionnaire the FM Net radio was rated as easy-to-use. The data from the Post-Test Questionnaire indicated that the FM-Net radio was the preferred device for selecting fields on the screen. Technicians ranked the FM-Net equivalent to the Itronix and Libretto devices for text entry. Users commented that they did like the lightweight nature and portability of the radio.

An interesting finding was that some users did not prefer the FM-Net for entering text (an intuitively obvious choice)—instead they found it equivalent to entering text with the Itronix and Libretto keyboards. Users expressed difficultly using a specific vocabulary in order to operate the radio. Another problematic issue was that the radio did not provide the user with any type of feedback or confirmation that the work order had been received. Users may have given the FM Net radio a better ranking if it provided some sort of visual feedback. That feedback could appear on a computer screen or be read back to the user via the radio. Based on technician's familiarity with radios the FM-Net radio could be the device of choice if these features could be incorporated into it.

Itronix Notebook

Results from the Post-Condition Questionnaire showed that technicians found the Itronix computer as easy-to-use. Data from the Post-Test Questionnaire showed that technicians preferred the Itronix display (equivalent ranking was given to the Libretto). They also preferred the Itronix computer for text entry (equivalent rankings were given to the Libretto and the FM-Net). Users commented that the computer's touch screen and stylus pen did aid them with maneuvering throughout the screen. The users also liked the ruggedized casing for the Itronix computer. Because the computer was ruggedized, the users did not feel the need to be as cautious with it as they were with the other equipment configurations.

The weight of the Itronix computer proved to be a problem. The users commented that the computer was too heavy or "bulky". Because of its ruggedization the Itronix computer was heavier than a typical notebook computer. If the Itronix computer was smaller and lighter users might rate it higher.

Libretto Miniature Computer

According to the Post-Condition Questionnaire, the Libretto was rated as easy-to-use—with all subjects rating it with a 1 or 2. Data from the Post-Test Questionnaire showed that technicians preferred the Libretto display (equivalent ranking was given to the Itronix computer). They also preferred the Libretto computer for text entry (equivalent rankings were given to the Itronix and the FM-Net). Post-Test Questionnaire data revealed that the Libretto computer was rated the preferred technology overall.

Users liked the size of the computer, and they liked that the Libretto was small enough to carry in their hand.

Some users indicated that the Libretto should be ruggedized. Another noted problem was with the sensitivity of the mouse-like device used for field selection. Users felt it moved too quickly around the screen.

Xybernaut: Head Mounted Display Configuration

The Post-Condition Questionnaire data showed that this configuration of the Xybernaut computer was the most difficult device to use. Two technicians rated ease-of-use in categories previously defined as unacceptable. That is, the Usability Test Plan clearly identified that ratings of 4 and 5 were unacceptable. The Post-Test Questionnaire data illustrated that the head-mounted configuration was the least preferred display, field selection device, and technology overall. Users commented that the various computer cables made it difficult to move around. The weight of the device also proved to be a problem for the users. Several users commented that the configuration was too "bulky". As for the display, users had difficulty with sunlight glare on the HMD. They also had difficulties adjusting the display's eyepiece so that the entire screen would be in focus.

Xybernaut: Wrist Mounted Display Configuration

The Post-Condition Questionnaire data showed the wrist-mounted configuration was easy-to-use; however, ratings were relatively low in comparison with ratings given to the other devices. The Post-Test Questionnaire data showed that the wrist–mounted configuration was ranked second to the FM-Net condition for field selection. The touch screen capability offered by this device appears to be conducive to field selection. Users indicated that they liked the touch screen nature of the wrist display and the use of the stylus pen.

This device was ranked second to last for technology overall. Comments provided by users were similar to the comments made for the head-mounted display configuration. Technicians did not like the weight and "bulkiness" of the computer and display. The cables made it difficult to move around.

Conclusions

This test compared five portable computing devices used for opening work orders on the flightline. Air Force Crew Chiefs and Communication/Navigation Specialists for the MC-130H aircraft participated in the study. These technicians identified that, among the devices shown, the Libretto computer was: 1) rated highest for overall ease-of-use, and 2) the preferred technology overall. This device is small and lightweight; it offers a full keyboard and full screen that technicians can carry in one hand.

Features of the other devices should be considered for a product such as the Libretto. The ruggedization of the Itronix device was seen as a benefit to technicians. Also, the touch screen offered by the Itronix and Xybernaut Wrist conditions was a desirable feature for selecting fields; however, for manual text entry it should be noted that physical keyboard is still preferred to an on-screen keyboard.

While the Libretto computer was the preferred device in this test, the potential for the FM-Net cannot be overstated. Timesavings offered by FM-Net are potentially substantial. According to this report, usability issues associated with this device were limited to providing feedback to the user on the text just entered (e.g., visual feedback) and expanding the vocabularies used. Should these usability issues be addressed, the potential exists for a device that offers not only performance improvement but also usability acceptance by aircraft flightline maintainers.

While previous studies conducted by the Air Force Research Laboratory have indicated that better performance is obtained when using a wearable computing system, usability comparison of these devices with portable computers clearly indicates that wearable systems are not as usable in conditions such as opening work orders from the flightline. If both improved performance and increased usability are to be considered for future systems, wearable computer research must focus on the usability issues identified in this study.

Recommendations

This Point-of-Maintenance test was the first in a series of spiral tests looking at the usability of computing systems on the Air Force maintenance flightline. The next major evaluation, P-O-Mx, Spiral 3, will require selection of hardware that takes into consideration results of this test, as well as the additional test requirements imposed by that test (e.g., presentation of Technical Manuals and additional forms entry). There are several considerations for selection of future hardware systems that can be made using the results of this first test. These recommendations are provided below:

- Hardware footprint should focus on small, lightweight devices. Devices that can be held with one hand are preferable. The size and weight of the Libretto and FM-Net devices were desirable to technicians in the first spiral. Note that while the Libretto was a full computer with a full screen, technicians frequently used a one-hand carry for the device.
- If any manual text entry (numbers and letters) is required, the device should offer a full, physical keyboard. In Spiral 1, technicians had some difficulty with the on-screen keyboard—especially with respect to screen real estate. Note, however, that physical key size does not necessarily have to be standard size. The smaller size of the Libretto keys did not pose any noted difficulties.
- Voice recognition should be offered wherever practical, and especially for field selection. Text entry is also still a prime area for entry of data by voice; however, vocabularies need to be further modified to allow wider acceptance of maintenance terminology.
- Information entered verbally (e.g., via radio) should allow technicians to confirm the information uttered. This confirmation could be through a display; however, it could also be provided to the technician through other means, such as verbal feedback of information entered.
- The computer should include as much ruggedization as is possible without compromising weight and size of the device. Many manufacturers currently use a 4-foot drop test. At a minimum, this criterion should be included where possible.
- A touch screen should be included for field selection. The device should offer a tethered pointing device. Technicians indicated the desire for this feature on the two devices in Spiral 1 that offered touch screen capability (Itronix and Xybernaut Wrist).

In addition to these recommendations, there was at least one area where data from the Spiral 1 test was inconclusive. This area should be further investigated in subsequent tests.

- The first spiral tested a full 640 x 480 VGA screen on the Itronix, Libretto, Xybernaut HMD, and the Xybernaut Wrist devices. Smaller screen sizes, such as ½ or ¼ screen VGA were not compared. Future tests should look specifically at determining the usability of these small screens for viewing Technical Manuals and forms completion on the flightline.

References

Department of Defense. (2000). Technical Manuals, USAF Series, AC-130H And AC-130U Aircraft. WR-ALC/LUTD, Robins AFB GA 31098.

Dumas, J.S., and Redish, J.C. (1993). *A Practical Guide to Usability Testing.* Norwood, N.J.: Ablex Publishing Corporation.

Friend, J. and Grinstead, R. (1992) "Comparative evaluation of a monocular head mounted display device versus a flat screen display device in presenting aircraft maintenance technical data," MS thesis, Air Force Institute of Technology, Wright-Patterson, AFB, Ohio.

Karat, C. (1994). "A comparison of user interface evaluation methods," J. Nielson and R. L. Mack, eds. *Usability Inspection Methods.* New York: John Wiley & Sons, Inc.

Masquelier, B.L. (1991). "Comparative evaluation of a monocular head mounted display device relative to portable computers for display of aircraft maintenance technical data," MS thesis, Air Force Institute of Technology, Wright-Patterson, AFB, Ohio.

Quill, L., Kancler, D., Revels, A.R., & Donahoo, C. (2001). *Point-Of-Maintenance, Spiral 1, Usability Test Plan.* Contract No. F33615-99-D-6001. Dayton, OH: University of Dayton Research Institute.

Carlson, E., Smith, S., Smith, B., Thomas, D., and Smillie, R. (1992). *Integrated maintenance information system (IMIS) F/A-18 interactive electronic technical manual (IETM) diagnostic demonstration.* Unpublished manuscript. Logistics Research Division, Armstrong Laboratory, Wright-Patterson AFB, Ohio.

Thomas, D. (1995). "Integrated maintenance information system: User field demonstration and test," Technical Report AL/HR-TR-1995-0034, Logistics Research Division, Human Resources Directorate, Armstrong Laboratory, Wright-Patterson AFB, Ohio.

Appendix A

Usability Test Questionnaires

Usability Test Questionnaires

PRE-TEST QUESTIONNAIRE

Date:_____

All information will remain confidential.

Name _____ Rank _____

AFSC_____ Time in Air Force _____ Time in AFSOC_____

Time on C-130s _____ Other Weapon Systems worked _____

Current Job Title_____ Years of Experience at Current Job_____

Previous Jobs with Years of Experience

_____

_____

_____

Circle the response that best applies to you.

1. Approximately how often do you open work orders on the C-130?

   Daily Weekly Monthly Rarely (less than once a month)

2. Approximately how often do you use a personal computer (on the job or at home)?

   Daily Weekly Monthly Rarely (less than once a month)

3. Approximately how often do you use open work orders on the CAMS terminal?

   Daily Weekly Monthly Rarely (less than once a month)

4. Have you previously participated in an AFRL field evaluation?

   Yes No

5. If you answered yes to the previous question, when and what did you evaluate?

   _____

[Insert device name here]

POST-CONDITION QUESTIONNAIRE

Date:_____

Subject # :_____

All information will remain confidential.

[insert device name here] at the aircraft:

1. Using the [insert device name here] to open work orders was:

Easy				Not at all easy
1	2	3	4	5

2. If you could change anything on the [insert device name here], what would it be?

_____

_____

_____

3. What features did you like about [insert device name here]?

_____

_____

_____

4. What features did you NOT like about [insert device name here]?

_____

_____

_____

5. What other comments do you have?

_____

_____

_____

POST-TEST QUESTIONNAIRE

Date:_____
Subject # :_____

All information will remain confidential.

1. Rank order your preference for a ***display*** device for opening a work order on the flightline (1 is first choice, 4 is last choice):
   Choices are: a) Itronix, b) Libretto, c) head-mounted, d) wrist-worn

1. _____
2. _____
3. _____
4. _____

Comments:

_____

_____

_____

_____

2. Rank order your preference for entering text when opening a work order on the flightline (1 is first choice, 5 is last choice):
   Choices are: a) voice, b) Itronix full keyboard, c) Libretto small keyboard, d) wrist-worn keyboard, e) wrist-worn on-screen keyboard

1. _____
2. _____
3. _____
4. _____
5. _____

Comments:

_____

_____

_____

_____

3. Rank order your preference for selecting fields when opening a work order on the flightline (1 is first choice, 5 is last choice):

Choices are: a) voice, b) Itronix, TrackPoint mouse, c) Libretto thumb pointer, d) hip-mounted pointing device, e) touch screen

1. _____
2. _____
3. _____
4. _____
5. _____

Comments:

_____

_____

_____

_____

4. Overall, rank order your preference for one ***overall technology*** for opening a work order on the flightline (1 is first choice, 5 is last choice):

Choices are: a) FM radio, b) Itronix computer, c) Libretto computer, d) Xybernaut wearable computer with head-mounted display, e) Xybernaut wearable computer with wrist-worn display,

1. _____
2. _____
3. _____
4. _____
5. _____

Comments:

_____

_____

_____

_____

4. What other comments do you have?

Appendix B

User Feedback Software Used In Test

User Feedback to Software used in Test

For the usability test, users were required to open work orders from the flightline. To perform this operation a Core Automated Maintenance System (CAMS) screen was mocked up in MS Visual Basic 6.0. This mock-up not only simulated CAMS screen data requirements, it also collected usability information used in the analysis.

While user feedback focused on evaluation of hardware, they also offered feedback on the software interfaces provided by the simulated CAMS screen. The CAMS screen was laid out similarly to the screen used in the field today. There were three primary differences between the mock-up and the actual CAMS screen.

First, some of the information required to fill out the CAMS screen was automatically generated in the mock-up. This information was gathered from a simple look-up table and inserted into the form. For example, upon entering CAMS the user is required to put in an ID number. This number was automatically inserted into the CAMS screen (Emp #) along with the user's Performing Work Center information (see Figure 9). TS/SCH, Job Duration, Symbol and Work Center Event (WCE) Symbol were also generated or defaulted to standard values. This reduced the amount of time required to enter data into the screen.

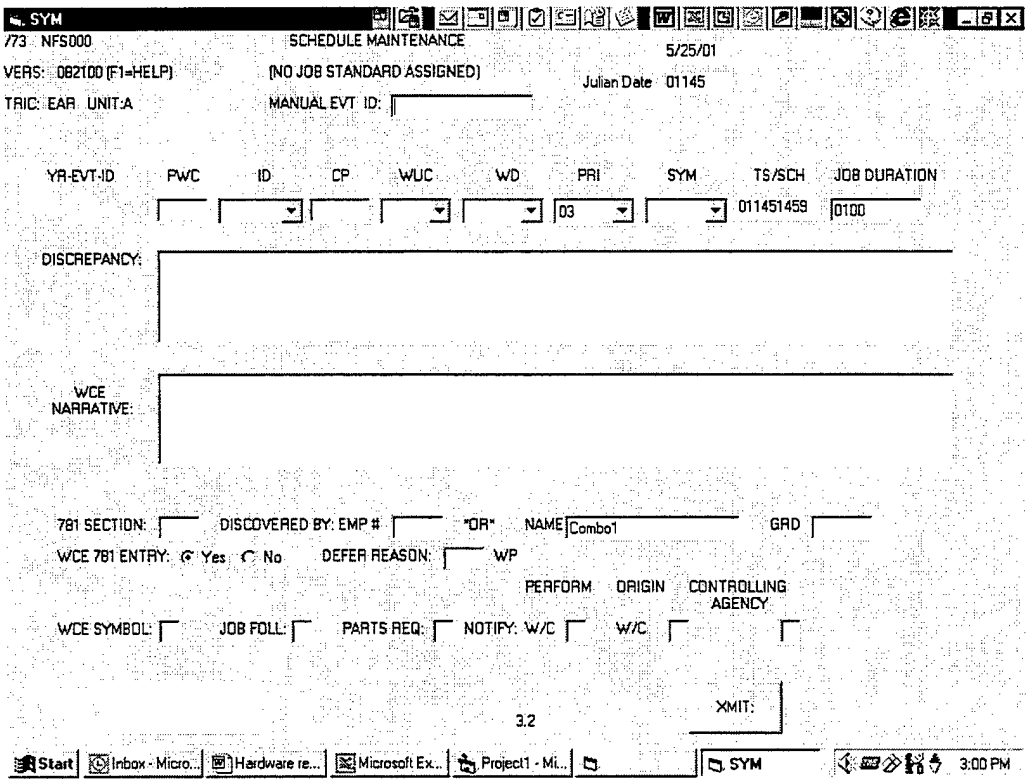


Figure 9. CAMS Screen Mock-Up

The other major difference between the mock-up and the actual CAMS screen was the existence of drop down lists offering appropriate information to users for these selected fields. Most of these drop down menus contained only a few choices. However, the Work Unit Code field contained information from the Technical Manual. This information was converted from PDF format into a table, and this list provided thousands of codes at the user's fingertips. This reduced the amount of errors committed by users when entering data into the fields.

The final difference (and one that several users commented on) was the automatic population of WCE narrative information and symbols. For example, the information is typically the same for the discrepancy and the first WCE narrative; therefore, when information was populated into the discrepancy field, the entry was duplicated into the WCE narrative field. This reduced both time and errors in entering data into the CAMS screen.

Users also expressed the desire to be able to maximize and minimize CAMS windows so they could come back to the screen at a later time.