CT-
TN
97/2

technical note technical note

COPY 1

FEDERAL AVIATION ADMINISTRATION
TECHNICAL CENTER LIBRARY
ATLANTIC CITY INT'L ARPRT. NJ 08405

Voice Technology Study Report

Richard M. Mogford, Ph.D., ACT-530
Armida Rosiles, Ph.D., ACT-530
Dan Wagner, ACT-530
Kenneth R. Allendoerfer, NYMA, Inc
Airway Facilities Human Factors Group

December 1997

DOT/FAA/CT-TN97/2

Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

U.S. Department of Transportation
Federal Aviation Administration

William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

DOT/FAA
CT-TN
97/2
c. 1

FAA Technical Center
00017177

DOT/FAA Mogford, Richard H.
/CT-TN9 Voice technology
7/2 study report
c.1

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof.

The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the objective of this report.

Technical Report Documentation Page

1. Report No. DOT/FAA/CT-TN97/2	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Voice Technology Study Report		5. Report Date December 1997 6. Performing Organization Code ACT-530
7. Author(s) R. H. Mogford, Ph.D., Armida Rosiles, Ph.D., and Dan Wagner, ACT-530, Kenneth R. Allendoerfer, NYMA, Inc., and the Airway Facilities Human Factors Group		8. Performing Organization Report No. DOT/FAA/CT-TN97/2
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		10. Work Unit No. (TRAIS) 11. Contract or Grant No. F2203D
12. Sponsoring Agency Name and Address Federal Aviation Administration Airway Facilities Requirements 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered Technical Note December 1996 14. Sponsoring Agency Code AFR-100

15. Supplementary Notes

This work was accomplished under the direction of Dan Wagner and Armida Rosiles, ACT-530, as part of the FAA Airway Facilities Human Factors R,E, & D Program. The FAA Headquarters Program Manager is Lawrence Cole, AAR-100, and the AF Requirements sponsor is Dr. Brenda Boone, AFR-100.

16. Abstract

This document presents the findings of a voice technology study that evaluated the potential of a speech-to-text and voice recognition system to support an Airway Facilities maintenance task. Researchers conducted the test at an Airport Surveillance Radar (ASR)-9 site at the William J. Hughes Technical Center. Thirteen Airway Facilities specialists completed the procedure twice, once with the voice technology system and again with a paper manual. The results showed no differences in task completion time or workload, suggesting that the voice technology system was no more time consuming or difficult to use than a traditional paper manual. The voice recognition rate was 86.6%. Questionnaire responses showed that users found the voice technology system understandable, easy to control, and responsive to voice commands. When asked to compare voice technology to the use of a paper manual for the Air Traffic Control Beacon Interrogator (ATCBI)-5 maintenance procedure, study participants indicated that the voice technology system made the maintenance task easier to perform, was more efficient and effective than a paper manual, and would be better for handling large amounts of information. Researchers concluded that this study resulted in a successful demonstration of voice technology for the Airway Facilities maintenance procedure. It was recommended that further, more extensive studies be conducted using voice technology systems in a wider variety of Airway Facilities environments and tasks.

17. Key Words Voice Recognition Voice Synthesis Airway Facilities Human Factors		18. Distribution Statement This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 29	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Acknowledgement

The Airway Facilities Human Factors Group consisted of the following individuals.

Federal Aviation Administration

Lou Delemarre, ACT-540

Contractors

Robert Cranston, SRC

Anton Koros, NYMA, Inc.

Jean MacMillan, Ph.D., BBN

Jack Oxford, NYMA, Inc.

Joseph Tortorelli, SRC

Table of Contents

List of Illustrations

Executive Summary

This document presents the findings of a voice technology (VT) study that evaluated the potential of a speech-to-text and voice recognition system to support an Airway Facilities (AF) maintenance task. ACT-530 selected an Air Traffic Control Beacon Interrogator (ATCBI)-5 antenna alignment procedure as the focus of this study. Researchers conducted the test at an Airport Surveillance Radar (ASR)-9 site at the William J. Hughes Technical Center. Thirteen AF specialists completed the procedure twice, once with the VT system and again with a paper version of the maintenance manual. Researchers counterbalanced the order of presentation with half of the participants using the paper manual (PM) first and the other half starting with the VT system.

The results showed no differences in task completion time or workload, suggesting that the VT system was no more time consuming or difficult to use than a traditional PM. The voice recognition rate was 86.6%. Questionnaire responses showed that users found the VT system understandable, easy to control, and responsive to voice commands. When asked to compare VT to the use of a PM for the ATCBI-5 maintenance procedure, study participants indicated that the VT system made the ATCBI-5 task easier to perform, was more efficient and effective than a PM, and would be better for handling large amounts of technical information.

Researchers concluded that this study resulted in a successful demonstration of VT for the ATCBI-5 maintenance procedure. They obtained positive responses from the participating AF specialists, even with their lack of familiarity with the VT system. Suggestions for improving the intelligibility of the speech output and user interface were made. It was recommended that further, more extensive studies should be conducted using VT systems in a wider variety of AF environments and tasks.

1. Introduction

1.1 Background

Voice technology (VT) allows computer systems to recognize speech input and create output in spoken form. The Federal Aviation Administration (FAA) is interested in exploring the potential of VT for increasing the efficiency and effectiveness of Airway Facilities (AF) system specialists in field settings. Accordingly, ACT-530 conducted a research program to evaluate a system incorporating voice recognition and synthesis capabilities to support a realistic maintenance task.

This test report describes a feasibility study conducted at the William J. Hughes Technical Center in December 1996. The test plan incorporated the results of the *Voice Technology Literature Survey* (MacMillan & Getty, 1996), which helped identify where VT could be best applied in the AF environment.

ACT-530 chose the air traffic control beacon interrogator (ATCBI)-5 antenna transmission system check as a suitable task for evaluating a prototype VT system. To complete it, the specialist follows a prescribed step-by-step procedure to test parameter values, consulting a technical manual for directions. While working, the AF system specialist needs information at the same time as being occupied with looking at and controlling test equipment. A VT system, which can deliver spoken instructions controlled by simple voice commands, could potentially allow the specialist to continue with the task without having to look at or manipulate written materials. Previous research indicated that VT has been useful in environments where there are heavy visual and manual dexterity demands such as on vehicle assembly lines (MacMillan & Getty, 1996).

1.2 Objective

The purpose of this study was to evaluate the potential of computer-based voice production and recognition technology in the AF operational environment. The following specific questions were posed:

1. Does the use of VT to complete the ATCBI-5 antenna alignment procedure result in an improvement in task completion time or subjective workload?
2. Do AF specialists find VT acceptable for use in the maintenance environment?
3. Are the current capabilities of VT sufficient to support the ATCBI-5 antenna alignment procedure based on the intelligibility of the voice output and reliability of the voice recognition software?

2. Technical Approach

2.1 Voice Technology System

Researchers created an interactive VT system using commercial off-the-shelf (COTS) technology. This system supplied spoken instructions for the ATCBI-5 antenna alignment task while responding to navigational voice commands. The VT software and hardware were installed in a Gateway 2000, 120 MHz Pentium laptop computer, which incorporated a Soundblaster sound card and *DragonDictate* (1995) voice recognition software. A *Text-to-Speech Software Development Kit Version 2.00* (1995) provided the text-to-speech function. Voice output and input were provided by a headset with an integral, directional microphone. The experiment developers drew ATCBI-5 instructions for the VT system and paper manual (PM) from the FAA publication entitled *Maintenance of Air Traffic Control Beacon Interrogator (ATCBI)-5 Equipment and Mode-S Collocated with Solid-State Radar Beacon Decoder (SSRBD)* (FAA, 1989). The text of the instructions was typed into the computer and is in Appendix A. Available voice commands for the VT system are in Appendix B.

2.2 Study Setting

The ATCBI-5 system was located at an Airport Surveillance Radar (ASR)-9 site on the grounds of the William J. Hughes Technical Center. The experiment staff moved the ATCBI-5 system into an outer equipment room in the ASR-9 building to reduce the effects of background noise. The test required a power meter, directional coupler, oscilloscope, crystal detector, and 50-Ohm terminations. Tables, equipment dollies, and power cords were also provided.

The experiment staff positioned a VHS video recorder on a tripod to record the participant's activities during evaluation trials. A mixer enabled sound recordings of both the participant's vocalizations and the computer voice output on the videotape.

3. Methods

ACT-530 employed the following approach for this experiment.

3.1 Participants

An AF subject matter expert (SME) recruited 13 AF maintenance specialists for the study. The specialists completed a Background Questionnaire that identified years in current position and familiarity with the ATCBI-5. There were 9 men and 4 women in the participant group. One participant's data were randomly removed from the data sets for some of the analyses to permit balanced comparisons.

Average number of years in current position for the thirteen participants was 9.6 (the range was from 2.5 to 22 years).[1] All but three participants were certified on the ATCBI-5, although all had received training on and had worked with the equipment. The average amount of time since last completing the ATCBI-5 procedure was 1.8 years (with a range from a few days to 6 years). On average, the participants used a personal computer 7.1 hours per week (with a range of 0 to 15 hours). Only one specialist had any previous experience with VT apart from the public telephone system.

At the ASR-9 site, a study director, VT developer, and AF SME conducted the experiment sessions. A technician from the Research Development and Human Factors Laboratory (RDHFL) set up the video recording equipment.

3.2 Briefing

Participants arrived at different times during the 10-day assessment period. Each group met at the RDHFL, and the research staff gave a briefing on the goals of the study. The staff answered questions, provided individualized schedules, and distributed the Background Questionnaire. They informed participants that their cooperation was voluntary and that they could withdraw from the test at any time. Researchers kept participant identities strictly confidential by assigning each a number to label questionnaires and forms. Participants attended the VT evaluation at the ASR-9 site. They also took part in other demonstrations and evaluations being conducted at the RDHFL.

3.3 Test Design

The study used a within-subjects design with maintenance procedure presentation method as the independent variable. There were two conditions, VT and PM. In the VT condition, participants used the laptop computer with voice input/output capabilities to accomplish the maintenance task. In the PM condition, they used a printed booklet to complete the ATCBI-5 maintenance procedure.

For the VT condition, the user wore the headset, and the computer speech synthesis software spoke the technical manual text through the earphone. The user's voice commands controlled the system through the headset microphone. No text maintenance materials were available to the participant during this condition. The maintenance procedures were equivalent for the two conditions.

Researchers divided the participants into two groups. Group A performed the PM test condition first, while Group B worked the VT condition first. Researchers provided participants with schedules indicating the time for their test run. One participant at a time completed each test session and experienced both experimental conditions.

---

[1] Some participants had been with the FAA longer than the time stated, but the question only addressed the current position.

After a general orientation, researchers gave each participant from Group A a brief introduction to the PM part of the test and mentioned that the VT condition would follow. The study director requested that no questions be asked during the trial unless it appeared that there was a problem with the test procedure or equipment. In such cases, the experiment staff assisted the participant but kept the intervention as brief as possible.

Following the introduction, the study director checked that all equipment was ready for the test trial. The participant started the experiment by reviewing the paper maintenance manual without completing any of the tasks. The staff turned on the video camera before the trial began and recorded the activities and vocalizations of the participant. Start and stop times and any problems or errors encountered were also recorded. Researchers asked the participant to fill out a Workload Questionnaire after completing the trial.

Following a short break, the VT condition began. The participant trained the system on a limited vocabulary set and used it to complete a preliminary review of the maintenance procedure. Ensuring that the ATCBI-5 and voice equipment were at a baseline configuration, the study director then activated the video recorder, and the session began. Participants completed workload and other questionnaires at the end of the trial.

Table 1 shows the Group A timetable. The Group B test involved a reversal of the VT and PM conditions so that participants experienced VT first. Scheduling of Group A or B participants was alternated between morning and afternoon sessions each day.

Table 1. Experiment Timetable

Duration (min)	**Activity**
10	Introduction to PM
45	Maintenance Procedure
15	Questionnaires
15	Break
10	Introduction to VT
30	Voice Recognition Training
45	Maintenance Procedure
20	Questionnaires

3.4 Data Collection

The following sections depict dependent measures of interest.

3.4.1 Background Sound Level

Research staff used a Radio Shack Sound Level Meter (Model Number 33-2055) to take sound level measurements in the outer equipment room. This was to establish a baseline for comparisons to other AF installations.

3.4.2 Time

The researchers measured the time for each participant to complete the test procedure from the start signal to the completion of the final step. The laptop computer recorded the onset of voice-triggered commands and duration of each maintenance step. It kept track of repeated steps and returns to previous steps. The laptop computer was also used by the study director to record the times of each step of the PM condition.

3.4.3 Workload

Participants recorded their perceived workload after each run using the National Aeronautics and Space Administration Task Load Index (NASA-TLX) form (Hart & Staveland, 1988).

3.4.4 Performance

Researchers initially had intended to count the number of errors made in the procedure. However, they decided that this was not practical because it was not possible to reliably attribute errors to specific causes. Two other possible measures of the performance of the VT system were to count (a) returns to a previous step, and (b) the number of times participants requested a repeat step to replay an instruction on the VT system. Rereading steps may have been related to problems with method of presentation.

Although researchers kept a record of returns to a previous step for both conditions, they could not determine if participants re-read an instruction page in the PM condition (the equivalent of a repeat step in the VT system). Therefore, they analyzed only the return to previous step data.

3.4.5 Voice Recognition Rate

Researchers determined the recognition rate of the VT system by calculating the number of times it failed to respond to a voice command, misidentified a command, or reacted to sounds or vocalizations that were not intended as commands. They captured these on videotape in the VT condition only.

3.4.6 Subjective Evaluation

Subjective data were gathered using

a. a Speech Recognition System Questionnaire given at the end of the VT condition,

b. a Usability Questionnaire (comparing the PM and VT conditions) given to each participant at the end of each test run, and

c. a Debriefing Questionnaire with open-ended items given at the end of each run.

4. Results

4.1 Data Reduction

Research staff extracted total completion time and number of steps from the laptop computer records. They then completed videotape analysis to determine the voice recognition error rate for the VT system. All other data (except for sound measurements) were gathered from questionnaire responses and participant debriefings. The data for one subject who experienced the VT system second were randomly removed from all data sets involving PM versus VT comparisons to permit a balanced analysis (equal numbers in the PM and VT first groups).

4.2 Sound Level

The sound level in the test area ranged from 54 to 70 dBA. (Typical speech is about 60-70 dB.) This represents a relatively low level of environmental noise. However, AF system specialists reported that they conduct many maintenance procedures under much noisier conditions. Informal tests of the VT system suggested that it may function well in a noisier environment.

4.3 Task Completion Time

There was a strong practice effect evident in the response time data. It took nearly twice as long for participants to complete their first run through the ATCBI-5 maintenance procedure (trial one, $M = 887$ sec, $SD = 223$ sec and trial two, $M = 499$ sec, $SD = 194$ sec), irrespective of whether the VT or PM condition came first. This was probably due to the level of unfamiliarity that some participants showed with the procedure. The experimental design incorporated counterbalancing of the order of presentation to compensate for this effect. Six participants experienced the VT condition first and seven were given the PM condition first. When the data were analyzed on this basis (with one subject removed), the means for completion time were VT, $M = 665$ sec, $SD = 222$ sec and PM, $M = 722$ sec, $SD = 343$ sec. A t test revealed no significant difference between the two times, $t(11) = -0.43$, $p = .6782$. This shows that VT was no more time consuming than the PM to complete the maintenance task.

4.4 Workload

Part of the NASA-TLX workload evaluation process required users to weight six components of workload in order of importance. The AF system specialists informed us that the order of the factors (from most to least critical) was: Performance, Mental Demand, Temporal Demand, Effort, Physical Demand, and Frustration.

Researchers compared workload ratings for the VT and PM conditions on each NASA-TLX sub-scale and for overall workload (as shown in Figure 1). They found that there were no statistically significant differences between the conditions on any of the scales. Average workload for the VT condition was $M = 5.2$, $SD = 1.5$, and $M = 5.3$, $SD = 0.95$ for the PM condition (1 = extremely low and 10 = extremely high). This indicates that workload for completing the ATCBI-5 antenna alignment procedure was in the moderate range.

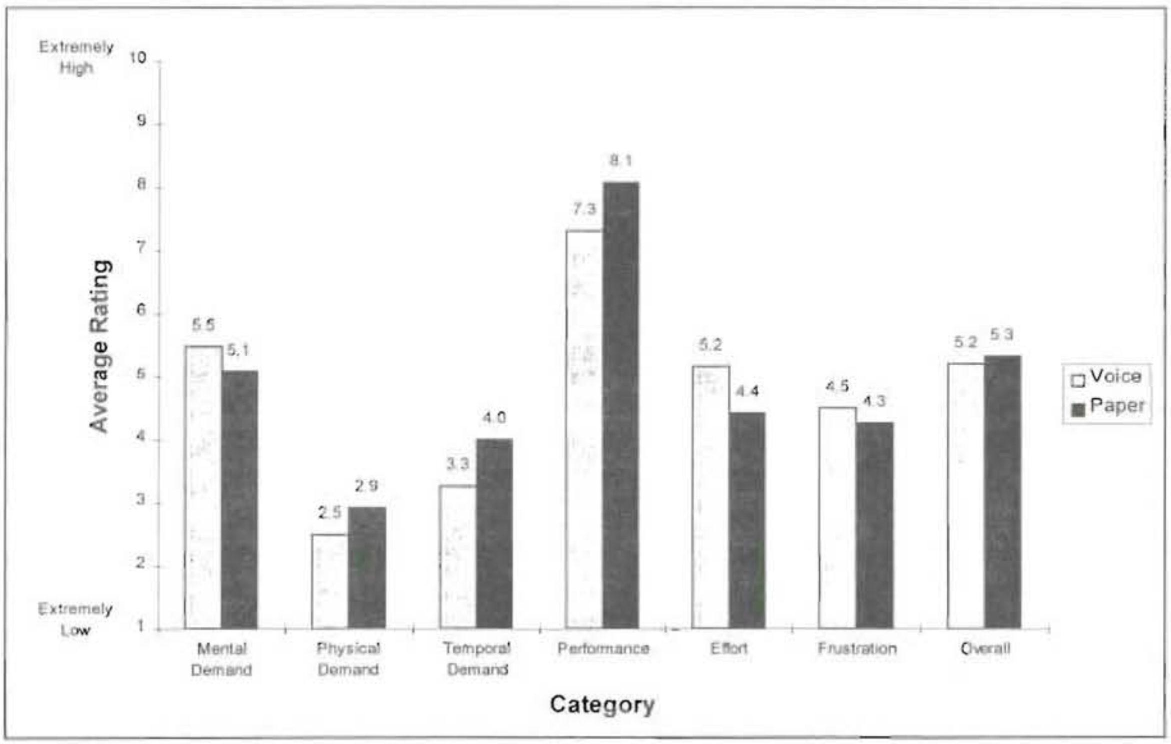

Figure 1. NASA-task load index workload ratings.

Most other NASA-TLX sub-scale ratings were also in the moderate range with lowest ratings for physical and temporal demand. Self-rated performance (as defined by success and satisfaction in performing the task) for the VT condition was $M = 7.3$, $SD = 2.1$ and $M = 8.1$, $SD = 1.7$ for the PM condition (out of a possible 10), suggesting that participants felt generally satisfied with their performance of the maintenance task under both test conditions.

4.5 Performance

In the VT condition, participants requested that the system go back at least one step nine times. In the PM condition, they turned back one or more pages 34 times. It is unclear why there was such a large disparity. It may be that the VT system allowed for improved comprehension of the instructions as compared to the PM condition, requiring less searching for information. On the other hand, users may have found it easier to page back through the PM and, therefore, did so more often.

4.6 Voice Recognition Rate

Researchers completed an analysis of the voice recognition errors made by the VT system to determine the recognition rate of the system. They separated errors into no response, incorrect response, wrong command, and wrong mode.[2] Successful voice recognition rates for individual participants varied from 75% to 100% with an average success rate of 86.6%.[3]

A breakdown of the error data showed that the "go to sleep" and "wake up" commands accounted for 40.6% of the "no" or "incorrect" response recognition errors. A further 32.4% were attributed to the most frequently used navigational commands, "continue" and "yes." Apart from legitimate commands that resulted in no response or an incorrect response from the system, 21% of errors occurred because the system was in the wrong mode. Either it was in a "sleep" mode and the user was trying to activate an unavailable action without first saying "wake up," or it had reverted to dictate mode where the normal commands were not valid.

4.7 Subjective Data

Researchers gave a Voice Technology System Questionnaire to participants immediately after their experience with the VT system. It posed the following questions, to which the participants agreed or disagreed on a 5 point scale (1 = strongly disagree and 5 = strongly agree):

1. I preferred to hear rather than read instructions while I was performing the antenna transmission system check.
2. Hearing rather than reading the instructions allowed me to complete the system check more easily.
3. Hearing rather than reading the instructions allowed me to complete the system check more quickly.

---

[2] The researchers did not include wrong command and wrong mode errors in the primary error analysis. They did not consider the use of incorrect commands by specialists as recognition errors. Wrong mode data were analyzed separately.

[3] Success Rate = (Correct Responses/(Correct Responses + No Responses + Incorrect Responses)) * 100%

4. I found the voice output of the system to be understandable.
5. I found it easy to control the system through speech commands.
6. The system recognized my commands.
7. I would use a spoken-instruction system for periodic maintenance tasks if it was available.

Figure 2 shows the results of the participants' ratings. Ratings were above 3 (a neutral response) tending to favor the VT system, except for question 3. Means and standard deviations are found in Table 2.

The results showed that the specialists found the VT speech output understandable. They also indicated that the system was easy to control using speech commands and recognized instructions reliably. There were also indications that hearing rather than reading instructions allowed an easier completion of the system test and that specialists would use such a system for maintenance tasks if it was available. Participants did not show a preference for hearing rather than reading instructions, and they suggested that the VT system might be a little slower for completing the ATCBI-5 procedure.

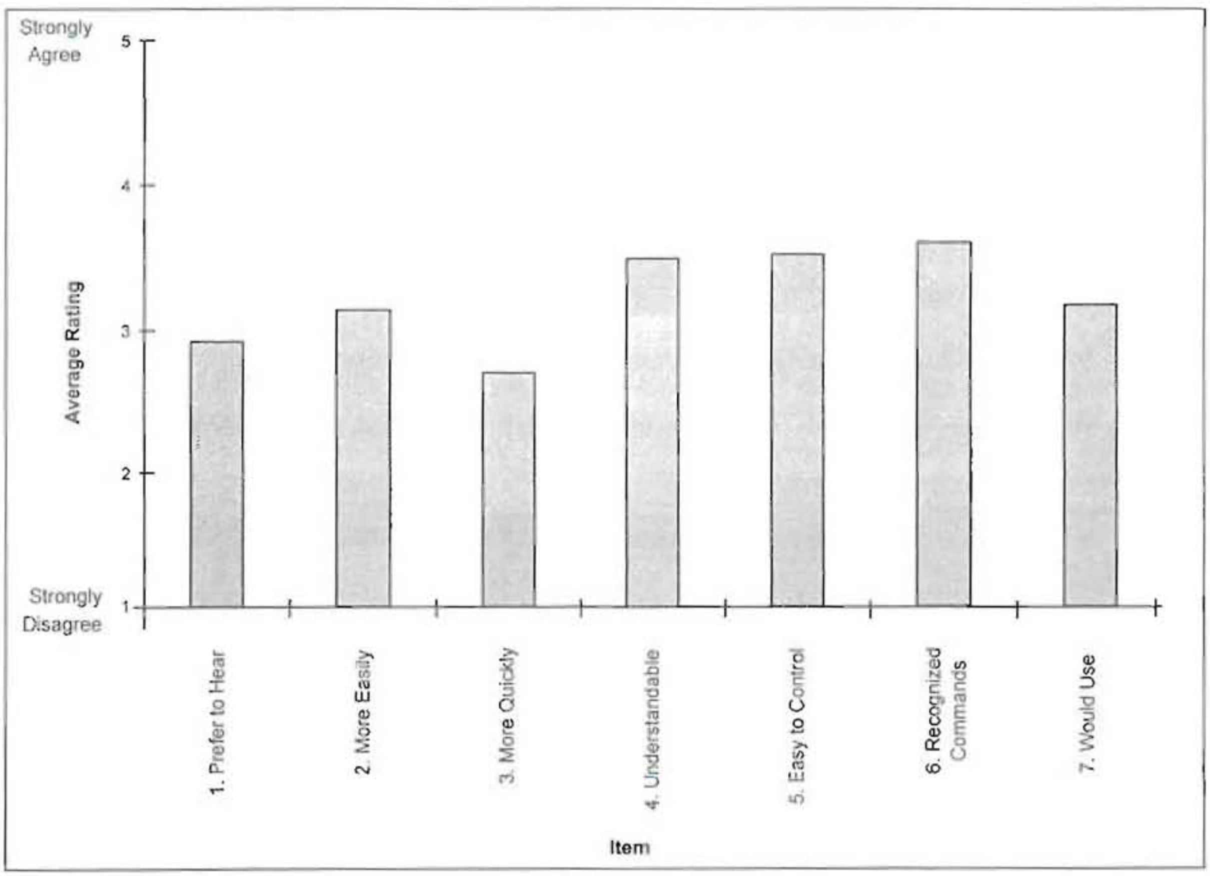

Figure 2. Responses to the voice technology system questionnaire.

Table 2. Statistics for Voice Technology System Questionnaire

Question	Mean	Standard Deviation
1	3.3	1.2
2	3.7	1.3
3	3.0	1.3
4	3.7	1.2
5	3.3	4.0
6	4.3	0.8
7	3.8	1.5

At the end of each session, participants filled out a Usability Questionnaire that requested a comparison between the VT and PM options on the following questions (using a 5-point scale with 1 = PM and 5 = VT).

1. With regard to the ease of performing an antenna transmission system check procedure, which method do you prefer?
2. With regard to the efficiency (time and actions) of completing the procedure, which method do you prefer?
3. With regard to the clarity of the instructions, which method do you prefer?
4. With regard to the ease of finding the information needed, which method do you prefer?
5. If you made a mistake in the procedure, which method made it easier to recover?
6. Overall, which method was most effective in helping you to complete the ATCBI-5 maintenance procedure?
7. Which method would be most efficient for handling the large amounts of technical information needed to support maintenance procedures?

Figure 3 shows the results of the participants' ratings. Ratings were above 3 (a neutral response) in favor of the VT system, except for questions 3 and 4. The means and standard deviations for each question are shown in Table 3.

Participants preferred the VT system for ease of performing the ATCBI-5 procedure and found it to be more efficient and effective than the PM for this task. They also thought that it would be more effective than a PM for handling large amounts of technical information. They tended to prefer the PM for ease of finding information. There was no strong preference for clarity of instructions and ease of recovering from mistakes.

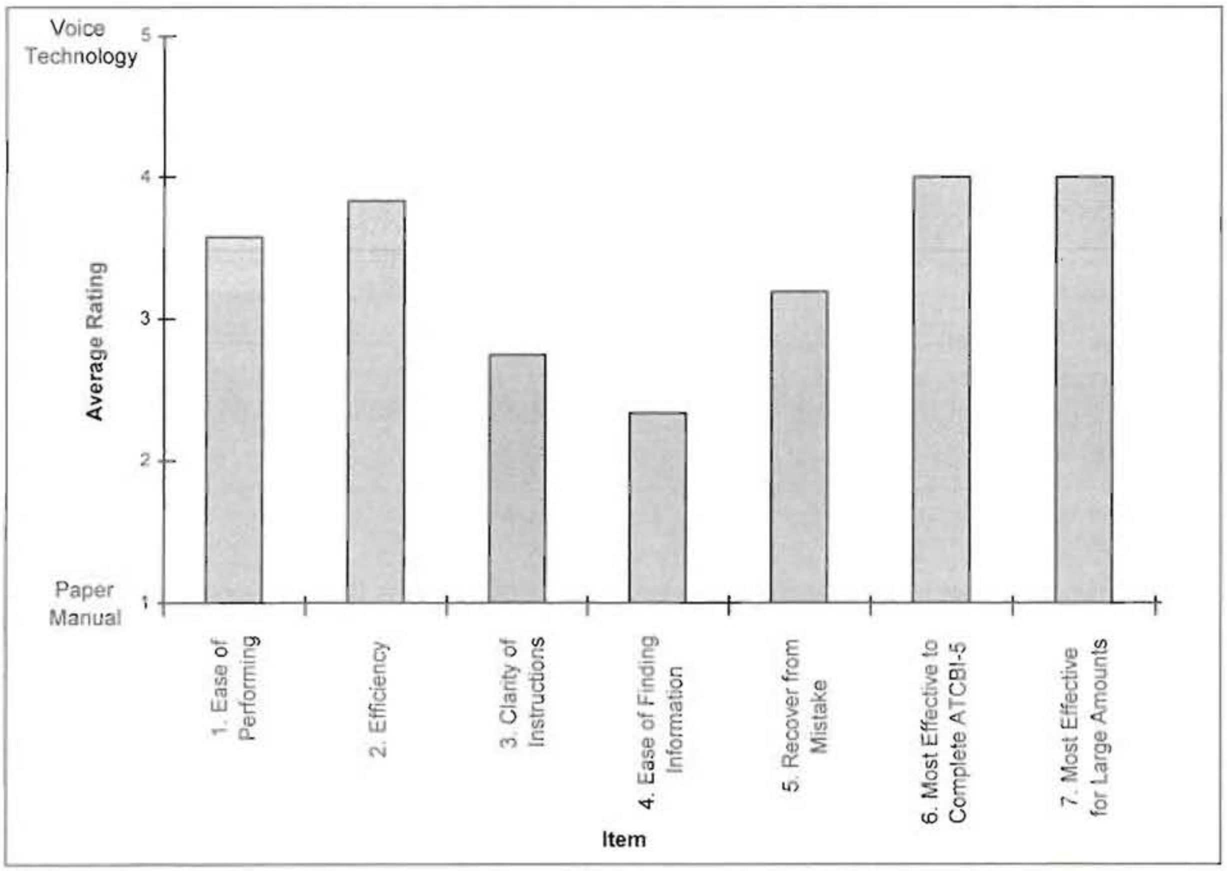

Figure 3. Usability questionnaire results.

Table 3. Statistics for Usability Questionnaire

Question	Mean	Standard Deviation
1	3.6	1.1
2	3.8	1.3
3	2.8	1.3
4	2.5	1.5
5	3.3	4.0
6	4.0	1.1
7	4.0	1.3

Researchers asked participants to make comments about the VT system in a final Debriefing Questionnaire. The questions with a summary of comments are listed below. (Comments from the Voice Technology System and Usability Questionnaire are also included.)

1. Overall, did you find the voice system useful?

   Participant responses to the system were very favorable. Eleven specialists answered "yes" to this question. One found the system "somewhat" useful, and another did not like it.

2. What was most useful about the system?

   Some participants liked the ability to start, stop, and repeat procedures and mentioned the ease and flexibility of the system. One user thought the VT system made instructions easier to interpret. There were many comments about the advantage of having hands and eyes free. With the VT system, the maintenance procedure can be completed while listening to the instructions rather than reading and then doing. One participant said that although the manual was still necessary for in-depth explanation of maintenance procedures, he was impressed by and preferred the VT system. Another veteran specialist thought that the voice method was twice as fast. Several felt the VT system permitted better concentration on the task and equipment because no reading was required. One participant said that hearing the instructions would be preferred if one was very familiar with the test procedure. Reading might be better while learning.

3. What was least useful about the system?

   Seven participants did not have any criticisms. One user (the specialist who answered no to question 1) found the system very cumbersome and preferred the PM. Another wrote that commands had to be repeated too often. The headset cable was cited as a problem in that it tended to impede mobility. Other concerns were that the user must listen carefully, that training and learning to use the system might be a problem, and that pausing the system was difficult. Some observed that if some speech output was missed on the VT system, the whole segment had to be repeated as opposed to re-reading just that section in a PM. One specialist preferred the PM because he thought it would be frustrating to formulate the right questions for the VT system to obtain a particular solution.

4. Do you have suggestions for improving the voice system?

   One specialist suggested having more diagrams and graphics on the screen to support the procedure. Several users requested improved control over the voice output for speed, frequency, or type of voice. There was some discussion of using a digitized human voice. Two participants reported partial hearing loss and suggested that more control over the voice output might have helped them adapt the system to their capabilities. An infrared headset was suggested as a way to free the user from cables connected to the computer. A headset with two headphones might help to reduce the effects of background noise. The ability to go back more than one step at a time in the procedure and an automatic go to sleep function between steps was suggested. The go to sleep function should have an auditory acknowledgment feature so that looking at the screen is not necessary. Including the printed manual text on the screen of the VT system would allow the user to quickly look at the information and skip ahead more quickly, if desired. Some ability to adapt the system for expert users who do not need to read or hear all the instructions may also be needed.

5. What is the most difficult thing about performing a periodic maintenance procedure such as an antenna transmission system check?

   The specialists mentioned

   a. going back and forth to the book while having to leave the equipment,
   b. setting up test equipment and carrying it to the location to perform the task,
   c. lack of explicit directions,
   d. climbing up and down ladders to read the book,
   e. manipulating test equipment,
   f. calculating the voltage standing wave ratio,

g. measuring pulse width, and

h. knowing what to do and then making sure that the measurements taken are within tolerances.

5. Conclusions

In this section, we review and interpret the results of the study concerning the three specific questions posed in Section 1.2.

1. Does the use of VT to complete the ATCBI-5 antenna alignment procedure result in an improvement in task completion time or subjective workload?

   The task completion time and workload data did not show any statistically significant differences between the VT and PM conditions. Workload was in the moderate range for the completion of the maintenance procedure, and participants seemed satisfied with their work in both the VT and PM conditions as shown by their high scores on the NASA-TLX performance scale.

   The finding of no performance time or workload differences is a positive result given that we had just introduced the participants to the VT system. The use of VT for the ATCBI-5 procedure was apparently no more time consuming or difficult than using a PM. Questionnaire responses regarding VT showed an overall positive response with mention of several specific advantages, particularly with regard to having their hands free.

   To reduce the effects of the relative unfamiliarity of specialists with the chosen AF task, future studies should include additional practice before the first run of the experiment. This will help reduce any learning or refamiliarization that may occur as they are first exposed to the maintenance procedure. While the counterbalancing strategy used in this study was effective for neutralizing the effect of practice on the results, additional preliminary training would help ensure better control over this potentially confounding factor.

2. Do AF specialists find VT acceptable for use in the maintenance environment?

   Participant questionnaire responses to the VT system were generally positive. They found the VT system easy to control, thought that it recognized their commands adequately, and said they would use such a system, if available, for maintenance tasks. When asked to compare the VT and PM options for use in the ATCBI-5 maintenance procedure, specialists indicated that the VT system was more efficient and effective and better for handling large amounts of technical information. Written responses were also positive with 11 of the 13 participants indicating they found the system useful. The primary advantage stated was the ability to work with hands and eyes free, a frequently mentioned benefit of VT systems used in other environments (MacMillan & Getty, 1996). It could be expected that acceptance of the system would improve even further with modifications to the quality of the voice output, incorporation of a wireless headphone and microphone system, interface improvements, and other features.

It is uncertain why specialists paged back more frequently using the PM. Without further information on this issue, we cannot draw a conclusion regarding any advantages or disadvantages of the VT system in this regard.

3. Are the current capabilities of VT sufficient to support the ATCBI-5 antenna alignment procedure based on the intelligibility of the voice output and reliability of the voice recognition software?

   A critical concern with VT systems is speech intelligibility and voice recognition accuracy rate. As noted in the participants' questionnaire responses, intelligibility of the voice output was generally acceptable, although there were several stated concerns regarding speed, frequency range, and type of voice. Two of the specialists had partial hearing loss, which may have introduced additional problems. Better control over the voice production software would help improve acceptance and allow those with specific preferences or difficulties (such as hearing loss) to adjust the system for their use. Digitized human voice should also be considered. If VT was implemented, specialists would presumably have their own system, which could be tailored to their individual preferences.

   Overall voice recognition accuracy was fairly high at 86.6%. Participants found this acceptable as evidenced by questionnaire responses giving the system a good rating for ease of control through speech commands and recognition of commands. This indicates that, even though the system made some errors, the users did not find that it significantly interfered with their work or their favorable impressions of its potential. With further refinements to the VT system and increased familiarity by the users, the recognition rate could be expected to improve.

   About 40% of the specific recognition errors were associated with placing the system in the sleep mode. There was no visual or auditory indication when the system was asleep, and this may have resulted in the users losing track of system mode status and using the wrong commands. Future revisions of the VT application should seek to rectify these problems, which will help reduce the voice recognition error rate. Participants also had problems when the system inadvertently entered dictate mode. This was a feature of DragonDictate in its COTS version and would probably not be needed for a dedicated AF maintenance support tool. Concerning the capabilities of VT, the system used in this study with some refinements is capable of supporting tasks such as the ATCBI-5 antenna alignment procedure.

In summary, the VT evaluation provided useful information on the application of text-to-speech and voice recognition products to support AF maintenance tasks. Although this initial study did not show reductions in completion time or workload as a result of using VT, there were few negative effects that might be expected from the introduction of a new system. A controlled study emphasizing situations where VT would be most applicable could be expected to demonstrate performance benefits.

With few exceptions, participants reacted to the VT system very positively. The AF system specialists appreciated the ability to concentrate on the chosen maintenance task without having to read and manipulate a PM. They believed that computer technology holds considerable promise in the management of the large amounts of technical information necessary for their work. Judging by the results of this evaluation, VT appears to have excellent potential for use in interactive applications for AF maintenance procedures.

6. Recommendations

1. Consider further research on the feasibility of using VT for AF maintenance applications in several types of settings or applications.
2. Consider exercising the potential of the computer for storage, display, and search of large amounts of text and graphics associated with AF manuals.
3. Allow for sufficient familiarization and practice with the selected maintenance procedures to reduce practice effects.
4. Conduct additional studies in realistic AF environments regarding background noise, equipment types, and location.
5. Evaluate the following recommended improvements to the VT system.
    a. Provide a wireless headset or microphone, or both, with two earpieces.
    b. Permit more control over the speed, frequency range, and type of voice of the text-to-speech system.
    c. Evaluate the feasibility of digitally recorded voice.
    d. Reduce recognition errors due to mode problems by improving the VT interface.
    e. Consider options for the display of text and graphics on the VT system computer monitor.

References

DragonDictate [Computer software]. (1995). Newton, MA: Dragon Systems, Inc. (Version 2.0).

Federal Aviation Administration. (1989). *Maintenance of air traffic control beacon interrogator (ATCBI-5) equipment and mode-S collocated with solid-state radar beacon decoder (SSRBD)* (DOT/FAA 6360.14A). Washington, DC: U.S. Government Printing Office.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload*. Amsterdam: North Holland.

MacMillan, J., & Getty, D. J. (1996). *Voice technology literature survey*. Cambridge, MA: BBN Systems and Technologies.

Text-to-Speech Software Development Kit Version 2.00 [documentation]. (1995). Woburn, MA: Lernout and Houspie Speech Products.

Appendix A
Antenna Transmission System Procedure Text

Note: The following steps were used for both the VT and PM procedures.

Step 0: You will need a power meter, a directional coupler, an oscilloscope, a crystal detector, and 50-Ohm terminations. RG-58U cable may be used as a test cable so that the power meter may remain inside the transmitter building during measurements of transmission line attenuation.

Step 1: These checks must be performed with the facility removed from service. Please remove the facility from service or stop here.

Step 2: High voltage of the primary radar must be turned off during measurements at the beacon antenna. Turn off the high voltage of the primary radar.

Step 3: If the omnidirectional antenna is suspension mounted, you may not be able to check incident power or voltage standing wave ratio (VSWR) at the input of the antenna. If this is the case, you should measure source VSWR.

Step 4: This step is necessary to prevent damaging the thermistor (also known as a bolometer). Make sure the coupling of the directional coupler and attenuation of the 50-Ohm test cable (if used) reduces the power level so that the rated limit of the power meter is not exceeded.

Step 5: Note that before making each radio frequency (RF) power measurement, adjust the power meter for a zero meter reading. Then connect it to measure the RF power level.

Step 6: With the thermistor connected, turn on the power meter and allow it to warm up and stabilize.

Step 7: If the power meter is warmed up, proceed to the next step.

Step 8: Source VSWR sub-procedure. Perform the following measurements at the directional and omnidirectional outputs. Connect the 10 decibel pad to the incident power port of the directional coupler.

Step 9: Connect the crystal detector to the 10 decibel pad.

Step 10: Connect the 50-Ohm cable, RG 58U, or equivalent, from the crystal detector to the oscilloscope with a convenient trigger, such as beacon sync, and observe the leading edge pulse.

Step 11: Adjust the scope so the leading edge of the detected pulse is referenced to a convenient graticule line.

Step 12: Measure the pulse width of the incident pulse.

Step 13: Disconnect the detector and 10 decibel pad from the coupler.

Step 14: Connect the detector to the reflected power port of the coupler and observe the reflected P2 or P3 pulse.

Step 15: If the leading edge of the reflected pulse is approximately coincident with the leading edge of the incident pulse, the reflection point is located near the source. If the leading edge of the reflected pulse is delayed from the reference graticule, measure the delay and determine the approximate location of the reflection point measured from the source. That is, multiply the time delay by the velocity, 492 feet per micro-second, to determine the location of the reflection.

Step 16: Check for significant multiple reflections. Multiple reflections are indicated when the leading edge of the reflected pulse is nearly coincident with the leading edge of the incident pulse and the trailing edge of the reflected pulse if delayed from that of the incident pulse.

Step 17: Disconnect the detector from the coupler.

Step 18: Measure the power levels at the incident power port and the reflected power port of the coupler.

Step 19: Subtract the dBm value of the reflected power meter reading plus coupling from the dBm value of incident power meter reading plus coupling and record this value.

Step 20: The source VSWR should be within a 1.5 to 1 ratio for initial systems and a 1.7 to 1 ratio for operational systems. The following page contains the conversion table of dB return loss VSWR. Locate the value found in Step 19 on the conversion table to determine the VSWR.

Step 21: This ends the source VSWR procedure.

Conversion Table - dB Return Loss to VSWR

Return Loss Pi−Pr(dB)	Percent Reflected Power 100 Pr/Pi(W)	VSWR	Return Loss Pi−Pr(dB)	Percent Reflected Power 100 Pr/Pi(W)	VSWR	Return Loss Pi−Pr(dB)	Percent Reflected Power 100 Pr/Pi(W)	VSWR
35.0	0.03	1.04	17.8	1.66	1.30	12.9	5.13	1.59
34.0	0.04	1.04	17.6	1.74	1.30	12.8	5.25	1.59
33.0	0.05	1.05	17.4	1.82	1.31	12.7	5.37	1.60
32.0	0.06	1.05	17.2	1.91	1.32	12.6	5.50	1.61
31.0	0.08	1.06	17.0	2.00	1.33	12.5	5.62	1.62
30.0	0.10	1.07	16.8	2.09	1.34	12.4	5.75	1.63
29.0	0.13	1.07	16.6	2.19	1.35	12.3	5.89	1.64
28.0	0.16	1.08	16.4	2.29	1.36	12.2	6.03	1.65
27.0	0.20	1.09	16.2	2.40	1.37	12.1	6.17	1.66
26.0	0.25	1.11	16.0	2.51	1.38	12.0	6.31	1.67
25.0	0.32	1.12	15.8	2.63	1.39	11.9	6.46	1.68
24.0	0.40	1.13	15.6	2.75	1.40	11.8	6.61	1.69
23.5	0.45	1.14	15.4	2.88	1.41	11.7	6.76	1.70
23.0	0.50	1.15	15.2	3.02	1.42	11.6	6.92	1.71
22.5	0.56	1.16	15.0	3.16	1.43	11.5	7.08	1.73
22.0	0.63	1.17	14.8	3.31	1.44	11.4	7.24	1.74
21.5	0.71	1.18	14.6	3.47	1.46	11.3	7.41	1.75
21.0	0.79	1.20	14.4	3.63	1.47	11.2	7.59	1.76
20.5	0.89	1.21	14.2	3.80	1.48	11.1	7.76	1.77
20.0	1.00	1.22	14.0	3.98	1.50	11.0	7.94	1.78
19.8	1.05	1.23	13.9	4.07	1.51	10.8	8.32	1.81
19.6	1.10	1.23	13.8	4.17	1.51	10.6	8.71	1.84
19.4	1.15	1.24	13.7	4.27	1.52	10.4	9.12	1.87
19.2	1.20	1.25	13.6	4.37	1.53	10.2	9.55	1.89
19.0	1.26	1.25	13.5	4.47	1.54	10.0	10.00	1.92
18.8	1.32	1.26	13.4	4.57	1.54	9.8	10.47	1.96
18.6	1.38	1.27	13.3	4.68	1.55	9.6	10.96	1.99
18.4	1.45	1.27	13.2	4.79	1.56	9.4	11.48	2.03
18.2	1.51	1.28	13.1	4.90	1.57	9.2	12.02	2.06
18.0	1.58	1.29	13.0	5.01	1.58	9.0	12.59	2.10

Appendix B
Voice Technology System Vocabulary List

The following is a list of the 19 application-specific words that were trained. They are divided by the screen on which they appeared. All screens had a green background except for the Conversion Table Screen, which had a white background. The System Vocabulary words were not trained but were available for control of DragonDictate functions.

Selection Screen

Antenna Transmission System
Change Name
Exit
Show Conversion Table

Yes/No Screen

Yes
No

Sub Procedure Screen

Source VSWR
Cancel Procedure

Response Form

Continue
Repeat Step
Previous Step
Cancel Procedure

Conversion Table Screen

Repeat Step
Close

System Vocabulary

Close Window
Command Mode
Dictate Mode
End Task
Go to Sleep
Oops
Wake Up