JSC 27617

DTO-675 “Voice Control of the Closed Circuit Television System

Report

National Aeronautics and
Space Administration

Lyndon B. Johnson Space Center
Houston, TX 77058

October 1996

Voice Command System Report for

STS-78

Prepared By:

George Salazar, Project Manager

Darilyn M. Gaston, Project Engineer

Dena S. Haynes, Project Engineer

Concurrence:

Tom Henricks, STS-78 Commander

Kevin Kregel, STS-78 Pilot

Acknowledgments

The authors wish to thank astronauts Nancy Currie, Pierre Thuot, Tom Jones, Kathy (K.T.) Thornton, Bruce Melnick, Bill Shepherd, Blaine Hammond, Leroy Chiao, Linda Godwin, Kevin Kregel, Tom Henricks, and Ellen Ocha for their important contributions to this DTO.

Special appreciation is expressed to astronaut Julie Payette for her time and effort in reviewing and providing valuable comments and suggestions to enhance this document.

1. Table of Contents

2. List of Tables

3. List of Figures

4. Acronyms and Abbreviations

Amp	Amplifier
CRA	Correct Recognition Accuracy
CCTV	Closed Circuit Television
DBSPL	Decibel Sound Pressure Level
DOE	Department of Energy
DTO	Detailed Test Objective
FDF	Flight Data File
HIU	Headset Interface Unit
IE	Insertion Errors
JPL	Jet Propulsion Laboratory
JSC	Johnson Space Center
MDF	Manipulator Development Facility
OCRA	Overall Correct Recognition Accuracy
OE	Omission Errors
Ops	Operations
PGSC	Payload General Support Computer
PLD	Payload
PTD	Push to Disable
PTT	Push to Talk
RMS	Remote Manipulator System
SE	Substitution Errors
SMS	Shuttle Mission Simulator
VCS	Voice Command System
VLHS	Very Lightweight Headset

5. Executive Summary

This report presents the results of 3-flight Detailed Test Objective (DTO)–675 "Voice Control of the Closed Circuit Television (CCTV) system "experiment flown on STS-78. The DTO is a follow-on flight of the Voice Command System (VCS) that first flew as a Secondary Payload on STS-41 and is scheduled for 2 additional flights to further investigate voice control technology on Shuttle. Several design changes were made to the VCS for the STS-78 mission. This report discusses those changes, reviews and analyzes the data collected, discusses the problems encountered and states conclusions drawn about the DTO.

This DTO collected additional information about the use of voice in space, particularly for controlling a spacecraft system in the actual environment. Unlike, the rather benign environments and applications of voice control on the ground, the shuttle presents a challenge to the use of voice control, particularly in the cockpit. The noise level can be as high as 95 dB and the reverberation of the crew cabin makes it difficult for noise cancellation microphones to work. Furthermore, the astronauts do not have the luxury of using the technology every day like in some ground applications, e.g. inventory inspection as done here on the ground. Consequently, the system must be able to accommodate limited training and usage, sometimes as little as once a month.

The VCS flown on STS-78 contains redesigns requested from the astronaut office:

- Develop a more robust recognition system than what flew in STS-41.
- Have the ability to perform one-word commanding (macro) for creating pre-defined camera scenes to aid the crewmember in simultaneous CCTV and Remote Manipulator System(RMS) operations.
- Aid the crew member in simultaneous CCTV and Remote Manipulator System (RMS) operations.

Overall, the VCS met the first objective although there were some recognition difficulties during some of the sessions. The second met with limited success. The following conclusions/recommendations resulted from this flight:

- Investigation into a better headset that is easy to adjust and has better noise cancellation capabilities is needed to take into account the dynamics of the cockpit. The present design and using the Shuttle headset requires the cockpit to have minimal noise in the background.
- Macro feedback needs improvement to help the user know that the macro is running and its percentage completion.
- Considerations for providing audio feedback **ONLY** to the user is an area requiring investigation. Viewing the display, the monitors, and the panel switch talk backs is distracting. Recognition with the VCS display located directly underneath the monitor should be investigated including having the ability to provide audio **ONLY** feed back to the user.

- Flight data files need to be ready early on to ensure astronauts are fully aware of what is in the document and they are comfortable with the content.

- CCTV training with the VCS should also include commanding the cameras by voice to observe earth views such as an approaching landmark. The commander had difficulty viewing approaching New Orleans using the VCS because of the questionable recognition queries he was receiving during voice commanding of the cameras.

- The data indicates that over the course of the 17 day flight no noticeable changes in the fundamental frequency and the first formant frequencies of the voice took place.

- In their opinion ,the commander and pilot recognize the potential for voice commanding on other systems. However, based on this flight, the VCS in its present implementation is not considered a viable operational system for the shuttle closed circuit television system.

The use of voice control during hands and eyes intensive tasks such as CCTV /RMS operations are considered an excellent application of this technology. Considering their experiences, STS-32 and 39 crew members have stated that a VCS would have significantly reduced the work load during complex RMS and rendezvous operations. Even though this flight did not provide the opportunity to apply voice control with RMS operations, it did however uncover further implementation improvements needed if this technology is to become operational. The remaining two flights of this DTO experiment will hopefully move the use of speech recognition for CCTV operations one step closer.

The results of this experiment have proven valuable in acquiring additional data on the use voice recognition in space. Efforts like controlling cameras with voice while operating a robot arm or dictating a report to a computer while doing gloved-box experiments are possible tasks that one may see on Station, the Moon, or Mars someday. The lessons, difficulties, and other experiment data gathered from the VCS DTO will hopefully help future spacecraft voice recognition implementers.

6. Introduction

Speech is a natural and convenient means for a human to communicate with another human being. What about communicating with a machine? If a machine could respond correctly to spoken commands through appropriate action, then simplification of information exchange between humans and machines could occur. Future space programs such as space Station, return to the moon, and going to Mars could benefit. In particular, if the system performs the equivalent of throwing several switches by speaking one word or if it allows the astronaut's hands and eyes to say focused on a complex task while he/she controls another task by voice, it would help reduce the work load. The purpose of the VCS DTO was to collect additional information related to the use of voice control in a spacecraft system in space. This report discusses the results of the follow-on flight of the Voice Command System.

Conclusions are not made about the recognition accuracy of the experiment or whether it was better than the system that flew on STS-41 since statistically the data and test subjects are insufficient and there was no consistency of users across the two flights of the VCS. However, the report does discuss the problems encountered, possible improvement to training and to the system, and what considerations are needed for using speech recognition for space flight.

This report is intended for a broad audience. There is sufficient detail to familiarize a person with the DTO and the VCS flight experiment. Persons already familiar with the VCS experiment can read the results section, conclusions, and recommendations first.

6.1. Background

The Voice command system (VCS) design allows control of the Shuttle's closed circuit television (CCTV) system by voice. The origin of the VCS flight experiment results of an investigation by JPL and JSC [2]. The idea of controlling the CCTV system by voice occurred because of the simultaneous operation of both the remote manipulator (RMS) robot arm and the CCTV system. The CCTV system contains over 30 switches for operation. During RMS operations, the shuttle astronaut must operate the arm and the CCTV system. To change camera scenes or monitor selection requires the crew member to stop the movement of the arm to make the CCTV adjustments. This disrupts the arm motion, diverts the crew member's visual attention ,and upsets their mental concentration on accomplishing the task. On Space Station, RMS and CCTV operations will require one crew member to operate both as well. Clearly, a need to look more closely at more advanced user interfaces to decrease the RMS operator workload is desired.

In October of 1990, the VCS flew on the five-day STS-41 mission [7]. Two astronauts used the VCS on flight days two, three and four. The VCS on that mission was speaker-dependent requiring the astronauts to train the system with their voices and storing their voice prints (or templates) on memory chips. Both averaged 95% or better during ground training. However, on-orbit one astronaut experienced difficulties with recognition of the 1-g voice prints getting recognition accuracy as low as 33% on flight day two. However, the recognition accuracy increased to 77% on flight day four. Retraining of the command words on-orbit resulted in his accuracy ranging from 45% on flight day two to 95% on the last day. The other astronaut's scores ranged from 72% on flight two to 82% on flight day three. Overall, one astronaut averaged 62% and the other 82%.

Video and audio data showed that microphone placement sensitivity was on of the major contributors to the problems one astronaut was having. In general, speaker-dependent systems tend to be sensitive to vocal characteristic changes of a specific speaker such as amplitude. VCS personnel found that the recognizer used required a considerable amount of time just to obtain good templates. Half of the 16 training sessions were dedicated to obtaining good templates. The STS-41 astronauts recommended improving the camera fine-tuning control feature of the VCS. They also recommended having the VCS create camera scenes through macro commanding. Macro commanding accomplishes the equivalent of several switch closures by the enunciation of a single voice command. This was strongly suggested by both astronauts. The astronauts recommended performing further in-flight investigations of voice technology with a capability of handling changes in the voice. VCS personnel redesigned the entire recognition electronics using a speaker-independent recognizer with adaptation capabilities and got approved for 3-flight DTO with STS-78 being the first flight.

7. Flight Experiment Description

7.1. Overview

The STS-78 VCS is a speaker-independent recognition unit designed to control the Orbiter's CCTV system by voice. The VCS allows an astronaut to operate selected CCTV functions such as pan, tilt, zoom etc. (except for camera power switching). The VCS installs in the A7 upper panel space in the aft flight deck. Key elements of the unit consist of a display unit to provide system status, a control panel with minimal switches, an interface board to the orbiter CCTV system, interconnecting cables from the VCS to the orbiter and a newly redesigned speaker-independent recognizer. A key feature of the redesigned recognizer is the ability to adapt to changes in the user's voice and to remember these changes for future improved recognition. Factors that affect the voice of the astronauts in orbit are stress, fatigue, microphone placement and background noise. In time, the VCS learns the user's voice changes due to factors affecting the voice as mentioned above. Figure 1 shows the block diagram of the interfaces between the VCS and orbiter.

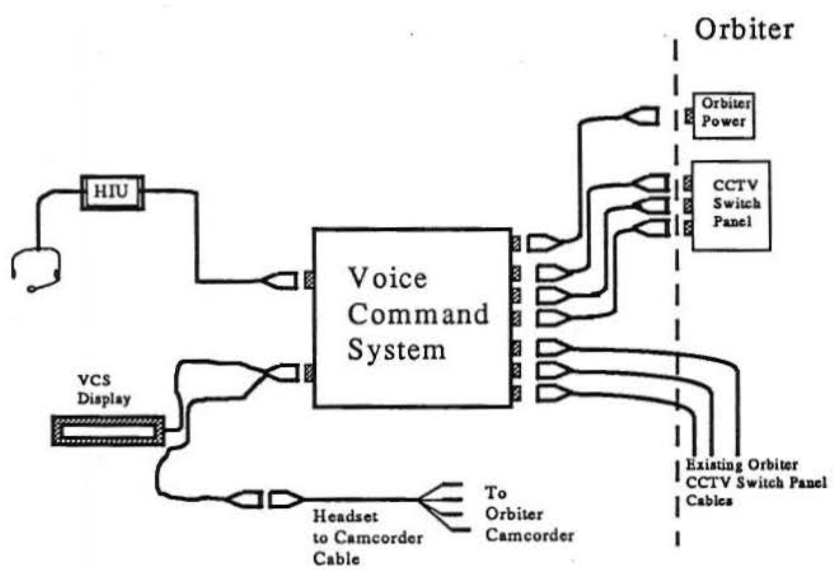


FIGURE 1. VOICE COMMAND SYSTEM INTERFACES WITH ORBITER

The new speaker-independent recognizer replaces the entire electronics related to the speaker-dependent recognizer of the original VCS design. The recognizer is a new electronic box that resides inside the VCS chassis. The recognizer contains: a processor, headset audio interface, solid state disk (SSD), CCTV interface control, power supply, and the speaker-independent recognizer board. Figure 2 shows a simplified block diagram of the VCS. The VCS processor orchestrates the entire VCS operation through the executive software resident in the solid state disk. Voice input to the VCS is received from the Shuttle headset through the headset interface unit (HIU). Signal amplification occurs automatically based on the setting of the feedback resistor comprising of a digitally controlled potentiometer. This potentiometer is under software control. The spoken command word is then sent to the recognizer where the word is decoded. The recognizer sends the results to the processor. If the decoded spoken word is a valid command the processor sends the appropriate CCTV discrete command to the CCTV I/F control and on to the orbiter CCTV controller for execution. Based on response messages from the recognizer such as "spoke too low", the processor changes the gain of the potentiometer attempting to obtain the correct amplification setting. The gain can change several times in a minute due to microphone placement or how loud the user is speaking at any time.

The VCS processor also sends display and audio messages to the VCS display and the headset earphone, respectively. These messages let the user know the status of the VCS such as recognized word decoded, what node they are in, and errors due to headset microphone placement. Audio from the headset and playback messages are routed to the camcorder for recording. Further, control of the headset mode either push to disable (PTD) or push to talk (PTT) is controlled by the processor depending on the user depressing the ICOM button(designated for headset mode control) on the HIU. Depressing the ICOM button toggles the headset mode either PTD-to -PTT or PTT-to-PTD.

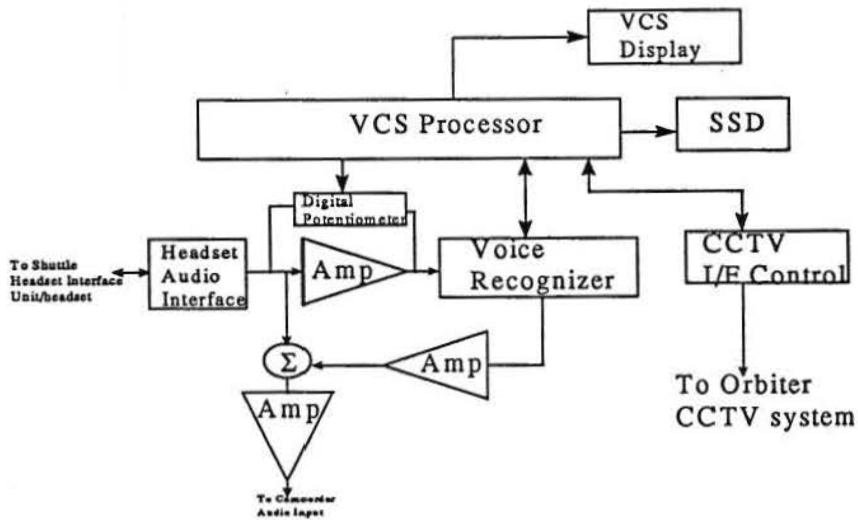


FIGURE 2. SIMPLIFIED BLOCK DIAGRAM OF THE VCS

7.2. System Features and Controls

The STS-78 VCS flight configuration had the capability to control the shuttle's four payload bay cameras , the mid and flight deck cameras and the two aft flight deck monitors. In addition, the VCS contained several macros for evaluation on this flight. Figure 3 shows the flight configuration of the VCS on shuttle. The camcorder recorded the use of the system. The audio recorded on to the camcorder helped correlate the recognition scores with the command words spoken. Connection of the Payload General Support laptop Computer (PGSC) to the VCS served as a contingency should recognition problems occur on-orbit. Up and down load of files through the PGSC such as a new vocabulary or real time recognition scores were some of the new capabilities built into the VCS.

The VCS allows simultaneous commanding of a given camera. For example, the astronaut can command a payload bay camera to pan left, tilt up, and zoom out simultaneously. Repeating the last command word (zoom out) stops the first two actions (pan left and tilt up) and continues with zooming out until another stop command is issued and recognized. The pan/tilt rate is selectable by voice. By simply saying "Change Rate," the pan/tilt rate toggles from either high rate (12 degrees per second) or low-rate(1.2 degrees per second) and vise versa. Indication of the changed rate appears on the display as either H "for high-rate " or " L " for low rate. Also, the CCTV system camera and monitor will visually indicate the changed rate as well.

The fine-tuning of a camera lens adjustment or pan/tilt position occurs through first selection by voice of either 1, 2, 3, 5, 10, or 15 degrees followed by a pan, tilt, zoom, focus, or Iris control command. The selected control such as pan or tilt moves for a duration of time based on the CCTV pan/tilt rate, and the number of degrees selected. For zoom, focus, or irises control the lens function rotates only for the duration of time computed based again on the number of degrees selected or duration of time before the command is stopped. Note that the CCTV system provides no feedback from the pan/tilt units other than visual on the CCTV monitors or out the window, nor does it have rates associated with zoom, focus, and iris control. The VCS computes the time duration and selects (emulates the control switch depressed) that switch function for the time computed.

The VCS incorporated six macros for the STS-78 mission as noted in Table 1. To run a macro requires first entering the node via enunciating the transition word "macro." Except for "stow cameras" the other macros require the user to calibrate the cameras by speaking "calibrate" before the macros are initiated. Calibrate homes the cameras to a known position. Again, since the CCTV system provides no feedback from the pan/tilt units, the VCS software needs to know where the cameras are before moving them to a pre-defined position.

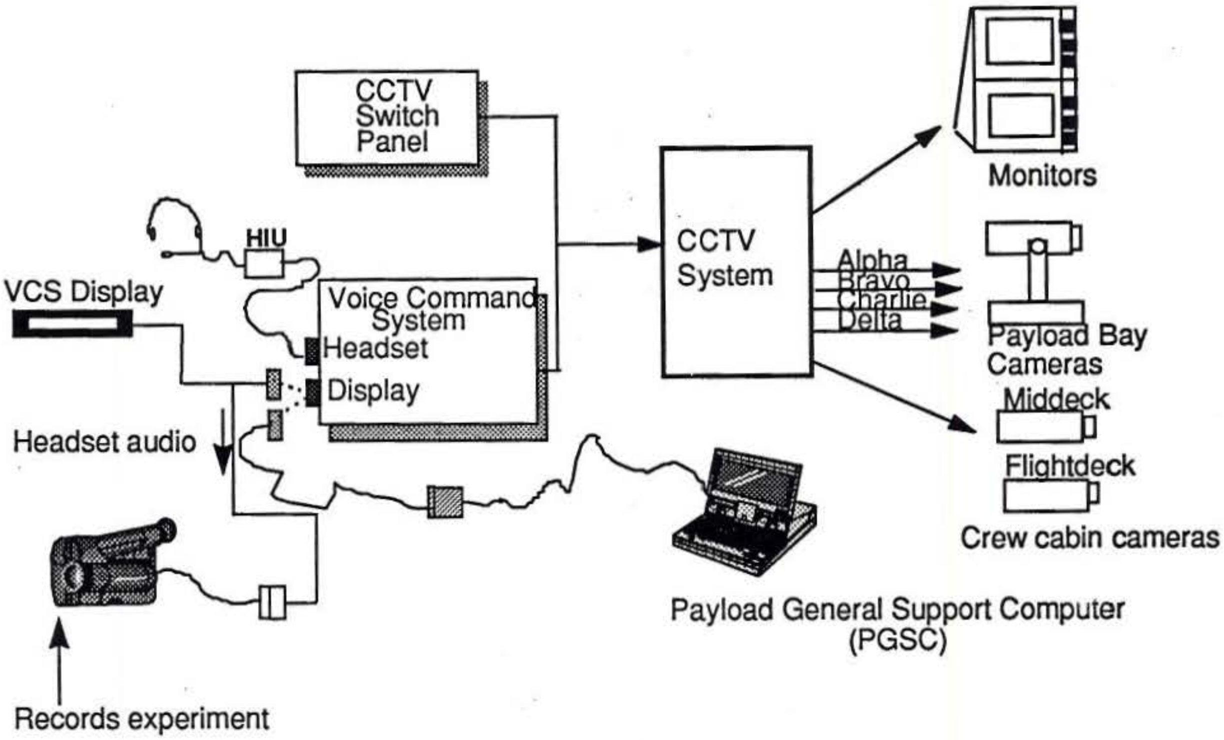


FIGURE 3. STS-78 FLIGHT CONFIGURATION OF VCS ON ORBITER

The VCS provides audio and visual feedback to the user. VCS system status displays on the 32-character VCS display. Monitors, cameras, node, and action commands selected appear on the display. In addition, the pan/tilt rate and the headset mode are displayed too. Audio messages are played out over the headset earpiece. Three audio tones are available indicating to the user results of a spoken command. A double tone indicates a transition from one node to the next. A single tone indicates it recognized the spoken command. A buzzer indicates that it did not recognize the spoken command. A fourth tone consisting of a dual low/high tone indicates to the user that the amplitude was too low. After three such spoke too low errors detected by the recognizer, an audio message is played out over the headset to check microphone placement.

When a questionable spoken command word occurs, the VCS prompts the user with the first choice word received from the recognizer. If the user says "yes," the system adapts the word by training that word and untraining the next choice word. If "no" is said, the VCS prompts the user with the next choice word received from the recognizer. If the user says "yes," the system adapts that word. If " no" is said, the system exits the query and awaits the next spoken commands with no adaptation taking place. Only spoken commands that exceed the acceptance level and delta level(difference in score from the first to second choice word) thresholds of 350 and 25,respectively, will cause a query to occur. Section 7.3 discusses this further.

Operating the VCS requires only four switches: power, reset, Active/Standby mode, and headset mode. The power switch provides power from orbiter to the VCS. The reset switch restarts the executive software should an error occur with the software of hardware. The reset light on the front panel indicates to the user when the reset switch should be depressed. The mode switch allows the user to place the VCS in active (listening for CCTV commands) or standby (non-listening) mode by toggling the momentary switch. This switch serves as a backup to the voice commanding of the function. The headset mode as mentioned previously is controlled by the ICOM button on the shuttle HIU. Depressing the switch results in the headset mode toggling from either PTT-to-PTD or from PTD-to-PTT.

A7 Panel Commands	Purpose	VCS Unique Commands	Purpose
Up/Tilt_Up	Tilts camera up	*Confidence_Check*	Trains all words
Down/Tilt_down	Tilts camera down	*Query_Check*	Trains "Yes"/"No"
Zoom_In/Close_Up	Zooms camera in	*User_Setup*	Loads requested parameters
Left/Pan_Left	Pans camera left	*Setup*	Transition word
Right/Pan_Right	Pans camera right	*Voice Command*	Transition word
Zoom_out	Zooms camera out	*Activate*	Transition word
Focus_near	Focus camera near	*Stop/Halt/Whoa*	Stops camera/pan/tilt movements
Focus_Far	Focus camera far	*Macro*	Transition word
Monitor_1	Selects monitor 1	*Configure*	Transition word
Monitor_2	Select monitor 2	*Standby/Relax/ Go_To_Sleep*	Put VCS into non-listening
Alpha	Select camera A	*Move/Adjust*	Transition word
Bravo	Select camera B	*Macro_1*	Reserved
Charlie	Select camera C	*Macro_2*	Downlinks test pattern
Delta	Select camera D	*Macro_3*	Views earth using cameras A,D
Middeck	Select Middeck camera	*Macro_4*	Scans PLB port side w/cameras A/D
FlightDeck	Select Flightdeck camera	*Macro_5*	Scans PLB STBD side w/cameras C,D
1_Left	Select Monitor1left side	*Calibrate*	All 4 PLB cameras placed into known position
1_Right	Select Monitor1right side	*Stow_Cameras*	All 4 PLB cameras into ascent/reentry position
2_Left	Select Mon 2 left side	*Yes*	Query response for choosing correct word spoken
2_Right	Select Mon 2 right side	*No*	Query response for not choosing a word spoken
Open_Iris	Opens the camera iris	*1-,2-, 3-,5-,10,15- degrees*	Number of degrees selected for a camera lens or pan/tilt movement command
Close_Iris	Closes the camera iris		
Change_Rate	Changes the pan/tilt rate		

Table 1. VCS command words and definations

7.3. STS-78 Vocabulary

The vocabulary for this flight retained a lot of what flew on STS-41. As Table 1 shows, much of the terminology related to CCTV system A7 switch function nomenclature is the same. What changed from STS-41 were the unique commands and the nodal structure. Several command words were added and many of the STS-41 commands deleted. The vocabulary is the result of several training sessions with astronauts in the manipulator Development Facility(MDF). Degrees were added instead of the "Easy", "Too much", and "A little more" command words used on STS-41. Those commands limited the fine-tuning of cameras and the pan/tilt positions. *Confidence_Check, Query_Check, Setup* and *User_Setup* were added to allow the user access the new features of the VCS.

User Setup lets the user configure the VCS prior to entering CCTV commanding. The user is first asked who they are. For STS-78, the VCS first would ask if the user was "Tom" or "Kevin." The user responds to the correct question by saying "yes." The VCS creates a file to store voice data and recognition scores in that file. Next, the user is asked if they want to enable voice recording. If the answer is "yes," then the digital representation of first 10 words spoken after leaving *User Setup* are recorded into that users file. Next, the system asked if they want to use the default recognition parameters (acceptance threshold=350 and delta value=25). The user simply responds "yes." However, if the user says "no," the VCS next prompts the person with a next set of parameters to accept or reject which are the tightened recognition parameters(acceptance threshold=300 and delta=30). If "no" is said, the final set of parameters are the relaxed recognition parameters(acceptance threshold=400 and delta=20) for consideration (explanation of these various parameters are given below). If "no" is said to all three parameter selection, the system uses the default recognition parameters.

"Query check" allows a user to specifically train the "yes" and "no" query response words. Three passes of each word occurs before returning the user to the standby node. "Confidence check" allows the user to train each word in the vocabulary should the user experience recognition difficulties. The algorithm ordinarily requires the spoken word to be one of the four top choices to allow for training the correct word and untraining the incorrect one. Should the word spoken not be one of the top four words after two tries, the system force adapts the word. This happens when the user's voice has changed dramatically from the vocabulary voice models.(In fact, this situation occurred with pilot Kregel in his last session and is explained in the results section). Each word adaption only influences the voice model for that particular word by 2%; that is, there is a 2% higher probability the recognizer will recognize the user saying that word again the next time he enunciates it . This is to prevent a user from skewing the vocabulary only to his or her voice.

The acceptance threshold and delta values for the vocabulary are set at 350 and 25, respectively. The values are loaded into the recognizer's memory during the system initialization of the VCS. An acceptance score of 0 implies that the word spoken is an exact match with the voice model for that word. The delta score is the difference in score between the first choice and the second choice word the recognizer determined from the spoken utterance. For an utterance to meet a correct recognized word, its score must be less than 350 and the difference in score between it and the next choice word be greater than 25.

Two additional sets of recognition parameters were included in the STS-78 flight software load: relaxed and tight. The relaxed parameters are set to an acceptance level of 400 and delta of 20. These parameters were intended to be used if the astronaut was experiencing too many queries where the first choice word was the correct word. The tight parameters are set to an acceptance level of 300 and delta of 30. These parameters were intended for the user experiencing too many substitutions. Selection of either tight or relaxed parameters occurs when the user responds with "yes" in "User Setup" to the specific request.

The acceptance of 350 was arrived at from initial Taguchi method analysis. The method arrived at an acceptance of 600 for best recognition. However, further analysis indicated that 350 would work best to minimize substitution errors. Of the rejections during testing 95% were due to delta violations not acceptance levels.

7.4. System Commanding

The CCTV commands that the VCS can control closely parallel the switch control functions on the CCTV switch panel. The VCS vocabulary of 59 command words corresponds to 23 of the 33 CCTV switch-panel related commands and 30 VCS-unique commands. As seen from figure 4, the command words are grouped into nodes. (Shown underneath each node name is the nominal VCS display message.) This grouping approach results in an increase in both search speed and accuracy of the recognition. Navigating node to node occurs by using transition words. When a transition word is recognized, the current set of node command words becomes inactive and the new set of command words associated with the spoken transition word becomes active.

Upon power-up or reset, the system enters the *Standby_Node*. The user must first enter the *Setup_Node* to configure the system if this is the beginning of the session. If it is a reset of the system, the user must still reenter the *Setup_Node* to continue collecting data by saying "yes" when prompted with "Continue?" on the VCS display. The *Setup_Node* allows a user to create a file for storing the digital representation of 10 spoken voice commands while actually commanding the system and recording recognition scoring data.(The reason for only 10 was determined during training. Recording of audio file while in recognition slowed the VCS down considerably and used a lot of memory. Therefore, 10 was determined to be an acceptable number of words to record.) Recognition data stored consist of the top four possible words and their associated score. In addition, operating system and recognition errors are also recorded. Once user setup is complete the system then moves the user back to the *Standby_Node*.

In the *Standby_Node* the astronaut can converse without inadvertently activating a CCTV system function. Only the node transition command sequence "voice command" followed by "activate" within three seconds will move the user into the *Configure_Node*. The VCS panel "active/standby" mode switch will perform the same function. All node transitions are displayed on the VCS display to alert the user in what node they are in. Also, a tone played out over the headset tells the user that they have just transitioned to another node.

The *Configure_Node* allows the user to select a monitor, a CCTV camera, or muxing the cameras onto the monitors. Afterwards, the user can action command the camera most recently selected. Commands such as "pan_left" or "tilt up" are recognized and the appropriate discrete commands are sent to the CCTV system. Should fine-tuning of the camera be required, the user can speak "Move/Adjust" transition word. This moves the user into the *Move_Node*. Here the astronaut can fine-tune a camera lens or pan/tilt position. The user first says the number of degrees desired e.g., "1-degree" or 10-degrees." Afterwards, the user speaks the movement command such as "pan_left." The selected movement function moves that number of degrees and then stops.

To initiate a macro command requires the user to transition to the *Macro_Node.* The user simply speaks "Macro." In the macro node, the user can enunciate any of the macro commands. However, for all macro commands except for stow cameras, "Calibrate" must be initiated first before the other macros are invoked. Calibrate moves the cameras to a known position so that the system will know how much to move the cameras to obtain the camera scene. Again, this is required because the CCTV system pan/tilt units do not provide feedback. VCS creates a home position by using counters within the software taking into account the pan/tilt rate. Stow cameras has its own home position built in to the software. Stopping of a macro can be done by saying any of the three stop commands ("stop," "halt," or "whoa"). The software will stop the macro command if any of these stop commands are the top four choice words returned by the recognizer to increase the likelihood that a stop command is recognized when enunciated the first time.(During vocabulary development, this was found to be important to several astronauts; that is, when he/she says "stop, halt or whoa" the system better recognize it the first time or user frustration of the system will occur.)

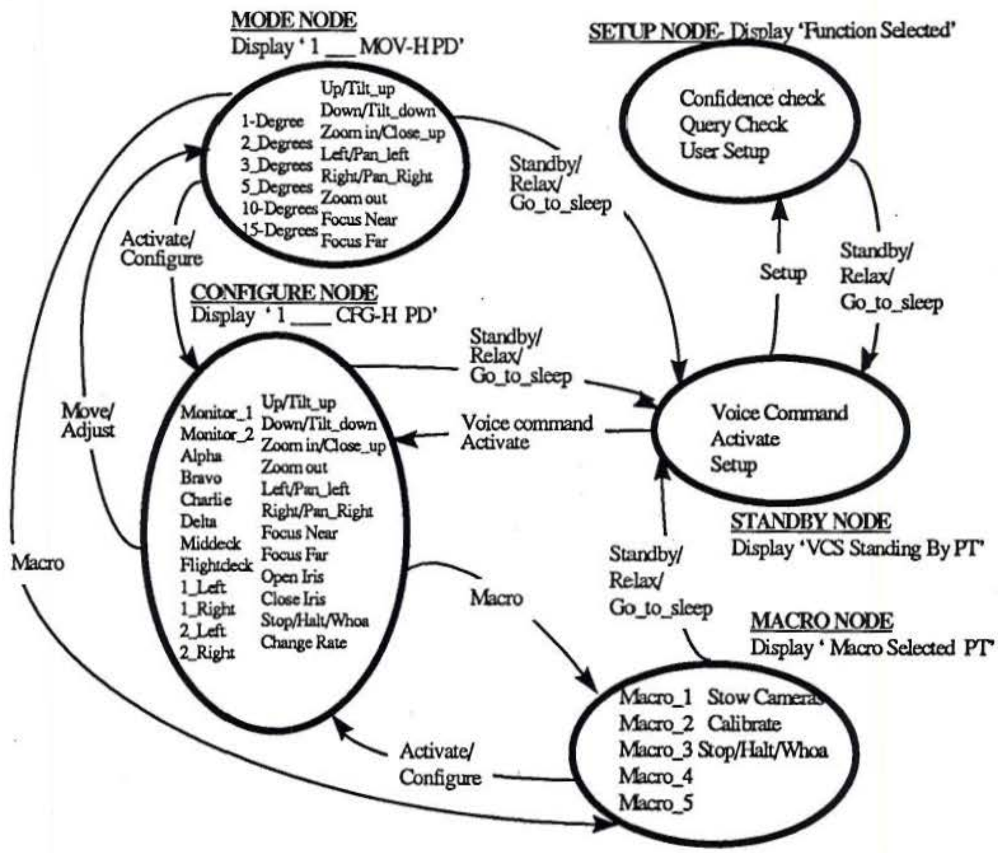


FIGURE 4. VCS NODAL STRUCTURE

8. DTO Training

8.1. Training Overview

The objectives of VCS training for the flight experiment were to train the astronauts to:

a. Learn and operate the mechanics of the VCS and the associated voice control CCTV tasks in a shuttle-like environment using a flight-like VCS system.

b. Speak in a normal voice to the system.

c. Understand the recognition adaptation process.

d. Learn to work with the headset mode(push-to-talk(PTT) or push-to-disable(PTD)) and become comfortable with the VCS audio/display feedback.

e. Value their participation in this experiment by helping to develop mission task procedures and improvements to the system.

Below recounts the training sessions (one session=1-hour). Initial training began with the STS-78 Commander Tom Henricks and pilot Kevin Kregel in January. Training should have started in early December. However, the government furlough pushed the start training date into mid January but the launch date remained the same.

8.2. Training Activities

The next several sections discusses various errors associated with speech recognition. The following words and definitions used are:

- Recognition accuracy = percentage of spoken command words correctly identified and processed
- Substitution error = legal commands spoken incorrectly identified as another word in the vocabulary and processed
- Omission error = legal commands spoken that were not recognized and processed
- Insertion error = illegal spoken commands (not in the vocabulary) that were incorrectly identified and processed as a legal command spoken

The first training session consisted of familiarizing the astronauts with the experiment (DTO). An overview of the VCS and its capabilities was discussed. A system demonstration of the device verification test unit (DVTU) in that session gave the astronauts their first hands-on introduction with the hardware they would be working with. Commander Henricks received his first training session in the manipulator development facility (MDF).The astronauts suggested combining the camera commanding and monitor selection nodes into one. VCS personnel set out to combine the nodes into one. However, due to lack of time, the two were combined without running the node through the vocabulary optimizer (called batching). The optimizer helps reduce the misrecognitions errors between words.

During February, Kevin had three training sessions and Tom two all in the MDF. VCS personnel noted that the astronauts were speaking more natural to the system. The first use of the combind node called now the *Configure Node* occurred. Some problems in substitution errors were noted. Both astronauts agreed that the new node was easier to use that the original two-node approach. The astronauts and VCS personnel discussed what macros to develop for the flight. A down link macro sending a test pattern to the ground was agreed to. One astronaut noted that using voice for control of the CCTV was easier than the panel approach of control. Both astronauts agreed that a 1-rejection before query would be better than the existing 2-rejection then query.

Voice prints were collected from both astronauts using the actual flight unit. They were collected to batch a new vocabulary taking into account the new node. The astronauts were getting too many substitution errors with the vocabulary that was being used. Again, it was not surprising since the original camera movement and monitor selection nodes had been combined into one without batching.

Only one session for each astronaut occurred in March. The astronauts evaluated the newly batched vocabulary. They were having some recognition problems. Kevin was queried several times. This may be attributed to not exercising the vocabulary(saying all words and adapting/training the problem words) after it was batched. The modified software with the 1-rejection before query was evaluated. Both astronauts liked it. Discussion of using PGSC with VCS began. Tom had suggested to have a "Stop" and a "Continue" command word for the macros.

April training consisted of 4 sessions each with Kevin and Tom. The first training session with Kevin indicated the vocabulary still needed further tweaking for him. A preliminary operation (OPS) checklist that would serve as their VCS operations manual on-orbit was given to the astronauts for review. Several comments were given on the check list regarding clarity of procedures. On April 15, a session on hookup of the camcorder occurred in the MDF with both astronauts. A Photo/TV checklist describing the steps was also given to them and reviewed. The primary focus was to show them how to hook up the very Lightweight Headset (VLHS) cable to the VCS audio cable and then from the VLHS cable to the Camcorder (CC) audio input. The session did not actually allow them to functionally hookup the camcorder and use it.

Still in April, a session occurred in the shuttle mission simulator (SMS) in building 5. Kevin experienced recognition difficulties. VCS personnel investigated the problem and found that the recognition errors were due to uplink audio messages coming through the aft flight deck speaker. Attempts to dampen the speaker's audio sound level failed. The combination of the 95 dbspl and the reverberation of the cockpit made it impossible to dampen the audio getting into the headset without turning the speaker off. This event resulted in calling out in the checklist to turn the speaker off in the flight deck during VCS operations. The headset does have some noise cancellation. However, the cancellation is only about 15 db.

In May, one session for each astronaut occurred. Kevin had a lesson on using the PGSC with VCS. The main emphasis was to show how software can be loaded into the VCS via the PGSC and how downloading VCS files into the PGSC for possible down link to the ground is performed. Cable connections were shown for connecting VCS to PGSC as well. Due to a software problem in the VCS, Tom did not have an opportunity to work with up/down loading of files to/from VCS/PGSC from/to PGSC/VCS. However, explanations of the procedures were given.

The final month of training in June allowed one session per astronaut in the MDF. The session focused on stepping through the OPS check list for VCS and making any last minute corrections. Kevin as well as Tom had omission error difficulty with "activate." They did not have an opportunity to work with macros. Overall, they had good recognition and were happy with the performance of the system. VCS personnel had spent some time exercising the vocabulary prior to the session to reduce the number of queries. Figures 5 and 6 shows the recognition results during ground training for both astronauts. Note that the data is for the new vocabulary encompassing the combined camera movement and monitor selection nodes only. The final vocabulary was not ready until the month of March. Due to hardware and software problems experienced in some of the sessions with the new vocabulary, recognition data for the vocabulary was recorded only for one session with Kevin, two for Tom.

Referring to figure 5 and 6,."Adapted and Correctly processed" are words that triggered a query, the word spoken appeared correctly on the VCS display during query, the word was adapted, and the command processed. "Correct Recognition" means the combined accuracy consisting of recognition accuracy(spoken words correctly recognized the first time and processed) and "Adapted and Correctly processed." For example, in figure 5, the correction recognition for pilot Kregel was 93% with 9% of the 93% score consisting of adapted words correctly processed. It was interesting that the users perceived adapted and correctly processed words spoken the same as words spoken the first time and correctly recognized.

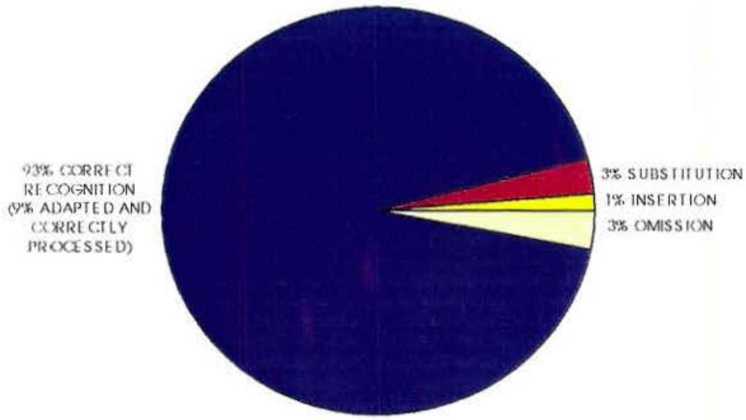


FIGURE 5. PILOT KREGEL- PREFLIGHT RECOGNITION RESULTS(ONE SESSION)

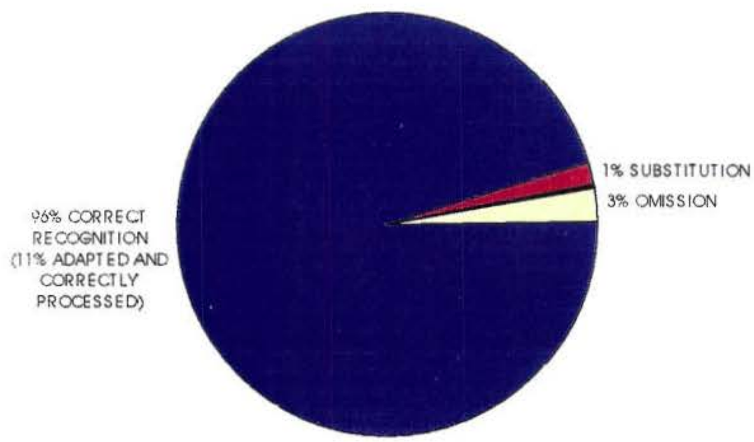


FIGURE 6. COMMANDER HENRICKS- PREFLIGHT RECOGNITION RESULTS(TWO SESSIONS)

9. On-Orbit Results

9.1. Overview

The launch of Columbia on its STS-78 mission occurred on June 20,1996 at 9:49 AM CDT. The mission lasted about 17 days. The VCS usage occurred on flight days 3, 10, and 16. Acquisition of recognition score data, video, and audio data for both commander Henricks and pilot Kregel occurred in each flight experiment session. Two types of audio data recording occurred. For each session, the VCS recorded ten spoken commands as digital audio files for voice analysis to compare against ground voice recordings. Also, the camcorder recorded the astronauts using the system and the spoken commands from the headset. The headset audio allows correlating the recognition scores to the spoken commands to determine recognition performance. For each spoken word the recognition data consists of the top four choice words and any operating system and recognition errors such as "spoke too low." In addition, the crew debrief provided additional information not obvious on the tapes. The recognition results figures that follow require qualifying the terms and how they are computed.

Table 2 below shows how the percentages for the accuracy and the various errors were obtained [5].

Accuracy Measurement %	*Equation*
Correct Recognition Accuracy (CRA)	$\frac{\text{\# of correctly recognized words}}{\text{\# command words spoken}} \times 100$
* Adaption Through Query (ATQ)	$\frac{\text{\# of words correctly adapted}}{\text{\# command words spoken}}$
Omission Errors (OE)	$\frac{\text{\# of omission errors}}{\text{\# command words spoken}} \times 100$
Substitution Errors (SE)	$\frac{\text{\# of substitution errors}}{\text{\# command words spoken}} \times 100$
Insertion Errors (IE)	$\frac{\text{\# of insertion errors}}{\text{\# command words spoken}} \times 100$
Overall CRA (OCRA)	CRA + ATQ

*This equation was added to the one in the reference to reflect an adaptive system.

Table 2. Recognition Performance Equations

Flight Activities

Flight day 3

Activities consisted of primarily exercising the vocabulary. Audio/video was recorded for pilot Kregel but not for the commander due to the only camcorder allocated for experiments was being used in the spacelab during the commander's session. Both astronauts had some difficulty with the activating sequence command words "voice command" and then "activate." However, part of the problem appeared related from speaking the word too quickly after "voice command." This may have been attributed to how it was documented in the flight data file. It called out to speak activate within 3 seconds after recognition of voice command. What was not conveyed was to pause at least half a second after "voice command." This situation occurred a few times during training. Commander Henricks had omission error difficulty with "Relax." "Go_to_sleep" appeared to work better for him in putting VCS into standby. Flight day 3 activities on the tape showed the VCS in use while uplink audio messages were occurring. On one occasion it clearly showed the VCS accepting the message as a command. However, during another uplink message the VCS rejected the utterance as noted by the sound of the rejection buzzer audio tone. The macro "calibrate" failed to successfully complete resulting in the VCS hanging up on the commander. He did however reset the system and successfully continued on. Both reported excellent results with the VCS on flight day 3 obtaining what they estimated recognition accuracy of better than 90 to 95% recognition.

Figures 7 shows the performance of VCS with pilot Kregel. Note that on the first session his overall correct recognition accuracy (included with adaption) was 95%. His estimation of accuracy was correct as stated in his flight day 3 report. Adaption of words appear to the user as part of correct recognition[9]. Commander Henricks had no audio or video with his first session again due to the use of the camcorder in the spacelab. As shown in figure 8 his confidence check of the vocabulary indicates the VCS did a good job of recognizing him with an overall correct recognition accuracy of 100% including 9% through adaption. Tom perceived his recognition as 90-95% on his first session as reported by him.

Flight day 10

Commander Henricks second VCS recorded session had audio and video. However, pilot Kregel had video but no audio until the last part of his session. Only one camcorder cable for recording audio was available and was shared with other experiments. Again, the commander got stuck in the macro "calibrate" having to reset the system again. Prior to the stuck macro, the commander reported his recognition was very good around 95%. Figure 10 shows that commander Henricks results closely matches his perceived recognition rate. However, after reset he reported recognition degradation. He reported that he perceived his recognition rate dropping to around 75%. It's not clear as to why this occurred. The commander did not place the camcorder in a position such that the placement of the microphone could be seen. He did not reenable recording of recognition data in the *Setup_node* after coming out of reset and therefore no actual recognition rate could be determined.

Commander Henricks attempted to perform an operational activity with the VCS and a payload bay camera. He chose to view New Orleans. However, he found using VCS frustrating and eventually used the CCTV switches to position the camera over New Orleans. This did show, at least for the present VCS configuration, that handling unplanned earth viewing is performed faster with the switch panel than with the VCS. Earth viewing tasks were not practiced during training in the MDF.

The video of pilot Kregel's session showed his microphone placement to be at the upper portion of the edge of his mouth and his nose. No audio was recorded with the video. He did a confidence check and recognition scores were recorded. Figure 9 shows the results indicating an effective recognition of 96.6%. On a couple of occasions, he pointed to the nodal structure chart trying to explain that he was being transitioned over to the *Move Node* without wanting to. What may have contributed to this was the microphone placement near the nose. VCS personnel had this problem occur when the users are breathing through their nose. The system would accept it as the "move" command. This never occurred with the astronauts during training. A spectrum and formant frequency analysis showed that at times the nostril breathing displays similar acoustic features such as 1st formant frequency. This clearly can be an annoyance. The word "Move" should be removed from the vocabulary and use instead "adjust" as the transition word. Towards the end of the tape, Kevin had audio. He mentioned he was going to turn on the flight deck speaker to assess the susceptibility of the VCS to uplink audio messages as compared to what occurred during the SMS training session. However, pilot Kregel had an incoming ground call and turned off the camcorder with no further recording of the session.

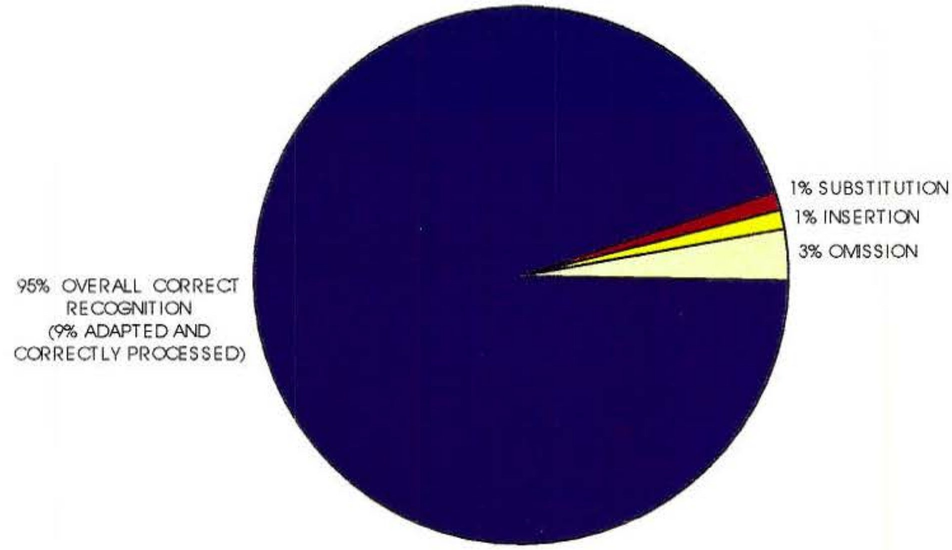


FIGURE 7. PILOT KREGEL- FLIGHT DAY 3 RECOGNITION RESULTS

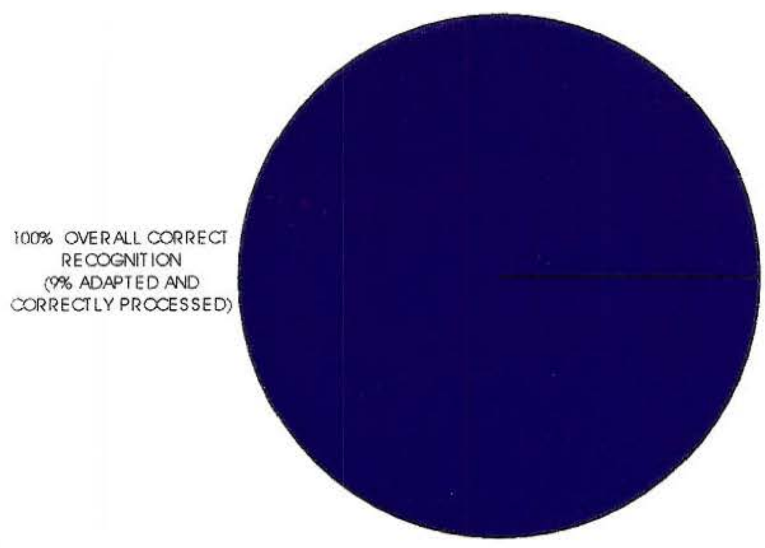


FIGURE 8. COMMANDER HENRICKS- FLIGHT DAY 3 CONFIDENCE CHECK

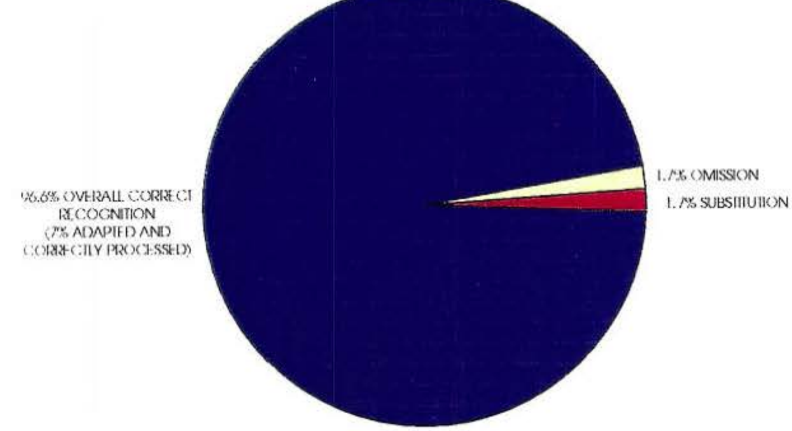


FIGURE 9. PILOT KREGEL -FLIGHT DAY 10 CONFIDENCE CHECK RECOGNITION RESULTS

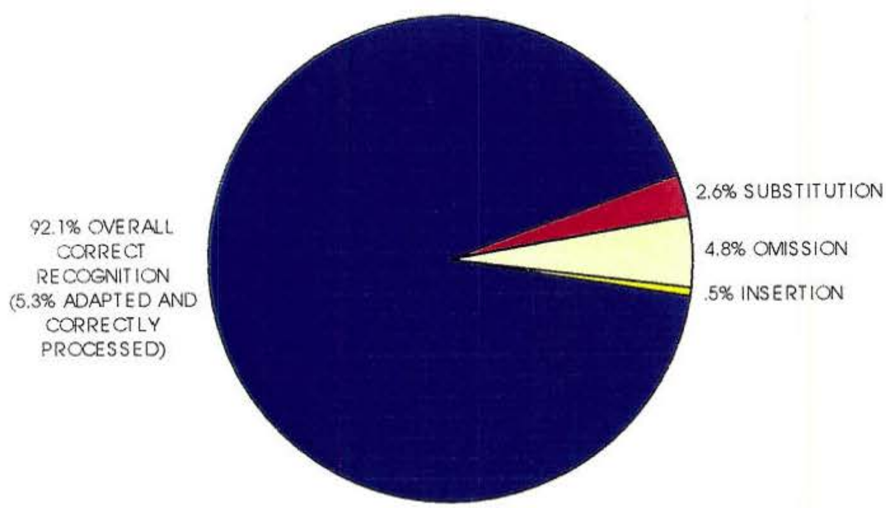


FIGURE 10. COMMANDER HENRICKS- FLIGHT DAY 10 RECOGNITION RESULTS

Flight day 16

In this session, the commander's recorded session had video but no audio and therefore correlation of recognition data and spoken commands was impossible. He did however do a confidence check of the vocabulary. Figure 11 shows the recognition results of that activity. He primarily exercised the vocabulary rather than performing CCTV tasks with the VCS.

The pilot's session had both audio and video. It is interesting that the pilot's microphone was placed very close to the mouth appearing to touch his upper lip. The audio revealed what appeared to be a distortion of the audio. His breathing was heard as quite loud. Recorded digital audio data indicated that the audio levels going into the recognizer were clipping at times which may have contributed to the recognition difficulty he was experiencing. Figure 12 shows his recognition performance rate that is clearly far different from his previous sessions. The number of omissions indicates that the voice as heard by the recognizer was quite different from what was in the voice model that made up the vocabulary. Pilot Kregel did perform two confidence checks but still had problems. Observations from the tape showed that after his second confidence check, the VCS recognized his commands better. Figures 13 and 14 shows the results of those two confidence checks. A considerable amount of forced adapts (words prompted to speak by the VCS but the spoken word was not one of the top three choices) resulted in 24% and 29% for his first and second confidence check, respectively, to occur. The pilot wanted to load a new vocabulary from the PGSC as part of malfunction procedures. However, he ran out of time. One drawback of the VCS flight data file was that no malfunction procedures were included.

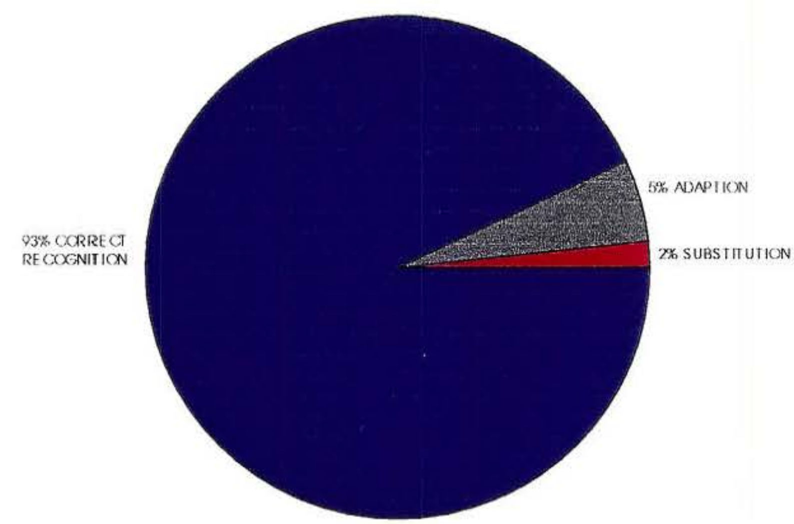


FIGURE 11. COMMANDER HENRICKS- FLIGHT DAY 16 CONFIDENCE CHECK RESULTS

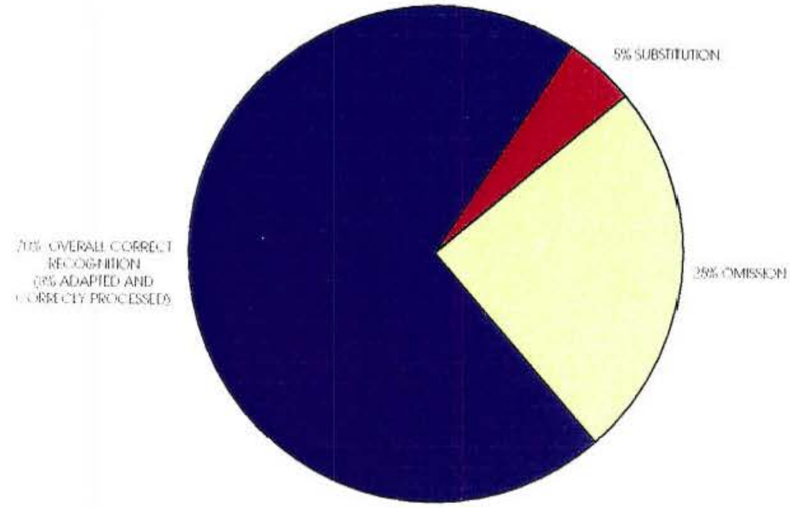


FIGURE 12. PILOT KREGEL-FLIGHT DAY 16 RECOGNITION RESULTS

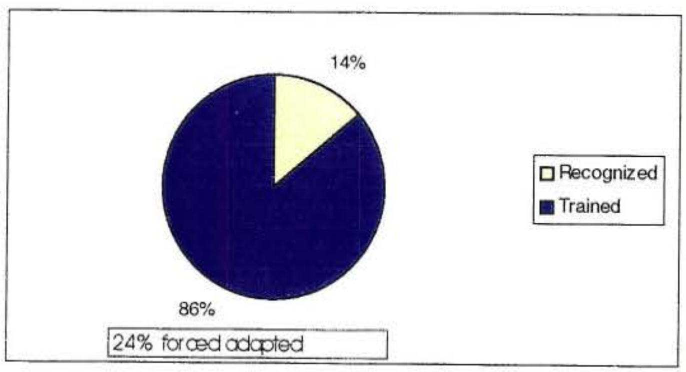


FIGURE 13. PILOT KREGEL- FLIGHT DAY 16 1ST CONFIDENCE CHECK

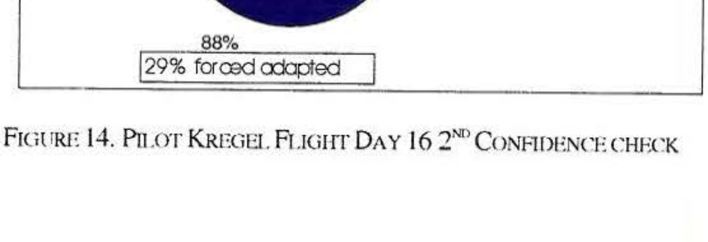


FIGURE 14. PILOT KREGEL FLIGHT DAY 16 2ND CONFIDENCE CHECK

In general, the recognition accuracy rates for all words were within the targeted recognition rate accuracy of better than 90%. Figure 22 shows the average of each word spoken by both astronauts over the course of the flight. Note that the command words "close_iris," "close_up," and "configure" show up as the most problem words, in reality only close-iris was the problem word. The other two were spoken very few times throughout the flight. Figure 23 shows the overall recognition from preflight to post-flight for both astronauts. Clearly, figure 24 for flight day 16 shows that pilot Kregel had difficulty using VCS.

Voice Analysis

Limited spectral analysis of the astronauts voices investigation was conducted to determine if the voice of the astronauts changed over the course of the mission. Spoken commands recorded for both the pilot and the commander were analyzed using a speech analysis program called Soundscope for the Macintosh. The words analyzed were: "Voice Command" and "Monitor _One" for the pilot and "Alpha' and "Monitor_Two" for the commander. These words were chosen because they were spoken in all sessions as part of the 10 words that were recorded(Unfortunately, they were not given a list of words to speak for the 10 words.). In addition, the fundamental frequency and the first four formant frequencies were determined using Soundscope. The formants arise from the various resonance in the vocal tract.[6]. The autocorrelation approach was used for determining the fundamental frequency. It is the average over the whole word. Measurement of pitch period took advantage of the word vowels. Vowel sounds are generated though the excitation of the vocal tract with quasi-periodic pulses of pressure with a definite pitch period.

Figures 15,16, 17, and 18 show the pitch periods, fundamental and the first through fourth formant frequencies for the selected word for the pilot. Figures 19,20, 21, and 22 show the pitch periods, fundamental and the first through fourth formant frequencies for the selected word for the commander. The pilots pitch periods are slightly higher during flight than on the ground. However, the data does not indicate that micro-gravity significantly affects the voice, at least for the duration of this mission. However, this conclusion is based on only limited data from two astronauts. It does not appear that change in the voice was the problem with the pilot on his last session.

Some astronauts have said that the pitch is perceived to have changed in listening to other astronauts while on-orbit. Perhaps the perceived change may be due to fluid shifts/no sinus drain.

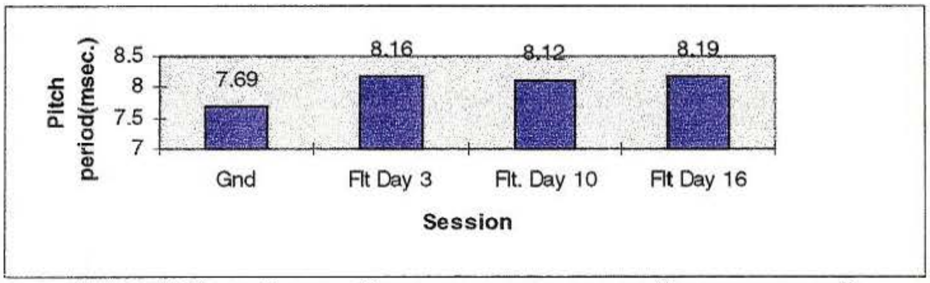


FIGURE 15. PILOT KREGEL- PITCH PERIOD FOR THE WORD "VOICE COMMAND"

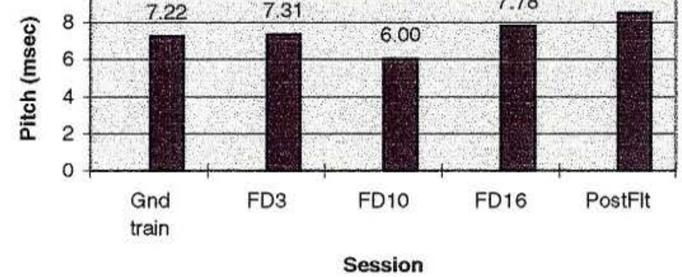


FIGURE 16. PILOT KREGEL- PITCH PERIOD FOR THE WORD "MONITOR ONE"

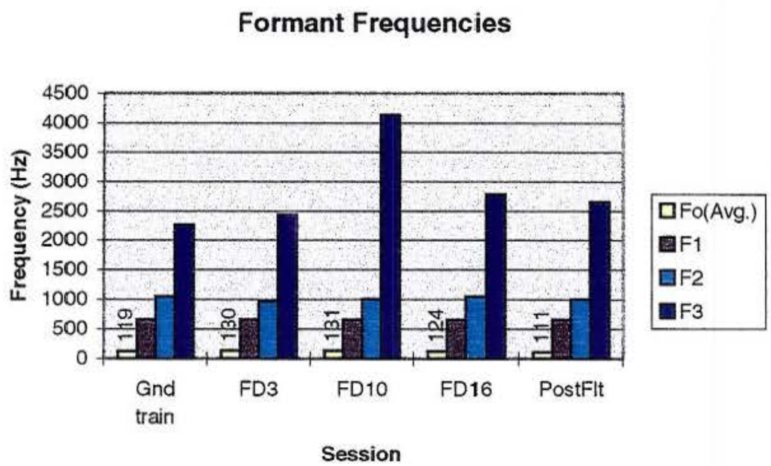


FIGURE 17. PILOT KREGEL- FUNDAMENTAL AND THE FIRST FOUR FORMANTS FOR "MONITOR ONE"

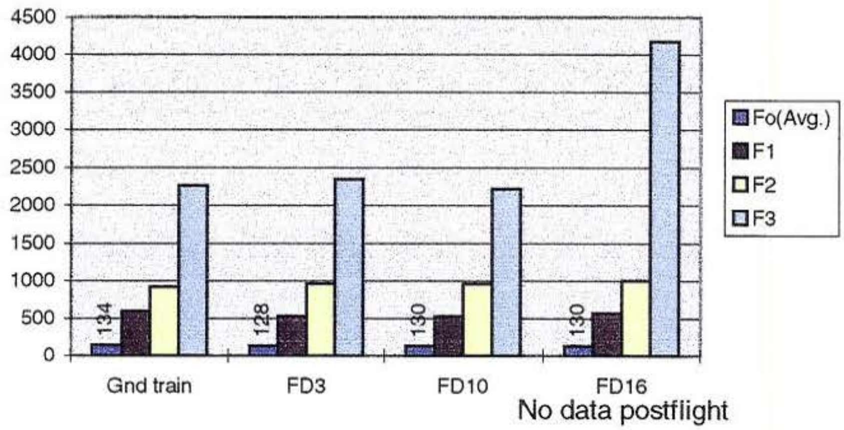


FIGURE 18. PILOT KREGEL - FUNDAMENTAL FREQUENCY AND THE FIRST FOUR FORMANTS FOR "VOICE COMMAND"

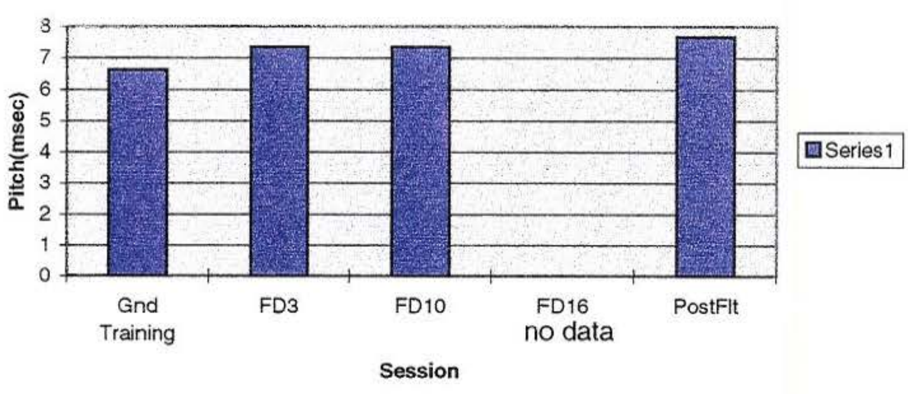


FIGURE 19. COMMANDER HENRICKS- PITCH PERIOD FOR "ALPHA"

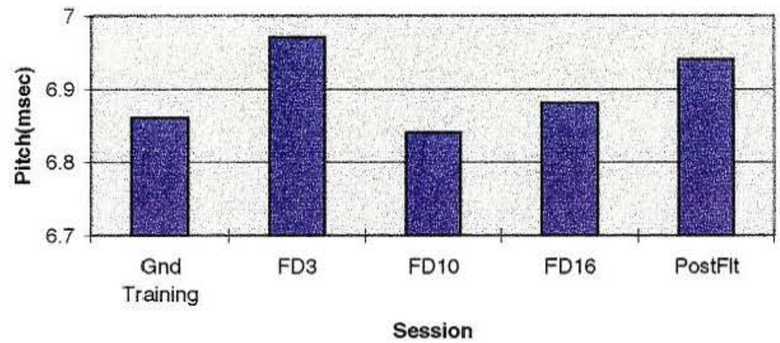


FIGURE 20. COMMANDER HENRICKS- PITCH PERIOD FOR "MONITOR TWO"

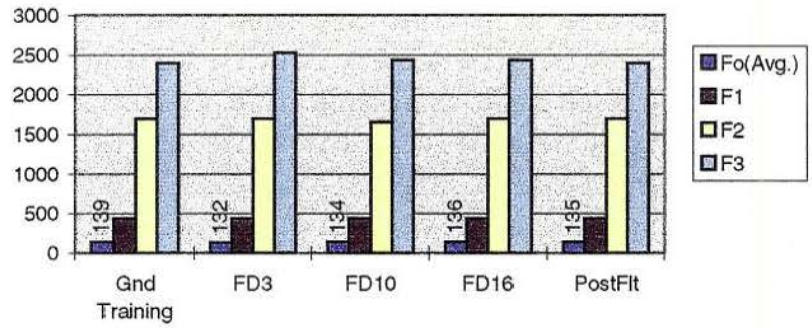


FIGURE 21. COMMANDER HENRICKS-FUNDAMENTAL FREQUENCY AND THE FIRST FOUR FORMANTS FOR "MONITOR 2"

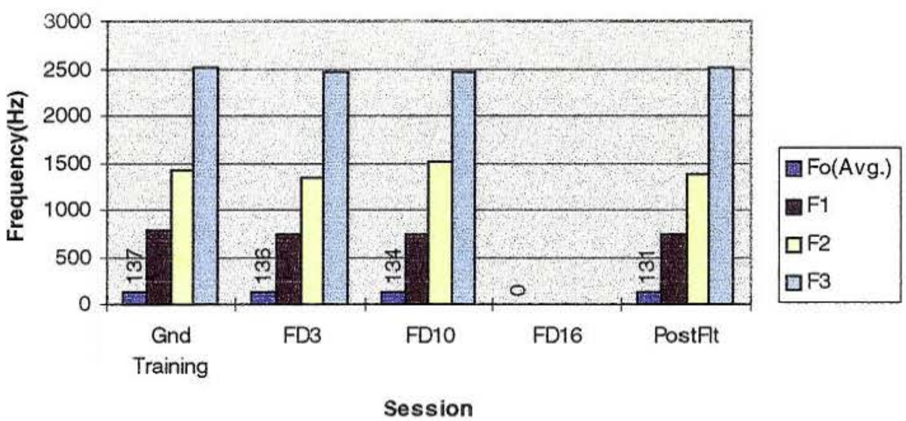
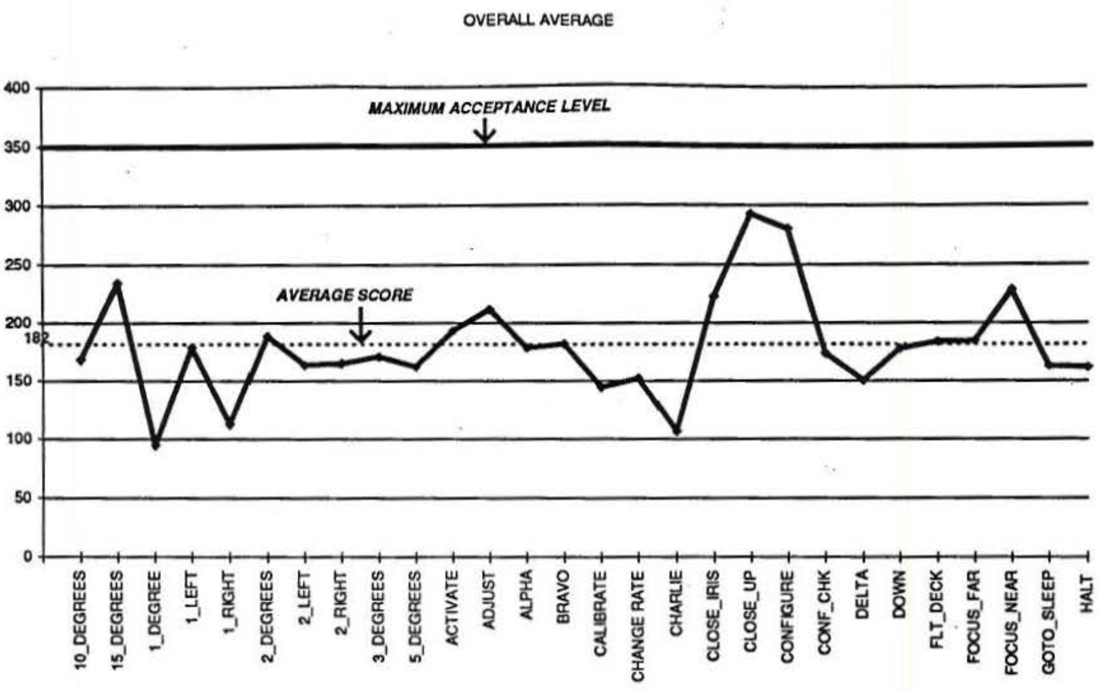


FIGURE 22. COMMANDER HENRICKS-FUNDAMENTAL FREQUENCY AND THE FIRST FOUR FORMANTS FOR "ALPHA"



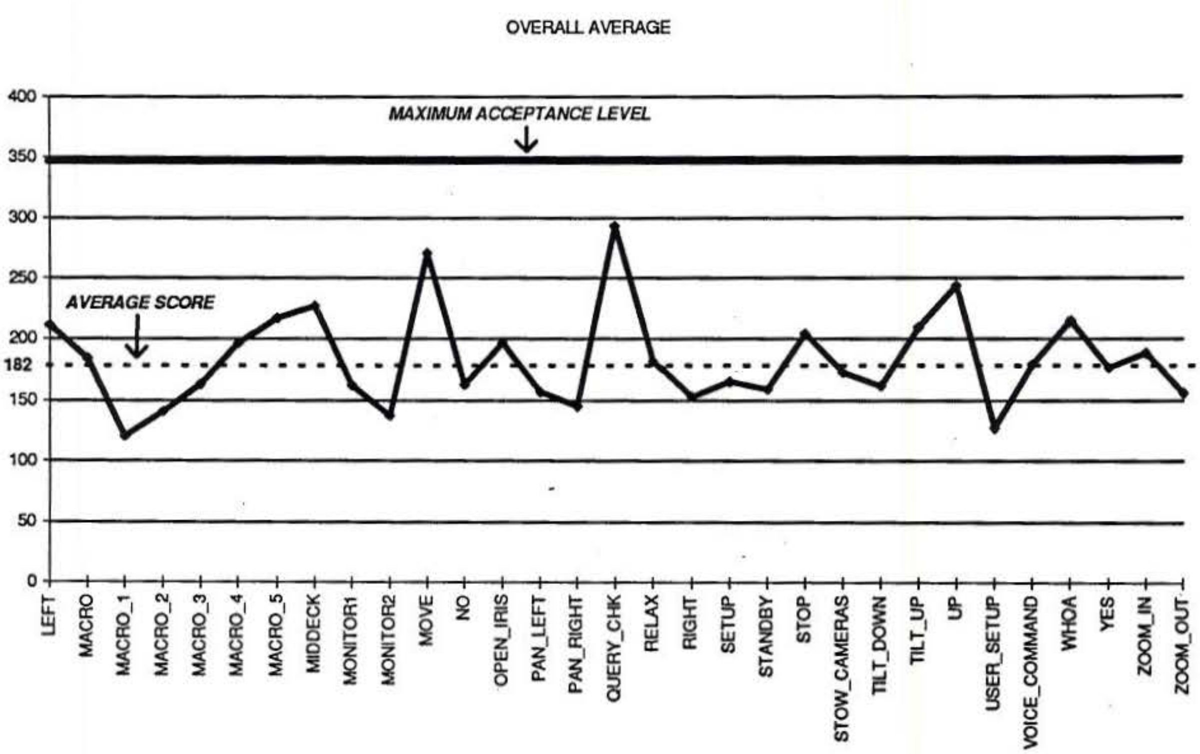


FIGURE 23. OVERALL AVERAGE RECOGNITION SCORES FOR THE VOCABULARY FOR THE FLIGHT

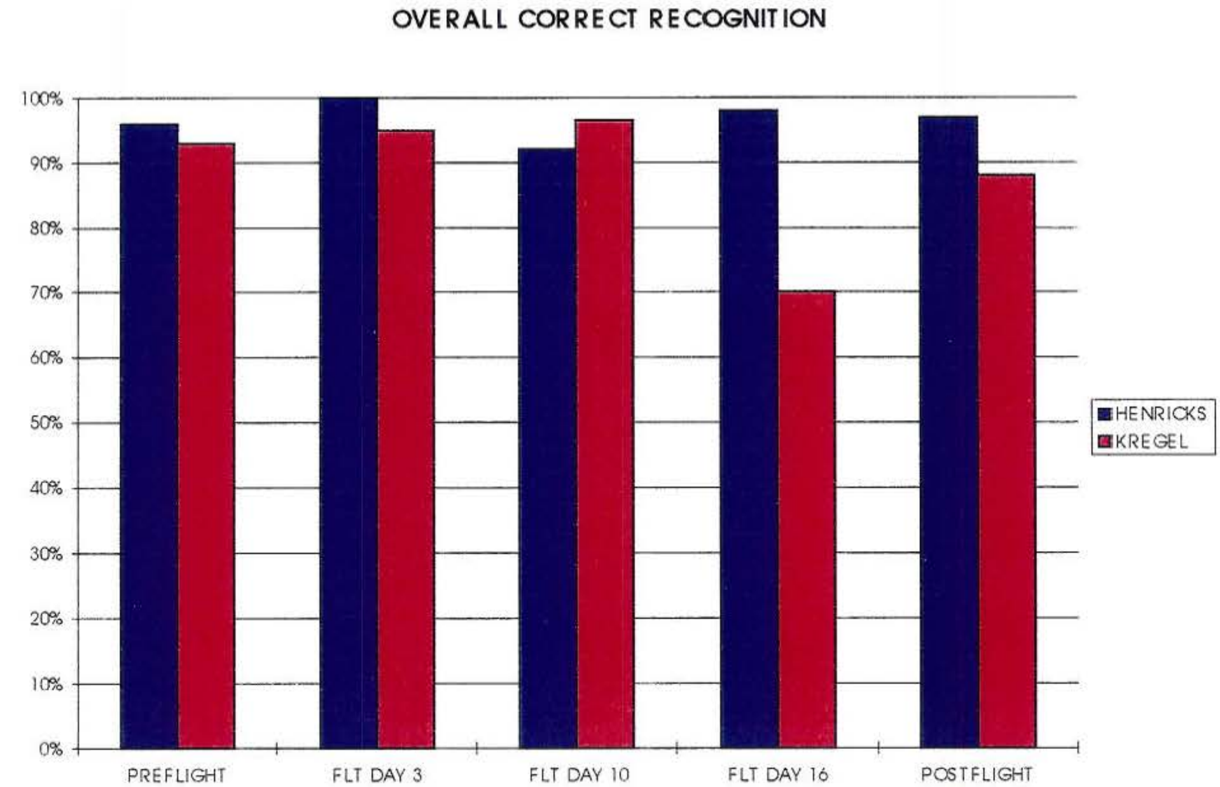


FIGURE 24. OVERALL RECOGNITION ACCURACY OF THE ASTRONAUTS FROM PREFLIGHT TO POSTFLIGHT

10. Discussion

The VCS experiment acquired valuable data on use of a speaker independent system for control of the CCTV system. Moreover, it gave additional data points for what are the strengths and weaknesses of voice as an input mechanism for human-machine systems. It is believed that the improved robustness in recognition compared to the unit that flew on STS-41 was achieved as requested by the astronaut office. However, further improvements still need to be made to increase the recognition accuracy. Inconsistencies in performance between both astronauts show that VCS still had limited capabilities in handling microphone placement changes and microphone sensitivity to background noise. The hypothesis for developing an adaptive system was to track and update the vocabulary as changes in the voice took place even with inconsistent microphone placement. Flight day 16 for the pilot showed that the system could not handle the mic placed close to the mouth and touching the lip. However, the shuttle headset does not easily allow for adjustment of the boom. Lack of formal malfunction procedures in the flight data file made it difficult to troubleshoot the recognition difficulty. A specific note on microphone placement (thumb distance from mouth to microphone with the microphone placed between the center and edge of mouth) would have helped the astronauts.

Another request from the astronaut office was to incorporate pre-defined camera scenes to simplify the use of the CCTV system. Macros were developed for the redesigned VCS. However, the feedback to the user needs improvement. The astronauts never knew where the macro was in terms of completion. The macro feedback to the user should tell him what percentage of the macro is complete, whether the feedback is audio or on the display.

The astronauts would like to be able to "stop' and 'resume" a macro which is possible. Faster macro commanding of the CCTV pan/tilt units is required to speed up the process. This needs investigating.

The headset mode of configuring the headset for push-to-talk or push-to-disable (hot microphone) required using the HIU XMIT and ICOM buttons. The STS-78 astronauts would like to see that function done all by voice which is possible. One issue that needs further work is dealing with the dynamics of the cockpit when using voice recognition. With the 95 dbspl uplink audio messages, at times constant conversations with other astronauts, and the reverberation of the cabin, a voice recognition system must handle these situations if the technology is to become part of operational hardware, particularly in the cockpit. Further testing in the SMS will iron out many of these limitations. Again, a new headset may be required as opposed to using the shuttle's headset. The headset is hard to adjust and has very limited noise cancellation(15db). Already there are smart headsets on the market that can handle high levels of noise and easily allow boom adjustments.

The flight data files need to be ready as early by the third training session. The astronauts on this mission did not have enough time with the final version of the VCS flight data file. Malfunction procedures need to incorporated into the document too. In fairness, the government furlough did impact training by a month. Consequently, documentation was running behind and the exercising of the final vocabulary as well.

Training in using the VCS for earth views needs to be included as part of training. Training for this type of activity was never performed to support this mission and will require an investigation. Originally, the VCS was intended to only control CCTV camera viewing the payload bay, rendezvous, or berthing with RMS ops as well. However, earth ops views are part of CCTV ops and therefore a valid task to train for. Many of the problems encountered with recognition were related to the system rather than the technology.

The VCS DTO is a 3-flight DTO. STS-78 was one of the three that it was approved to fly. Further improvements to the VCS will be performed before it is flown again. Additional testing in the MDF and the SMS will insure that system problems encountered on STS-78 do not occur in future flights. VCS personnel would like to fly the VCS on a RMS mission. The VCS would be used to control the CCTV system while the operator performs **simulated** RMS operations .This task would show the usefulness and time-savings gained of macro commanding the CCTV system while the operator controls the RMS arm.

The flights of voice recognition on STS-41 and 78 have certainly provided valuable insights to using this technology for spacecraft control. With further testing and continual flight s of this technology , in time, very high reliable voice control of systems in a spacecraft will become a reality. Certainly, space station and programs beyond will benefit from voice control technology as the hardware, software, and training methods improve.

11. References

1. Baber, C. and Noyes J. "Automatic Speech Recognition in Adverse Environments." Human Factors, 1996 38(1), p.p. 142-155.

2. Bejczy, A. K., Dotson, R. S., Brow, J. W. and Lewis, J. L. "Voice Control of the Space Shuttle Video System." Proc. 17th Annual Conference on Manual Control(p.p. 627-640,1981).

3. Casali, S. P., Williges B. H. and Dryen R. D." Effects of Recognition Accuracy and Vocabulary size of a Speech Recognition System on Task Performance and User Acceptance." Human factors, 1990, 32(2), p.p. 183-196.

4. Draper, J. V. and Herndon, J. N. " Manual Voice and Automated Control of a Teleoperator Viewing System." Proc. Teleoperation and Control, July 1988. (DOE contract no. DE-AC05-840R21400).

5. Pallett, D. S. "Automatic Speech Recognizer Performance Assessment." National Bureau of Standards Special Report, Jan. 1985.

6. Rabiner, L., Juang, B. "Fundamentals of Speech Recognition." Prentice Hall, 1993.

7. Salazar, G. A. "STS-41 Voice Command System Flight Experiment Report." Tracking and Communications Technical Report No. EE-2-91-001(U) Available from EV2 at the Johnson Space Center.

8. Simpson, C. A. McCauley, M . E ., Roland, E. F ., Ruth, J. and Williges, B. H. "System Design for Speech Recognition and Generation." Human Factors, 1995, 27(2) p.p. 115-141.

9. Spine, T. M. , Maynard, J. F. and Williges B. H. " Error Correction Strategies for Voice Recognition." Office of Naval Research under contract number N00014-81-K-0143.