VOICE CONTROL OF THE SPACE SHUTTLE VIDEO SYSTEM

A. K. Bejczy and R. S. Dotson
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

J. W. Brown and J. L. Lewis
Spacecraft Design Division
Johnson Space Center
Houston, TX 77058

SUMMARY

The feasibility and utility of controlling the Space Shuttle TV cameras and monitors by voice has been investigated. The voice control application concept is related to task scenarios where the operator uses both hands to control the 50-foot (16-meter) manipulator of the Space Shuttle. The use of computer-recognized voice commands allows the operator to effectively press the control buttons of the Shuttle TV cameras and monitors by voice while he manually controls the Shuttle manipulator. The pilot voice control system developed at the Jet Propulsion Laboratory (JPL) to test and evaluate the feasibility of controlling the Shuttle TV cameras and monitors by voice commands utilizes a commercially available discrete word speech recognizer which can be trained to the individual utterances of each operator. Successful ground tests have been conducted with this pilot application system at the Johnson Space Center (JSC) Manipulator Development Facility (MDF) using a simulated full-scale Space Shuttle manipulator. The test configuration involved the berthing, maneuvering and deploying a simulated science payload in the Shuttle bay. The handling task typically required 15 to 20 minutes and 60 to 80 commands to 4 TV cameras and 2 TV monitors. The best test runs have shown 96 to 100% voice recognition accuracy. The main conclusions of the tests are: (i) the application concept offers potential for enhancement of Shuttle operations; (ii) additional development is needed to achieve operational accuracy and reliability over a broad user population; (iii) the use of computer-recognized voice commands can contribute to a better man-machine system interaction; (iv) human acoustic characteristics and training have a major impact on system performance. As a conclusion it was decided to conduct further application tests and to promote the development of a prototype flight voice command system for future Space Shuttle applications.

I. INTRODUCTION

Efficient on-line decision making for manipulator control requires that the operator have an easy access to the relevant information sources. This is particularly important when the task requires frequent changes in the setting of a video system which contains several TV cameras and monitors in order to obtain the necessary information for manipulator control. In a fully manual control mode, where both the manipulator and video system are manually controlled, the operator can often attend either the video system control panel or the manipulator hand controllers. He cannot do both at one time. This is equivalent to a strictly sequential hand control of the manipulator and video system.

Altogether seven TV camera mounting locations exist in the cargo bay and on the manipulator of the Space Shuttle. Two TV monitors, located in the Shuttle cockpit, can be used in a split screen mode. Hence, up to four scenes can be displayed at one time. From the TV control panel in the Shuttle cockpit any camera can be linked with any monitor, and the pan, tilt, focus, iris, zoom and some internal electronic parameters of the cameras can be controlled. The control panel contains altogether thirty three pushbuttons and switches. (Figure 1.)

The RMS (Remote Manipulator System) operator normally uses both hands to control the motion of the Shuttle manipulator as shown in Fig. 2. The left hand controls the three translational motions, the right hand controls the three orientation motions of the manipulator. The video system control keyboard is under the left arm of the operator. (A few keyboard switches and pushbuttons are visible in Fig. 2.)

When simultaneous manual operation of the RMS and video system is impractical, the manual control of the Shuttle video system requires the execution of a complex multi-step process:

a. Decide which TV camera and monitor should be changed and how.
b. Stop manipulator motion, set RMS brakes on, and take hands off the manipulator hand controllers.
c. Turn visual attention to the video system control keyboard.
d. Find the appropriate buttons and switches on the keyboard.
e. Activate the appropriate buttons and switches and verify the success of this action on the keyboard.
f. Turn visual attention back to the TV monitors.
g. Verify the success of the desired information change on the monitors; if not satisfied repeat the process from step c. If everything is all right, proceed with step h.
h. Release the brakes, put hands back to the manipulator hand controllers, and continue the control task.

This process causes a disruption of RMS motion, diverts the operator's visual attention and manual work, and distracts his mental concentration from the manipulator control tasks. All these can contribute to lengthening the whole operation and to increasing operator workload.

The complex process of manual contr of the Shuttle video system during manipulator operations can be considerably simplified by using a computer-based discrete word voice command system for controlling the TV cameras and monitors. Since, in effect the buttons are "pushed by voice" and the switches are "turned on/off by voice", the entire video system control process is reduced to the following simple steps:

a. Decide which TV camera and monitor should be changed and how.
b. Say the appropriate word(s).
c. Verify the success of the desired information change on the monitors, and proceed with the manipulator control task if everything is all right, otherwise repeat step b.

It can be hypothesized that voice control of the TV cameras and monitors does not disturb the operator's visual attention and manual control work, and minimizes mental distraction from the control task. Consequently, the potential of voice control for enhancing the Shuttle RMS operation was investigated in this experimental study.

A pilot voice control system was developed at JPL to test and evaluate the feasibility and utility of controlling the Space Shuttle video system by computer-recognized voice commands during manual control of the Shuttle manipulator. The voice control system is briefly described in Section II. Alternative control vocabularies are presented in Section III. Control tests conducted at the JSC MDF using the simulated full-scale Space Shuttle manipulator are described in Section IV. The test results and conclusions are summarized in Section V.

II. VOICE CONTROL SYSTEM DESCRIPTION

The pilot voice control system developed at JPL to demonstrate and evaluate Space Shuttle application concepts utilizes VDETS, a commercially available discrete word speech recognizer. VDETS is essentially a trainable acoustic pattern classifier that produces a digital code as an output in response to an input utterance. VDETS is implemented in a Nova 2 minicomputer.

The basic software used in conjunction with VDETS includes a LINC Tape Operating System (LTOS) and the VOICE Executive. LTOS allows one to edit programs and save them on a LINC tape, to store voice reference templets on a LINC tape, and to execute Nova machine language programs. The VOICE Executive is a Nova machine language core-image program that assembles user VOICE programs into Nova machine code with embedded calls to the VOICE Executive. The VOICE programming language allows one to define and develop application vocabularies and syntaxes and to perform training and recognition. The VOICE Executive is completely interrupt driven to accommodate real time response to external events.

The voice control system must be trained to each individual operator whose voice pattern templets are then stored on LINC tape for recall before using the system in the recognition mode. Training typically consists of repeating the vocabulary words set seven times as it is displayed on the self-scan display unit. The operator wears a headset with a noise cancelling microphone and adjusts the volume control to accommodate his normal speaking voice. In the recognition mode, the self-scan display shows the word recognized by the system in response to the operator's utterance.

The voice command system was connected to the TV camera and monitor control circuits through a programmable interface for which a Motorola 6802 microprocessor was employed. Whenever an operator said a command word, the programmed VDETS would send an ASCII code to the interface. The interface microprocessor would then send the data out over a parallel line to a hardware decoder which energized one of the 52 wires connected to the video system control circuits. The 6802 microprocessor also performed some simple timing and logic functions. For example, some of the switches are momentary contact switches, while the camera movement toggle switches must be held in the "on" state until a "stop" command is heard.

The video system voice control was implemented so that the commands voiced by the operator did not require verification before execution; they were executed immediately. The effect of a misrecognized command was immediately visible on the monitor. The operator needed only to voice new commands to correct for misrecognition.

The voice control system ran in parallel with the manual control keyboard so that, if required, the operator could always revert to the manual control of the video system. The main elements of the voice control system together with the overall system implementation are shown in Figs. 3-4. Performance was recorded on a printer.

III. ALTERNATIVE CONTROL VOCABULARIES

Several different combinations of vocabulary words both with and without syntax restrictions were developed and tested. Figure 5 shows a vocabulary and syntax which closely follow the words and organization of the keyboard. As seen in Fig. 5, the actual TV camera and monitor control words are arranged in five groups corresponding to the grouping of buttons and switches of the keyboard shown in Fig. 1.

In general, the syntactic organization of command words serves the purpose of increasing word recognition accuracy. The syntactic organization limits the number of words to a subset of the total vocabulary that the speech recognition system has to look up for identification of a spoken command word. Figure 6 shows a vocabulary with a multilevel syntax. As seen in Fig. 6, one can construct many subsets of the vocabulary which only contain two, three or five words. But increased syntactic grouping of words increases the application rules that the operator must remember and follow. Note also in Fig. 6 that some of the subset words are very short, e.g., "far", "in", "out", etc. Very short words have higher misrecognition probability than the longer words. The words in Fig. 6 are "natural" in the sense that they closely follow the names or functions of the keyboard buttons and switches.

The training experiments have shown that the operators prefer simple vocabularies with minimum or no syntactic restrictions. Following this desire, two vocabularies were constructed shown in Fig. 7 and 8. Note that many vocabulary words shown in Fig. 7 and 8 are concatenated words, e.g., "zoom-in", "tilt-up", "focus-far", etc. The use of concatenated words increased recognition accuracy by 6 to 8% and provided smoother and faster operation performance. The use of a concatenated word requires only one voice command (e.g., "zoom-in") for an action instead of two words (e.g., "zoom" and "in"). But some of the words shown in Fig. 7 and 8 are rather lengthy. In some cases it was necessary for the operators to speak at an unnaturally fast speech rate to get the entire utterance within the 1.5 second window that the speech recognition system allows for each spoken word. If the utterance lasts longer than 1.5 seconds, the recognition accuracy can be poor.

The vocabulary which was used during the tests at the JSC MDF is the simplest one without syntax shown in Fig. 8. It only contains two words ("stop" or "reverse") which logically must follow the action commands like "iris-open", "pan-right", etc.

IV. CONTROL TESTS

Control tests were conducted at the JSC MDF in January 1981 to evaluate the feasibility and utility of controlling the Shuttle TV cameras and monitors by voice during manual control of the Shuttle manipulator. The task configuration chosen for the tests was that of handling a Plasma Diagnostic Package (PDP) payload mock-up by the manipulator in the Shuttle bay. The PDP was berthed to and deployed from a retention mechanism.

The task started with the manipulator holding the PDP payload mock-up above the aft cargo bay area (Fig. 9). It was then docked to the retention mechanism in the aft bay, the time recorded, then deployed from there, moved and docked to a similar retention mechanism in the forward cargo bay area. The task ended when the payload was removed back to a starting position above the cargo bay. The windows of the cockpit were blocked so that the operators were forced to rely upon the TV cameras and monitors for visual feedback from the task area. The manipulation task typically required 15 to 20 minutes.

Altogether 48 test runs were performed by four operators, 32 runs with voice control of the video system. Table 1 summarizes the average number of video system control commands in both manual and voice control modes. The average number of voice commands in Table 1 does not include the misrecognized command words. Table 1 shows that the average number of manual and voice commands varies from operator to operator. It is interesting to note that the average command number variation between operators in the voice mode is less than in the manual mode.

The control tests were performed after six, seven, eight, and nine training passes for each operator. Where all the training was done at approximately the same time, six training passes seemed to give the best results. Any more than this seemed to corrupt the training patterns. The training was performed by repeating the whole vocabulary sequentially rather than repeating each word individually. When the tests were performed on a subsequent day from the training, two extra update training passes seemed to give the best results. The standard procedure was to save seven primary training passes on the LINC tape for each operator, and then update these seven passes just before the system was used in the recognition mode during the control tests, disregarding the prior updates.

During the tests the typical mode of operation was to first position the cameras and then concentrate on payload docking. This was true even with the voice system, although near the end of the tests three operators were able to combine a certain amount of camera movement with payload movement as they became more comfortable with the system.

The voice command system was used not only to select the various cameras and monitors, but also to control the camera movement and lens parameters (pan, tilt, focus, iris, zoom). The most troublesome part of the test was to control camera movement. Here the accuracy was most important since timing is critical in order to stop the movement at the right time to achieve the desired results. In most cases the operators preferred to control camera movement in "low-rate" setting. This was also the preferred setting in manual control mode. "High-rate" setting was typically used for coarse movement control.

V. RESULTS AND CONCLUSIONS

The best individual test runs have shown a recognition accuracy from 96% to 100% As seen in Table 2, there is relatively large recognition accuracy variation between the individual operators. Three of the four operators underwent familiarization training with the voice command system at JPL two months prior to the tests at JSC. Their recognition accuracy during the tests at JSP was consistently better than the recognition accuracy of the fourth operator who learned the use of the voice command system one day before the tests.

The two "accuracy" columns in Table 2 refer to two methods of computing recognition accuracy. In the first column the accuracy is computed without the rejected words. In the second column the accuracy is computed by taking account of the rejected words. That is, rejected words were counted as errors. Each percent number belonging to an operator in the columns of Table 2 is the result from four individual test runs.

Table 2 indicates that voice recognition accuracy also depends on the vocabulary to some extent. The vocabularies JSCN04 and JSC002 in Table 2 correspond to the vocabularies shown in Figs. 7 and 8, respectively. But, as seen in Table 2, the recognition accuracy of the best scoring operator (operator B) was insensitive to both vocabulary variation and accuracy computing method.

Several off-line recognition tests (without manipulator control) were also performed at JSC with four naive users who had never previously used a voice recognition system. Their average recognition accuracy was about 90%. It is interesting to note that among the four primary operators and four naive users there were altogether three female and five male subjects, and the average recognition accuracy of the female subjects was 8-9% higher than the average recognition accuracy of the male subjects. It is also noted that only one female was used for the on-line tests, and her recognition scores were nearly perfect, ranging from 96% to 100%. Of course, these data don't have statistical significance since the test subject population was too small.

The duration of each test run with the voice command system steadily decreased as each operator became more familiar with the system. The average time per task in voice control modes was still about 10% longer than in manual control mode during the tests which should be regarded as introductory. It is felt that this time duration average will be reversed where (i) the operators gain more experience with the voice command system and (ii) the accuracy of the voice recognition system is improved. It should be kept in mind that all operators had several years extensive experience with the manual operation of the video system. As seen in Table 3, however, there was a large variation between the average time performance of the four operators even during the manual operation of the video system.

After becoming more familiar with the system, the operators were impressed with its potential and enthusiastic about it even though they felt that the recognition accuracy should be improved. In general, there was an agreement among the operators that at least 95% average total recognition accuracy is needed with a 50-word vocabulary in order for the operators to feel comfortable with the voice command system during real-time operation. In the total recognition accuracy the rejected words are counted as errors; see last column in Table 2.

A few interesting general remarks emerged after the tests:

1) Command words should be added to the vocabular that will (i) restore camera and monitor to the condition prior to a recognition error, and (ii) allow an operator to name a selected camera position once it has been set up so that it may be re-invoked with a single word instead of repeating a complete command sequence.

2) Though the commands voiced by an operator did not require verification before execution, the operators often felt it reassuring to look at the self-scan display of the recognized words. This display, however, should be a small device and placed very close to the TV monitors.

3) The operators would like to be able to issue commands other than "stop" or "reverse" while a camera is moving. This capability would speed up the operation.

Though the control tests were not meant to test and evaluate a particular voice recognition system, it should still be mentioned that the VDETS*) system performed very well even in the presence of acoustic and electrical noise.

The main conclusions of the test are: (i) the application concept offers potential for enhancement of Shuttle operations; (ii) additional development is needed to achieve operational accuracy and reliability over a broad user population; (iii) the use of computer-recognized voice commands can contribute to a better man-machine system interaction; (iv) human acoustic characteristics and training have a major impact on system performance. As a conclusion it was decided to conduct further application tests and to promote the development of a prototype flight voice command system for future Space Shuttle applications.

Acknowledgment

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under NASA Contract NAS7-100. The programmable interface for connecting VDETS to the video system control circuits was developed by H. C. Primus of JPL.

Note: A seven-minute narrated movie is available which shows the control tests at the JSC MDF using voice control of the Space Shuttle video system.

*) VDETS is carried by Interstate Electronics, Anaheim, CA.

Table 1. est Summary.

- 4 TRAINED OPERATORS
- 48 TEST RUNS (EACH 15 TO 20 MINUTES)
- 32 RUNS WITH VOICE COMMANDS

	AVERAGE NUMBER OF COMMANDS	
OPERATOR	MANUAL MODE	VOICE MODE
A	57	62
B	80	70
C	47	63
D	66	52

Table 2. Voice Command Recognition Summary Accuracy.

OPERATOR	VOCABULARY	ACCURACY W/O (%)	ACCURACY W (%)
A	JSCN04	96	90
B	(Fig. 7)	97	95
C		89	86
D		80	78
AVERAGES		91	87
A	JSC002	92	89
B	(Fig. 8)	97	97
C		86	83
D		79	72
AVERAGES		89	85

Table 3. Average Task Durations.

OPERATOR	MANUAL		VOICE
A	14:36	(Minutes:Seconds)	21:12
B	20:56		20:19
C	11:25		14:74
D	25:12		25:30

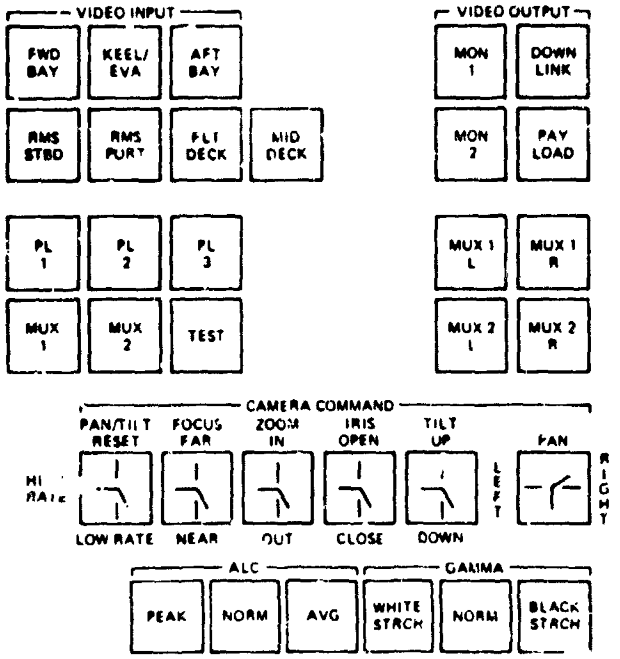

Figure 1. Space Shuttle TV Camera and Monitor Control Keyboard.

Figure 2. Space Shuttle Cockpit Control and Information Environment for Manipulator Operation.

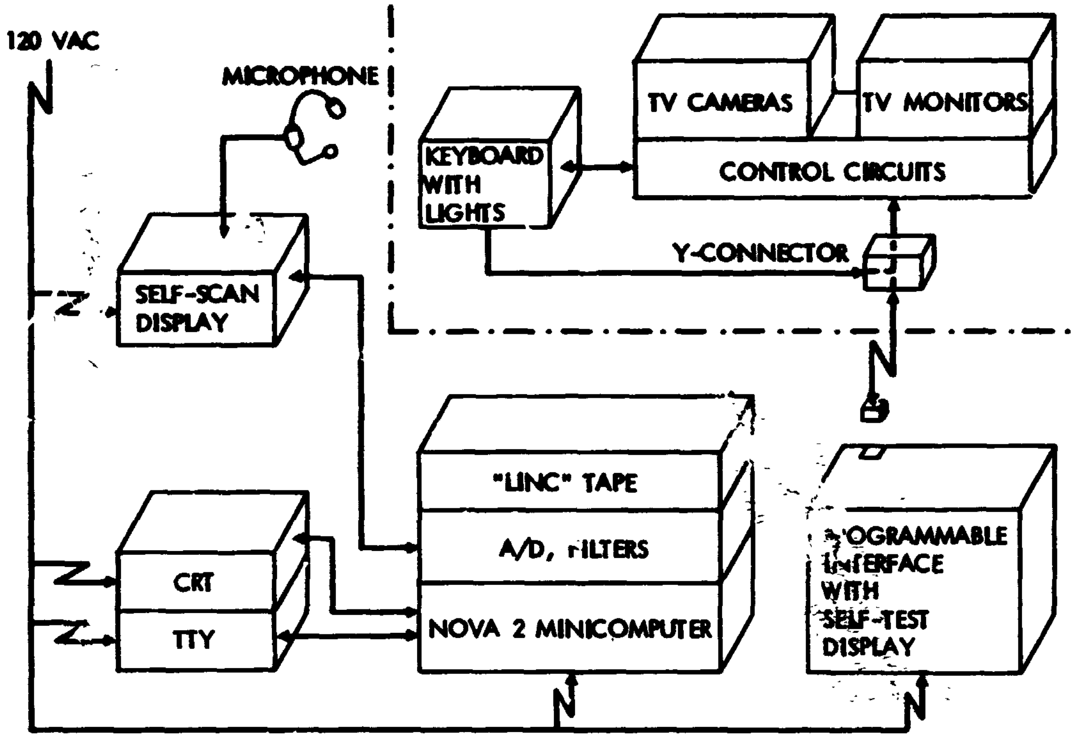

Figure 3. Main Elements of the Voice Control System and Overall System Implementation.

Figure 4. Operator Uses Voice Control of Video System During Manual Control of Shuttle Manipulator.

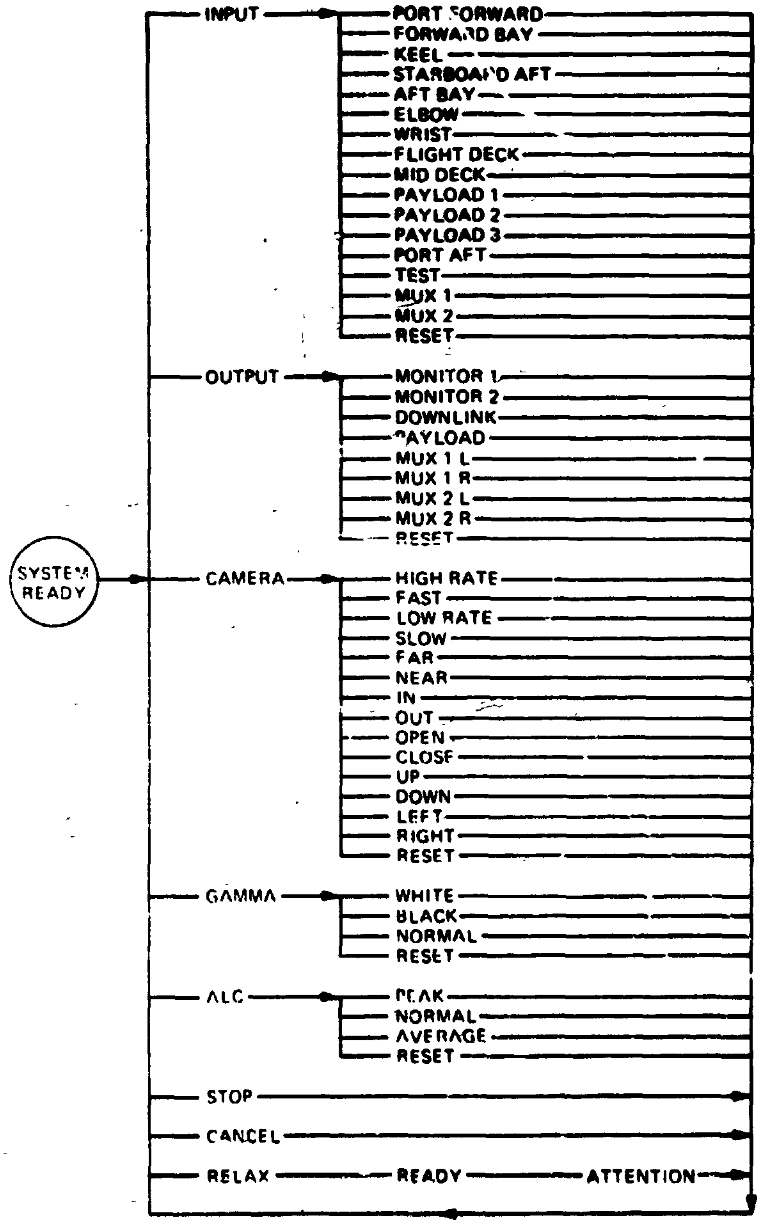

Figure 5. Natural Vocabulary with Simple Keyboard Syntax.

ORIGINAL PAGE IS
OF POOR QUALITY

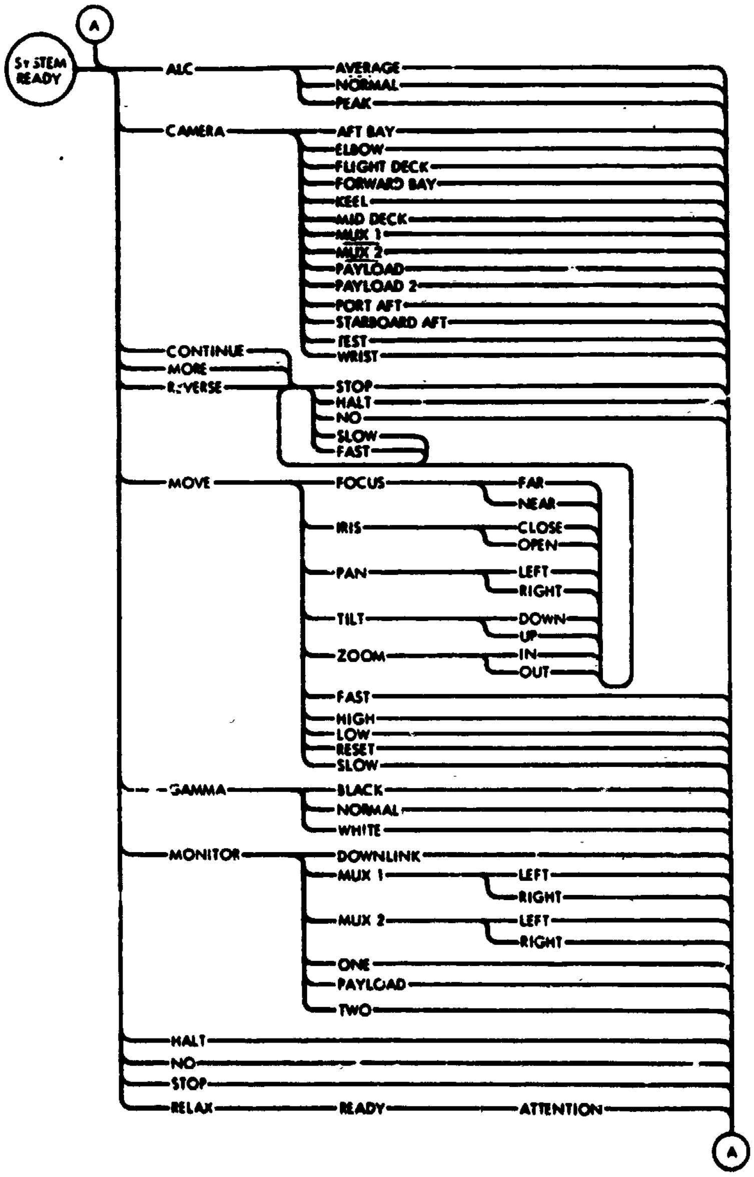

Figure 6. Natural Vocabulary with Multilevel Keyboard Syntax.

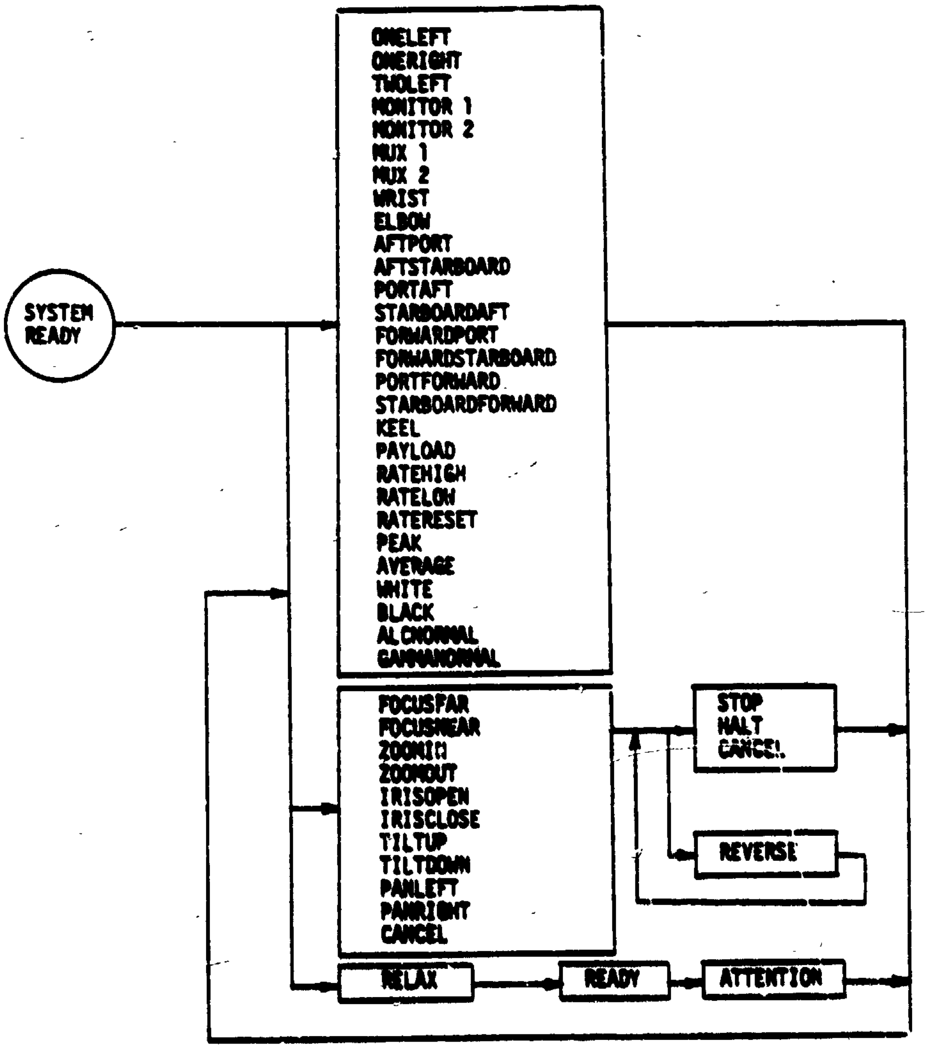

Figure 7. Vocabulary with Concatenated Words and Minimum Syntax.

PORT-FORWARD
FORWARD-PORT

AFT-STARBOARD
STARBOARD-AFT

MUX-1
MUX-2

AFT-PORT
PORT-AFT

FORWARD-STARBOARD
STARBOARD-FORWARD

RMS-PORT
ELBOW

MONITOR-1
MONITOR-2

MUX-1-LEFT
MUX-1-RIGHT
MUX-2-LEFT

MUX-2-RIGHT

RATE-HIGH FOCUS-FAR ZOOM-IN IRIS-OPEN TILT-UP PAN-LEFT
RATE-LOW FOCUS-NEAR ZOOM-OUT IRIS-CLOSE TILT-DOWN PAN-RIGHT

REVERSE STOP

PEAK ALC-NORMAL AVERAGE WHITE GAMMA-NORMAL BLACK

RELAX READY ATTENTION

Figure 8. Reduced Vocabulary with Concatenated Words and without Syntax.

Figure 9. Task Scene for Voice Control of the Shuttle Video System During Manual Control of the Shuttle Manipulator.