EXAMINATION OF THE VOICE COMMAND CONCEPT FOR APPLICATION TO AIR FORCE COCKPITS

[redacted]

JOHN M. REISING

FOREWORD

This report describes the results of an analysis conducted to examine the applicability of the voice command concept to aircraft cockpits. The work was accomplished under the USAF Integrated Flight Control-Display program (Project 6190). The work is specifically related to Task 21 of this project entitled, "Advanced Integrated Tactical Fighter Cockpit Development".

Mr. William F. Swartz, of the Bunker-Ramo Corporation, made significant contributions in the overall organization and planning of this work effort. In addition Mr. John Rumer, also of the Bunker-Ramo Corporation, provided many valuable inputs in formulating the description of the automatic speech recognition process.

This report includes research performed between 24 September 1971 and 10 December 1971. This report was submitted by the author March 1972.

This technical report has been reviewed and is approved.

JOHN H. KEARNS, III
Acting Chief
Flight Deck Development Branch
Flight Control Division

ABSTRACT

Voice Command holds potential for off-loading many manual switching tasks from the pilot. This report surveys the state-of-the-art in voice command, discusses cockpit applications, and identifies primary manufacturers/government agencies doing research. Such factors as the functional specifications of an airborne voice command system and concern for the pilot in relation to voice command are stressed. An analysis of the overall situation leads to the conclusion that voice command is on the threshold of application into cockpits, given that certain restrictions are acceptable.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
	Aircraft Complexity and the Pilot	1
	Simplifying the Pilot's Job through Voice Command	4
II.	SPEECH AND AUTOMATIC SPEECH RECOGNITION	5
	How We Speak	5
	Acoustical Properties of Speech	6
	How Automatic Speech Recognition Works	7
III.	COCKPIT APPLICATIONS	10
	Weapon Systems	10
	Emergency Procedures	10
	System Status	12
	Communications	12
	Checklists	13
IV.	SYSTEM CONSIDERATIONS	14
	Equipment Considerations	14
	Human Factors Considerations	17
	Pilot Factors Considerations	20
V.	MANUFACTURERS AND GOVERNMENT AGENCY RESEARCH IN AUTOMATIC SPEECH RECOGNITION	23
	Manufacturers	23
	Research by Government Agencies	25
VI.	ANALYSIS OF THE POTENTIAL OF VOICE COMMAND	28
	Where the Voice Command Concept Stands	28
	Functional Specifications of an Airborne Voice Command System	30
	Progress Towards an Airborne Voice Command System	31
	Setting the Development Strategy	34
	What Not to do in Implementing Voice Command	34
	A Research Program for Implementing Voice Command	35
VII.	SUMMARY	39
VIII.	CONCLUSIONS	50
	REFERENCES	51

SECTION I

INTRODUCTION

1. AIRCRAFT COMPLEXITY AND THE PILOT

The F-111 is a far cry from the Spad of WWI fame; the capability of the F-111 exceeds that of the Spad by several orders of magnitude. However, this increased capability was not achieved without a corresponding increase in the number of avionic subsystems on board the aircraft. Figure 1 shows the increase in the number of subsystems on-board manned weapon systems for a 60-year time frame.

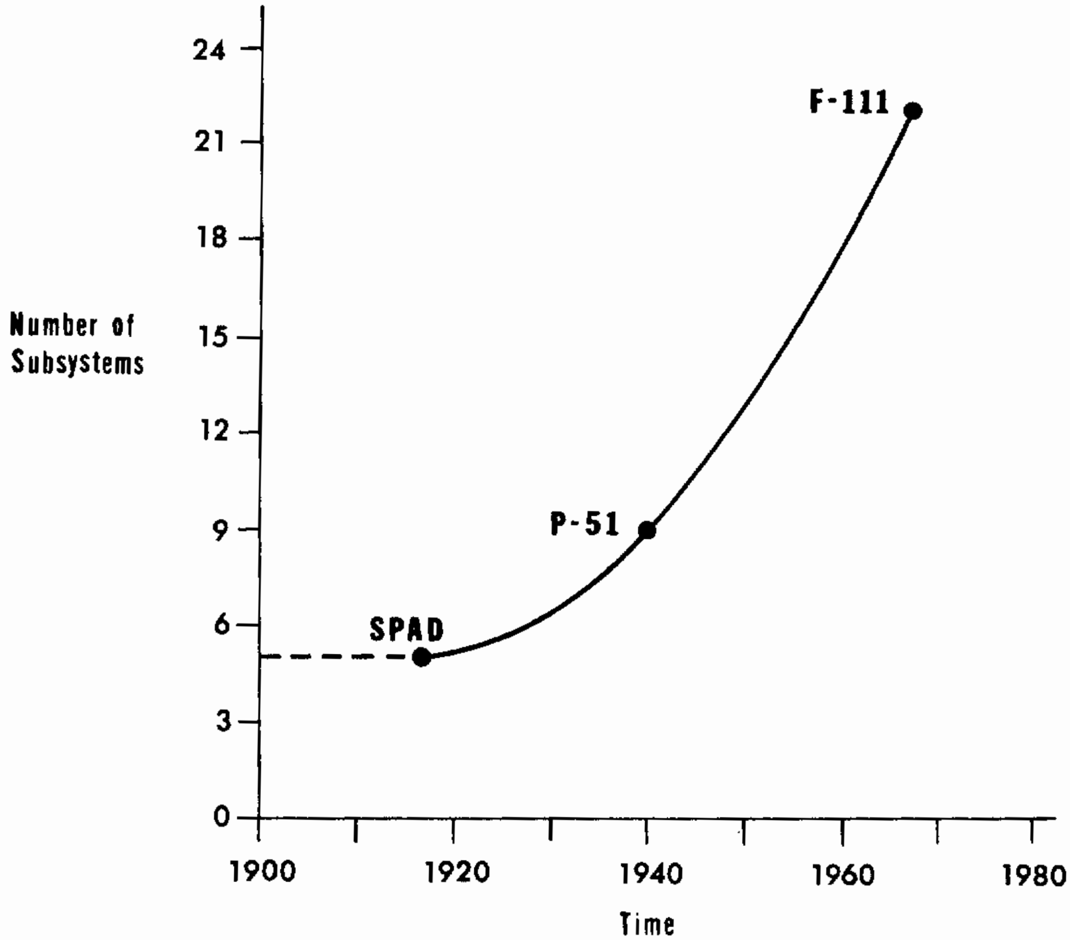

Fig 1. Number of Subsystems on-board Three Aircraft as a Function of a 60-year Time Period.

Figure 2 shows that the Spad of the early 1900s had five subsystems. The F-111, today's most sophisticated weapon system, has 22 subsystems. And, not only are there more subsystems, but they are infinitely more complicated. This continuing increase in the number and complexity of subsystems has also resulted in more annunciations, more switches, and more procedures. The increase in switches and controls is illustrated in Figure 2.

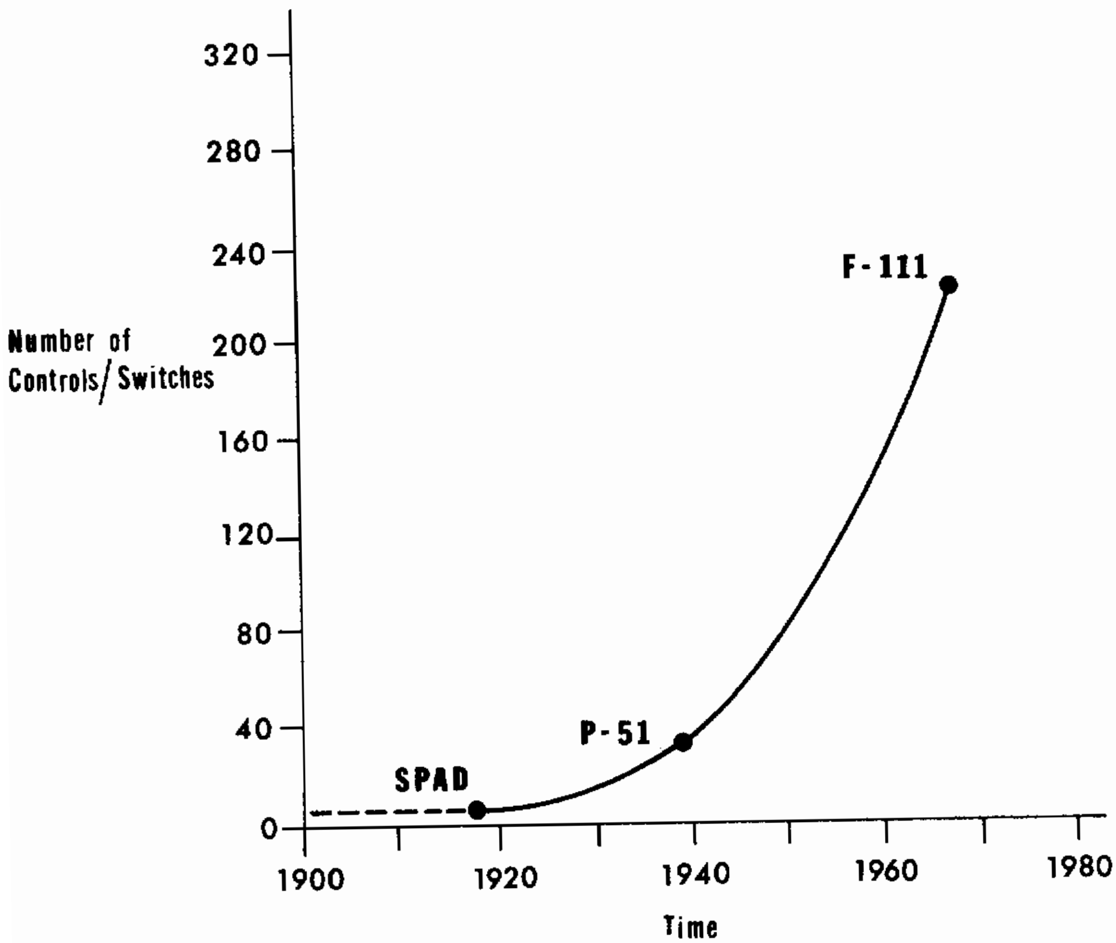

Fig 2. Number of Controls/Switches for Three Aircraft as a Function of a 60-year Time Period

The Spad had less than 10 switches, but the F-111 has more than 220.

The complexity of the decision-making involved in operating today's intricate subsystems can be illustrated in comparing the weapons delivery system in the P-51 and F-111. In the P-51 the delivery system consisted of a 5-position rotary selector knob and the weapon release mechanism on the control stick. The F-111, on the other hand, has a delivery system that contains eight separate control/display panels that the pilots must consider before they can deliver ordnance properly. In addition, there are literally dozens of possible weapon delivery modes available to the weapon system operator, and the selection must be made during those portions of the flight profile that are most stressful.

The increased complexity has been found to have direct impact on mission effectiveness. Detailed analysis of masses of data collected in Southeast Asia (SEA) reveals that, in a number of cases, pilots were not able to hit their ground targets or fire missiles at enemy MIG's because the weapons systems were so complex; the pilot had to do so many things in a relatively short period of time that the weapons did not fire or were fired incorrectly. These were not isolated cases but happened time and again.

Communications systems operations have also contributed tremendously to the pilot's work. SEA pilots have reported numerous communication problems. Probably the most overworked pilots in the USAF, in terms of communications, are those performing search and rescue (SAR) missions. Data collected from SAR pilots shows that the pilot and copilot often must be concerned with the operation of five communication radios in addition to five navigation radios. Listening to a single mission tape will show the tremendous workload imposed upon SAR pilots.

What does all this mean? It comes down to this: Aircraft have become increasingly more sophisticated in terms of number and kinds of subsystems plus the accompanying controls/displays and procedures. Each of these means more things for the pilot to do, until he has become saturated with the manual manipulation of knobs, switches, and buttons. Even with the outstanding training today's pilots receive, there is a limit on the number of subsystem manipulations we can require them to perform.

A Flight Safety Officer at the F-105 Combat Crew Training School relates his observations about the problems encountered by UPT graduates entering CCTS. The T-38 is the only high-performance, fighter-type aircraft that these pilots have flown, but in his opinion, operating the F-105 weapon delivery system, alone, is more complicated than flying the T-38. Operating the weapon delivery system is so demanding that flying duties are neglected to such an extent that aircraft have been lost.

If this little story has a moral, it is that we are faced with a very real, operational problem. Something has got to be done to make it simpler and easier for the pilot to use the capabilities of today's superaircraft.

2. SIMPLIFYING THE PILOT'S JOB THROUGH VOICE COMMAND

One way of simplifying the pilot's job would be by providing for operation of cockpit subsystems by means of voice command. The subsystems would be activated and manipulated, at least in part, by the pilot merely telling them what he wanted them to do. This would relieve the pilot of many of his manual duties. For example, to change radio frequencies on the UHF radio, he would simply say the new frequency - "UHF 259.2" - and the radio would change to the new frequency. A weapon system could be set up in a similar manner. The pilot could select the desired weapon by saying its name, then giving the command - "Phoenix, select." The weapon could be armed (or even fired if thought desirable) by voice command alone, e.g., "Phoenix, arm; Phoenix, fire!"

There is an obvious advantage in using voice command. The voice command mechanism can be employed without hampering any of the existing command sensory mechanisms. That is, the pilot's hands and feet, eyes and ears are still free to process information and control the aircraft. We have merely off-loaded some of the tasks onto a previously untapped command mechanism.

Before we begin equipping the USAF aircraft with voice command systems, however, we must make some serious considerations and decisions. This report is prepared to provide information on some of these factors:

a. The concept of voice command and state-of-the-art of related hardware.

b. An overall assessment as to where we stand with respect to application.

c. Who is doing what in the voice command area.

d. Conclusions and recommendations as to what needs to be done.

The overall objective of this report is to provide a foundation concerning the application of voice command to crew station design problems so that decisions can be made based on adequate knowledge.

SECTION II

SPEECH AND AUTOMATIC SPEECH RECOGNITION

Before one can understand how a voice command system functions, one must understand automatic speech recognition techniques. However, understanding how automatic speech recognizers work requires knowledge of the physiological and accoustical nature of speech.

Speech and speech recognition are highly technical areas, each filled with its own specialized jargon. This report is not intended to present a detailed analysis of these two areas. (More detail may be found in the references.) But some examination is necessary to understand the capabilities and problems involved in voice command systems. We will begin with a superficial examination of the physiology of speech.

1. HOW WE SPEAK

The basic organs involved in the process of speech are shown in Figure 3.

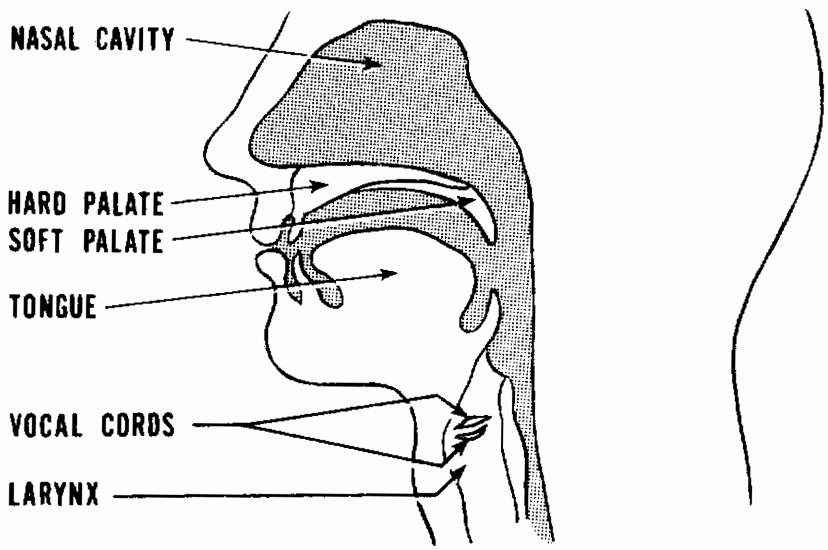

Fig 3. Speech Mechanisms. Note the position of the larynx and vocal cords. (From Munn, 1969, p. 583).

The heart of the speech system in the human being is the vocal cords. According to Munn (Reference 4, pages 582-583):

> These are two membranes stretched across the interior of the box-like larynx at the level of the Adam's apple. Without these one can still speak, but only in a whisper. The vocal cords add "voice" to what is said.
>
> We normally cause the vocal cords to vibrate by forcing air up the windpipe, using the lungs as a type of bellows. When the opening between the vocal cords (the glottis) is only a small slit—because the membranes comprising the vocal cords are tightly drawn — there is rapid vibration, resulting in tones of high pitch. A larger opening, with the membranes under less tension, produces lower tones. With a wide open glottis there can be only whispering. This whispering is the only mode of speech available when disease necessitates removal of the larynx.

The human vocal apparatus produces a wide variety of sounds. Some of these (voiced sounds), as noted above, are produced by the vibration of the vocal cords as air passes over them. Unvoiced sounds, on the other hand, are generated by a change of size in the vocal tract or by an abrupt release of pressure in the mouth. There are only approximately 40 unique basic sound elements. called phonemes; for example, the sound of the "c" in cool and the "k" in keep are both described by the phoneme /k/. One approach to speech recognition is to design a machine to recognize speech based on phonemes.

2. ACOUSTICAL PROPERTIES OF SPEECH

Acoustically, speech can be thought of as sound waves that are modified by the vocal cords, tongue, soft palate, lips, and nasal cavities. "Thus, the wavefront issuing from the mouth and nose is a complex waveform varying with each change in position of the organs of speech" (Reference 1, p. 58). The energy concentrations, known as formants, are one of the keys in the analysis of these complex waveforms, and the formants are produced by resonances of the vocal tract. Most sounds can be identified as acoustical events by analyzing the first four formants. A visual presentation of sounds (and hence of formants) can be obtained by means of a sonogram. A sonogram is illustrated in Figure 4.

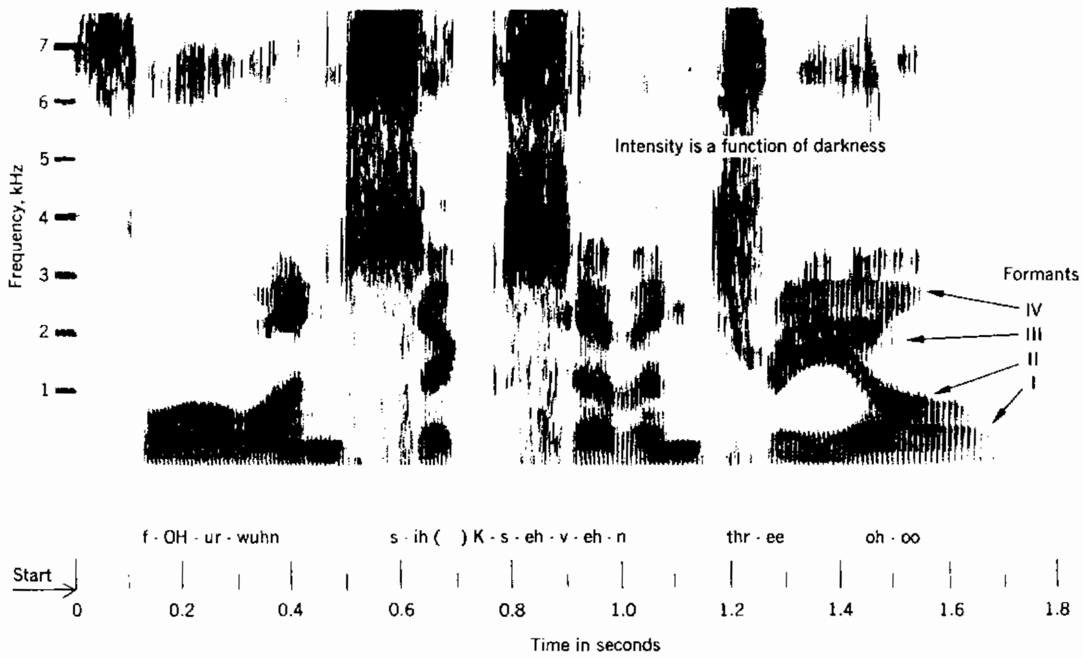

Fig 4. Sonogram of a male voice saying the digits 4, 1, 6, 7, 3, 0. (From Clapper, 1971, p59.)

The formant characteristics of the sound waves displayed on the sonogram can then be analyzed to identify the unique qualities of a given sound.

3. HOW AUTOMATIC SPEECH RECOGNITION WORKS

The essential functions of a voice recognition device are conversion of the spoken word into an analog signal, sampling of this signal to extract sufficient significant characteristics of the utterance to produce a representative voice pattern, and comparison with previously stored voice patterns for identification of the word.

a. General Description of a Representative System.

The following is a description of the voice command system manufactured by Scope, Inc., which is representative of hardware approaches used to construct a discrete word voice recognition system. Figure 5 is a block diagram of the Scope device, which is essentially an acoustic pattern classifier that produces a BCD output in response to a spoken command. The system consists of an analog signal processing subsystem, multiplexer, analog/digital converter, digital processor, and an output register. The blocks in Figure 5 labeled command detector, coding compressor, classifier, estimator and pattern memory are actually functions of the digital processor which consists of a CPU, Read Only Memory (ROM), and Core Memory.

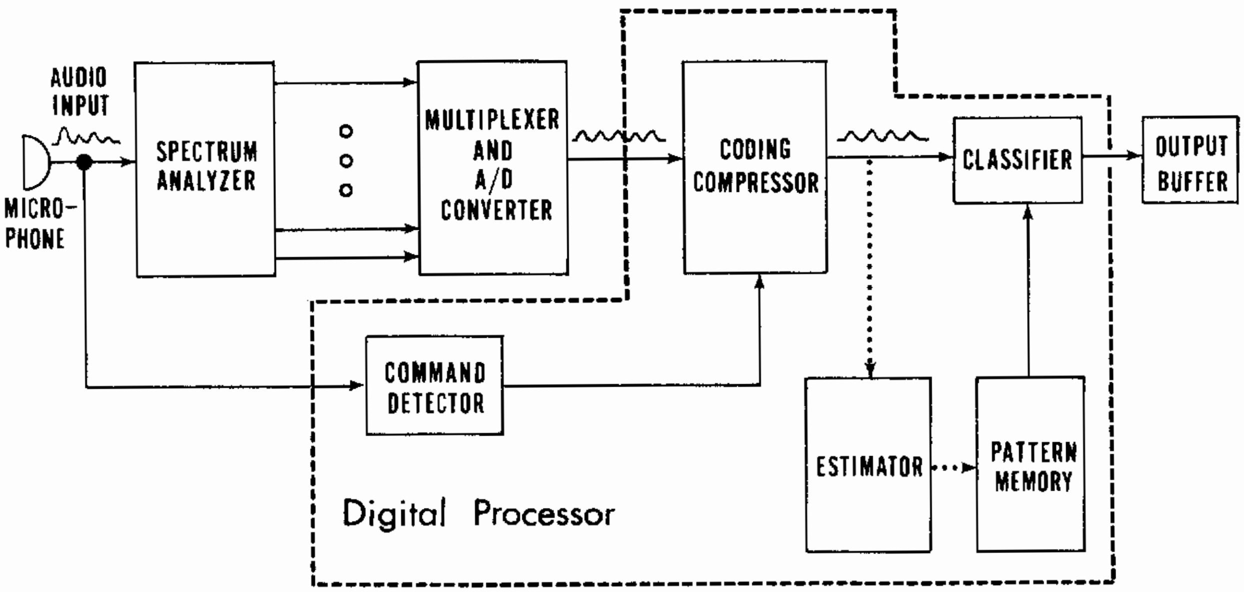

Fig. 5 Block Diagram of an Automatic Speech Recognition Device. (From Glenn & Hitchcock, 1971, p. 86).

The system has a maximum vocabulary of 24 discrete words of up to one-second duration. A pause of at least 250 milliseconds is required between words. Adaptation, or training, to each of several speakers is necessary. This is accomplished by having each user repeat each word 5 times; the average acoustic pattern is stored in the digital processor core memory for comparison during pattern recognition.

The combination of the spectrum analyzer and multiplexer/A-D Converter performs the function of feature extraction, or selection of significant characteristics from the total acoustic content of a spoken word which are used to construct a representative voice pattern.

b. Specific Description

The acoustic content of a spoken word is first converted into an analog signal at the audio input (see Figure 5) by a microphone and amplifier. This signal is then input to a spectrum analyzer consisting of an array of 16 bandpass filters covering the range from 200 Hz to 5KHz. Each filter section consists of a bandpass filter, detector, and lowpass filter which provides a slowly varying DC level proportional to the energy in each frequency band of the spectrum.

This power spectrum is then sampled and digitized. The 16 filter section outputs are sampled 60 times a second by the multiplexer and the level quantized to a 4-bit description by the analog to digital converter. Thus, the original utterance is output from the A/D converter as a string of 4-bit binary numbers, each representing the amplitude of one of the frequency bands at some instant of time. The total number of bits generated will vary with the length of the utterance. A word of one-second duration would produce 16x4x60=3840 bits.

Since the length of time to utter a particular word varies significantly between speakers and even between two utterances by the same person, the system includes a coding compressor. This function, (performed by the central processor) reduces the spectral data generated from each spoken word to a 120-bit binary pattern, regardless of the length of the original utterance. Raw binary data is compressed by preserving significant spectral variations while eliminating redundant data. The computer algorithm used to accomplish this function is essentially an arithmetic process based on information theory which preserves changing data and eliminates constant data. The function of the command detector is to inhibit coding compression if the audio input signal level is below a predetermined threshold.

The recognition decision process--performed by the classifiers, estimator, and pattern memory elements of the figure--consists of forming the correlation products between the 120-bit output of the coding compressor and each of the stored 120-bit reference patterns obtained during the training process. The index of the stored word pattern yielding the highest correlation value is then selected as the output (BCD code from the output buffer register), provided the correlation value exceeds the reject threshold for that word pattern.

SECTION III

COCKPIT APPLICATIONS

The major idea throughout this report is voice command as a useful tool for the pilot to use. Therefore we must necessarily discuss cockpit applications. Voice command holds potential for various applications in the cockpit. The applications discussed in this section will consider the use of voice command in conjunction with (1) weapon systems, (2) communications, (3) checklists, (4) emergency procedures, and (5) system status.

1. WEAPON SYSTEMS

As the data gathered in SEA reveals, the weapon systems of many of today's fighter/fighter-bomber aircraft are both complex and time-consuming to set up. One of the obvious applications of voice command is to the manipulation of weapon systems.

Besides eliminating the need for the pilot to fly "head down" while trying to deal with the knobs and buttons required to set up the system, the use of voice command would also allow the pilot's hands to be free for other tasks. In addition, to set up the weapon system, the pilot could go through one series of steps to select a weapon and another series to arm it. With voice command the pilot would merely go through two steps ("Weapon, select; Weapon, arm") to select and arm the weapon. If desired, he could also fire the weapon by voice command. If he chose the wrong weapon or wished to disarm it, he would merely say, "Weapon de-select" or Weapon disarm" and the system would carry out the command. Each of the words keys a number of functions, but from the pilot's point of view, the use of voice command greatly simplies this task by off-loading the work onto the hardware. To prove that this idea isn't something out of Buck Rogers, the Navy has conducted a study to examine this very concept.

2. EMERGENCY PROCEDURES

Another application of voice command in the cockpit could be to carry out emergency procedures. Basically, emergency procedures can be classified as a checklist. But because of their special implications on the performance and safety of the pilot and his aircraft, we will consider emergency procedures as a separate topic.

Voice command could be used with existing emergency procedures and also with new types of emergency procedures that could be implemented in the not too distant future. Current emergency procedures are written in the Dash 1 and in the flight crew's abbreviated checklist. In an emergency situation the pilot's master caution light goes on as well as a light on his annunciator panel which tells him of the particular failure (e.g. DC GEN OHEAT). The pilot must judge the seriousness of the failure in view of the particular situation (takeoff, cruise, combat, etc.) and either ignore the annunciator light or take the appropriate action as outlined in the abbreviated checklist.

Voice command could fit into these current emergency procedures in the following way. When the failure occurs and the annunciator light comes on, the pilot, after looking at the emergency procedure in the checklist, would decide either to ignore the light, or execute the procedure. If he decided to ignore the light, he would say nothing; if he decided to execute the procedure, he would merely say "DC Gen Oheat Execute," and the procedure would be completed automatically. An END annunciator light would tell him when the procedure was completed.

A more desirable format for executing emergency procedures automatically could be implemented by linking voice command with a CRT. When a failure occurred it would be displayed on the CRT along with two options and their impact. Option 1 would be the "no action" choice and would mean that the pilot has judged that there is no appropriate action he wishes to execute at this time. Along with option 1 would be displayed the impact of choosing it (e.g. loss of the DC generator in 30 minutes). Option 2 would consist of the emergency procedure as listed in the Dash 1. Its impact would also be listed (e.g. loss of windshield gen.) If the pilot chose option 2, as each step of the procedure was executed it would disappear from the CRT. If a step could not be completed, it would flash on and off. Voice command would be the means by which the pilot started the execution of the options; as he read the failure and the options, he would merely say "option 1 (or 2) execute".

Note that we have kept the pilot as an active component in the decision making process but have freed him from the routine task of carrying out the steps demanded by the particular option chosen. This idea fits in very well with recent data presented in Interceptor. In an examination of accidents in fighters and fighter-bombers, it was found that over 50% of the emergency procedures carried out were incorrect. "There are indications that many pilots do not practice precise emergency procedures enough so they can accurately perform these during a real emergency." (Ends, 1971, p.5). By utilizing voice command coupled with a CRT, it is our belief that these errors could be reduced significantly. Not only are "carrying out" errors reduced, but the pilot can still concentrate on flying the airplane instead of reaching to various places in the cockpit to complete the emergency procedure.

3. SYSTEM STATUS

In the course of his mission, the pilot has a continuing need to check on the status of his aircraft's systems. Besides routine checks on such things as fuel flow and engine rpm, he also has need for information relative to the combat situation. For example in a dogfight, the pilot may want to know how many g's he is pulling. To obtain this information, he would say "g's status", and the information would be presented to him either aurally or on an HUD. An additional feature could be the continuing presentation of the information. Using the above example, the information would be presented, say, every 5 seconds (the time intervals between presentations would have to be worked out) until he would stop the presentation by saying "terminate". A noncombat situation where voice command could be used to give the pilot system status information might be in the final approach. One of the things a pilot needs to know during this time is his airspeed. He could request this information as he captures localizer, for example, and it could be given to him periodically during final approach. He would call up the information by saying "airspeed status," and stop it by saying "terminate."

4. COMMUNICATIONS

The pilot is forever talking to someone, whether he is in the U. S. or Vietnam. In the search and rescue (SAR) mission, for example, there is such a clamor for information that one wonders how the crew can do anything but handle the communications task. As far as the U. S. environment is concerned, anyone who has flown IFR realizes he must communicate and change frequencies innumerable times. For example, in a typical cross-country, the pilot may have to contact clearance delivery, ground control, tower, departure control, traffic control centers, sectors within the center, approach control, tower, and ground control. In addition numerous navigation radio frequency changes must be made.

It certainly appears as though voice command could be used quite effectively for frequency change purposes. As was illustrated earlier in this report, if the pilot merely had to call the particular radio (e.g., UHF comm., VHF comm, Tacan) and then state the new frequency, his workload would be alleviated tremendously. The advantages of voice command that were cited for weapons systems also apply to radios. The pilot can fly heads up and still use his hands for other tasks.

5. CHECKLISTS

Checklists are currently accomplished in one of two ways: in single-place aircraft, the pilot both reads and carries out the checklist items; in two-place (or larger) aircraft, they use a "call and challenge method. One crew member reads the item and another completes it and responds that it is completed. For example in the T-39, the co-pilot reads "Gear Electric Reset," and the pilot presses the gear electric reset button and replies, "Reset." The voice command system could serve as a second crew member in a single-place aircraft. As the pilot reads the item into the voice command system, the system would carry out the item and provide feedback when the item was completed.

It may not prove feasible from an engineering point-of-view to have an entire Before Engine Start Checklist performed by voice command, but it certainly appears feasible for parts of it. For example, all the command and navigation radios could be turned on or the electrical systems checked by voice command. In addition, shorter checklists could be done in their entirety by voice command. Using the Before Takeoff Checklist for the F-100 as an example, the pilot would merely say the words, "Before Take-off Checklist", and the voice command system would carry out all the items except those few items the pilot may wish to do himself (e.g. Takeoff trim check) or those items that would be difficult to automate (e.g. "as required" items).

SECTION IV

SYSTEM CONSIDERATIONS

The man-machine aspects of the voice command concept must be considered in any of the cockpit applications discussed in the previous sections. Once we adopt this basic philosophy, we must consider certain implications:

(1) The unique characteristics of the machine; these characteristics will be examined under Equipment Considerations.

(2) Since a man will be operating the equipment, we must take into account man's particular capabilities and limitations; the particular attributes that a human operator contributes to the system will be explored under Human Factors Considerations.

(3) Since the pilot will be operating the system, all the factors unique to the airborne environment must be looked at; we will look at these under Pilot Factors Considerations.

1. EQUIPMENT CONSIDERATIONS

There are limitations to the capabilities of today's automatic speech recognizers. In this subsection we will discuss some of the problems that must be considered in conjunction with current state-of-the-art voice command systems. Specifically we will examine continuous vs. discrete speech, multiple vs. single speakers, vocabulary, trainability, ambient noise, and recognition accuracy

a. Continuous vs. Discrete Speech

The recognition of continuous speech by automatic means is beyond the capabilities of existing systems. As was pointed out above, the automatic speech recognizer works by comparing a particular utterance with a stored pattern in its memory. For discrete utterances, this is a relatively simple problem, once the pattern is stored. Because the acoustical qualities of discrete utterances are relatively constant over time, a particular word, e.g., "fuel" will match its stored pattern much more closely than it will match the pattern of another word, e.g., "checklist". However, with continuous speech, other problems arise.

Acoustically, continuous speech is not merely the stringing together of a series of discrete utterances. One of the major problems with continuous speech recognition is that the phoenemes, the basic distinguishable speech units of any language, do not remain invariant. Phonemes can vary depending upon the context in which they occur. The various forms of phonemes which result are called allophones. Basically, allophones result because of the different sounds which can precede or follow the phonemes. For example, the word "three" has the phoneme/i/ for the sound of the double e at the end of "three". In discrete words this sound, within the context of the word "three", will remain relatively invariant since the letters, t, h, and r always precede it and silence follows it. Let's now look at this same word in the context of continuous speech. The t, h, and r still precede the phoneme/i/ but the sounds that follow it could be any of the other 40 phonemes of the English language. If "three" were followed by the word "degrees," the allophone would be different from one resulting when followed by the word "left." The problems that the automatic speech recognition system has recognizing continuour speech are, therefore, of a magnitude higher than those involved in recognizing discrete speech.

b. Multiple vs Single Speakers

The heart of the automatic speech recognition system is the comparison of a given utterance with the pattern it has stored in its memory. If the machine is trained to recognize the voice of a single speaker, the pattern match and subsequent recognition rate will be quite good, regardless of the speaker's particular accent. As the machine is trained to accept more speakers, it will receive additional patterns, which are not exactly alike because of speaker differences, but which it has to recognize as the same word. If all the speakers have the same regional accent, e.g., Midwest American English, the patterns for a given word would remain fairly close; however, if speakers from different regions train the machine, recognition problems increase rapidly. For example, suppose the Midwesterner trains the machine on the word "data"; the final "a" sound is pronounced like the "u" in the word "up". The machine stores the pattern accordingly. Now let's have an Easterner say the word data; in most cases he will put an "r" sound on the end and "data" will be pronounced as "dater" (rhymes with later). Now the machine has two quite different patterns for the same word. It is easy to see that the recognition problems increase manyfold as new speakers, each with his own regional accent, are added to the system.

The conclusion to be reached about present systems is that as number of speakers the system must recognize goes up, recognition accuracy goes down. In addition, as the differences among speakers increase recognition rates also decrease.

c. Vocabulary

Vocabularies of very large size, e.g., 10,000 words are not practical with today's systems. In fact the capacity of most systems does not exceed a few hundred words. However, tying the recognition system into a large, general-purpose, digital computer would permit an increase in vocabulary size.

The main problem with large vocabularies has two aspects. The first is that the voice recognition machine's storage space for the acoustical patterns is limited. It cannot store many words before the system's storage capacity is exceeded. More storage capacity is one solution, but this quickly reaches a practical limit in terms of size and weight when considered in the context of an airborne avionics package. The second problem centers around similar sounding words as vocabulary size grows. Specialized vocabularies of a few hundred words can be constructed with a minimum of overlapping sounds. For example, if the words "send" and "end" appeared in the same vocabulary, "end" could be replaced by "terminate". But, for larger vocabulaties, alternate words which still make sense are difficult to find.

One approach to increasing vocabulary size is syntaxing, which imparts a certain order to the words. Through this ordering, the machine creates subsets of comparisons depending upon the preceding word. For example, if one purpose of our vocabulary was to provide for the change of radios through voice command, the vocabulary might contain the words "UHF", "one", "two" .... "nine", "zero" in addition to the other words that made it up. To change the UHF radio to a new frequency, the pilot might say, "UHF two, nine, one, six". The UHF would then change to the frequency 291.6. For the purposes of discussion let's say the vocabulary was 60 words in size. Through syntaxing when the machine heard the word "UHF", it would compare the pattern with the 60 stored patterns and recognize it. However, because "UHF" had been said, and because "UHF" is always followed by a digit, the machine would not have to compare the pattern for the first digit, e.g., "two" with 60 patterns. It would only have to compare it with 10 patterns (1 through 0), since the only words said after "UHF" are digits. This reduction in the number of pattern-comparisons greatly increases the effective size of the vocabulary; i.e., there may be 60 words in the vocabulary, but the machine does not have to concern itself with all 60 patterns all the time.

d. Trainability

The machine recognizes a word by comparing the utterance with a series of stored patterns and matches it to the pattern it most closely resembles. The pattern for a given utterance, e.g., the word "fuel", is formed by the speaker training the machine to recognize it. He does this by repeating the word a number of times. (Actually the machine forms a composite pattern composed of multiple utterances of the same word). The number of times the speaker must repeat a given word varies, but training repetitions per word generally range from 1 to 10. Training quickly becomes a factor if the vocabulary is of any size at all. The author had experience with Scope Electronics Voice Command System (training required five repetitions per word), and it is a very tedious task to train the machine on even a limited vocabulary of 10 words. Training time is a special problem in the USAF operational environment where maintenance prevents the pilot from flying the same aircraft each time, thereby requiring him to train the recognition system in each separate aircraft he files.

One solution to this problem is to tape the vocabulary on a cassette. Once an aircraft was assigned to the pilot, the crew chief could place the cassette in the aircraft's recognition system, and the system could be trained while the pilot was being briefed on his mission.

e. Ambient Noise

Since the machine in the listening mode can "hear" everything that is said, environmental noise must be taken into account. To pick out the pattern for a particular word, the machine must be able to hear it. In high ambient noise situations, this can be a problem. However, at least one manufacturer (Scope Electronics) claims to have a 99% recognition rate in an environment with an ambient noise level of 85 db (the sound of a passing truck or bus at 20 ft.) In addition, the use of noise suppression and directional microphones have reduced the effects of noise on the voice recognizer. To sum it up, it does not appear that ambient noise poses any special problems up to 85 db. Higher noise levels would have to be examined.

f. Recognition Accuracy

As is true for other things in this imperfect world, automatic voice recognizers are not 100% accurate. The recognition accuracy degrades as: (1) vocabulary size increases; (2) speaker differences increase; and (3) number of speakers increases. The basic question is, "How much is enough?" Obviously 100% would be ideal, but that is presently unattainable. The accuracy of today's systems, for a small number of speakers, varies from about 88% to 99%. It certainly seems feasible at this time to expect a recognition rate of approximately 95% for one or two speakers using a small vocabulary (e.g., 70 words).

To put this accuracy in perspective, remember that we certainly don't have 100% accuracy in the present way we manipulate aircraft subsystems. In other words, errors constantly occur in changing radio frequencies, setting up weapon systems, etc., when the pilot must push buttons, turn knobs and move selectors.

2. HUMAN FACTORS CONSIDERATIONS

In this subsection we will look at the factors that we must take into account because a man is in the loop. Even though these factors are spoken of in terms of pilots and cockpits, they are not unique to the airborne environment and would have to be considered in whatever environment the voice command system was to be used. Therefore, they are discussed here rather than under Pilot Factors, which deals with those factors that are unique to the airborne environment. The factors discussed here are vocabulary size, word arrangement, operator training, and accuracy requirements.

a. Vocabulary Size

Present voice recognition systems are limited to a vocabulary size of a few hundred words. In addition to this, we must define the minimal vocabulary size that the pilot can master in using voice command. This minimum will vary, of course, depending on the number and nature of the functions the machine is assigned. As the number of functions increases, obviously, the minimum vocabulary size will increase. This is true because the language used in the manipulation of various systems cannot be identical. Even adding one more communication radio (e.g., HF) to a system already set up for voice control of communications would require at least the name of the new radio.

Another factor which will affect minimum vocabulary size is similarity of the various systems we wish to control by voice. To use the example of the addition of the HF radio to the voice controlled communication system, we can see there is a great deal of overlap in vocabulary e.g., the digits 0 through 9 will be used for all the radios. But if we now add the weapon system, a relatively dissimilar system, we will have to add various weapon names (Phoenix), modes (air to air) and control words (select, arm, fire). The point is that the size of the minimum vocabulary will depend on what functions we wish to place under voice control. Once this has been decided we must be sure that the pilot has an adequate vocabulary to carry out the commands in a natural and easy manner. We certainly do not want to restrict vocabulary size such that the pilot must use combinations of words (e.g., "Phoenix execute" instead of "Phoenix fire") that are not natural and easy for him to use.

b. Word Arrangement

We must constantly remind ourselves that the system will work only if the pilot is willing to use it. In the past, there have been "hot shot" systems developed that were going to be the pilot's salvation, yet they ended up being deactivated. This could be due to a number of reasons (maintenance, reliability, etc), but one of the primary reasons is that the pilots considered them more trouble than they were worth. That is, they felt that the system added to, rather than decreased their workload. One of the principal factors in determining pilot acceptance of voice control could be word arrangement.

The words must be arranged in sequences that are easy and natural for the pilot to use and which would lead to a minimum or error. For example, the word arrangement should not be to first select the weapon by the words "Phoenix, select" and then arm the weapon by the words "arm, Phoenix," reversing the word order will lead to errors because of what is called negative transfer. This kind of problem must be avoided if the system is to be effective.

c. Operator Training

An untrained operator would find it difficult to operate the voice command system effectively. Training is necessary to optimize such things as pronunciation and determining inter-word interval. If words are slurred or not pronounced consistently, the recognition accuracy decreases. In addition, the words cannot be spoken continuously; there must be a finite interval (not less than 250 milliseconds) between consecutive words. These requirements dictate that the pilot will have to be trained in the use of voice control equipment to get used to both input and feedback sections, and learn the vocabulary so that it becomes second nature to him. If this amount of training appears excessive, remember that learning how to manipulate instruments and weapons systems is a significant portion of pilot training. The training time for voice command does not appear to affect overall training time to any great degree.

d. Accuracy Requirements

Recognition accuracy ranges from 88% for a completely untrained operator to 99% for a highly trained one. The question we must answer is, "What is the minimally acceptable recognition accuracy?" The answer is that it depends on the system. Ideally we would like 100% accuracy, but this, at present, is not possible. If one or more systems with multiple commands are to be operated by voice control, we must decide which commands would be most critical if they were not recognized correctly. The commands are ranked on a criticality basis, and those with the highest ranking would be given more training than those of lower ranking (as words are repeated [training], recognition accuracy increases). For example, if the command to change radio frequency were inserted incorrectly, it would certainly be a nuisance and possibly even dangerous, but not nearly so dangerous as the mininterpretation of "Phoenix, fire" during a combat situation. Therefore, in general, weapon system commands would be more highly trained than communication commands.

Within a given system, priorities can also be established. All commands to a weapon system may not have the same criticality. If "Phoenix, select" is misunderstood, the pilot would probably have adequate time to repeat the message; however, "Phoenix, fire" should be understood the first time, so it would receive more training.

3. PILOT FACTORS CONSIDERATIONS

It is often said that the pilot, and especially the fighter pilot, is a different breed of cat. This observation seems justified since there must be something special about people willing to learn the very complex task of flying jet aircraft and spending many hours in the artificial environment of the cockpit, which is certainly not a life of luxury. The unique characteristics of the pilot and his airplane affect the usefulness of the voice command system and, consequently, must be considered if the system is to be effective. The factors that we will examine are vocabulary composition, feedback. degraded modes, and environmental effects.

a. Vocabulary Composition

Besides negative transfer another fact we know about learning is that meaningful material is remembered more easily than "nonsense" material; basically, voice control requires the pilot to remember a series of words and phrases. With a vocabulary composed of words familiar to him, he is not likely to mix the words up or to waste valuable time thinking, "Now what words do I need to set up this thing?" The words would be written down somewhere in the cockpit, but having to read the words defeats one of the purposes of voice control -- keeping the pilot in a "head up" mode. Therefore, the word commands must be designed around the pilot's vocabulary for the particular situation so that he can transition to voice control with a minimum of effort.

b. Feedback

The central issue here is, once the pilot has spoken his command into the system, how does he know: (1) that his command is being executed, and (2) that it is being executed correctly. The pilot must be provided with some indication that his orders are being carried out. The two most feasible options are visual feedback on a Head Up Display (HUD) and aural feedback through the headset. These options are optimal because they allow the pilot to remain head up, which is one of the significant advantages of voice control. Which option to use would depend on the particular functions to be placed under voice control.

A HUD presentation offers the advantages of allowing the pilot to keep his eyes outside the cockpit and of not being hampered by environmental noise. It has the disadvantage of presenting yet another bit of information to be processed visually; there is a constant danger of presenting so much information on a HUD that the pilot's visual senses become overloaded. In addition, the pilot must look through the HUD and, consequently, he does not have total freedom of head movement. An aural presentation of feedback would serve to offload some information from the visual senses and permit unrestricted head movement. In addition, aural feedback can easily be arranged so as to prevent overloading the aural sense by establishing a message priority hierarchy, thereby allowing only one message at a time to come through his headset. The major disadvantages of an aural feedback system would be ambient noise and having to establish the message priority.

If the head up condition were not of primary importance, feedback could be readily provided by CRT. This means of feedback is superior to an annunciator panel since it provides considerable flexibility and takes up relatively little panel space. However the problem inherent in CRT presentation (e.g., washing out under high ambient illumination) must be considered. If for some reason a CRT presentation is not feasible, a series of annunciator lights could provide feedback.

c. Degraded Modes

What happens if the pilot gives the command "Weapon arm," and the system fails? The degraded mode issue must be considered very carefully before a voice control system can be introduced into the cockpit.

The most obvious answer to the degraded mode is to keep present systems intact and merely add voice command. This means that the pilot would still have at his disposal all the buttons, switches, dials, and knobs that he currently uses to set up the system. Then, in the event of voice control failure, he would merely have to resort to his old arming procedure. This approach is intuitively appealing but presents some serious problems, not the least of which is determining where in the arming sequence the system failed. Each voice command ("Phoenix, arm") triggers a number of steps; if the system failed after completion of several steps in the arming procedure, the pilot must be given some feedback to indicate which step failed. This information could be displayed on a CRT; it could flash the fail signal and then list the steps to be completed manually. As the pilot then flips the appropriate switches, the steps are compled and the item disappears from the CRT.

d. Environmental Effects

This problem deals with the effects of such things as noise and stress on recognition accuracy of the voice command system. Based on the data that Scope gathered in a study performed for the Navy, cockpit noise does not appear to be a problem so long as the pilot uses the microphone in his oxygen mask. Since the fighter pilot has his oxygen mask on for virtually his entire flight, ambient noise in the cockpit should not be a major problem.

Another environmental factor that must be considered is the effect of oxygen flowing through the mask on the speech recognition equipment. The oxygen flow does affect accuracy but, fortunately, it is a constant effect which can be filtered out. "The flow of oxygen into the pilot's mask is accompanied by a stationary acoustic spectrum that can be automatically discriminated from human speech. This is the only interferring effect noted in the data collected at various altitudes from 0 to 40,000 feet." (Reference 3. p.49). The conclusion regarding oxygen flow, therefore, seems to be that it does affect speech recognition, but the effect is constant and can be dealt with.

A third environmental factor is the forces on the pilot and, hence, the system. "Under the stress of forces, heavy breathing sounds are generated by the speaker. These sounds cannot be easily discriminated from human speech and thus must be accommodated in the recognition process" (Reference 3, p. 49). The stress effects pose a real problem which must be solved before speech recognition equipment can be used in operational fighter/fighter-bomber aircraft.

SECTION V

MANUFACTURERS AND GOVERNMENT RESEARCH IN AUTOMATIC SPEECH RECOGNITION

1. MANUFACTURERS

The following is a list of manufacturers currently working in the area of voice command. The format of the listing is as follows: (1) the name and address of the manufacturer; (2) the person contacted at that particular establishment; and (3) an estimate of their work to date.

The list is representative of those working in this area, not definitive. Time constraints prevented our contacting all the manufacturers working in the voice command area. Also, the contacts were made by telephone and consequently, the information is not nearly as detailed as a face-to-face visit might have provided. Each manufacturer was queried about a trainable 75 word, two-speaker, discrete speech recognizer.

Adaptronics Inc.
Westgate Research Park
7700 Old Springhouse Road
McLean, Virginia 22101

CONTACT: Mr. Lewey O. Gilstrap
(703) 893-5450

STATUS: Work is centered around the theory of automatic speech recognition. They are working primarily on the continuous speech problems, using a phoneme recognition technique. When asked about the required system, they said they could work out the basic designs in 9-10 months and a prototype in an additional 6 months. These requirements are the minimum for a practical airborne voice command system.

Boeing Aircraft Corp.
Seattle, Washington

CONTACT: Mr. Leo Hickey
(206) 655-4334

STATUS: The only work they are doing in this area is for FGR on the IPACSS and STOL Programs.

Culler-Harrison Co.
Goleta, California

CONTACT: Mr. Richard Levee
(805) 967-7613

STATUS: They presently have no marketable voice command system. They expect to have prototype hardware by 1 Jan 72. They are currently working on an eight-speaker phrase recognizer that does not require training. They said they could build a machine that would fulfill the basic requirements.

International Business Machines
Raleigh, North Carolina

CONTACT: Mr. G. L. Clapper
(919) 549-6385

STATUS: They presently have no commercially available equipment. They can build a system to meet the basic requirements if a contract is let.

McDonald-Douglas
St. Louis, Missouri

CONTACT: Mr. Toby Watkinson
(314) 232-6743

STATUS: They are not engaged in this activity at this time.

North American Rockwell
Columbus, Ohio

CONTACT: Mr. Ben Schohan
(614) 231-1851

STATUS: They are not engaged in this activity at this time.

Perception Technology
Winchester, Mass.

CONTACT: Dr. Beninghof
(607) 729-0110

STATUS: They currently have, in the breadboard stage, a discrete word recognizer giving 95-98% accuracy. One very positive feature is that it is speaker independent; anyone off the street can use it without first training the machine to accept his voice. The negative aspect is that its vocabulary is presently limited to 19 words.

Philco-Ford Corp.
Willow Grove, Pa.

CONTACT: Mr. Charles Teacher
(215) 659-7700

STATUS: They presently have no commercially available equipment. They could build a system to meet the basic requirements in 6 to 12 months.

Radio Corporation of America
Camden, New Jersey

CONTACT: Mr. Jerry Richards
(609) 963-8000

STATUS: They presently have no commercially available equipment. They said they could build a prototype to meet the basic requirements in 6 months.

Scope Electronics
1860 Michael Faraday Drive
Reston, Virginia 22070

CONTACT: Mr. John Steves
Mr. Mike Hitchcock
(703) 471-5600

STATUS: They have a commercially available Voice Command System which may meet the basic requirements.

Threshold Technology, Inc.
Cinnaminson, New Jersey

CONTACT: Mr. Marvin Herscher
(609) 829-8900

STATUS: They have a voice command system (VIP 100) which should be available in the near future. This package may be able to meet the basic requirements.

2. GOVERNMENT AGENCIES

AIR FORCE

Air Force Aerospace Medical Research Laboratory
WPAFB

CONTACT: Dr. Hans Oestreicher
255-3603

Dr. J. Ryland Mundie
255-3673

STATUS: They have two operating automatic speech recognizers, Dr. Oestreicher's Astronaut Maneuvering Unit (AMU) and Dr. Ryland's package which works on a pitch pattern classification scheme. The AMU is a rather primitive device which was merely designed to demonstrate the feasibility of voice command - and it does that. Dr. Ryland's device is very interesting and seems to have potential, but it presently is at the prototype stage.

Air Force Avionics Laboratory, WPAFB

CONTACT: Mr. John Mysing
255-3826

STATUS: Previous work involved a contract with RCA to develop a speech-controlled radio-channel selector. The package was very noise-sensitive. Currently working with Scope on the effects of stress on speech recognition. Supporting the Navy in their work on stress and g-forces affecting speech recognition. Currently setting a one-year target date for an airborne speech recognition system.

Rome Air Development Center
Griffiss AFB

CONTACT: Dr. Bruno Beek
Mr. James Brech
88-587-4150 (autovon)

STATUS: Their work is centered around the use of automatic speech recognition in speaker identification, and appear to be very competent in this area. Their lab contains a prototype Voice Information Processor (VIP-100) from Threshold Technology, Inc., as well as a number of other speech recognition devices.

GOVERNMENT AGENCIES

ARMY

U. S. Army Electronics Command
Ft. Monmouth, N. Jersey

CONTACT: Mr. J. Pruesse
(201) 535-1952

STATUS: Their work centers primarily around the speaker identification and verification problem, very close to the work being done at RADC. They have not been concerned with g-forces, etc; however, they do have a contract with Signatron to investigate the effects of psychological stress on speech recognition by hypnotizing vietnam veterans and recreating combat situations for them. While hypnotized, the veterans will read the phonetic alphabet into an automatic speech recognition system.

NAVY

Naval Air Systems Command
Washington, D. C.

CONTACT: Hr. Howard Arnoff
(202) 692-7487

STATUS: A program at the Naval Air Development Center, Warminster, Pa., will place Scope's Voice Command System in a similator at the end of a centrifuge and test its performance under various g-forces, vibration, buffeting and temperature. The program will start in the near future.

DoD

Advanced Research Projects Agency, DoD.

CONTACT: Dr. Allen Newell
Carnegie Mellon Univ; Pittsburg, Pa.
(412) 612-2600

STATUS: This agency funded an ad hoc committed, chaired by Dr. Newell, to investigate the feasibility of designing a sophisticated speech command system by 1973 and a refined model by 1976. Work has begun, and an advisory committee, headed by Dr. J. C. R. Licklider, is coordinating the effort.

SECTION VI

ANALYSIS OF THE POTENTIAL OF VOICE COMMAND

This section deals with the analysis of crucial issues pertaining to how voice command can be introduced into the cockpit in such a way that it will be useful. There are many aspects to this issue, some of which will be examined in this section:

a. Where the voice command concept stands.
b. Functional specifications of an airborne voice command system.
c. Progress toward an airborne voice command system.
d. Setting the development strategy.
e. What not to do in implementing voice command.
f. A research program for implementing voice command.

1. WHERE THE VOICE COMMAND CONCEPT STANDS

Basically the state-of-the-art can be stated as follows: *We are on the threshold of having the technology in hand for applying the voice command concept to Air Force cockpits, if we are willing to accept certain restrictions.* Many of the restrictions are based on the equipment the author has had access to: Scope's and a prototype VIP 100 at RADC. Discussions at RADC indicated that a number of specialized pieces of equipment may eliminate one or more of the restrictions, but not a significant number of them. This specialized equipment presents its own unique problem (e.g., excessive computer size).

The restrictions, paradoxically, can work to our advantage in the application of voice command to the cockpit. Our knowledge of how voice command can be used in the cockpit is still rather limited. Thus, the limitation of the hardware prevents our being overwhelmed by the scope of all the possible applications. Present voice command systems will not operate on continuous speech, large vocabularies, or a large number of speakers; thus, the number of possible applications is limited. This forces us to consider voice command in very specific circumstances; e.g., as the weapon system of a single-seat, fighter aircraft. This is a bonus because it restricts us to a problem we can "get a handle on." We will avoid falling into the trap of having the capabilities of the system far exceed our knowledge of the particular problems we are trying to solve.

The following restrictions limit the capacity of a voice command system at the present:

a. Laboratory Environment

The present equipment has not been tested for its adaptability to stress, temperature, or other factors encountered in the airborne environment. Therefore, present equipment is strictly ground-based. This is not to say, however, that current voice command systems could not be packaged for use in aircraft.

b. Limited Vocabulary

The current vocabularies are limited in two ways: (1) size is limited to one or two hundred words: and (2) composition is also limited since similar sounding words ("send" and "end") decrease recognition accuracy significantly.

c. Discrete Words

Continuous speech recognition is not within the scope of today's equipment. The equipment that the author has examined limits specific words to a one-second length, with a minimum between word interval of 250 milliseconds. Interestingly, most of the researchers have given up on discrete speech (for lack of application) and have moved on to continuous speech.

d. Trainability

Most machines require a number of repetitions of each word before the system will recognize it. The number of repetitions varies, but a representative range is between five and ten per word. Some specialized systems do not require training, but they have their own unique problems (e.g., low recognition accuracy).

e. Limited Accuracy

Accuracy ranges from around 88% to 98%, depending on the operator's familiarity with the system. It appears that with current equipment a highly trained operator can achieve recognition accuracies within the range of 93% to 98%. To attain the higher accuracy ranges, a rejection mechanism seems to be necessary, which works as follows. If the system cannot recognize a particular word, it does not try to match it to the closest word pattern available, but rejects it as nonsense and requires the operator to repeat the word. This greatly reduces the number of incorrect pattern matches and therefore errors.

f. Noise Elimination

Voice command systems cannot operate in a high noise environment (90db or higher); steps must be taken to reduce the amount of noise entering the machine. A number of methods have been devised. In fighter aircraft, using the microphone inside the oxygen mask seems to be effective. Where oxygen masks are not used (Forward Air Controller), noise suppression microphones may be feasible.

2. FUNCTIONAL SPECIFICATIONS OF AN AIRBORNE VOICE COMMAND SYSTEM

This subsection deals with the characteristics that the system must have in order to function in the cockpit at a minimal level. These specifications are:

(1) Discrete speech,
(2) One or two speakers,
(3) 100-word vocabulary,
(4) Trainability,
(5) Oxygen mask microphone inputs, and
(6) Cockpit environment.

These are the minimum specifications for an airborne voice command system; they do not describe an optimized system. The purpose of implementing this basic system would be to "get us off the ground" in voice command; the system's growth potential could be incorporated at a later date. This system can be put into an aircraft in the near future; a more elaborate system can come later.

a. Discrete Speech

It is customary (and even encouraged) for the pilot to speak in short, discrete phrases in performing his communications tasks. Therefore, from the pilot's point of view, requiring him to use discrete rather than continuous speech for the voice command system poses no special problems. In addition, discrete speech appears to be adequate for performing voice command functions, e.g., setting up a weapons system.

b. One or Two Speakers

Initial application of voice command will provide the most payoff in the single-seat aircraft. Because only one man is in the cockpit, he has a tremendous number of manual chores (switch throwing, knob turning) as well as complex decision-making tasks. These pilots have to work hard to get the mission accomplished, a fact verified by data collected in SEA. Therefore, the single-seat aircraft in general and the fighter/fighter-bomber in particular (because of the weapon system complexity) appears to be a logical starting point for application of voice command. Mission requirements in two-seat aircraft often tax the crew to their limit. Data from SEA show that the crew of two-seat aircraft have a great number of tasks to do in accomplishing their mission. Voice command could certainly play a vital role in these aircraft also. Since the most payoff would be gathered from the application of voice command to the single or two-seat fighter/fighter-bomber, the system would have to recognize at most two speakers during any given time period.

c. 100-Word Vocabulary

The airborne voice command system will be required to recognize a very specific set of words which are unique to the functions to be placed under voice command. The few words needed to make the system function are the <u>only</u> words it must recognize. Since the number of functions placed under voice command initially will probably be limited to one or two, a vocabulary of 100 discrete words appears to be sufficient to significantly reduce the manual switching tasks.

d. Trainability

The system must be able to be trained to recognize the unique voice characteristics of the pilot and/or copilot. The accuracy of automatic speech recognizers that do not require training is currently much too low to be of practical value in the cockpit. The training is accomplished by the pilot and/or copilot repeating each word in the vocabulary a given number of times.

e. Oxygen Mask Microphone Inputs

The system must be able to accept voice inputs from an oxygen mask microphone of the type employed in fighter/fighter-bomber aircraft. In a FAC aircraft where a boom microphone is used, the voice command system will have a more difficult time to recognize commands because of higher ambient noise. However, since fighter and fighter/bomber are the prime areas of concentration, the oxygen mask microphone is an adequate initial specification.

f. Cockpit Environment

The system must be able to function accurately and reliably within the temperature, stress, and noise levels of the airborne environment. This also implies that it must be able to recognize words despite the effects the airborne environment has on the pilot's voice, e.g., distortion under g-forces.

3. PROGRESS TOWARD AN AIRBORNE VOICE COMMAND SYSTEM

The Information Processing Techniques Office of the Advanced Research Projects Agency (ARPA) of DoD recently sponsored an ad hoc study group to determine the feasibility of developing an automatic speech recognition system to the following specifications. The initial specification describes a system that can be demonstrated in 1973; the final specification describes a system to be demonstrable in 1976. The contracts to begin this work have been let, and an advisory group, headed by Dr. J. C. R. Licklider, has been set up to monitor progress and coordinate the various research efforts. Periodic progress reports will be made.

	Initial Specifications (1973)	Final Specifications (1976)
	The system should:	The system should:
(1)	accept continuous speech	accept continuous speech
(2)	from many	from many
(3)	cooperative speakers	cooperative speakers of the general American dialect
(4)	in a quiet room	in a quiet room
(5)	over a telephone	over a good quality microphone
(6)	allowing moderate tuning of the system per speaker	allowing slight tuning of the system per speaker
(7)	but requiring only natural adaptation by the user	but requiring only natural adaptation by the user
(8)	permitting a vocabulary of 10,000 words	permitting a slightly selected vocabulary of 1,000 words
(9)	but with strong syntactic	with a highly artificial syntax
(10)	and semantic support	and a task like the data management or computer status tasks (but not the computer consultant task)
(11)		with a simple psychological model of the user
(12)		providing graceful interaction
(13)	tolerating less than 10% semantic error	tolerating less than 10% semantic error
(14)	in a few times real time	in a few times real time
(15)	on a dedicated system with 10^8 instructions per second	

(16)*	(16)
(17)*	(17)
(18)*	(18)
(19) be demonstrable in 1973 with a moderate chance of	(19) be demonstrable in 1976 with a moderate chance of success

*Parameters not specified

The interesting thing about the specification is that even the 1973 system encompasses most of the requirements needed for implementation in USAF aircraft: discrete speech, one or two speakers, 100 word vocabulary, trainable, oxygen mask microphone inputs, and cockpit environment. For example, ARPA Requirement 1 - Accept continuous speech, goes beyond the discrete speech requirement that we need. Requirement 2 - From many speakers, - exceeds the requirements of the one or two speakers needed for the fighter/ fighter-bomber aircraft. Requirement 5 - Over a telephone, dovetails quite nicely with the airborne requirements, since microphones in today's aircraft have at least as good transmission qualities as do today's telephones. Requirement 6 - Allowing moderate tuning of the system per speaker - fulfills the trainability requirements. Requirement 8 - Permitting a vocabulary of 10,000 words - far exceeds the 100 word vocabulary we require.

There are two specifications that do not meet the requirements for the cockpit environment: Requirement 4 - In a quiet room; the noise level in a cockpit is certainly above this, but noise-suppression microphones may be able to get around this restriction. Requirement 13 - Tolerating less than 10% semantic error; this seems to present problems, although we are concerned with errors at the syntactical level, a much less sophisticated level than the semantic. Therefore, this may not be a serious stumbling block.

Overall it appears as if the 1973 system could meet the requirements of the airborne environment (given that the system can be packaged to fit into, and function in, fighter/fighter-bomber aircraft). It offers promise of a very effective voice command system.

4. SETTING THE DEVELOPMENT STRATEGY

In the cockpit of aircraft designed for the year 2000, all the functions may be hooked into a voice command system. For the near future, however, this is not practical, and is probably not desirable, either. Therefore, for aircraft designed for the near future, we face the problem of what functions to place under voice command. Section III discussed possible cockpit applications of voice command, but did not assign any priority as to which functions could most benefit from it. In deciding which of the tasks should be assigned, there are a number of things to keep in mind.

The overriding consideration is always what will benefit the pilot most. Therefore, pilot inputs as to the worth of new crew station technology are highly valuable. In order to obtain these inputs in relation to voice command, pilots at the Air Force Flight Dynamics Laboratory, the Instrument Flight Center, and the Fighter Weapons School have been briefed on the concept and have expressed the belief that voice command has potential value as a new control device within the crew station. Once this is established as the guiding principle, the next step is to ask, "What are the key systems that enable the pilot to do his job?" A corollary to this is, "Are these also the systems which contribute most to the work the pilot must perform?" The SEA data can be used to help answer these questions; this data base is very large and covers comments by operational pilots functioning in a combat situation. The next step is to decide, on the basis of this information, the system or systems to be placed under voice control. The third step is to construct the vocabulary and test it; this involves deciding what words to include, the arrangement of the words, and the "naturalness" of the words when spoken by the pilot. The testing would involve a simulated mission in which pilots employ the particular vocabulary decided upon in order to optimize it.

Let us take an example to show how this process would work for fighter/fighter-bomber aircraft. The overall guideline is "What is the job of the pilot in this type aircraft?" His primary role is to destroy either airplanes or ground targets. Now the question is, "What is the key to accomplishing his job?" The answer to this would obviously be his weapon system. Even if he has perfect flight control, communications, etc., he could not fulfill his primary job if his weapon system were defective. Therefore, this is the first area to concentrate on. If an examination of this system reveals that there are no particular problems, the next step would be to proceed to the next most important system, and so forth. Let us suppose that there are problems connected with the weapon system. The next question is, "How can voice control be used to eliminate the problems?" One way could be by automating many of the tedious steps currently performed by the pilot. The final step would be constructing the vocabulary and testing the system.

5. WHAT NOT TO DO IN IMPLEMENTING VOICE COMMAND

One way of getting voice command into the cockpit is to buy a system, plug it into the aircraft, and see how it performs. This approach will not work for a number of reasons. First, we lack an adequate data base of voice characteristics of pilots in the airborne environment. From the small amount of data that is available, we know that it is much more difficult to recognize words spoken in this environment. Therefore, if an existing system is purchased and placed in the aircraft, we will have to trust to the competence of the machine to deal with the distortions imposed on spoken words in the airborne environment. Given the present state-of-the-art in automatic speech recognizers, their ability to handle these distortions is doubtful.

The second, and probably most important reason is that it shows a lack of concern for the pilot. The pilot must use the equipment if it is to be successful. Ultimately, he is the one to determine whether voice command is to be effective or not. No matter how much "scientific" data is furnished to show what a great breakthrough voice command is, if the pilot finds it cumbersome to use or is really not convinced of its worth, the system is doomed. Therefore, before equipping aircraft with the voice command system, we must do some very careful groundwork. This will be discussed in the next section.

The price of not doing this groundwork will be another engineering success and operational failure. This will benefit no one. In fact, it may actually detract from the pilot's efficiency by adding a "nuisance" item that he must pay attention to but which really doesn't aid him in completing his mission.

6. A RESEARCH PROGRAM FOR IMPLEMENTING VOICE COMMAND

The following three-phase research program is proposed as a means of examining the application aspects of an airborne voice command system:

(1) PHASE I — Analysis of the functional requirements of voice command.

(2) PHASE II — Optimization of the Pattern Recognition Algorithm.

(3) PHASE III — Inflight Investigation and Evaluation of the Voice Command Concept.

a. The Research Program

PHASE I — Analysis of the Functional Requirements of Voice Command

In Phase I, the factors which are directly related to the application of voice command will be examined in detail, including aircraft, mission, and pilot characteristics. The resultant data can be used to determine whether or not the voice command concept answers the real world needs. We recommend that AFFDL assume prime responsibility for accomplishing this step, since this laboratory is applications oriented. This phase of the program has four parts:

(1) Define the Environment The first step is to determine the overall goal of the work. This would require us to study the application of voice command in the airborne environment for fighter/fighter-bomber (possibly FAC)

aircraft for the near future (1980-85) time frame and define in as much detail as possible the functions of the pilot (or pilots) of this aircraft.

(2) Assign Functions to Voice Command Following the guidelines discussed previously, decide which functions are to be voice command.

(3) Construct the Vocabulary The particular words the pilot will use in conjunction with the systems assigned to voice command must be incorporated into a vocabulary, keeping in mind such factors as size (approx. 100 words) and composition (avoid like sounding words, e.g., "send" and "end").

(4) Specify Functional Design With the information generated by preceding steps, conduct tradeoff studies in terms of design alternatives for determining those system requirements which best meet the application criteria. Consideration of the state-of-the-art, and projections for further developments will be taken into account to ensure an orderly growth of the capability over the life cycle of technological development.

PHASE II Optimization of Pattern Recognition Algorithm.

Once the characteristics of the mission are known and the vocabulary has been developed, the next step is to purchase equipment and optimize the pattern recognition algorithm. During this phase, AFFDL would enter into cooperative efforts with other organizations such as Air Force Avionics Laboratory and/or RADC to deal with optimization. However, AFFDL/FGR would still remain as the focal point for supplying the application type inputs. A general outline of how the optimization might be accomplished is given below.

(1) Purchase Hardware. Two alternatives are available to accomplish this step: (a) purchase "the front end" (spectrum analyzer, multiplexer program, minicomputer). The software can be changed to allow for optimization of the algorithm, and the minicomputer allows for a minimum of packaging problems for flight testing. The drawback to this approach is that either the minicomputer must be programmed in machine language, a difficult and tedious task, or software must be developed for an intermediate language, which is expensive. (b) This is like the first alternative but a medium-sized, laboratory-type, general-purpose digital computer would be substituted

for the minicomputer, which would allow programming in a readily available intermediate language which is easy to use and known to many engineers and behavioral scientists. The drawbacks are the expense and the difficulty in packaging for flight testing. Of the two alternatives, the second appears the more feasible for conducting the in-house program.

(2) Test Recognition Algorithm It is unlikely that the algorithm will be optimized initially. The testing will involve an iterative approach and will be repeated until the desired degree of algorithm optimization has been achieved.

PHASE III Inflight Evaluation and Investigation of Voice Command Concept.

Since the ultimate goal of the voice command research program is to have a viable airborne voice command system, the effectiveness of the concept must be evaluated in the inflight environment. There are two steps to this phase.

(1) Package Optimized Voice Command System for Airborne Testing. Once the voice command algorithm has been optimized, the system must be packaged small enough to fit into fighter/bomber aircraft and function adequately despite the rigors of the in-flight environment.

(2) Demonstrate Inflight. The final step is to operate the optimized voice command system in the aircraft. Any problems arising during the flying would then be eliminated through further testing and development.

b. The Frequency Distortion Problem

The airborne environment is a hostile one as far as voice command is concerned. Airborne conditions cause frequency distortions in spoken words; also, breathing patterns add noise to the system. These frequency distortions must be eliminated if the voice command concept is to function in this environment.

Two current programs, one by the Navy and the other by the Air Force Avionics Laboratory are being conducted to examine the frequency distortion problem; however, both of these programs are ground-based, utilizing simulators. Use of the simulators approximates but does not duplicate the airborne environment; therefore, the data obtained from these programs may not be representative of the frequency distortion problem in flight.

The application of voice command to aircraft depends upon being able to handle the frequency distortion problem. Certainly any applications work is dependent upon the confidence that voice command is going to function adequately in flight. We need in-flight data to use in analyzing the frequency distortion problem, and this analysis can then serve as the basis for a go, no-go decision on whether to proceed with the program.

The organizations conducting the current work efforts apparently do not have access to the in-flight environment; to correct this deficiency, the Air Force Flight Dynamics Laboratory could undertake an in flight program immediately, before Phase I of the program is started. Because of its relationship with the Instrument Pilot Instructor's School (IPIS) AFFDL/FGR is in a position to make a significant contribution to study of the frequency distortion problem. A program conducted in conjunction with IPIS, utilizing their aircraft, could obtain a data base quickly and inexpensively, in the in-flight environment where voice command must function. This effort could provide the confidence needed to proceed with the applications program. Further study and ultimate resolution of the problem, of course, would rest with those responsible for the development of the system. If such a plan is adopted, the study could be conducted as follows:

(1) Obtain the Data Base (Recordings). A sample vocabulary developed during a Navy research effort in voice command would be used. IPIS would furnish the aircraft and pilots, while AFFDL would coordinate on conducting the experiment. Recording equipment would be furnished by AFFDL.

(2) Analyze the Recordings. Determine the quantitative characteristics of frequency distortion resulting from stress so that an optimum design can be produced. This might be accomplished in one of several ways: (a) take the tapes to RADC and work with RADC in the analysis, Utilizing their laboratory; (b) take the tapes to competing manufacturers for analysis, or (c) perform it in-house, but this would require the acquisition of equipment (e.g., spectrum analyzer). AFFDL, in all probability, would not perform the work in-house but would launch a cooperative effort with other agencies to perform the analysis of the recordings.

SECTION VII

SUMMARY

INTRODUCTION

Today's aircraft, as exemplified by the F-111, are a far cry from the Spad of WWI fame; the performance capabilities of the F-111 exceed those of the Spad by an order of several magnitudes. Unfortunately, coupled with this increase in performance capability has been a corresponding increase in the number of subsystems, e.g., Terrain Following Radar, needed to provide for this augmented performance capability. As subsystems proliferate so do the controls, displays, and procedures which accompany them. The ultimate result of all this is that aircraft have become increasingly more sophisticated in terms of number and kinds of subsystems, and accompanying controls/displays and procedures resulting in more things for the pilot to do. Consequently, the pilot has become saturated with the manual manipulation of knobs, switches and buttons. Even with the outstanding training today's pilots receive, we have reached a limit on the number of subsystem manipulations we can require them to perform.

In order to alleviate this problem of having the pilot overwhelmed with manual tasks, the power of the computer must be brought into play. By automating many of the manual sequences, the pilot is freed to fulfill his primary role as a decision maker. The long suit of the man in the cockpit is his ability to act as an integrator and manager of all the information presented by the various subsystems. It is certainly a waste of his unique capabilities to use him as a simple switch thrower. The computer holds potential for freeing him from much of the routine and repetitive tasks required in present day cockpits.

Before the computer can be used effectively, a means must be provided for interfacing with it. Voice command is ideally suited for this purpose. Utilizing voice command the pilot would merely have to "tell" a particular subsystem what was needed, and the rest of the task would be carried out automatically by the computer. Through the voice command computer combination, the pilot will be able to fulfill his true role as the manager of his crew station.

The purpose of this report is to discuss the many factors that must be considered before voice command and its application to Air Force cockpits can become a reality. Specifically the report will discuss:

1) How Voice Command Works
2) Possible Cockpit Applications of Voice Command
3) Voice Command System Considerations
4) Manufacturers and Government Research
5) Analysis of the Potential of Voice Command
6) Voice Command System Development
7) Research Program for Airborne Voice Command
8) Conclusions

HOW VOICE COMMAND WORKS

The heart of the voice command system is the automatic speech recognizer. This device, actually a series of devices, takes the complex acoustical waveform issuing from the mouth of the speaker and processes it in such a way so as to "recognize" the various phrases uttered by the speaker. There are a number of ways to perform the automatic speech recognition task. Since discussion of all the speech recognition options is beyond the scope of the report, a representative system is examined. The system, as designed by Scope Electronics Inc., can be represented by the following block diagram.

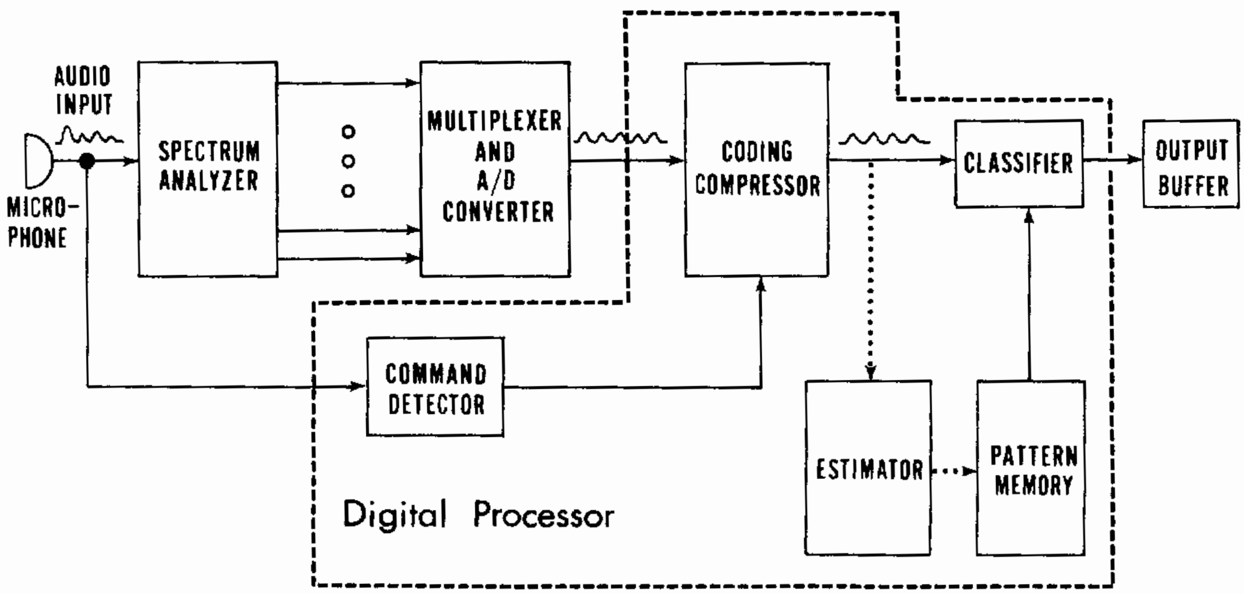

Fig 6. Block Diagram of an Automatic Speech Recognition Device. (From Glenn & Hitchcock, 1971, p. 86).

The Scope device is essentially an acoustic pattern classifier that produces a BCD output in response to a spoken command. The system consists of an analog signal processing subsystem, multiplexer, analog/digital converter, digital processor and an output register. The blocks in the above diagram labeled command detector, coding compressor, classifier, estimator and pattern memory are actually functions of the digital processor which consists of CPU, Read Only Memory (ROM) and Core Memory.

The system has a maximum vocabulary of 24 discrete words of up to one-second duration. A pause of at least 250 milliseconds is required between words. Adaptation, or training, to each of several speakers is necessary. This is accomplished by having each user repeat each word 5 times; the average acoustic pattern is stored in the digital processor core memory for comparison during pattern recognition.

The combination of the spectrum analyzer and multiplexer/A-D Converter performs the function of feature extraction, or selection of significant characteristics from the total acoustic content of a spoken word. The characteristics then are used to construct a representative voice pattern.

POSSIBLE COCKPIT APPLICATIONS OF VOICE COMMAND

The Air Force's main interest in voice command is in its possible benefit to the pilot. Therefore, the heart of the issue comes down to cockpit applications. Voice Command holds potential for various applications in the cockpit, such as weapon delivery, communications, emergency procedures, checklists and system status.

Weapon Delivery. As the data gathered in SEA reveals, the weapons systems of many of today's fighter/fighter-bomber aircraft are complex and time consuming to set up. Voice command could play a key role in alleviating some of the pilot's work in this area. Besides eliminating the need for the pilot to fly "heads down" while trying to deal with the knobs and buttons required to set up the system, the use of voice command would also allow the pilot's hands to be free for other tasks. In addition, the number of steps needed by the pilot to set up the weapon system could be greatly reduced.

Communications. A second area in the cockpit where voice command could be applied is communications. The pilot is forever talking to someone; it makes no difference if he is in the U. S. or Vietnam. It certainly appears as though voice command could be used quite effectively for frequency change purposes. As was illustrated earlier, if the pilot merely had to call the particular radio (e.g., UHF comm., VHF comm) and then state the new frequency, his workload would be alleviated tremendously. The advantages of voice command that were cited for weapons systems also apply to radios. The pilot can fly heads up and still use his hands for other tasks.

Emergency Procedures. This is a third area where voice command could be employed. Voice command could fit into these current emergency procedures in the following way. When the failure occurs, e.g., DC generator overheats and the annunciator light comes on, the pilot, after looking at the emergency procedure in the checklist, would decide either to ignore the light or execute the procedure. If he decided to ignore the light, he would say nothing; if he decided to execute the procedure, he would merely say "DC Gen Oheat Execute", and the procedure would be completed automatically. An END annunciator light would tell him when the procedure was completed.

Checklists. Closely akin to emergency procedures are the checklists which are carried out in the course of normal aircraft operation. Checklists are currently accomplished in one of two ways. In single placed aircraft, the pilot both reads and carries out the checklist items. In two placed (or larger) aircraft, the "call and challenge" method is used. One of the crew members reads the item and another crew member completes the item and responds that it is completed. The voice command system could serve as a second crew member for the pilot in a single placed aircraft. As the pilot reads the item into the voice command system, the system would carry out the item and provide feedback to the pilot when the item was completed.

System Status. The final area of voice command application that the report considers is system status. In the course of his mission, the pilot has a continuing need to check on the status of his aircraft's systems. Besides routine checks on such things as fuel flow and engine rpm, he also has need for information relative to the combat situation. To obtain this information, he would request it, and the information would be presented to him either aurally or on a HUD. An additional feature could be the continuing presentation of the information.

VOICE COMMAND SYSTEM CONSIDERATIONS

A vital aspect of the cockpit applications just discussed is the man-machine relationship. The man-machine system consists of three essential subparts: equipment factors, human factors, and pilot factors.

Equipment Factors. The equipment factors concern the unique characteristics of the machine components of the system. There are limitations to the capabilities of today's automatic speech recognizers. The recognition of continuous speech by automatic means is beyond the capabilities of existing systems. A second limitation of the machine is that its recognition accuracy degrades as a function of the number and dissimilarity of the speakers it must recognize. Automatic speech recognizers also are limited as to the size of vocabulary. In fact the capacity of most systems does not exceed a few hundred words. Finally, high ambient noise (above 90 db) impedes the performance of the machine unless remedial measures are employed.

Human Factors. The second system consideration aspect that must be taken into account is human factors. These must be considered because a man is in the loop. Just as the machine has problems, so does the man. As far as vocabulary is concerned, the man is not concerned with the upper limit (ideally the machine should recognize the 100,000 or so words in the man's vocabulary) but rather with a minimum vocabulary. Because the machine can recognize so few words, the problem is that of defining the minimal vocabulary size necessary for the pilot to utilize voice command in the cockpit. This minimum, of course, will vary depending on the number and nature of the functions the machine is assigned.

A second factor that must be considered because a man is in the system is that of word arrangement. It is critical that the words are arranged in sequences that are easy and natural for the man and which lead to a minimum of error. We would not have a word arrangement that first selected the weapon by the words "Phoenix, select" and then armed the weapon by the words "arm, Phoenix". From what we know about the population sterotypes of pilots, we can predict that reversing the word order will lead to errors because of what is called negative transfer. Reversing the word order is the kind of problem we must avoid if we want the man to use the system effectively.

Since men are not born with innate knowledge, operator training must be considered. An untrained operator has a difficult time operating the voice command system accurately. Training time is necessary in order to optimize such things as pronunciation and inter-word interval. If words are slurred or not pronounced consistently, the recognition accuracy decreases. In addition, the words cannot be spoken continuously; there must be a finite interval (not less than 250 milliseconds) between consecutive words.

The accuracy requirements needed for the man to operate the system effectively are a fourth topic in the human factors area. The question to be answered is "What is the minimally acceptable recognition accuracy for cockpit use?" The answer is that it depends on the system we wish to place under voice control. Ideally we would like 100% accuracy on all words, but this, at present, is not possible. If either a single system with multiple commands or a number of systems with multiple commands are put under voice control, we must decide which commands result in the largest system degradation if they are not recognized correctly. Once the commands have been ranked on a criticality basis, those with the highest ranking would be trained more thoroughly than those of lower ranking. (Remember that as words are repeated more often [training], recognition accuracy increases).

Pilot Factors. The unique characteristics of the pilot and his airplane impinge upon the usefulness of the voice command system and, consequently, must be considered if the system is to be effective. These characteristics are discussed under the heading of pilot factors. The first characteristic considered is vocabulary composition. Besides negative transfer another fact that is known about learning is that meaningful material is remembered a lot better than "nonsense" material, and basically, when using voice control in the cockpit we are asking the pilot to remember a series of words and phrases. This means that with a vocabulary composed of words that are familiar to him it is much less likely to have the situation in which he wastes valuable time thinking, "Now what the hell were the words to set up this weapon system?" Or to have the situation where he gets the words mixed up.

The second issue crucial from the pilot's point-of-view is feedback. The central issue here is, once the pilot has spoken his command into the system, how does he know: 1) that his command is being executed and 2) that it is being executed correctly. The pilot must be provided with some indication that his orders are being carried out. The two most feasible options are visual feedback on a HUD and aural feedback through the headset. The reason why these options are optimal is that they allow the pilot to remain heads up, one of the significant advantages of voice control. Which of the two to use would depend on the particular functions to be placed under voice control.

Since equipments do not operate with 100% reliability, the third factor that the pilot must consider in conjunction with voice command is degraded mode operation. What happens if the pilot is ready to arm his weapon by means of voice command and the system fails? This is the heart of the degraded mode issue and must be considered very carefully before a voice control system can be introduced into the cockpit. The most obvious answer to the problem is to keep present systems intact and merely add voice command to them. This means that if we consider the weapon system, the pilot would still have at his disposal all the buttons, switches, dials, and knobs that he currently uses to set up the system. And in the event of voice control failure he would merely have to resort to his old arming procedure. This approach is intuitively appealing but presents some serious problems, not the least of which is the pilot's determining where he is in the arming sequence. Remember that each voice command "Phoenix, arm") triggers a number of steps to occur. If the system failed after completion of several steps in the arming procedure, the pilot must be given some kind of feedback to indicate which step failed.

Environmental factors constitute a fourth crucial issue with which the pilot must be concerned. This issue deals with the effects of such things as noise and stress on the voice command recognition accuracy. Based on the data gathered in a study performed by the Navy, cockpit noise does not appear to be a problem so long as the pilot uses the microphone in his oxygen mask. Another environmental factor that must be considered is the effect of oxygen flowing through the mask on the speech recognition equipment. The oxygen flow does affect accuracy but, fortunately, it is a constant effect which can be filtered out. A third environmental factor is the effects of g-forces on the pilot and, hence, the system. The stress effects pose real problems which must be solved before speech recognition equipment can be used in operational fighter/fighter-bomber aircraft.

MANUFACTURERS AND GOVERNMENT RESEARCH

There are a number of manufacturers and government organizations engaged in automatic speech recognition research. On the industrial side the leader appears to be Scope Electronics, which has developed and is currently marketing a discrete word recognizer. Threshold Technology, Inc. may also have a promising discrete word recognizer, but it is presently still in the prototype stage.

On the government side the Army, Navy, Air Force, and Advanced Research Projects Agency (ARPS) of DoD are all currently engaged in automatic speech recognition research. Of these programs the ones most directly related to voice command are those conducted by the Naval Air Systems Command, and the Air Force's Avionics Laboratory. These programs are directly concerned with developing automatic speech recognizers to be employed in the airborne environment.

ANALYSIS OF THE POTENTIAL OF VOICE COMMAND

Basically the state-of-the-art can be stated as follows. *We are on the threshold of having the technology in hand for applying the voice command concept to Air Force cockpits, if we are willing to accept certain restrictions.*

Restrictions. There are six restrictions that must be taken into account with existing equipment.

1) Laboratory Environment. The present equipment has not been tested for its adaptability to stress, temperature or other factors encountered in the airborne environment.

2) Limited Vocabulary. The current vocabularies are limited both in terms of size and composition.

3) Discrete Words. Continuous speech recognition is not within the scope of today's equipment.

4) Trainability. Most machines require a number of repetitions of each word before the system will recognize it.

5) Limited Accuracy. Accuracy ranges from around 88% to 98% depending on the operator's familiarity with the system. In reality, it appears that with current equipment a highly trained operator can achieve recognition accuracies within the range of 93% to 98%.

6) Noise Elimination. Voice command systems cannot operate in a high noise environment (90 db or higher), unless steps are taken to reduce the amount of noise entering the machine.

Functional Specifications. Once these limitations are realized the minimum functional specifications for an airborne voice command can be discussed. There are six of these specifications.

1) Discrete Speech. It is quite customary (and even encouraged) for the pilot to speak in short, discrete phrases in performing his communications tasks. Consequently, discrete speech fulfills the current needs of an airborne voice command system.

2) One or Two Speakers. Because of the complexity of the aircraft and small crew size, fighter/fighter-bomber aircraft appear to be the best targets for the initial application of voice command. Since these aircraft have either one or two seats, the voice command system would have to recognize, at most, two separate speakers.

3) 100 Word Vocabulary. Since the number of functions placed under voice command initially will probably be limited to one or two, a vocabulary of 100 discrete words appears to be sufficient to significantly reduce the manual switching tasks.

4) Trainability. The system must be able to be trained to recognize the unique voice characteristics of the pilot and/or copilot.

5) Oxygen Mask Microphone Inputs. The system must be able to accept voice inputs from an oxygen mask microphone. This is the kind of microphone employed in fighter/fighter-bomber aircraft, and the voice command must be able to interface with it.

6) Cockpit Environment. The system must be able to function accurately and reliably within the temperature, stress, and noise levels of the airborne environment.

VOICE COMMAND SYSTEM DEVELOPMENT.

Progress toward development of such a system as just illustrated is well along. The Information Processing Techniques Office of the Advanced Research Projects Agency (ARPA) of DoD recently sponsored an ad hoc study group to determine the feasibility of developing an automatic speech recognition system whose specifications encompass virtually all of the functional specifications just discussed. Overall it appears as if the system targeted for completion in 1973 could meet the requirements of the airborne environment (given that the system can be packaged to fit into - and function in fighter/fighter-bomber aircraft). It offers promise of a very effective voice command system for Air Force applications.

Development Strategy. Even if the system is functioning in 1973, it is still necessary to set up a development strategy in order to guarantee that the pilot and the system will function together as a unit. The overriding consideration is always what will benefit the operational pilot most - not what might be most easily mechanized or what appeals to us personally. Once this idea is firmly in mind as the overall guiding principle, the next step is to ask, "What are the key systems that enable the pilot to do his job?" A corollary to this is, "Are these also the systems which contribute most to the amount of work the pilot must perform?" Based on the information available the next step is to decide on the system or systems to be placed under voice control. Following this decision, the third step of constructing the vocabulary and testing it is carried out.

What Not To Do. With the concern for the pilot as the overall guiding principle, what not to do in implementing voice command becomes rather clear. One way of getting voice command into the cockpit is to buy a system, plug it into the aircraft, and see how it performs. This approach will not work for a number of reasons. First, we lack an adequate date base of voice characteristics of pilots in the airborne environment. However, from the small amount of data that is available, we know that it is much more difficult to recognize words spoken in this environment. The second, and probably most important reason why this approach won't succeed is that it shows a lack of concern for the pilot. No matter how much "scientific" data is furnished to show what a great breakthrough voice command is, if it is cumbersome for the pilot to use or if he is really not convinced of its worth, the system is doomed.

RESEARCH PROGRAM FOR AIRBORNE VOICE COMMAND

The following three phase research program is proposed as a means of examining the application aspects of an airborne voice command system:

PHASE I: Analysis of the functional requirements of voice command.

PHASE II: Optimization of the Pattern Recognition Algorithm.

PHASE III: Inflight Investigation and Evaluation of the Voice Command Concept.

PHASE I: Analysis of the Functional Requirements of Voice Command.

Phase I will analytically examine in detail the factors which are directly related to the application of voice command, taking into account aircraft, mission, and pilot characteristics. The resultant data are required for interlacing the real world need with the voice command concept. It is recommended that FGR assume prime responsibility for the accomplishment of this step as it is applications oriented.

PHASE II: Optimization of Pattern Recognition Algorithm.

Once the application work has progressed sufficiently, i.e., once the characteristics of the mission and, consequently, the vocabulary are known, the next step is to purchase equipment and optimize the pattern recognition algorithm. During this phase FGR would enter into cooperative efforts with other organizations such as Avionics Laboratory and/or RADC to deal with the optimization. However, FGR would still remain as a focal point supplying the application type inputs.

PHASE III: Inflight Evaluation and Investigation of Voice Command Concept.

Since the ultimate goal of the voice command research program is to have a viable airborne voice command system, it is necessary to evaluate the effectiveness of the concept in the inflight environment.

A Qualification: The Frequency Distortion Problem:

The airborne environment is a hostile one as far as voice command is concerned. Airborne conditions cause frequency distortions to spoken words; also breathing patterns add noise to the system. These frequency distortions must be eliminated since the usefulness of the voice command concept depends on its ability to function in this environment.

FGR because of its relationship with the Instrument Pilot Instructor's School (IPIS) is in a position to make a significant contribution in the study of the frequency distortion problem. A program conducted in conjunction with IPIS, utilizing their aircraft, could obtain a data base actually gathered, quickly and inexpensively, in the inflight environment while duplicating exactly the environment in which voice command must function. This effort by FGR will be of "quick fix" nature and will be conducted to obtain the confidence needed to proceed with the applications program. Further study and ultimate resolution of the problem of course resides with those responsible for the development of the mechanization aspects.

CONCLUSIONS

There is a viable research program for implementing voice command. The program is tied intimately to the role of FGR in voice command. One of the fundamental issues in flight deck development over the next 20 years will be directed to drawing upon the power of the computer to enhance the operational capability of manned weapon systems. Voice command is one of the means by which the pilot can interact with the computer. As an organization, the Flight Deck Development Branch needs to be concerned with the application of voice command in the cockpit, in terms of the pilot, the man-machine interface, the mission, etc.

It is essential that work be undertaken on examining the application aspects of voice command because:

the concept addresses one of the great issues confronting flight deck development.

the voice command concept may not enhance operational capability unless consideration of the application aspects is undertaken.

SECTION VIII

CONCLUSIONS

During the next 20 years, the power of the computer will be used to enhance the operational capability of manned weapon systems. Voice command is one means by which the pilot can interact with the computer. Certainly, many disciplines, technical specialties, and organizations will be involved because voice command is a man-machine problem. As an organization, the Air Force Flight Dynamic Laboratory needs to be concerned with the application of voice command in the cockpit, in terms of the pilot, the man-machine interface, the mission, etc.

At present, no organization is concerned with the application aspects. Organizations are more concerned with the engineering mechanization aspects. A vacuum exists with respect to determining the utility of the voice command concept in Air Force cockpits. The effects of this vacuum are apparent in that research has moved toward continuous speech recognition (which is a problem of at least a magnitude more complex) because nobody has expressed a requirement for a discrete speech recognition capability.

It is essential that work be undertaken to examine the application aspects of voice command because:

(1) the concept addresses one of the great issues confronting flight deck development.

(2) the voice command concept may not enhance operational capability unless consideration of the application aspects is undertaken.

REFERENCES

1. Clapper, G. L. "Automatic Word Recognition," IEEE Spectrum, August 1971, 57-69.

2. Ends, E. J. "Pilot Error...Where Does It Start?" Interceptor, July 1971, 4-6.

3. Glenn. J. W., Gordon, R. N. and Moschetti, G. Voice Initiated Cockpit Control and Interrogation (Vicci) System Test for Environmental Factors. Final Report Prepared Under Contract N00019-70-C-0426, Naval Air Systems Command, Washington, D. C.

4. Munn, N. L., Fernald, N. D. and Fernald, P. S., Introduction to Psychology (2nd ed.) Boston: Houghton Mifflin, 1969.

BIBLIOGRAPHY

Lindgren, N., "Machine Recognition of Human Language; Part 1 - Automatic Speech Recognition." IEEE Spectrum, 1965 (March), 114-136.

Beck, B., Vonusa, R., and Grech, G. (operating committee), Workshop on Automatic Pattern Recognition Problems in Speech, Rome Air Development Center, Rome, New York, 1971.

Kelley, T. P., Martin, J. T. and Garger, J. R., Voice Controllers for Astronaut Maneuvering Unit. AFAL-TR-68-308, Air Force Avionics Laboratory, Wright-Patterson AFB, Ohio, 1968.

Erman, L. D. and Reddy, D. R., Implications of Telephone Input for Automatic Speech Recognition, Budapest, 1971, 85-88.

Neely, R. B. and Reddy, D. R. Speech Recognition in the Presence of Noise, Seventh International Congress on Acoustics, Budapest, 1971, 177-180.

Newell, A., Barnett, J., Forgie, J., et al., Speech-Understanding Systems: Final Report of a Study Group, Computer Science Department, Carnegie-Mellon University, Pittsburg, 1971.

Oestreicher, H., "Mechanisms of Speech Recognition." In Proctor (ed.) Biocybernetics of the Central Nervous System. Little, Brown and Company, 1969, 127-138.

BIBLIOGRAPHY (Con'd)

Preusse, J. W., Word Recognition and Speaker Authentication Using Amplitude Independent and Time Independent Word Features, Research and Development Technical Report ECOM-3439, United States Army Electronics Command. Fort Monmouth, N. J., 1971.

Scott, P. B. and Richards, J. R., Speech Controlled Radio Channel Selector, AFAL-TR-266, Air Force Avionics Laboratory, Wright-Patterson AFB, Ohio. 1971.

Uncl
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Air Force Flight Dynamics Laboratory Air Force Systems Command Wright-Patterson Air Force Base, Ohio	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

EXAMINATION OF THE VOICE COMMAND CONCEPT FOR APPLICATION TO AIR FORCE COCKPITS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

John M. Reising

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
	52	4

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. 6190	AFFDL-TR-72-108
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	AFFDL(FGR) Air Force Flight Dynamics Laboratory Wright-Patterson Air Force Base, Ohio

13. ABSTRACT

Voice Command holds potential for off-loading many manual switching tasks from the pilot. This report surveys the state-of-the-art in voice command, discusses cockpit applications, and identifies primary manufacturers/government agencies doing research. Such factors as the functional specifications of an airborne voice command system and concern for the pilot in relation to voice command are stressed. An analysis of the overall situation leads to the conclusion that voice command is on the threshold of application into cockpits, given that certain restrictions are acceptable. (U)

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Voice Command						
Voice Recognition						
Automatic Speech Recognition						
Voice Control						
Cockpit Voice Control						
Cockpit Controls						

☆U.S.Government Printing Office: 1973 — 759-495/605

Security Classification