(1)

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS
N.A.

3. DISTRIBUTION/AVAILABILITY OF REPORT
N.A.

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
APTEC INC.

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
DCASMA PHOENIX

6c. ADDRESS *(City, State, and ZIP Code)*
7442 E. SWEETWATER
SCOTTSDALE, ARIZONA 85260

7b. ADDRESS *(City, State, and ZIP Code)*
7B THE MONROE SCHOOL
215 N. 7TH STREET
PHOENIX, ARIZONA 85034-1012

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
DEPT. OF THE ARMY

8b. OFFICE SYMBOL *(If applicable)*
AMSMC-PCS(D)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DAAA21-82-C-0184

8c. ADDRESS *(City, State, and ZIP Code)*
U.S. ARMY AMMCCOM
AMSMC-PCW-D(D)
PICATINNY ARSENAL, NJ 07806-5000

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*
SCIENTIFIC AND TECHNICAL REPORT
VOICE ACTIVATED GUN TURRENT CONTROL

12. PERSONAL AUTHOR(S)
FRANK TONEY

13a. TYPE OF REPORT
FINAL

13b. TIME COVERED
FROM 10/87 TO 8/88

14. DATE OF REPORT *(Year, Month, Day)*
20 AUGUST 1988

15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
VOICE RECOGNITION
ARTIFICIAL INTELLIGENCE

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The objective of the research project was to determine the feasibility of using a Low Memory Requirement Voice Recognizer to activate gun turrent control. Such a system would enable a single individual to manually pilot the helicopter while activating gun functions by voice command.

To develop an effective voice recognition system, major limitations of existing voice recognition technology needed to be overcome. The major limitations of current voice recognition systems are: (a) they are normally capable of recognizing only one speaker and one mode of speaking, (b) recognition time is slow when working with sizeable vocabularies, (c) they require extensive hardware with resultant excessive weight and size limitations (for aircraft applications), and (d) they are difficult to interface with, and access artificial intelligence/expert software systems.

(24)#1

Most of the preceding limitations occur because existing voice recognition technology recognizes a word by segmenting the word into several thousand parts and making a memory

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
FRANK TONEY

22b. TELEPHONE *(Include Area Code)*
(602) 242-0798

22c. OFFICE SYMBOL

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

~~89 1 09 252~~

template of the parts. Then, when a word to be identified is spoken, that word is also segmented into several thousand parts and matched with the word templates in memory. All of this segmenting and storing of thousands of bytes of information consumes and requires large amounts of processing and memory capability.

The approach taken by this research project to solving the voice recognition limitations, was to develop a new method of recognizing words. Instead of electronically duplicating a word, this project developed hardware and software that generates an unique electronic description of a word, utterance, or phrase -- similar to the way the Dewey Decimal System describes a book in the library. The result was a reduction in memory required from approximately 4,000 bytes to 128 or less bytes per word. Several inherent advantages of the system result:

1. Essentially real time recognition of approximately 700 utterances per 8K memory device
2. Capability to recognize different speaking voices from the same pilot or different individuals.
3. Easier differentiation of utterances from ambient helicopter noise.
4. A design that lends itself to use of intelligent software to identify unclear commands
5. Capability to recognize phrases and sounds as easily and with as few bytes of memory as are required for single words.
6. Ability to store unrecognized sounds for later identification.
7. Ability to adjust recognition sensitivity to fit the application needs.
8. Reduced hardware costs (under $200) and component size (6"x10"x1").
9. A totally independent system -- no host computer required.
10. Capability for pilots to have a personal plug in PROM containing their speech history. PROM can be removed and reinstalled when the pilot changes aircraft.

In conclusion, the system represents an opportunity to significantly improve the interface between the helicopter crew and the aircraft and its armament systems. It appears to solve or minimize the inherent problems of conventional voice recognition. The system appears to offer a practical and workable method of improving crew effectiveness, response time to multiple targets, and accuracy versus time saved. It has potential in other military and commercial applications.

SCIENTIFIC AND TECHNICAL REPORT
VOICE ACTIVATED GUN TURRENT CONTROL

"THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION"

HELICOPTER APPLICATIONS THAT WOULD BENEFIT FROM VOICE COMMUNICATIONS

Implementation of voice communications with helicopter armament could increase both pilot and gunner effectiveness. A voice command system offers potential to enable one person to both pilot the helicopter and to verbally control armament functions. In the event the gunner or pilot were disabled, the remaining crew member could perform the dual functions of piloting the aircraft and operating the armament. When encountering multiple targets, the pilot could supplement the efforts of the gunner. If the pilot could effectively perform piloting and gunnery activities, overall crew effectiveness, speed of response, and vehicle and armament control would be improved. Ultimately, the gunner position might be eliminated. Helicopter weight, size and complexity could be reduced.

a. HEADS UP & EYES OUT APPLICATIONS -- In many helicopter applications, the pilot is most effective when he is looking at a subject not related to armament operation. He might be concentrating on terrain, targets, or other critical activity. In many cases, when an operator diverts his attention and refocuses his eyes to monitor gauges or perform cockpit functions, performance can be significantly reduced. Capability to give voice commands provides the pilot capability to maintain eye contact with his objective while controlling armament or helicopter auxiliary functions with his voice.

b. HIGH CONCENTRATION SITUATIONS -- In situations such as maneuvering in close confines, flying at low altitude, or making an instrument approach to land, it is impractical to divert attention from the task at hand. Pilots often can barely scan critical functions such as air speed. Less critical items such as fuel level and system vacuum are often ignored. For the pilot to divert his attention to control armament can immediately reduce his aircraft control capability. Voice control enables the pilot to maintain attention on the task at hand while providing capability to give voice commands.

e. "BUSY HANDS" -- Helicopter pilot's hands are occupied maneuvering the craft and performing functions such as keying in information on a keyboard and changing radio frequencies. It is generally felt the pilot can not effectively be given more tasks that increase the work load on his hands. The capability to give the machine voice commands to control armament would leave the pilot's hands free to fly the aircraft.

PROBLEMS WITH CONVENTIONAL SPEECH RECOGNITION TECHNOLOGY THAT THIS PROJECT ADDRESSES AND RESOLVES

The preceding discussion of applications indicates the theoretical benefits of using human voice to control armament functions. Unfortunately, there are several inherent weaknesses of conventional speech recognition technology which have limited its effectiveness and utilization in human/machine interface applications.

a. LACK OF PRACTICAL COMPREHENSION. A problem in many applications is that conventional voice recognizers can normally only be used by one person at a time and their accuracy declines dramatically when the person's voice changes.

For example, a pilot commanding "Activate" in a normal voice could be understood but not his co-pilot or gunner commanding "Activate". Probably more critical to actual conditions, is that if the same operator were shouting "Activate" because of a problem, the recognizer couldn't recognize him. In other words, existing technology has not achieved even the comprehension capabilities of a 3 year old human. Most conventional recognizers have to be retrained for each speaker on a periodic basis. For example, they often cannot recognize an individual speaker if he has a cold without retraining.

The capability to recognize the same utterance as spoken by different voices is of major importance in situations where an operator encounters varying degrees of stress. Voices can be altered due to a number of factors such as stress, ambient conditions (vibration, noise), or health. For example, it is reported that pilots in battle situations may scream into the microphone. While experiencing severe G forces or vibrations the pilot's voice may also be altered. For any recognizer to be truly practical, it must be functional in virtually all user speaking modes.

Compounding the problem is the fact that in some of the more severe situations, voice changes are difficult to safely test, evaluate and record. Consequently, it would be beneficial if the recognition device could have the capability to recognize new forms of speaking the same phrase or word based on experience.

b. REQUIRE TOO MUCH HARDWARE AND COMPUTER CAPACITY. To achieve a vocabulary capability to handle multiple operators under varying speaking conditions, existing units require a host computer or comparable hardware to provide memory storage and computational support. This results in a device that is too big, too heavy, possibly too expensive; and hence, is generally considered impractical for machine and vehicular applications.

c. TOO SLOW. IBM has stated that their "state-of-the-art" voice recognizer has a retail price of $15,000 to $20,000 and may take 2 or 3 seconds to recognize a single word. Other systems, by limiting vocabulary capability, have gotten this lag time down as low as 1/3rd of a second. A few prototype devices have reportedly achieved near real time recognition by restricting vocabularies to around 300 words. Most users feel the response time is too slow when coupled with a reasonable vocabulary and capability to handle multiple operators and speaking conditions -- particularly in critical situations where an immediate and precise machine response is critical to performance.

Conventional speech recognizers tend to be slow and require so much hardware because there is such a large mass of data for the microprocessor to search before it can make a match between the spoken word and the words in its memory. At a rate of roughly 4,000 bits per second, it doesn't take long to use up a lot of memory.

d. DIFFICULT TO ACCESS AND INTERFACE WITH EXPERT SYSTEMS UNDER FIELD CONDITIONS. Existing systems generally make a "template" of each word or phoneme by cutting it into several thousand bits. When a word is spoken, the computer compares the spoken word's template of a few thousand bits with all the ones stored in memory until it finds the spoken word's duplicate. At roughly 4,000 bits per second, conventional word recognizers consume excessive amounts of memory. This large amount of memory and hardware required creates several problems that make it difficult for most systems to interface with conventional word recognition software.

A. Word templates don't lend themselves to easy indexing and access by interfaced programs. Each spoken command results in a different template when spoken by different people or by the same person in differing stress conditions.

B. Most existing systems are limited to individual words. Many aircraft related vocabularies as well as expert and artificial intelligence systems lend themselves more to complete phrases or utterances.

TECHNICAL APPROACH

This research project represents an effort to resolve the inherent weaknesses of conventional voice recognition technology.

The hardware developed as a part of this project produces a unique and quantifiable electrical description of an audio signal, whether it be a word, utterance, phrase or sound. A typical word or phrase can be described with less than 128 bytes of information compared to 4000 bytes per word for most conventional systems.

a. FEWER PARTS -- The hardware developed for this project is not bound by numerous filters as are currently used to separate frequencies on conventional recognizers. The number of components and hence expense, is significantly reduced. Hardware for the unit is low cost since common, off the shelf components are used. All components are available in mil spec. In addition, the efficiency of the concept reduces microprocessor and memory size requirements to further eliminate much of the expense found in other systems.

b. DESCRIBES A SOUND, DOESN'T DUPLICATE IT -- The recognizer is unique also because it won't regenerate the word or sound. Conventional recognizers normally must acquire enough information that they could actually reproduce the sound.

c. REAL TIME RGNITION -- The alpha numeric representation of a word, utterance, or phrase is recognizable by eye with near "real time" or instantaneous recognition time. It is somewhat similar to the way a Dewey Decimal System number not only refers to the location of a book, but describes it as well. For example, most librarians can look at a Dewey Decimal Number and know basically what type of book is being described.

SUMMARY OF OBJECTIVES AND RESULTS

In general, the primary objective of the project was to "prove" that a voice recognition system can identify human words, utterances, phrases and other sounds represented with 500 bytes or less of information: and, as a result, that the major problems associated with conventional speech technology in vehicular applications, could be overcome.

In the researcher's judgment, the objectives have been achieved. Words, utterances, phrases and noises can be sufficiently represented with 128 bytes of memory that they can be subsequently recognized with accuracy. Consequently, the major problems of conventional voice recognition can be overcome.

Following is an outline of each specific objective of Phase I and the result of the investigation.

OBJECTIVE: Demonstrate that the hardware and software can recognize a normal helicopter gunnery vocabulary as spoken by two people under different conditions. Specifically, the system should have a range of understanding and instantaneous recognition of up to 300 utterances as spoken by different individuals or by the same individual under varying stress levels and ambient conditions.

RESULT: Objective Achieved. The system can recognize, on an essentially real time basis, approximately 400 utterances per 8K memory device (Electrically Alterable ROM). ROMs could be added to increase vocabulary correspondingly. In a conventional voice recognition system, an 8K memory device could store about 2 - 3 words.

Since the description of an utterance serves as an indexing function to a word, meaning, or knowledge concept stored in memory, it is possible for several descriptions to refer to the same meaning. This is the basic advantage that enables the unit to Work with different operators under varying stress levels and methods of speech. The voice commands could be from an individual from Alabama or from a person with a Boston accent. All the module knows it that each utterance refers to a certain meaning in its memory. In fact, one operator could give an instruction in English and an other operator could speak Spanish with equal system comprehension and effectiveness.

OBJECTIVE: Demonstrate system's ability to differentiate voices from ambient helicopter noise.

RESULT: Objective Achieved. The system has proven accurate in conditions where helicopter ambient noise has been simulated. This was accomplished using a tape recording of helicopter ambient noises. It is recognized that there are many limitations in this type of test. The next phase of the project will require testing in actual helicopter operating environments.

The system theoretically is inherently more efficient at isolating commands from ambient noises. The spoken command is represented by a clearly identifiable alpha numeric description amid a background clutter of meaningless symbols.

OBJECTIVE: Demonstrate the system's ability to use logic to interpolate unclear commands.

RESULT: Objective Achieved although additional research is required to capitalize on the system's capabilities. The system has the capability to use logic or intelligence to interpolate unclear or poorly heard commands. This capability will require additional definition. For example, how much "guessing" do we want the device to do when interpreting commands to a helicopter gun?

The nature of the alpha numeric description of the word also should allow development of predictive software. For example, when a person screams or yells a command in a high stress combat situation, his voice description is altered. However, the description of the screamed word will hopefully maintain a reasonably constant relationship with the pilot's normal command voice. If this relationship can be determined, it could be used to predict the possible description of other commands in high stress situations. As a result, it would be more useful in dangerous situations such as battle conditions that are difficult to duplicate in the more normal testing environment.

OBJECTIVE: Demonstrate ability to close relays and/or perform other remote control functions. Be compatible with common communications protocols.

RESULT: Objective Achieved. The system outputs ASCII characters, the most common communications protocol. ASCII provides capability to communicate with most expert software systems. The system can also be modified to respond to commands by opening and closing relays, generating B-C-D output, and activating potentiometer responses.

OBJECTIVE: Reduce hardware costs by using off-the-shelf componentry.

RESULT: Objective Achieved. Total hardware costs for the prototype are under $250. Comparable mil spec component cost is estimated to be approximately $600. All components are readily available in mil spec.

OBJECTIVE: Be self contained i.e. no host computer required. compact size approximately 5"x7"x1.5", and low weight (less than 1 pound).

RESULT: The system is self contained and requires no host computer (although one can be used if desired). Dimensions of the wire wrapped prototype are approximately 5"x10"x3". The production version is estimated to be approximately 6"x10"x1".

OTHER RESULTS AND CONCLUSIONS OF THE RESEARCH PROJECT

The system has evidenced other advantages which should prove of benefit in broader applications.

1. The technique of generating descriptions by the hardware allows a complete phrase, utterance, or sound to be identified as easily as an individual word. This provides the advantage that a pilot can use a complete phrase such as "Lock On" or "Missile Activate" with no excess use of module memory. It also gives the capability to convert ambient sounds to ASCII or a written description. This advantage should enable the system to monitor noises and respond; for example, "Explosion, location 4".

The length of the phrase the system can recognize is currently limited to two seconds. This is a result of current design rather than a system limitation. The length of phrase recognition can be increased if necessary.

2. Theoretically, the system will concantenate (string together) words at a much faster rate than conventional systems. Consequently, speakers should be able to use a more normal speaking speed when communicating with the system.

3. The system stores unrecognized commands. This enables the operator to review and input the proper response for the unrecognized command. For example, a pilot might encounter a high G loading for the first time. If the system didn't recognize his altered voice, it would store the commands until the pilot could enter them into the system at his convenience. As mentioned previously, once the system has defined the way an individual's voice changes in a particular situation, it can use predictive programming to identify the voice in the future --

even though it may never have heard the individual speak in the same manner before.

4. The production module as installed in a helicopter could have the capability to be used by any pilot or gunner without retraining the system to recognize the new voice. This would be accomplished by having each pilot or gunner maintain his own personal plug-in PROM. The PROM would contain all the pilot's commands and queries. Theoretically, the pilot could build his library of various speaking modes over his entire career. Whenever he transferred to a new aircraft, he would simply unplug his PROM from the aircraft he was leaving and plug it into the new aircraft.

5. Use of artificial intelligence techniques. The system has inherent advantages in using and interfacing with artificial intelligence and expert software systems. By using a 128 byte description of an utterance a large amount of memory is made available for other uses. This could be used for increased vocabulary, incorporation of more sophisticated recognition software, and operating instructions for various systems and software interfaces.

Phase I research barely scratched the surface of decision making, expert, or artificial intelligence software. Time constraints resulted in efforts being directed primarily toward proving the validity of the basic voice recognition hardware and software.

Please note, specific advantages as related to utterance recognition need additional development. Some of these are:

a. The system has the capability to have its recognition sensitivity adjusted. Recognition sensitivity is the degree of precision with which a speaker must duplicate a previous spoken input before the system can recognize the word. Additional research is required in this area to determine the optimum range of sensitivity for helicopter gunnery applications. For example, the software can be written to be quite sensitive for security applications. In other words, it would only recognize the voice of one person and then it would have to be spoken precisely. However, too sensitive of a setting would be impractical in normal or battlefield operating conditions where the pilot's or gunner's voice can change over a wide range due to stress, vibration, or other external factors.

At the other end of the spectrum, the system can be adjusted to the point where it applies a broad range of logic in trying to understand unclear commands or queries. Theoretically, the system could be adjusted to the point it will recognize most people speaking the same word. This could be accomplished by using a range of word descriptions or comparative word values. However, the limits of this capability will require careful consideration. For example, in a gunnery application if the pilot gives a command that sounds similar to "fire", how much reasoning should the system apply before it actually fires the gun?

b. Future iterations of the system could provide the •capability to anticipate commands; and hence, shorten recognition time. For example, at present the system would close a relay to fire a gun at the completion of the command "fire". However, it might be advantageous for the system to initiate the relay closure midway through the word for a more "real time" response. In other words, does the pilot want the gun to fire at the beginning, the middle, or the end of the word "Fire"?

To accomplish firing before the end of the command "Fire", the system would be programmed to know that normally after the command "lock on", the command "fire" or "knock off" would normally be the next commands expected. Since the beginnings of these commands are distinctly different, the system could be set up to close the relay at an earlier point in the speaking of the command.

The concept is similar to the way a pilot normally prepares himself for the next expected radio transmission. If he is approaching to land he can expect to be transferred to the Tower frequency. Most pilots will have looked up this frequency in advance. When the command is received, even if it is not clear, they have a high degree of confidence that they are switching to the correct frequency.

b. The system has the ability to modify behavior or responses based on errors made or new information received. It can have the ability to chose or predict the most likely meaning of an utterance from data that is imprecise, incomplete, and not totally reliable. For example, if a pilot's speaking voice is gradually changing such as from age or a cold, the system could be adjusted to compensate for the change.

c. The real time recognition capability, ease of indexing the binary description of the utterance, and the availability of excess memory, make possible interfacing with, and utilizing, many artificial intelligence/expert software systems on a practical basis. Of particular importance, it makes possible an easy and economical method to input data under field conditions. The system should interface with several expert systems and lends itself to reasoning concepts dealing with classification of data, making judgments based on incomplete data, information and resource constraints.
utterance.

PROBLEM ENCOUNTERED

In general, the system performed as expected. The only problem encountered was related to voice volume -- It was determined that voice volume and amplitude extremes can significantly affect the word description. Originally, we used an audio leveling circuit to produce the same audio signal regardless of input level.

It was determined that holding the input level constant removes critical information from the word. Consequently, the circuit was removed from the system.

This characteristic is most evident when comparing a whisper and a normal speaking voice. A person can speak more and more quietly until he must switch to a whisper. It is somewhat similar to a pipe organ where a certain amount of air must blow over the reed before a tone is produced. Until that point is reached, the pipes will only hiss.

After the hardware was modified the system accounts for different speaking amplitudes. Nevertheless, if a person were to communicate in whispers, the system would need to be trained in the specific whispered commands. It is expected there will be none of these in helicopter applications.

CONCLUSIONS AND RECOMMENDATIONS

1. The system appears to resolve the major problems of conventional voice recognition systems.

2. It would appear to potentially improve pilot/gunner operating effectiveness, speed, response to multiple targets, and control of the aircraft.

3. The system appears practical, workable, and economical -- there are no reasons to believe it could not be applied in helicopters. Remaining questions deal with refining the system rather than proving the concept.

4. It is proposed that Phase II be entered with the objective of producing a preproduction prototype for testing on a simulator or aircraft.