N [redacted]
N93-72619
57-32
176343

VOICE INTEGRATED SYSTEMS

CDR MIKE CURRAN, PH.D.[1]

NAVAL AIR DEVELOPMENT CENTER
WARMINSTER, PENNSYLVANIA

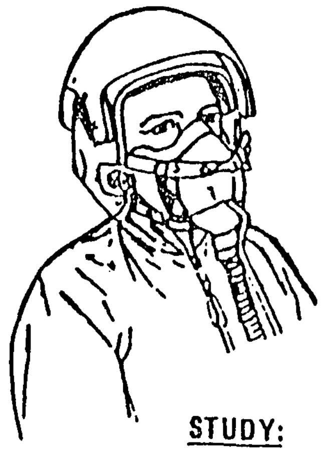
VRAS - A Voice Recognition and Synthesis System

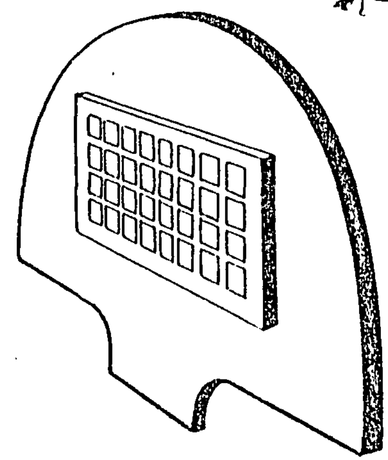
The program at NADC was initiated to determine the desirability of interactive voice systems for use in airborne weapon systems crew stations. To accomplish this effort, a voice recognition and synthesis system (VRAS) was developed and incorporated into a human centrifuge. The speech recognition aspect of VRAS was developed using a voice command system (VCS) developed by Scope Electronics. The speech synthesis capability was supplied by a Votrax, VS-5, speech synthesis unit built by Vocal Interface. The effects of simulated flight on automatic speech recognition were determined by repeated trials in the VRAS-equipped centrifuge. The relationship of vibration, G, O_2 mask, mission duration, and cockpit temperature and voice quality was determined. The results showed that: 1) voice quality degrades after 0.5 hours with an O_2 mask; 2) voice quality degrades under high ($\pm$ 0.3G) vibration; and 3) voice quality degrades under high levels of G. The "voice quality" studies are summarized in Figure 1. These results were obtained with a baseline of 80 percent recognition accuracy with VCS.

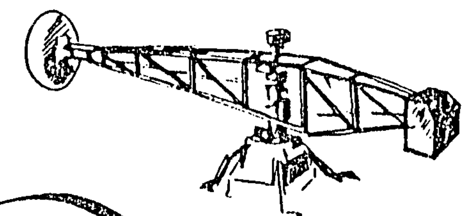
The next phase of the development program called for improvement of the VCS system. This was accomplished in two ways. A consistent bit was incorporated into the process wherein reference patterns are established to improve recognition accuracy. Improved recognition accuracy was noted. A syntactical handler was developed to facilitate use of the isolated word VRAS system and to assist simultaneously in the understanding process. The developed syntactical handler was tested with teletype input and was operational with 100 percent accuracy in real time.

The major components of the VRAS system and its general operation are shown in Figure 2. We see that the spoken words, originated by a speaker, are analyzed and sent to the "Statement Understanding" component. Once the meaning of the statement is understood, it is forwarded to the "Message Handling" unit which is responsible for all exchanges of information between the VRAS system and the system computer to which it is interfaced. Having accomplished the intent of the speaker's statement, the appropriate reply is created by the "Response

[1]The opinions expressed here are those of the author and do not necessarily reflect the official policy of the United States Navy.







STUDY:

EFFECTS OF:

COCKPIT TEMP.	COOL AVERAGE WARM
VIBRATION	0 .15 .30
G	1. 2.0 4.0
O_2 MASK	YES NO
MISSION DURATION	90 MIN.
WORDS SAID	1 – 28

FINDINGS:

VOICE QUALITY DEGRADES AFTER ½ HR. WITH O_2 MASK
VOICE QUALITY DEGRADES UNDER HIGH (.3 G) VIBRATION
VOICE QUALITY DEGRADES UNDER HIGHER LEVELS OF G*

* MAY BE ATTRIBUTABLE TO MASK SLIPPAGE

Figure 1. Summary of Voice Quality Studies

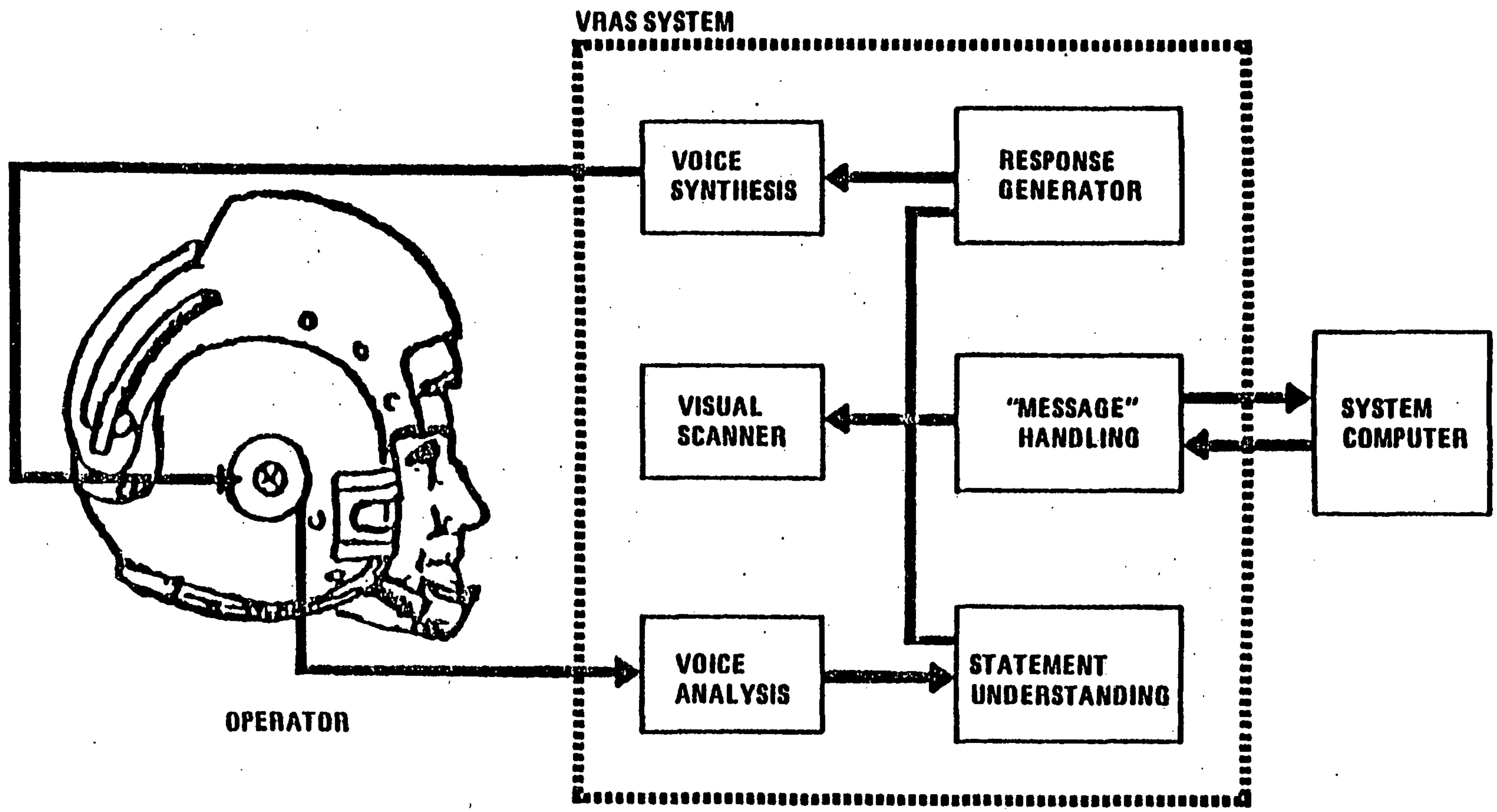


Figure 2. Voice Recognition and Synthesis (VRAS) System

Generator" and then given to the original speaker via a "Voice Synthesis" unit. The "Visual Scanner" permits visual feedback to the speaker from a variety of other VRAS units which allows the speaker to monitor visually what is being said and understood by those various VRAS components. The VRAS system also includes a printer, card reader and disc drive for logging out all communications, inputting vocabulary data, and storing different speakers' trained words for word recognition purposes.

An overview of the VRAS system, and the statement understanding approach it employs, is presented in a paper entitled "VRAS - A Voice Recognition and Synthesis System" which appears in volume VII of the 1976 SID International Symposium Proceedings. This paper was authored by Dr. Robert J. Wherry, Jr., who originated and developed the VRAS system.

This system permits the use of medium-sized vocabularies (250 words) and highly flexible statement formats. Among the unique concepts featured in the VRAS system are: 1) a "universal" statement format, 2) the use of "truth" logic to eliminate words which can no longer be appropriate in the sentence being said; and 3) a "dictionary of meaning" which permits the VRAS user to communicate a given message in a large variety of different sentences. Since the syntactical handler only requires a recognized word or phrase as input, it can be used with recognition devices other than the Scope VCS. The flexibility of the VRAS system allows for the development of a syntactical handler to accomplish any specified application within a month. The value of the syntactical handler is that it will allow the user to vary the syntactical arrangement of words during data entry without affecting recognition accuracy. Thus, the natural quality of speech as a data entry means is preserved.

The development of the VRAS facility and the VRAS concept has resulted in a powerful approach for accomplishing voice recognition and synthesis. However, since the the programming language used was at the assembly level, and since the computer employed was a Raytheon 704, only the Naval Air Development Center could readily utilize the VRAS capability. Because of the use of assembly level language, changes and improvements to VRAS have proved extremely time consuming and costly. Because of the use of the Raytheon computer the VRAS approach has not been readily applicable to the requirements of other identified voice development efforts. To correct these deficiencies a work effort has been completed which developed, tested and documented a FORTRAN IV packaging of the VRAS program. Program modifications or transferral to other computer systems or recognition and synthesis equipment have become simplified and readily implementable since all coding, except for the specific equipment interface routines, are in ANSI Fortran. Eight types of programs are required to ensure flexibility and inter-system compatibility, as well as to accomplish the VRAS syntax processing function. Seven system implementation programs are interfaced through a supervisory program that provides the few instructions required to operate the system. The programs, as shown in Table 1, can be run independently or through the supervisory program. A more detailed description of the use of each program, and how it is accessed, is provided in a report which is available for general distribution.

"Unlimited" Vocabulary Recognition and Understanding

The thrust of the previously described VRAS program was to concentrate on understanding the meaning of what was being communicated rather than merely on the particular sequence of words which was employed. While the VRAS system does permit the use of medium-sized vocabularies, and while it does permit a relatively flexible sentence structure, the greatest single drawback to the use of real-time voice recognition and synthesis today is still the limitation on vocabulary size and sentence structure. To understand the nature of this limitation the two major approaches to word recognition or "voice analysis" must be presented. One approach involves an analog-to-digital conversion of the input voice signal and a frequency analysis using bandpass filters to record what the voice signal was during a given period of time. Correlating the obtained and expected activity levels for the different bandpass filters over time permits word recognition to occur. A second approach also uses time samples of activity for a bank of bandpass filters. In this case the patterns of activity in the filters are compared against a set of phonetic features to determine the presence or absence of various kinds of sounds (fricatives, stops, etc.). When using small, tailored vocabularies, both approaches tend to do a very good job of correctly identifying the actual word being said by the speaker.

The concept of permitting the speaker to use an "unlimited" vocabulary - any legitimate English word - has been rejected as an unrealizable near-timeframe objective for voice recognition and synthesis systems because of the difficulty in word recognition for a relatively few words. It is not merely that as the number of words in the vocabulary increases the more probable it is that two words will sound alike; it is more than this problem, which we might call the "word confusion" problem, which has discouraged the development of truly "large vocabulary" voice recognition and synthesis systems. For example, for each word in the vocabulary, its "recognition vector" (the way the speaker has previously said that word) must be stored, and if that word is to be used as a synonym for another word, then its definition must also be stored. With the VRAS system, using the Scope device, 256 bits of storage for each word in the vocabulary were required just to store the recognition vector. Assuming 16 bit computer words, each recognition vector would require 16 computer words; a vocabulary

TABLE 1

VRAS SYSTEM PROGRAMS

VRAS	VRAS System Interface Program allows access to all other programs in the VRAS System.
TRAINING	Trains VRAS to a specific speaker and vocabulary.
PARSE	Processes sentences entered by speaker.
VOCAB	Lists the current vocabulary and subsidiary programs.
CONFUSIN	Enables the user to determine an appropriate recognition correlation threshold and a list of possibly confusing words.
RAWDATA	Enables the user to obtain listings of both the short (processed) and long (unprocessed) recognition vectors.
DEMO	Demonstrates the VRAS training and recognition logic.
RETRAIN	Allows the user to add additional repetitions of the words in the vocabulary to the composite recognition vectors stored on the disc.

of only 2000 words would require 32,000 words just to store the recognition vectors. Another accompanying problem with large vocabularies is the increased processing time required for the additional comparisons to be made when trying to determine which word has been said by comparing the "spoken vector" with the various "recognition vector." Thus, present word recognition technology cannot handle "unlimited" vocabularies for three very good reasons: 1) large vocabularies require too much storage, 2) large vocabularies require too much processing time, and 3) large vocabularies permit too many words which tend to get confused with each other because they produce too similar recognition vectors.

While the above reasons would, at first consideration, seem sufficient to reject the concept of an unlimited vocabulary voice recognition and synthesis system, it will be seen that an alternative approach to word recognition may be possible. The alternative approach, which will be referred to as the "word-part" approach, is based on the concept that the vast majority of words used by speakers are various combinations of relatively few prefixes, stems, and suffixes. If an incoming word can be analyzed into its component word-parts, not only can the word be correctly recognized, but also its appropriate meaning can be established without reference to a "dictionary of meaning."

Just as the stem of the word has its own meaning, so also do the prefixes and suffixes. It, therefore, becomes possible to conceive of a new word recognition approach which analyzes each spoken word into its component prefixes, suffixes, and stems; to determine the meanings of these components; and to use those component meanings to determine what the speaker is saying without ever attempting to recognize the whole word or to have a definition of the whole word stored in memory.

This new approach to an "unlimited" vocabulary voice recognition and synthesis system will be pursued during fiscal years 1978 and 1979 as an exploratory development effort which should complement and extend the previously described VRAS development program.

Integrated Applications of Automated Speech Technology

Progress in isolated word recognition, syntactical handling, and other speech technology areas provides evidence to suggest that the initiation of a Navy Advanced Development Program is justified. However, developments and progress in separate speech technology areas can only achieve their true potential if and when they are successfully integrated into total system applications. It is noteworthy that several such integrations have been achieved. The Naval Training Equipment Center's Ground Controlled Approach Controller Training System has utilized speech recognition to effect control of an aircraft/pilot simulation, and to provide the basis for the objective performance measurement of the trainee's behavior. The Department of Transportation's Automated Command Response Verification System has demonstrated the integration of automated speech technologies (AST) in an operational ship-safety role. In these and other government applications AST has done more than make life a little simpler. A number of applications have successfully demonstrated the power of AST-based systems in solving problems that could not have been addressed before the emergence of these technologies. Therefore, it seems that an advanced development program which synergistically draws upon the results of past and present AST efforts is a reasonable and worthwhile next step.

However, before such a program can be initiated and successfully pursued, several information gaps must be resolved. Specifically, if AST is to be applied to the areas of airborne crew station design, performance measurement and training simulation, several new methodologies must be developed. They include: 1) a method for identifying high payoff applications of voice interactive systems in terms of the enhancement of both operator and system performance; 2) a methodology for assessing the technical feasibility of AST for each proposed application; and 3) a methodology for integrating the above information sources and generating a rationale for mutually supportive basic research, and exploratory and advanced development requirements.

The Integrated Applications of Automated Speech Technology was an exploratory development program initiated in fiscal year 1977 to develop these methodologies. This program will be completed in early fiscal year 1978. The program objective is to develop methodologies for integrated applications of automated speech technologies for Navy system development, training, and operational settings. The program approach includes five major tasks: 1) review government applications of AST; 2) perform crew station design analyses; 3) examine performance measurement capabilities; 4) examine training applications; and 5) prepare a program/implementation plan.

The objective of the review task was to review critically the present applications of AST, and their supporting data, to establish a baseline of present progress from which the Navy can draw to plan AST applications. The completed review identified present capabilities and advancements, as reflected in successful system applications of AST, for type of speech recognition (i.e., isolated and limited continuous), vocabulary, recognition accuracy, syntactical handling, and user acceptance.

The crew station design, performance measurement, and training applications tasks have addressed documentation available for the Navy P-3C anti-submarine aircraft weapon system to develop the desired methodologies. For the crew station design task the "on station" mission segment was examined for each crew member by considering tasks to be performed and subsystems to be exercised. After consideration of the variables that affect the application of voice technology to crew station design, a four dimensional rating system was developed. The dimensions included: the technical feasibility of implementing voice for accomplishing the task; the utility of implementing voice to accomplish the task; time/accuracy requirements for the task; and the impact of unassessed variables such as aircraft noise and mission duration. Using this rating strucutre, each task was reviewed utilizing the four dimensional requirements, and assigned a four digit code of numbers corresponding to the four task requirement. For each task the four digit code was reduced to a one digit code corresponding to initial AST payoff. Previously obtained criticality and frequency ratings for each task were applied to this initial AST payoff rating to obtain an overall AST payoff rating. Finally the ratings for the tasks were converted to a matrix format. As an example of this process, Table 2 shows tasks by subsystems for the P-3C Pilot. The AST ratings for all detailed tasks to be included within a matrix cell were treated statistically to determine a single AST potential payoff rating for each matrix cell. Table 3 summarizes the most promising voice applications areas for both the P-3C Pilot and TACCO. It should be recalled that the objective of this task was not to identify voice applications for the P-3C, but to develop the methodologies required to identify high payoff applications of voice technology.

As of this time the performance measurement and training applications tasks are not completed.

The last task of this effort involves preparation of a program/implementation plan. The general approach for integrating various information sources and generating a rational for research implementation requirements is presented in Figure 3. Neither the various tradeoff analyses nor the methodology for effecting the integration of the various information elements have been completed. Again the promise of this task is that the approaches developed for the generation of the trade-off analyses and the integration of the information sources will be incorporated when attempting to apply voice technology to other airborne systems/subsystems. The identification of technology voids and problems will serve as the basis of an interlocking technology base program covering the full spectrum of basic research through advanced development.

VIST - Voice Interactive Systems Technology

VIST is a new advanced development program being initiated in fiscal year 1978. It is viewed as the application or implementation of the products obtained from the previously described AST exploratory development effort. The objective of the program is to demonstrate the

TABLE 2

VOICE TECHNOLOGY PAYOFF FOR P-3C PILOT

POSITION: P-3C PILOT

MISSION SEGMENT: ON STATION

POTENTIAL PAYOFF CODE:

1. GREEN HIGH
2. BLUE MEDIUM
3. YELLOW LOW

SUBSYSTEMS

GENERIC TASKS	PROPULTION	FUEL	ELECTRICAL	HYDRAULIC	ECS	AIR FRAME	FLT CONTROLS	FLT INSTRUMENTS	COMMUNICATIONS	NAVIGATION	ECM	ASW:	SEARCH STORES	ON-TOP POS. IND.	BT & SEA NOISE	MAD	LLLTV	PHOTOGRAPHIC	CASS	DATA HANDLING & DISPLAY	SEARCH RADAR	ARMAMENT	CREW
MONITOR INDICATOR	1	2							1									2				2	
MONITOR SITUATION																							
PERCEIVE DATA																							
FORMULATE PLAN																							
DETERMINE SOLUTION																							1
ENTER DATA																							
RECEIVE DATA									3											3			
COORDINATE DATA																							
ACTIVATE CONTROLS	3								2	3			2					2		3		2	
ADJUST CONTROLS									2											3		3	
PERFORM MANEUVER																							

TABLE 3

P-3C PILOT/TACCO SUMMARY RESULTS

PRELIMINARY ANALYSIS OF MOST PROMISING AREAS FOR VOICE APPLICATION

A. TASKS (IN ORDER OF PROMISE)

1. MONITOR INDICATORS (ALERTS)
2. ACTIVATE SWITCHES (FUNCTION SWITCHES)
3. ENTER DATA (KEYBOARD)
4. ADJUST CONTROLS (KNOBS)
5. RECEIVE DATA (CRT TABLEAUS)
6. COORDINATE DATA (COMMUNICATION)

B. SUBSYSTEMS (IN ORDER OF PROMISE)

1. COMMUNICATIONS
2. PROPULSION
3. SEARCH STORES
4. PHOTO
5. DATA HANDLING
6. ARMAMENT
7. CREW

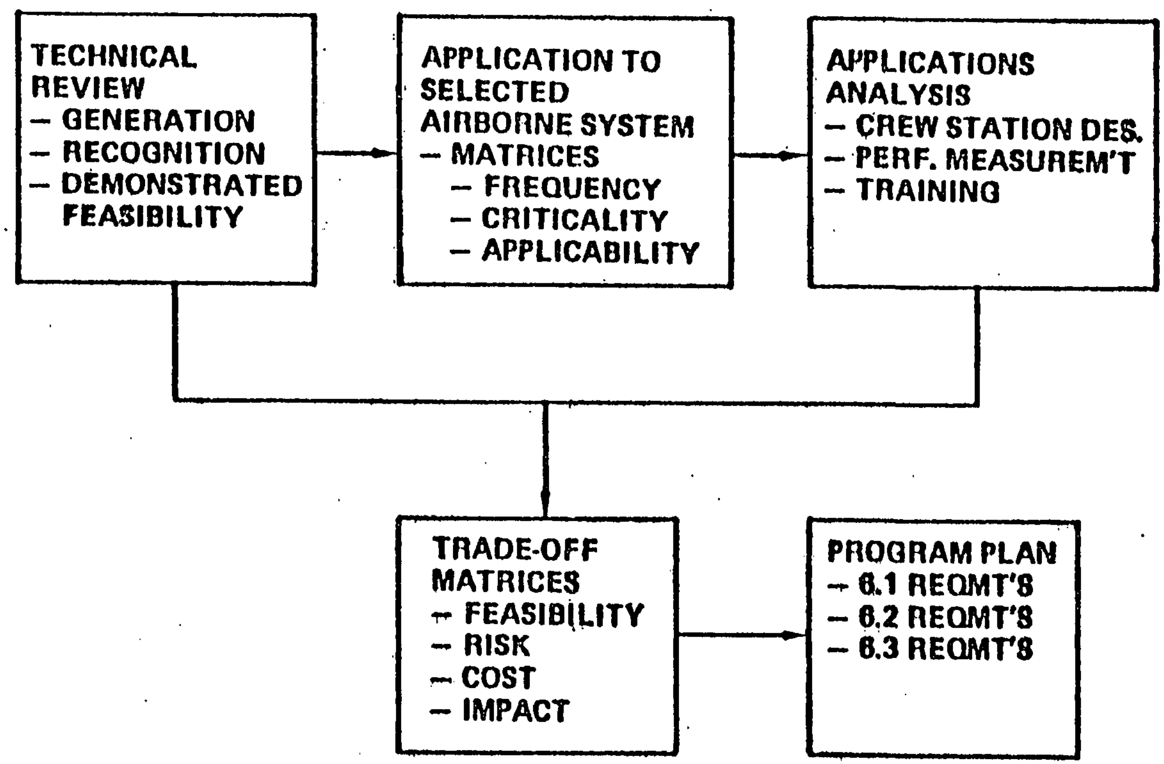


Figure 3. AST Program Plan Rationale

applicability of voice-based technologies to the specific areas of airborne controls/displays design, performance measurement, and training simulation. The general purpose of the effort is to apply AST to obtain a carefully determined voice interactive system (VIS) which, when coupled with a multi-task simulator will: 1) reduce operator loading by sharing operator display/control functions between the visual/motor channels and the verbal/auditory channels, and 2) allow direct measurement of operator performance which will result in more precise control of the training and skill monitoring processes, and a more meaningful index of operational readiness.

The general approach of this program involves: 1) incorporation of the results of the exploratory development Integrated Applications of Automated Speech Technologies effort to develop a strategy for demonstrating the suitability of applying voice to accomplish program objectives; 2) design, development, and exercising of a voice system and simulator to allow for performance of selected airborne tasks, and the provision of a capability for recording and processing operator/system voice transactions; 3) exercising the voice system and simulator to evaluate proposed task applications for cost effectiveness, contribution to system effectiveness, and operational acceptability; and 4) generation of detailed system specifications for the implementation of voice applications by program managers.

A five year development effort is spelled out in the detailed program plan. Major milestones include: 1) determination of voice system and simulator functional requirements; 2) determination of a configuration for the demonstration system; 3) preparation of the detailed work breakdown structure; 4) preparation of the detailed implementation plan; 5) development and integration of the demonstration system; 6) demonstration/evaluation of the system to determine the adequacy of each implemented task to fulfill functional requirements; 7) performance of required cost/benefit/effectiveness analyses; 8) generation of detailed design specifications for each major component of the voice system, e.g., a detailed design specification for a voice recognition element; 9) generation of a taxonomy of selected airborne tasks which can serve as a guide for the utilization and application of voice technology; and 10) transition to engineering development. This last effort calls for the identification of target or candidate systems, and system tasks, which promise high payoff for the application of voice technology. The advanced development program will also include integration of the prototype voice system with selected platforms to provide intermediate demonstrations of the utility of voice technology for target or candidate systems, and to provide for an orderly transition to engineering development.

The generalized product of this effort is an intermediately-validated voice interactive system which has demonstrated its capability for providing a solution for problems of airborne control/display design, performance measurement, and training simulation. Upon completion of the VIST program a data base consisting of: 1) a set of detailed design specifications, and 2) a set of identified airborne tasks will be available to program managers responsible for the development of major airborne systems/subsystems. Such managers can match the tasks to be performed by the operators of their system or subsystem against those tasks which have been shown to contribute to enhanced system effectiveness by the application of voice technology. More important, program managers will have available a vehicle to accomplish the implementation of voice technology for their particular purpose. That is, they will have available detailed design specifications for voice interactive systems which, if identified as system requirements, will provide sound assurance that the developed system will best reflect the maximal contribution of voice technology.

In addition, the voice integrated system and simulator capability developed during the advanced development program can serve as an in-house capability to be exercised to support a program manager in determining whether a contractor-developed voice system adheres to system design and performance requirements. A program manager will have available for his use a tool to accomplish independent system verification and validation. This approach has repeatedly demonstrated its value in the development of large-scale software and hardware systems.

While the VIST program is directed toward implementation in airborne systems, the processes and products of this program should be readily utilizable for accomplishing the incorporation of voice technology in complex operator stations in submarines, surface ships, and ground-based installations which require operator and system interfacing.

BIOGRAPHICAL SKETCH

CDR Mike Curran, Ph.D.

Commander Mike Curran is Block Program Manager for human factors engineering exploratory development, and Head, Technology Development Branch, at the Naval Air Development Center (NADC), Warminster, PA. He received his B.A. from U.C.L.A. in 1961, his M.A. in 1964 from George Peabody College, and his Ph.D. from Texas Tech University in 1974. In addition to his involvement in Navy technology development Commander Curran has served as coordinator of all crew systems support for the LAMPS MK III Weapon System at NADC. While at the Pacific Missile Test Center, Pt. Mugu, CA., he was Head, Human Factors Engineering Branch (1970-1972), and also conducted both exploratory development simulation studies of new airborne display concepts and supported the human factors engineering test and evaluation of the A-7E and F-14A Weapon Systems (1968-1972). From 1963-1968 Commander Curran was stationed at the Naval Aerospace Medical

Research Laboratory, Pensacola, FL., where he contributed to the development of an on-line, interactive test system to be employed as a research tool and as an approch for implementing secondary selection procedures. He also pursued a long-term program concerned with the effects of psychological stress on performance.

DISCUSSION

CDR Mike Curran, Ph.D.

Q: Bob Hilgendorf, Reconnaissance Strike System Program Office at Wright Patterson: You mentioned that your FY78 VIST Program is a 6.3 effort. Was your FY77 program where you were developing some of these methodologies a 6.3 effort also?

A: No, maybe it was confusing to you. The Integrated Applications of Speech Technology, i.e. the five tasks I described, will flow into our advanced development program.

Q: Hilgendorf: Okay, so it's a 6.2 effort?

A: Yes.

Q: Hilgendorf: Who is doing the work?

A: Who is doing the integrated applications work? Logicon and Boeing on a teamed contract.

Q: Hilgendorf: Okay, you were talking about the generation of some systems specifications as a fallout to your 6.3 program.

A: No, as the product, not the fallout.

Q: Hilgendorf: Okay, as a product. Do you have any time factors as to when you have pitched to your people that you are going to be producing these specification?

A: I would love to give that one to Cdr. Lane. Depending on the availability of funding, what is the current program schedule, Cdr. Lane?

Cdr. Lane: Fiscal Year 81 or 82.

Q: Hilgendorf: Okay, then my last question is even though the decisions concerning the applications to certain airborne platforms are supposed to be made in an orderly fashion, do you have any teasers as to what systems you're talking about besides perhaps P-3C?

A: My own private opinion?

Q: Hilgendorf: Yes, that's good enough.

A: V/STOL-C. And only if you know that there is a V/STOL-A and a V/STOL-B. It's a pretty safe bet. We can make an impact on V/STOL-C, if there continues to be a Navy V/STOL Program. That time frame is approximately 1990.

Q: Hilgendorf: Okay, I don't want to comment any further.

Q: John Allen, MIT: I wonder in the application you mentioned where you talked about computing the meanings of words from route words and prefixes and suffixes, what kind of vocabulary you have in mind and do you have a basis for determining those meanings from the constituent morphemes of the word?

A: Yes. I'll make one statement. Since this effort is in the processing of contracting now, I can't go into any more detail. We are not going to be dealing with the total unlimited English language dictionary of 80,000 words. Our effort will restrict itself to what we call the aviation dictionary. We think aviators do not utilize more than 10,000 words, whether they're Navy or Air Force. We have already gone though a College Dictionary and know there is a large number of words that are never used in aviation. We're talking, initially, of an unlimited vocabulary of 10,000 words. Those words, of course, include prefixes and suffixes. You may be talking about 1,000 to 2,000 parts, or what we'd call word roots.

Q: Ken Woodruff, Systems Research Laboratories: Cdr. Curran, you talked about the Boeing effort...

A: Boeing - Logicon effort.

Q: Woodruff:....tried to develop a methodology for determing high payoff applications and I believe they are using the P-3C as their test bid for that. You did not get into the criteria by which they are making those payoff decisions. Would you care to comment on that? The reason I ask the question is the P-3C is a multi-crew vehicle. We have a severe problem within the Air Force community that we are going to all single seat aircraft. The priority structure might change considerably if that were your consideration.

A: Even though Boeing and Logicon are here, I think I can comment on that since our first decision in initiating this effort was what platform we were going to use. Obviously, in the Navy we have other than one-seat airplanes. But our decision regarding a candidate platform to develop these methodologies involved do we use fighter, do we use attack, do we use multi-engine? By the way, Commander Hanson is here. This effort is jointly sponsored with the Office of Naval Research. There was considerable debate as to what type of platform we should examine. Should it be ASW, fighter, should we use helo or what? I think by browbeating or consensus, we decided on ASW because we in the Navy at NADC think we know something about ASW. Boeing has a wealth of experience in P-3's and Logicon is close to the Navy environment and quite expert in voice systems applications. So we chose something we are familiar with, but also something that is well-documented. The P-C Charley is a well-decumented aircraft in terms of what the tasks are. The other approach was we could have chosen a future platform, and tried to anticipate tasks. It would have been more difficult. So, that's why we chose the P-C Charley looking at all operators and all tasks.

Q: Unidentified Questioner (in distance): Question not recorded.

A: You are going to get into the complex meaning of the technology payoff, the utility. I brought copies of the rating process which describes in detail how we came up with the final "AST Payoff Rating." And I'll be happy to share it with you. Any other questions?

Again, just let me emphasize just one thing. You saw final one-digit ratings and heard my verbal description which was cursory. You can realize that with the P-C Charley we had available a Coarseware documentation on criticality and frequency for operator tasks. That information, combined with Boeing and Logicon's best judgment of what the technical feasibility of the application was, gave us the worth of the proposed. Any other questions?