AUTOMATIC SPEECH RECOGNITION FOR ELECTRONIC WARFARE VERBAL REPORTS

by

D.W. Moore

Project Report submitted to the Faculty of Virginia Polytechnic Institute and State University in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

in

SYSTEMS ENGINEERING

APPROVED:

R.J. Beaton

B.S. Blanchard

W.W Farley

24 June 1994
Blacksburg, Virginia

Automatic Speech Recognition for Electronic Warfare Verbal Reports

by

Dana William Moore

Systems Engineering

(ABSTRACT)

The goal of this project is to develop a top level design for an automatic speech recognition system that identifies the type of emitter report given by Electronic Warfare (EW) operators during their shipboard duties. The verbal report is one of the overt actions used to measure an operator's task proficiency. The operational requirements and maintenance concepts are defined in parallel with a feasibility study of computerized speech recognition systems. A functional analysis of the operator's task actions and the requirements for speech recognition in this task situation identify the elements necessary for the speech recognition system. To determine the functional allocation for the speech recognition system, a study was conducted to determine the baseline accuracy of a commercial system. It was determined that the PC-based speaker-independent recognition system was capable of identifying the type of emitter report given by the EW operator. The accuracy of the system was greater than 75% so the design will continue with the PC-based system used for the automatic speech recognition function.

ACKNOWLEDGMENTS

I would like to acknowledge my advisory committee for their support during this project. A special thanks to Dr. Beaton the Principle Investigator for the Shipboard Assessment of Required Proficiency (SHARP) contract.

I also want to thank other members of the SHARP program for their help: Mike Snow and John O'Shea. I wish to thank Woody Farley as well. He was not only an advisor, but provided immeasurable support and guidance during this project. The commitment of these individuals to the project made the project much easier.

Finally, to my wife and children who have missed a lot in the last two years, I wish to thank them for their patience when mine ran thin and for their support when I was immersed in this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

The goal of this work is to identify a commercial automatic speech recognition system to functionally recognize the type of radar reported by an Electronic Warfare (EW) operator. The requirement for an automatic system stems from a U.S. Navy project to measure the proficiency of EW Operators during shipboard operations.

This project examines the top level system requirements and the current technology for automatic speech recognition systems. The system requirements were analyzed through functional analysis techniques. Commercially available systems were reviewed and two systems were purchased for evaluation. One system was evaluated in terms of voice recognition accuracy to determine if it could be used in shipboard studies.

II. BACKGROUND

2.1 Development of Needs Analysis

The Shipboard Assessment of Required Proficiency (SHARP) program, was established by the U.S. Navy to set human performance standards and, in particular, performance standards for functional skills used in ship self-defense activities. The SHARP program goal, is to develop performance standards for shipboard combat systems across the fleet. The U.S. Navy does most of its training in formal courses that teach the skills required to complete assigned tasks. This skill base is continually declining over time because the skills learned are not practiced or reinforced consistently during afloat operations. There is a major concern that the reduction in personnel skills may degrade ship self-defense readiness.

The operator and supervisor proficiency are now subjectively rated by personnel assessments that are inconsistent among the different fleets. Vitro Corporation, as the prime contractor for the program, contracted Virginia Polytechnic Institute and State University (Virginia Tech) Displays and Control Laboratory to develop and validate a methodology to objectively measure operator proficiency from a human engineering prospective. Two contracts, identified in this project as SHARP I and SHARP II, were given to Virginia Tech to produce a method to quantify operator proficiency in a ship-borne environment.

During SHARP I, Beaton, Farley, McGee, and Snow (1994) developed the Proficiency Assessment Method (PAM) to measure operator proficiency. The PAM was developed through task analysis of the operator's actions on the console and the emitter events occurring on the console displays. The initial work was done by Dyess (1992) and Moscovic (1992) to identify the functional flow and timing for operator's actions. SHARP I testing validated the operator actions for functionally recognizing radar guided anti-ship missiles, euphemistically called Hostile-Air-Homing (HAH) emitters. The operator actions that provided the best proficiency measure were: time to bring the emitter parameters into view and listen to the audio tone and time to verbally report the type of emitter to the supervisor. The verbal report time provided the highest correlation of operator actions to his proficiency.

Figure 1 shows the top level functional flow of the functional recognition of HAH emitters. The functional recognition for a HAH emitter starts with the display of an icon on an AN/SLQ-32(V) display, and it ends with the operator providing a verbal report to his supervisor. To automate the PAM for use during shipboard activities, the type of emitter in the verbal report must be recognized and the time the operator made the report must be known.

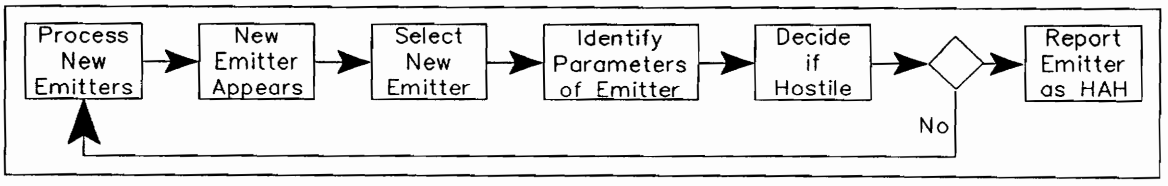

Figure 1. Top level HAH event functional flow.

The complete functional flow diagrams for the HAH recognition task are shown in Green and O'Shea (1994) with a description of each operation in this task. The measurable actions performed by the operator are physical events such as; pushing buttons, moving hands between areas on the console, and speaking into a microphone. The data extraction (DX) system, captures the AN/SLQ-32(V) console emitter events and parameters and the operator actions of key and button presses. Each action or event has a time-tag associated with the function. Because the joystick operation and verbal report are not captured by the DX equipment, the data for SHARP I were collected with a video system.

Video cameras captured both the operator's hands on the keyboard console and the emitter events on the console display as well as the audio. The hand and display actions allowed the joystick and verbal report functions to be time synchronized to the DX timed events. Researchers manually reviewed each verbal report to check for the correct response and typed the time the report occurred into the PAM database. This process of analysis was very time consuming and impossible to accomplish in real-time with the equipment that was available. A method was required to automatically capture the time and recognize the verbal report.

The verbal report generally provides three pieces of information in addition to the time. This information is; the emitter classification or type, the bearing either true or relative and other parameter information describing the emitter. The type of emitter and parameters may be redundant information. PAM requires; the time the verbal report was stated and whether the emitter was correctly identified as the emitter that the operator has in close control. The verbal reports often follow a semi-structured format for example " Steady scan, Bearing 235, Inbound missile." Discussions with subject matter experts (SMEs) and review of the SHARP I data shows that the semi-structured report is used 80 to 90 percent of the time for the HAH emitter, but for other events this percentage varies greatly between operators.

On the ship, the structure of the report is dependent on the enforcement of communication discipline. The unstructured reports contain the same information, but in different order or using words that may not follow the protocol developed for the report. This causes difficulty in recognition and misinterpretation by the supervisors and would be more difficult for the speech-recognition system.

The speech-recognition requirements for the Vitro SHARP II contract are: (1) identification of commercial vendors of voice recognition systems, (2) conduct an engineering evaluation of commercial systems for use in the SHARP laboratory and shipboard studies, (3) selection and development of a single voice recognition system (including hardware, software, and all supporting interface equipment) for use in the SHARP program, and (4) demonstration of the selected voice timing system for the SHARP proficiency standards requirement (Beaton 1993).

This report provides a structured method of providing an engineering evaluation to identify an automatic speech recognition system for use on a ship. Figure 2 shows the general methodology used in this project to identify the top level system requirements and to identify the feasibility of allocating the speech recognition function to a computer. The design process begins with the customer identifying a need, either real or perceived, for the system to serve a useful purpose. In this case, the need is to provide an automatic speech recognition system that will work in conjunction with the components required for the PAM model implementation.

All of the requirements are derived initially from the systems operational requirements and system maintenance concepts. In this project, a feasibility study is required to identify the technological feasibility of computerized speech recognition system and to identify the terminology to properly specify the system requirements. The trade-off studies are used to determine the functional allocation of the speech recognition system and to provide input to the production system design. Feedback as the system develops is critical to the process.

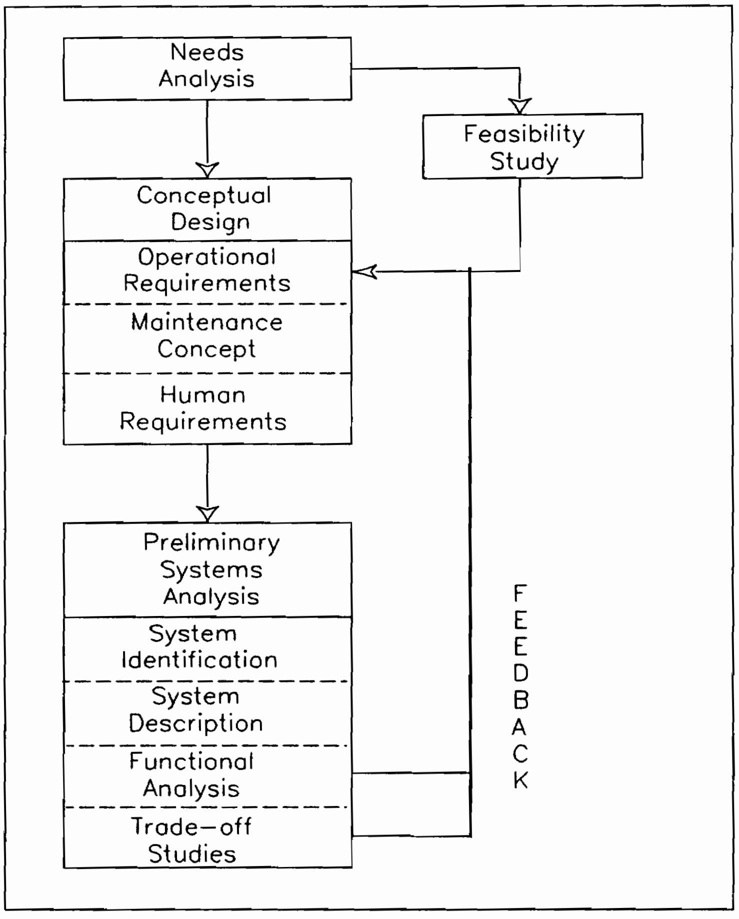

Figure 2. Methodology used in project.

The restriction that applies to this report is that it must be unclassified. Because the operational verbal reports are classified to the secret level, simulated phrases are used in the system specification and actual feasibility testing. The development and demonstration of the production system will require classified documentation. This project will remain unclassified and the detailed design will be accomplished in a follow-on project. The requirements can be specified except for the actual lexicon in the verbal report and the parameters used for the emitters.

In order to identify the system requirements, it is necessary to define the system. Figure 3 shows the system components and the interfaces between the system and the environment. The operator is not shown in the figure, but is a critical component of the system. The general description of the components follows, with more in-depth descriptions are provided later in the report. The interrelationships for each of the components are shown by the arrows in Figure 3.

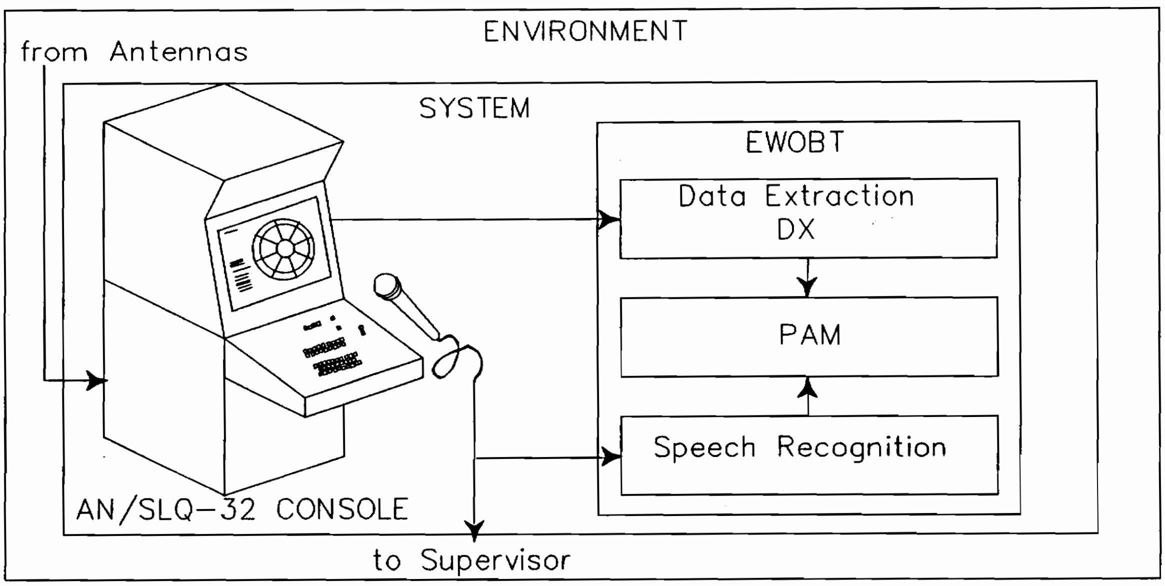

Figure 3. System components except operator.

The AN/SLQ-32(V) console provides the operator with the emitter parameters and bearing used to recognize and track an emitter. The console has a keyboard, fast action buttons (FABs) and joystick to select different options on the screen in the center of the console. The console also contains a speaker to listen to the emitters and a microphone to pass the verbal reports to the supervisor. The Electronic Warfare On-Board Trainer (EWOBT) is a ruggedized 486 DX computer. The EWOBT is new and part of the Navy's plan to reinforce training in the shipboard environment. It will stimulate the AN/SLQ-32(V) so the operator can train on the ship. The data extraction (DX) is either a card within the computer or a rack mounted component. The PAM software program, presently in the design stage with testing to begin shortly, is hosted in the EWOBT. The speech recognition subsystem must interface with the operator through a microphone or input device and with the EWOBT through an output device to provide the time and emitter type to the PAM model. The speech recognition is not necessarily a part of the EWOBT but must interface and have time synchronization with the EWOBT.

The operator, while not shown in the figure, is a critical component and the reason for the proficiency model. The operator's actions, including the verbal report, must be correctly measured in order to gauge his proficiency in the real-time operation of the AN/SLQ-32(V) console. A system, to automatically recognize the type of emitter the operator is reporting, is essential for the PAM to operate in a real time environment.

Speech recognition, being a relatively new technology, required a review of the current literature in order to determine the different capabilities and terminology used to identify the functions necessary for the system definition and requirements phase. A feasibility study was necessary to identify current technology and to understand different terms used for speech recognition. The study covered the general application of speech recognition on computers and how the recognition capability is implemented. The study was prepared in parallel with the conceptual design.

III. FEASIBILITY STUDY

3.1 Speech recognition review

3.1.1 Speech

Speech is a stream of utterances that produce time varying sound pressure waves of different frequencies and amplitudes. Speech recognition takes these sound waves or acoustical waveforms and derives a corresponding sequence of discrete units such as phonemes, words or sentences. Computer-based speech recognition systems often model human speech recognition, but the computer-based systems do not have the flexibility and capacity of humans to understand speech.

3.1.2 Computer based systems.

Computer based systems use a two step process for speech recognition. In the first stage, the computer receives speech input, the signal is converted from an analog signal to a digital signal in a digital signal processor (DSP). The DSP conversion produces a digitized representation of the acoustic signal. Most systems use a vector quantization (VQ) representation, this is a two dimensional array of the time and frequency components of the acoustic signal. The VQ representation is used as algorithms have been produced that reduce the amount of data storage and computation time. In the second stage, the digital signal is compared to digitized speech patterns stored in databases. Much of the focus in speech recognition is how to efficiently search these databases to increase the recognition accuracy and decrease recognition time. The database searches use statistical algorithms to find the most likely match for the vector representations to the stored patterns. The databases can also contain contextual, grammatical and syntactical rules and the search would produce the most likely combinations of the patterns to form words, phrases or sentences. More rules require a longer time to recognize the speech, but produce more accurate recognition.

The vector pattern databases normally contain acoustic parameters or vectors of words or smaller units of speech called phonemes. For a word based system, each word must be in the databases to be recognized. If the vocabulary is 10,000 words, the database must contain the same number of vector representations.

A phoneme is a discrete unit of speech analogous to letters in a written word. There are 13 phonemes for vowel sounds, 25 phonemes for consonant sounds and a number of diphthongs (sounds like oy in boy and ou in about) in the English language

(Sanders and McCormick 1993). Word recognition is accomplished by combining the phonemes into words. The database of phonemes is much smaller than the word based systems, but generally requires more computational resources to combine the phonemes correctly into words. The phoneme based recognizers often replace word based recognizers as training requires a relatively few number (40 - 50) of phonemes.

In order to add a word to the phoneme based system, the products have text-to-phoneme conversion algorithms that allow direct input of words from a keyboard. For a word that is spelled phonetically, it is typed directly and the system does the phoneme conversion and insertion to the data base. To add a word such as "light" that is not spelled phonetically, the phonetic spelling "lite" is used. Some systems require the phoneme representations be typed as in "L IH T" which then converts the phonemes to the vector representation.

Because there is no standard terminology among companies, it is difficult to compare one type of system with another. Rabiner and Juang (1993) claim that the recognition of phoneme based systems is lower than word or pattern based recognizers, but they were comparing a large mainframe systems to the capabilities of workstations. PC systems, using phonetic recognition, claim recognition capabilities of 95 percent which is comparable with word based systems.

BBN Systems and Technologies identified some of the advantages of a phoneme based approach as: a) results are based on probability likelihood algorithms rather than comparison measures of the similarity of acoustic templates, b) time and acoustic variability leading to noise immunity are modeled automatically, based on training, c) additional training can improve performance, d) new words can be easily added without the requirement for a large number of speakers to train each new word and e) speech can be speaker-dependent, speaker-adaptive or speaker-independent.

3.1.3 Speech types.

There are three speech types: discrete or isolated, continuous and connected. In discrete speech, the system recognizes the word or phrase only as one unit. Words or phrases can not be spoken together but must be spoken discretely, delineated at the beginning and end by a silence or identifiable end points. The silence that brackets the utterance, provides a start and an end point, allowing the system to divide the utterance into separate units for easier recognition. Discrete recognizers work well for command and control functions to direct computer driven operations such as industrial and computer assisted settings to control machine actions. Many systems for Windows and multimedia control, use discrete word recognizers. These systems are inexpensive and work well for command driven applications using menu type formats.

Continuous speech is natural or conversational speech with words spoken in a flow. Speech, when presented in a continuous stream, is more difficult to recognize as there are no pauses between the words or phrases. The recognizer must "guess" where one word stops and another begins. This guessing is the statistical analysis of the combination of the digital vector representations to produce the most likely word or the most likely words to produce a grammatically correct sentence. Once the words are recognized, they can be compared to context or grammatical rules to see if they make sense in forming a grammatical sentence or follow a syntactical form. When words are spoken in a continuous stream, the acoustics of neighboring words are often changed which must be included in the syntactical rules. Search algorithms and grammar modeling can improve the recognition capability in continuous speech systems but at the cost of more computational power.

Acero (1993) identifies the Hidden Markov Model (HMM) algorithms as the most widely used statistical model for continuous speech recognition. An HMM is a statistical model that uses two transitions between states to search quickly through a database. Each transition has two sets of probabilities; the first is a transition probability of going to the next state and the second is a probability density function which defines the conditional probability that a word is correct. This type of statistical search pattern reduces the search time and improves recognition accuracy. Neural-network algorithms are generally used for grammatical and syntactical database searches providing the contextual information for word identification in continuous speech.

Connected speech is a combination of both discrete and continuous speech. Some companies use connected and continuous synonymously. Connected speech is a continuous phrase that has a set format or syntax. Silence or some other recognizable end points surround this syntax. The syntax is fixed in its format but may contain different combinations of words. Figure 4 shows a representation of a connected speech syntax. The sample utterances show some of the combinations that would be recognized and could be used to produce different outputs for a computer application. The syntax could be in the format of a operator's verbal report with type, bearing and classification information. The syntax in Figure 4 has about 250 distinct utterances using the four object words and the ten digits in different combinations with the commands open, close, and check.

Example Syntax
S>{open|close|check} OBJECT (number) DIGIT OBJECT = window, file, device, channel DIGIT = number 1 through 10
Sample Utterances open window three close file number four check channel five

Figure 4. Connected word syntax example.

Key-word spotting is a term used to describe the recognition of individual words in continuous speech. Key-word spotting is a subset of continuous speech recognition where only specific or key-words are recognized. For a particular application, a word or phrase would initiate an action and this action could be employed to turn on the recognizer or to switch vocabularies for application specific words. For instance, "Computer turn on the radar screen", computer is the key-word to tell the recognizer to listen as an action will follow. Radar also a key-word turns on the radar screen and switches the speech recognition system to the radar specific word vocabulary.

It does not matter if the keyword is surrounded by silence or other words. As each word arrives it is compared to a small set of key-words. If there is no match in the initial comparison, the recognizer rejects the word and goes to the next word. When the comparison is close the recognizer searches further into the database in order to recognize or reject the word. If the word is recognized as a key word, a program can be set-up to identify the key-word and to initiate a computer action. Key-word recognition reduces the computational time by reducing the number of words that must be recognized. Figure 5 shows a graphical representation of how key-word spotting works.

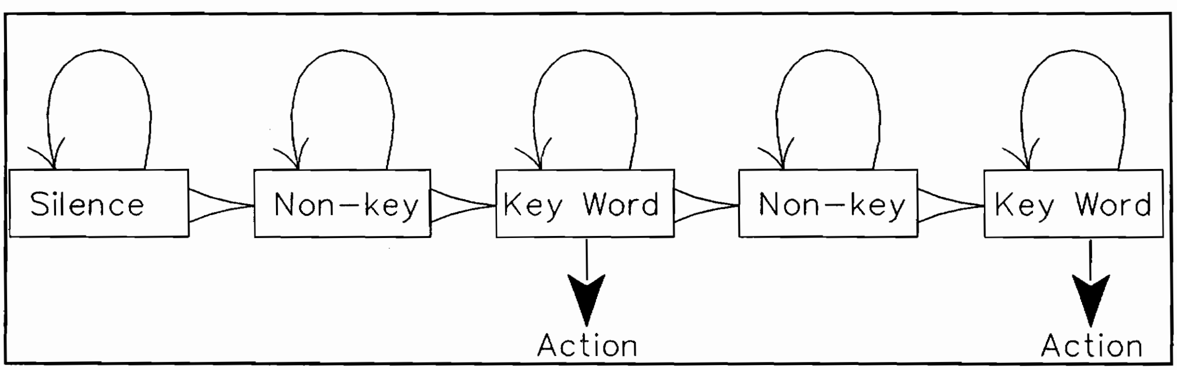

Figure 5. Key-word spotting representation.

3.1.4 Speaker dependency.

There are three types of speaker dependencies; speaker-dependent, speaker-independent and speaker-adaptive. Speaker-dependent systems require individuals using the system to train or enroll each word or phrase by repeatedly speaking the words and phrases until the system recognizes the individual's voice pattern. This recognition can be very good for these systems because the voice patterns are matched to an individual.

Speaker-independent recognizers will recognize speech from all users as training is done by the speech recognition company with a large number of speakers. These people produce the variability required for the system to recognize all speakers. Most of the PC systems, with connected or continuous speech, use phoneme based systems since relatively few phonemes have to be trained.

In speaker-adaptive systems, a generic vocabulary is supplied with the system. The user adapts the system to his or her voice by enrolling each word or phrase typically one to five times. This type of system may recognize other personnel but the individual who trained the system will get much better recognition accuracy. More enrollment generally improves the recognition accuracy and many systems, dependent or independent speaker, use a speaker-adaptive system.

3.1.5 Lexicon.

The lexicon includes both the word in the application dictionary and the format or structure of the phrases for the application. This lexicon must be defined for each application that the speech recognition system is expected to recognize. The dictionaries, provided with the recognition systems, can contain words in the hundred of thousands but the number of words available for each application is limited to a finite number. The number of words between for each application is fifteen and five thousand depending on the capabilities. For key-word spotting the number is generally small fifteen to one hundred, but for discrete speaker-dependent applications that the dictionary can be large. Each of the words to be recognized must be included in the application dictionary and written in the system's syntax then compiled in order for the system to recognize the word or phrase.

The connected speech system requires not only each word be included in the dictionary, but the format of each phrase must also be specified in a syntactical format. The syntax may be written to allow a large number of combinations and permutations of phrases but the larger the syntax possibilities the more time that is required for recognition. Each product requires a different method of specifying and compiling which is much like computer programming to specify the most "elegant" and efficient syntax. The identification of the lexicon and format by the user is very important early in the design process in order to identify the format and number of words and phrases that must be recognized.

3.1.6 Noise.

A large problem with all speech recognition systems is the degradation associated with the environmental noise and input noise. Acero (1993) identifies some of the factors that influence recognition as: input level, additive noise, spectral tilt, physiological differences, and interference by the speech of other speakers. Different techniques increase the robustness such as: using different microphones, using arrays of microphones, inserting ambient noise into the quiet training data and using robust algorithms. All these methods increase the noise immunity or robustness of the discrete recognizers (Das et al 1993)(Hansen 1993)(Acero 1993). However; all of these techniques have had only limited success in operational settings. Noise rejection or robustness is currently being addressed with tailored or customized systems.

The noise is closely related to the input devices used for speech. Systems often require input devices or microphones that are specific or very similar to those used in training the system as inputs from different microphones will adversely effect recognition accuracy. Systems requiring highly directional inputs and noise cancellation may not work well with narrow frequency range microphones typically found in telephone and intercom systems.

3.1.7 Board Requirements.

As previously stated, speech-recognition is accomplished in two stages; an acoustic signal conversion to a digital vector representation and a database search to compare digital input signals to stored data. PC systems use two methods to implement this recognition, an audio card and a DSP board. In the first case, digitizing occurs on a sound card and the search algorithm processing on the host computer. This can slow down both the computer and the speech-recognition functions. The second method uses digital signal processing (DSP) chips on a board to provide both functions with the board optimized for speech-recognition. Large vocabularies and continuous speech require a large amount of processing time so it is best to use a DSP board in the computer to decrease recognition time and to reduce the workload on the host computer.

3.1.8 Application Development.

Speech recognition systems were generally designed to work with specific applications such as; dictation, control of the software applications, telephone operations and operation of industrial processes. Application kits are now available to allow developers and users to write and compile programs for different applications. The developers write Application Program Interfaces, usually in Visual Basic or "C" language, to satisfy their particular application requirements. These development kits vary in complexity from the ability to add new words to being able to write programs that allow speech recognition for any computerized application. The development kits decrease the development time for new applications or changing the vocabularies.

IV. CONCEPTUAL DESIGN

4.1 Development of System Requirements

The design to integrate components into a system requires the definition of the functional, support and operational requirements to determine where, how and when the system will be used and supported. The operational requirements and the maintenance concept provide the framework to identify the requirements for the life cycle of the system. The speech recognition system will be integrated with existing components so many of the requirements will match similar components within the system.

The system requirements define the top level operational and support areas providing a common baseline for all of the design efforts. This report shows the macro level functions that the operator performs to provide the verbal report to the supervisor. The speech recognition functions are then decomposed to identify each function a speech recognition system must perform. These individual functions identify the information needed for the detailed design. The functional allocation is not accomplished until the functions have been clearly defined and the decision is then made how to allocate the function; machine versus operator, PC computer versus main-frame, hardware versus software. The trade-off study in this project will determine if the automatic speech-recognition function can be allocated to a PC-based system.

4.1.1 Operational Requirements

4.1.1.1 Mission Profile. The AN/SLQ-32(V) system is deployed on over 400 ships of different classes. Each ship has one or more missions to perform which require different mission profiles. A typical mission profile, modified from Dyess (1992) for an Aegis class CGN Guided Missile Cruiser, is shown in Figure 6. The life cycle for the ship is twenty-five years with four mission profile cycles and a possible mid-life update after the second mission profile increasing the two-year refit time. For the two-year refit period, the ship is in dry-dock for refueling, major repairs and modifications. The mission segment at the bottom of the figure shows the time spent in training, operations and maintenance for each eight and a half-month mission segment.

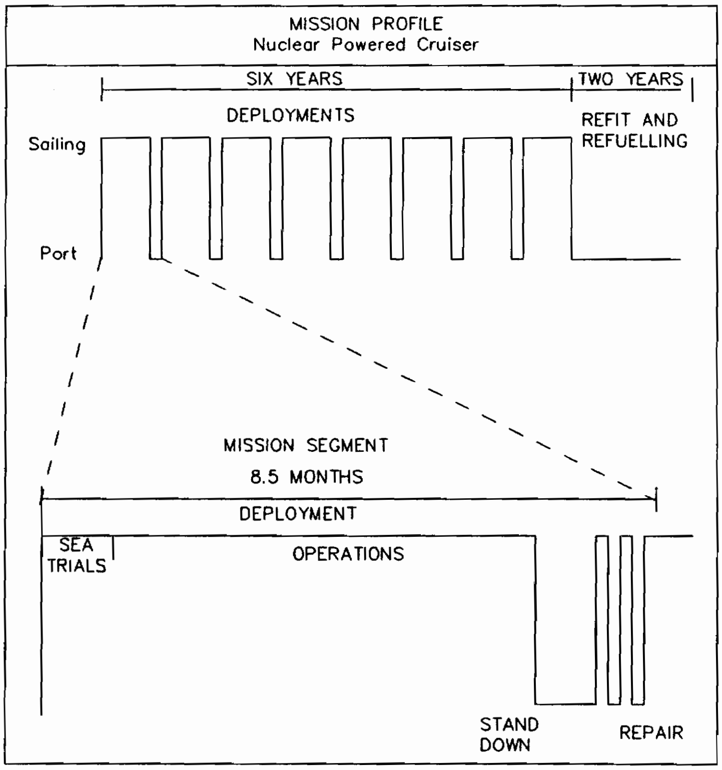

Figure 6. Operational profile of a CGN Guided Missile Cruiser.

4.1.1.2 Operational Life Cycle. Table 1 provides the breakdown of the operating hours within each mission segment. These data are essential for the system designers to determine the life cycle support and administrative requirements for the components of a system. The operational life of the AN/SLQ-32(V) is programmed until the year 2020. The EWOBT equipment is assumed to have the same life cycle. The speech recognition system will operate with the AN/SLQ-32(V) and EWOBT so has the same mission operational characteristics. The speech recognition component is expected to be available for January 1996.

Table 1 - Mission Segment Profile for 8.5 Month Operating Cycle.

1. Mission Breakdown	Percentage of time
Operations	71
Maintenance and calibrations	23
Training	8
2. Operating hours during mission segment	
Sea Trials	24 hours/day x 15 days
Deployment	24 hours/day x 180 days
Maintenance	8 hours/day x 40 days
3. Total operating time for the mission segment is 5000 hours	

4.1.1.3 Environmental and Power Requirements. The power and environmental requirements for militarized performance are identified in Gumble (1988) and the specifications for computers are shown in Table 2. The AN/SLQ-32(V) meets the militarized or class three requirements for ship-borne operation. The EWOBT as defined in the Systems Manual (1992) is a "ruggedized commercial computer designed for the rigors of onboard ship and ashore commands". The EWOBT system meets the commercial requirements shown in Table 2, but does not meet all of the ruggedized specifications. A computerized speech recognition system must meet the requirements for a commercialized computer system. The environmental parameters provide the system designers with the minimum design requirements in selecting a production system.

4.1.2 Maintenance Concept.

The AN/SLQ-32(V) uses a typical three level maintenance concept; operational, intermediate and depot. Blanchard and Fabrycky (1990) provide a detailed description of each level's breakdown. Basically, organizational maintenance is performed on-board the ship. The maintenance is limited to adjustments, fault isolation with built-in-test and removal of line replaceable units. Intermediate level repair is card or component changes to the system. Little intermediate maintenance is conducted on the AN/SLQ-32(V), it is mainly a supply support function between the depot and organizational levels. The depot level repair and overhaul is completed at a Naval Repair Depot or a contractor's facilities. This is the detailed repair to correct all the faults to the component level.

All EWOBT repair is conducted at depot level. The speech recognition system would also be repaired at depot levels as the ships do not have the capability of repairing integrated circuit boards nor the test equipment available on each ship. The speech recognition system requires integration into the supply network through an Integrated Logistic Support (ILS) analysis. It is assumed that any speech recognition system hosted on an EWOBT would be integrated with the EWOBT Repair Facility in Virginia Beach VA.

Software upgrades generally fall under the maintenance concept for smaller systems. Upgrades provide a means of meeting changing operational requirements but they must follow the configuration management control guidelines. The software modifications for the speech recognition system must be coordinated with any changes to the PAM and EWOBT software and vice versa. The software configuration management must be controlled for all of the software within the system and through the different system interfaces. For instance, if the software is set-up for male operators and at a future date female operators are introduced, the software may have to change to accommodate this change in the interface requirements. Small changes in the software can have large effects on the life cycle costs and operational capabilities. The software development creates a new system with all new support, training, and deployment requirements of a hardware system.

Table 2 - Environmental requirements for computers for different applications.

Computer Class	Class 1 (Commercial)	Class 2 (Ruggedized)	Class 3 (Militarized)
Temperature (^{o}C) Operating Not Operating	 10 to 40 -40 to 70	 0 to 50 -40 to 70	 -50 to 85 -50 to 85
Humidity (%)	10 to 90	5 to 95 (non-condensing)	5 to 95
Altitude (feet) Operating Not operating	 -1000 to 8000 40,000	 -1000 to 10,000 40,000	 -1000 to 50,000 50,000 +
Vibration Operating Not Operating	 0.8g rms /5-300Hz 1.2g rms / 5-300Hz	 3.5g rms /5-2000 Hz 5.4g rms /2-2000 Hz	 4.5g rms /5-2000 Hz 9.0g rms /5-2000 Hz
Shock (11 millisec) Operating Not Operating	 10 g 15 g	 20 g 30 g	 20 g+ 30 g+
EMI/RFI	FCC Class 15 part J	Mil-Std 461B	Mil-Std 461B
Tempest	Cabinet level	Chassis level	Chassis level
EMP	Cabinet level	Chassis level	Chassis level
Radiation Hardening	No	No	Optional
Sand/Dust protection	Cabinet filters	Chassis filters	Sealed Unit
Salt Fog protection	No	Yes (conformal coating)	Yes
Environmental Stress Screening	Typically none	Temp: 0-50 ^{o}C Vibration: $0.10g^2$/Hz	Temp: -50 to 85 ^{o}C Vibration: $0.04g^2$/Hz
Quality Standards	Variable	Mil-Std 45208	Mil-Std 9858

4.1.3 Human Requirements.

Woodson et al (1992) point out some of the environmental conditions that must be understood when designing ship-borne systems that humans will operate. The ship is constantly in motion often causing seasickness that will interfere with task performance and cause safety problems. The equipment vibration and noise due to the close quarters requires adequate control. This is true, not only in the working spaces, but also in the living and sleeping spaces. The weapons systems on-board cause a special hazard and noise conditions that must be controlled, however; the performance of the weapons should not be compromised. The atmospheric conditions are important to maintain the crew in an efficient and healthy state.

Space is always critical on a ship. Much can be done to increase crew effectiveness if the requirements are kept in mind during ship design. The main areas of concern are head clearance, cramped living and working conditions, passage way clearance and space organization. The operators work within the confines of the CIC and the speech recognition system must function in this environment. The environment is relatively noisy and the operators are continually under the stress of shipboard living. The system should recognize the speech from the operator under all of the operational conditions.

At the present time, all of the operators are male with a high school education and are between the ages of 18 and 40. Each operator goes through a number of training courses to provide the skills and knowledge necessary to operate and maintain the AN/SLQ-32(V) system. The operators generally maintain four hour watches on the AN/SLQ-32(V), but may stand watch as long as six to eight hours. All of the operators speak English and are trained in communication and radio protocol. The operators perform may tasks besides the functional recognition of the HAH emitter. These include tasks such as; classification and identification of other emitters, launching countermeasures, passing reports to other ships, updating libraries with new parameters and reporting changes in the emitters to his supervisor.

There are a number of system requirements for an automatic speech recognition system. An automatic computerized speech recognition system must operate with the EWOBT computer. The input should be from a microphone currently on the AN/SLQ-32(V) console. The recognition can be for either a structured report or a non structured report but the accuracy requirements are not identified at this time as the PAM specifications have not been identified. The trade-off study used an accuracy of seventy-five (75) percent which is explained later. The system must operate in the shipboard environment and be integrated into the logistical support requirements, similar to the current components of the design system. The system must recognize reports from all the different operators and must not require the verbal reports to be changed to facilitate recognition of the emitter type. The timing component of the verbal report will not be addressed in this report, but is required for the production system.

V. PRELIMINARY SYSTEM ANALYSIS

5.1 System Identification

Blanchard and Fabrycky (1990) define a system as a set of interrelated components working together for a common purpose or objective. Figure 3 showed the components of the system with the exception of the operator. The interrelationship of the components must be understood in order to design the interfaces and subsystems incorporated into the system as well as the interfaces between the system and the environment. The environment to system interfaces allow the flow of information, material or energy across the system boundaries. The system for this project is the AN/SLQ-32(V) Electronic Warfare Display and Control Console (DCC), the EW operator, the EWOBT computer, the PAM software, the data extraction and speech recognition subsystems.

In order to understand the function of the system, it is necessary to look at how and why the system works within its environment. The AN/SLQ-32(V) system's primary function is to provide support to the ship's self-defense detection capability against low flying anti-ship missiles. The operator provides the knowledge and skills necessary to identify the different emitters from the parameters and alerts displayed on the DCC. The operator's proficiency, for correct and timely reporting of the emitters, is measured by the PAM system through the speech recognition and data extraction systems.

In order to provide insight into the requirements of the operator interface with the AN/SLQ-32(V) console, a description of anti-ship missiles and incidents are discussed to demonstrate the time constraints and criticality of the correct recognition of the emitters. This overview is provided to show the time, stress and operational conditions the operator and the speech recognition system would be exposed to in a missile engagement. The anti-ship missiles are the main reason that the U.S. Navy wishes to measure the operator proficiency so the missile is not misrecognized or missed by an operator who is below a standard proficiency level.

5.2 Anti-ship Missile Development

5.2.1 The threat.

Anti-ship missile development has changed the face of naval tactical warfare by increasing the vulnerability of warships as no other naval weapon has previously accomplished (Baranauskas 1988). The weapons are launched from over-the-horizon and approach the ship at sea-skimming altitudes making detection very difficult and requiring a very fast response. The newer generation missiles can reach supersonic speeds. Most of the missiles use radar guidance for navigation with infrared used for backup.

The typical anti-ship missile flight profile launched from an aircraft shows the task facing an EW operator sitting at a DCC. The aircraft will fly, into the area under the radar horizon, to a range of 15 to 25 miles. The aircraft will then "pop up", get a radar fix on the ship and drop back down below the horizon. (The AN/SLQ-32(V) will recognize and display this emitter as a hostile airborne targeting threat.) The position of the ship is fed to the missile inertial navigation system and the missile is launched in a "fire and forget" mode. The missile drops down to a sea-skimming altitude (approximately eight feet), flying under its own guidance system and a radio altimeter. As the missile approaches the ship it rises slightly to scan the horizon and locks its terminal homing radar onto the target. (The AN/SLQ-32(V) will display the emitter as a missile or hostile air homing threat). From the radar horizon formula, the EW system would be able to pick out the missile emissions between ten and fifteen kilometers. The missile flight time for a ten kilometer lock-on (Mach .93) is thirty-five seconds and for fifteen kilometers the time is fifty seconds under ideal conditions. The operator must recognize the emitter as a threat and report the emitter to his supervisor, as quickly as possible, so defensive measures can be taken to defeat or counter the missile.

There are three significant incidents that show the nature of anti-ship missile warfare. A synopsis of each is included to show the pressure and time constraints that an EW operator is exposed to during a missile attack and how the proficiency of the operator can greatly influence the HAH detection. The incidents are; the Falklands Islands War in 1982 between Great Britain and Argentina, the USS Stark attack in 1986 by an Iraqi Mirage and the USS Vincennes downing of an Iranian Airbus in 1988. Each of these incidents indicate the criticality of the EW operators and why the Navy is concerned with any degradation in the operator proficiency. These incidents show the operational environment of the CIC and the constraints imposed on the speech recognition system to understand the operator's verbal report.

5.2.2 Falklands Islands War.

The Falklands Islands provided the first significant use of these missiles in a combat situation. The missiles used by the Argentineans were the French built Exocet. Of the six missiles known to be fired, four hit ships, destroying two ships; HMS Sheffield and HMS Atlantic Conveyor and damaging one; HMS Glamorgan. Only one of the missiles that hit the ships detonated, the resultant damage from the other missiles was caused by the unexpended fuel exploding (Walker 1983). An interesting observation from the Sheffield is the passive radar detection system (similar to the AN/SLQ-32(V)) detected the missile, but automatically classified it as friendly, because the British owned Exocet missiles.

The Secretary of the Navy, John F. Lehman, Jr. pointed out at the time that it was not surprising that one lone ship on picket duty with inadequate anti-missile defense and air cover could be destroyed by a missile of the Exocet capabilities. He went on to say that "if an American ship had been the target of the Argentine attack , it would not have gotten anywhere near the kill range." (Walker 1983). This statement proved not entirely to be accurate as is shown in the next incident.

5.2.3 USS Stark Incident.

As a result of the Iran-Iraqi War, the supply of world oil was threatened and President Reagen in 1986 ordered the reflagging of eleven Kuwaiti super tankers. This allowed the ships to come under the protection of the US Navy. It was reasoned that the U.S. Navy could take action to stop the destruction of merchant shipping by providing an escort for the tankers through the Persian Gulf and out to open sea. On 17 May 1986, the USS Stark was struck by two Exocet missiles, fired from an Iraqi Mirage F-1 aircraft, resulting in 37 dead and 21 wounded sailors (Waters 1988).

Figure 7 adopted from Dyess (1992) and Committee on Armed Services (1987) represents the time and distance involved in an air engagement with a sea skimming missile. While the aircraft was known to be approaching the area for over an hour, the USS Stark acquired the aircraft on radar about fifteen minutes before the missiles hit. The EW operator detected the aircraft radar in search mode five minutes later. Shortly after detection, the EW operator observed a shift in the Mirage radar signal from search to tracking mode. Tracking mode is used to identify the ship's location for the missile inertial navigation and guidance system. The EW operator verbally reported the lock-on by the aircraft to the Tactical Action Officer, who is in charge of the Combat Information Center (CIC). The time between the lock-on and the impact of the first missile was approximately three minutes. The second lock-on by the aircraft happened approximately ten seconds before the impact of the first missile (Committee on Armed Services 1987).

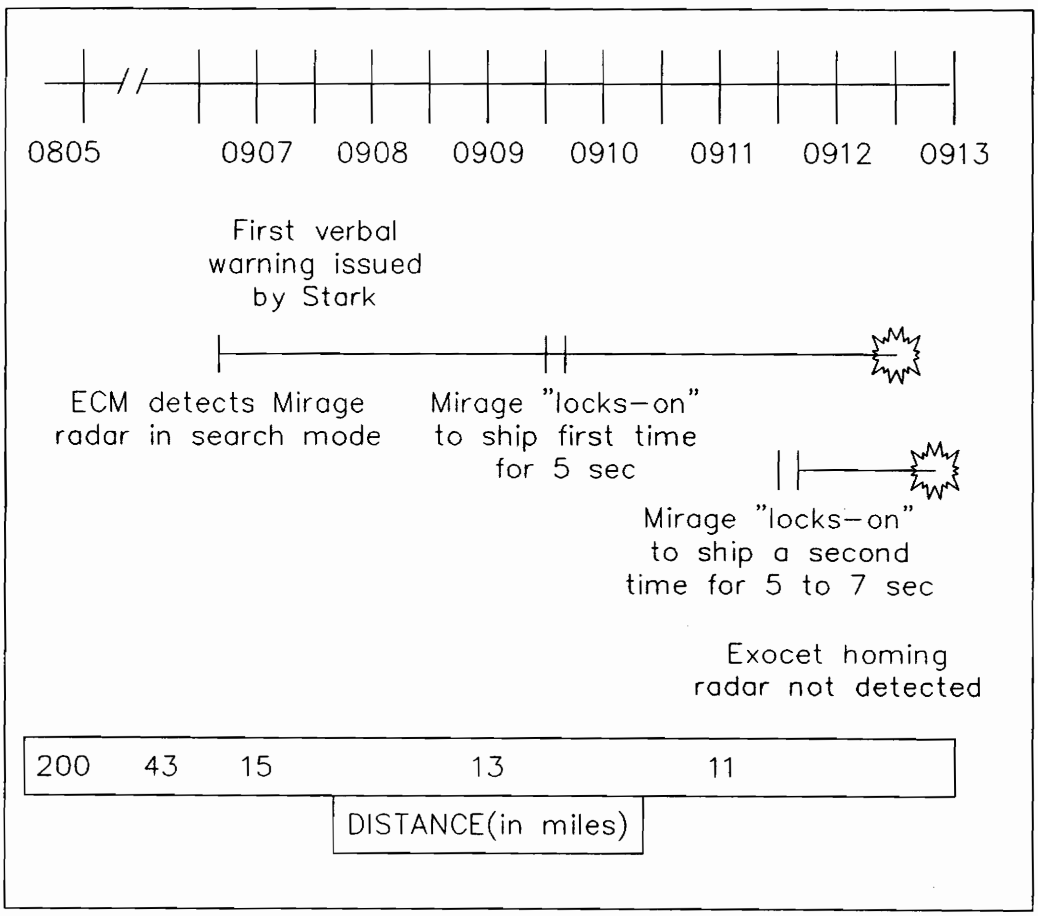

Figure 7. Time Line of USS Stark missile engagement.

The emissions from the Exocet missile were never detected by the EW operator. There are two reasons the operator did not detect the missile; the AN/SLQ-32(V) may not have discriminated the emission because the missiles approached the ship from the bow where there is an antenna blind spot, and the operator missed the missile icon on the display because he had inhibited the audio alarms that warn of a missile attack.

5.2.4 USS Vincennes Incident.

This incident deals with the incorrect identification of a civilian airliner and subsequent destruction with a loss of life of over two hundred personnel (Committee on Armed Services 1988). The aircraft was incorrectly identified as an Iranian F-14 that was in an attack profile against the Vincennes. Jaska (1988) describes the incident in detail but I will cover only the salient points

An Iranian Airbus flying from a combined military/civilian airport was climbing to a commercial airway, that would mean flying directly over the USS Vincennes position. The Vincennes, at the time, was engaged in a battle with Iranian gunboats. The Vincennes' identification supervisor misinterpreted the identification friend or foe (IFF) code on his display and the Airbus was misidentified as an Iranian F-14 aircraft. The aircraft was repeatedly warned on the international distress frequency to disengage and not over-fly the Vincennes. The Airbus continued to climb as it approached the ship, but it was reported by the radar operator as descending, possibly because of the transposition of the altitude and range displays. The EW operator detected no emissions from the aircraft which would identify it as a commercial carrier and the commanding officer concluded that his ship was being attacked and launched missiles to destroy the aircraft. There are a number of other factors that lead up to the incident that increased the belief that the ship was under attack such as; an intelligence briefings that indicating that an attack was being planned, Iranian gunboats attacking the Vincennes' helicopter and an Iranian P-3 aircraft used for radar targeting approaching the Vincennes location during the incident.

For this report the most important observation is that the EW equipment is often the only method of identification for long range targets. The EW operator is often the sole individual that can identify an aircraft by the type of emitters used in the different modes. The current technology sensors can guide the weapons with pinpoint accuracy but can not identify what is being destroyed. In the investigation, the committee on the Armed Services found that aircraft, both military and civilian, often flew in the area without any navigation or search radar because of the high volume of chaff present in the air.

5.3 System Description

5.3.1 AN/SLQ-32(V)

This section describes the AN/SLQ-32(V), EWOBT, PAM and EW operator verbal report as these subsystems interrelate with the speech recognition system. The AN/SLQ-32(V) is a passive electronic detection and countermeasures system used in a support role for the ship defense. The system, generally referred to as the "Slick-32", uses instantaneous frequency measurement for emitter detection and lens-fed multiple beam antennas to discriminate electronic radar signals in all radar bands (Jaska 1988). The system contains a library of different emitter parameters used to automatically identify the emitter types. The system can provide bearing but not range data as detection is passive.

The AN/SLQ-32(V) has five versions all manufactured by Raytheon and each version has expanded capabilities from the previous versions. Jane's Radar and EW systems (1991-1992) provides a listing of the functions of the different versions and which classes of ships have the different versions. While the system capabilities vary, the DCC's are the same externally with operator input/output devices consisting of a cathode-ray display, light emitting diode indicators, an alphanumeric keyboard, joystick and numerous fast action buttons (FABs). The system has been deployed on about 400 U.S. naval vessels as well as a number of allied ships.

The display on the DCC contains the area of polar display for a graphical representation of the emitter type and the bearing of the emitter. Around the polar display are alphanumeric information on system status, emitter parameters and library information. The operator uses this information with the emitter audio signals to recognize the type of emitter. The operators proficiency is a function of the time to functionally recognize the emitter type from the information presented for each emitter.

The EW operator controls an array of communication systems for both intra and inter ship communications. The type of communication systems on each ship varies with the mission of the ship and the layout of the Combat Information Center (CIC). These communication channels are the method of communication between the operator and supervisor. Generally, link 14 is used within the EW module in the CIC (Gaster 1993). Verbal communication is essential in order to provide the individuals with situational awareness information required to operate their systems.

5.3.2 EWOBT

Electronic Warfare On-Board Trainer is a ruggedized 80486/DX2-50 personnel computer designed as a multimedia training system. The hardware and software specifications are shown in Table 3. The EWOBT program provides the hardware and integrated courseware to train and assess electronic warfare personnel on their own equipment at sea. The system, when used with an embedded training device (EDT), can stimulate any version of the AN/SLQ-32(V) including the Operator Training Equipment (OTE). The EDT card provides simulated audio and video information to create "real world" scenarios of the electromagnetic environment.

Table 3 - Configuration and specification for EWOBT

Component	Model/Version/Specifics
CPU	80486/DX2-50
Disk System	>60MB Removable Hard Disk Drive 1.2MB 5 1/4" Floppy Disk Drive 1.44MB 3 1/2" Floppy Disk Drive
RAM	16.0MB (with maximum of 32.0MB)
Cache	256KB Memory Cache
Speed	50MHz CPU Clock Internal
Graphics	Super Video Graphics Array (SVGA) VESA Local Bus (VLB)
Max Resolution	1024 x 768 Pixels (SVGA)
Max Colors	16.7M
Monitor	NEC Miltisync 3FG
Keyboard	121-key Enhanced Keyboard
Trackball	AWKTRACKBALL (Serial Trackball)
Sound	Stereo Sound Board with 4KHz to 44KHz Sampling rate.
Software	MS-DOS 5.0 and MS-Windows 3.1
Miscellaneous	SCSI Interface for NEC CDR-84 Drive EMI/RFI Filter

A data extraction card (DX) is used to extract the emitter events and actions the operator performs on the DCC. These data are stored in the EWOBT but the hard drive is removable, because of the classified nature of the EW work, so is relatively small. The speech recognition hard drive memory requirement should be kept as small as possible but must be less than fifteen megabits. The central processing unit (CPU) is adequate for most operations but may slow down with to many additional requirements. The speech recognition system should use minimal EWOBT CPU resources.

5.3.3 PAM

Proficiency Assessment Methodology is based on human factors engineering techniques of task analysis and operational sequence analysis. The task analysis provides the means of understanding the operator actions and decisions and the operational sequences to measure the time relationships among the actions and the alternate paths through the sequences. The PAM model was found to be sensitive to the changes in operator proficiency on the HAH function recognition task for SHARP I (Beaton et al 1994). In other words, the PAM model will measure the operator proficiency from the time relationships of operator actions that are included in the task analysis.

For the HAH task, the terminal action for functional recognition was a verbal report from the EW operator to his supervisor. The goal of the SHARP project is to measure operator proficiency in an operational environment, automatically and in real-time. In order to accomplish this goal, a method of recognizing the emitter type in the verbal report and time-tagging when the report occurs must be automated. This requires a voice recognition system capable of capturing the verbal report as well as measuring when the report occurred.

The real-time implementation of PAM has not been finalized as the study is still underway, so the input from the speech recognition and the DX system are not specified. The sixty percent correct recognition is an estimate of the number of emitters that need to be correctly identified to provide an accurate measure of the operator proficiency. Exact information is required for the production system design but will not be available until the completion of the SHARP II study.

5.3.4 Verbal Report

The verbal report for functional recognition conveys three pieces of information, type of emitter, bearing or direction and description. Discussions with EW operators and supervisors indicate that this is generally true for all verbal reports. The lexicon for identification may change to a code, but the information is the same. In the case of HAH, operators in SHARP I reported the functional recognition of a missile as scan type, bearing number and description (i.e. Delta Scan, Bearing 123, Missile Homing). The order of the information within the report often changes, but the information remained the same. The voice recognition system must be able to capture the emitter type or description in order to functionally recognize the emitter type.

The operational verbal reports are classified, so before any shipboard testing can be completed the operational lexicon and format are needed to ensure that the emitter types can be recognized. This project did not investigate the actual verbal reports because of security constraints on the author. The number and time between verbal reports has not been defined as all of the testing has been performed in laboratory settings. The time between reports may vary from ten seconds to ten minutes because of the variability of the number of emitters and the emitter types. The speed and accuracy trade-off in any recognition system will have to be identified and may vary among the different operational conditions. This report will specify a ten second recognition time as that is the approximate time between emitter changes for the Stark attack.

During the Vincennes incident, the investigation by the Committee on Armed Services (1988) found there was extensive noise and confusion in the CIC, with personnel shouting to be heard above the noise. This comment was reiterated in the SHARP I report where it states that at General Quarters the conditions in the CIC were equivalent to a roomful of people yelling at one another (Beaton 1991). If a speech recognition system is to operate in this environment it should be robust enough to operate in an area of high noise. This noise includes both equipment and human conversational noise.

VI. FUNCTIONAL ANALYSIS

The functional analysis in the system engineering process is a logical and systematic approach for the design and development of a system. The method used for the analysis is a functional flow or task analysis of the system functions. The process takes the top level requirements and decomposes these requirements for operational and support functions into specific qualitative and quantitative design requirements. Each level of the analysis shows in finer detail the functions to be performed to support, operate or maintain the system. The process is iterative and provides a graphical presentation of each of the input and output requirements.

6.1 Functional Flow Diagrams

The functional flow diagrams depict the interrelationships among the system functions and the interfaces at the system boundaries. The top level flow shown in Figure 8 is the operation of the AN/SLQ-32(V) system to produce a report by the operator. The DCC console is a mature component of the system and the remainder of the system interfaces with the console and the operator. The functional flow shows the operation of the system. Each level in the flow diagram relate to previous levels or functions to provide the functional relationships between the functions and the different levels within the design. This functional relationship is used throughout the design to ensure that the designers from the different disciplines work toward a common goal or objective.

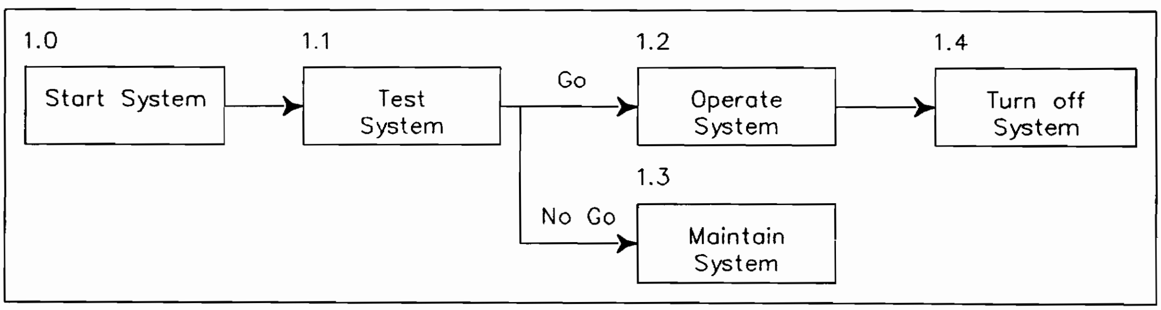

Figure 8. Operational Functional Flow

Figure 9, taken from Dyess (1992) shows the operator actions on the DCC console of the AN/SLQ-32(V). The operator's actions for this project provide the input to the PAM software and are part of the emitter functional recognition task. This functional flow formed the basis for the PAM model. Dyess' work did not include the verbal report function in his flow description so this task was defined from the SHARP I testing and laboratory study.

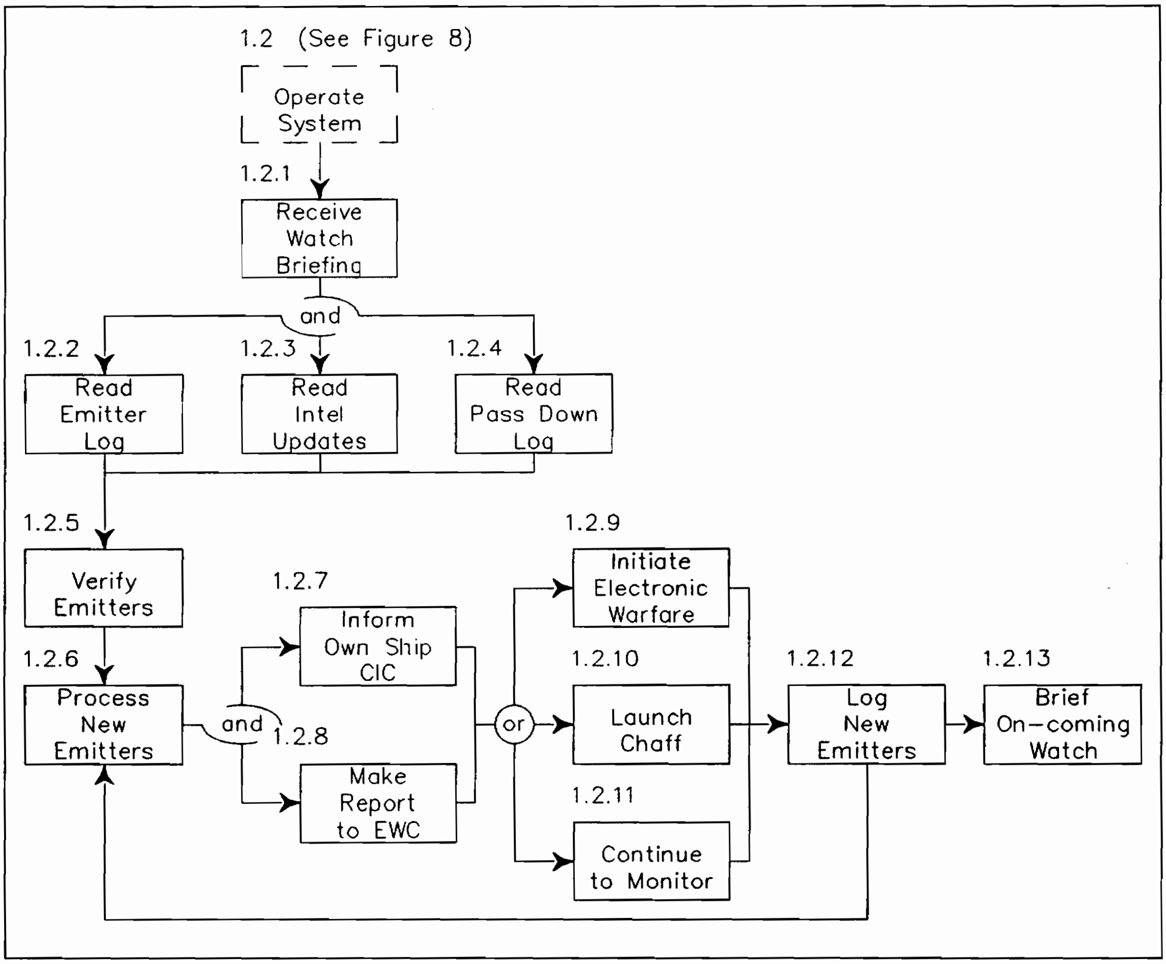

Figure 9. Functional flow diagram for system operation.

Figure 10 shows the functional flow relationships for the verbal report and the automatic speech recognition system. The structure shown in Figure 10 identifies the operator functions of a new emitter appearing on the screen at the top level through to the functions required in a speech recognition system. The decomposition of the speech recognition functions provides the detail required to specify the testing and design requirements for an automatic speech recognition system. The function has not been allocated to operator or machine at this point as further testing is required to ensure that the computer can recognize the type of emitter in the verbal report. The trade-off study will identify the allocation of the function and if a PC computer can be used for the automatic speech recognition function.

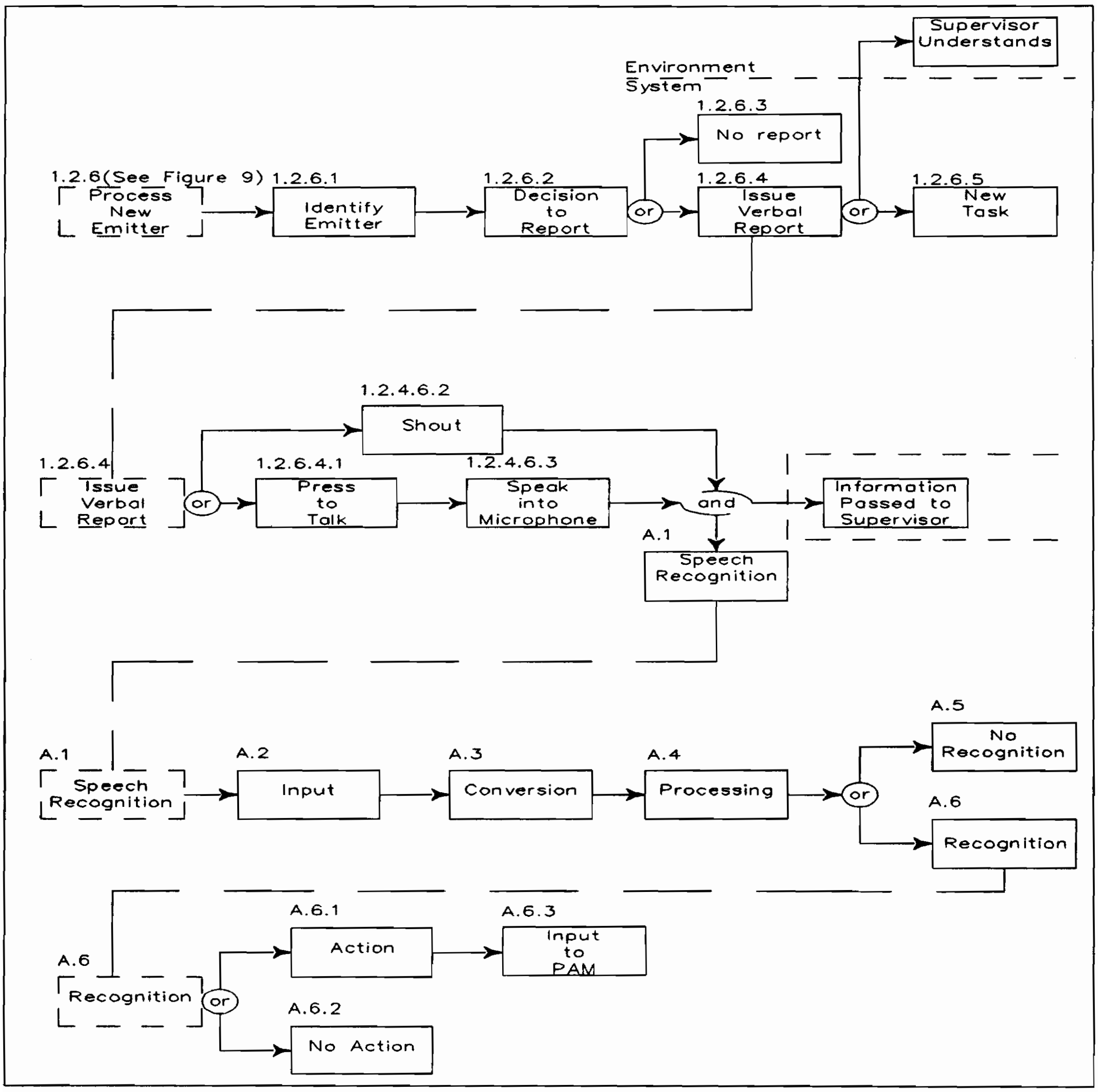

Figure 10. Functional Flow for Verbal Report

6.2 Functional Allocation.

Engineers generally allocate the functions to machines wherever possible, but this reliance on automation has often caused problems. The Navy recognizes this problem and insists on keeping the operators in-the-loop for the EW systems because of the criticality of the function. The speech recognition system and PAM is not critical so the speech recognition can be allocated to a computerized system. The problem is whether or not the computer can actually recognize the type of emitter the operator is reporting to his supervisor.

This project looks at some of the products that are available for speech recognition from commercial vendors to see if a speech recognition function is feasible with a computer. The literature review during the feasibility study showed that automatic speech recognition with a computer is theoretically feasible and systems are on the market that will recognize human speech and provide an electronic output compatible with other computers. The goal of SHARP II is to determine if speech recognition can be accomplished with the equipment currently used for the PAM data collection. In essence, can a PC-based system be used to recognize the verbal reports and evaluate commercial systems for use in the SHARP II testing.

Prior to the functional allocation of speech recognition to a PC computer based system, it was necessary to determine if products are available to recognize the emitter type in a verbal report. To accomplish this task, a number of companies were contacted to find what PC products were available prior to April 94. Twenty companies that advertised speech recognition were contacted for information.

VII. SPEECH RECOGNITION PRODUCTS

7.1 Speech Recognizer Requirements.

The operational requirements provide the criteria for selecting and testing a system. For the SHARP II study the requirements are not as strict as the testing will be conducted within a controlled environment. The speech recognition system must have the capability of working within the following operational conditions: PC-based, recognition from a large number of people, capable of recognizing verbal reports, ability to add new words and ability to develop an application interface with the PAM software. These requirements are from the design concepts and the functional flow diagrams. Other requirements stem from the need for the system to be non-intrusive and operate within the system. These requirements are the ability to: use inputs from the microphones or intercoms, recognize key-words in the verbal reports, and reject or suppress noise.

The large number of operators across the fleet makes speaker independence an essential requirement. There is insufficient time to train or adapt the system to recognize individual operators. The stress of the task will change the operator speech and the independent-speaker capability should provide better recognition in stressful situations

The verbal report is mainly spoken in a semi-structured manner. The AN/SLQ-32(V) uses a press-to-talk intercom system so the report is constrained by silence when the switch is not pressed which will allow the use of connected speech systems. The recognition system must identify the type of emitter within the formatted reports. Because the reports are not always structured, the system should recognize the descriptions and types of emitters as key-words. Therefore, key-word spotting is a necessary requirement to capture key-words in unstructured reports.

The total number of words in the report may be very large, with 1000 to 10,000 combinations. However, the number of distinct words will generally be about 100 to 300. The simple requirement for three types of emitters used almost four hundred words with only thirty unique words implemented. With large vocabularies, recognition is slow and prone to error because of the similarity between words. The system must have the capability of recognizing 100 unique words in the verbal report. The key-word spotter must have the ability to recognize 25 active words and should be able to recognize 100 words. The basis for the twenty-five word minimum number comes from the SHARP I contract where four personnel used the words (missile, delta, homing, inbound, locked-on, steady, homer) as either types or descriptions to describe the HAH emitter. Other emitter types would increase the number of key-words to as large as one hundred.

As the system is to be non-intrusive, the speech recognizer must work with the EWOBT computer. The EWOBT has a slot available for a DSP or audio card and could provide the data collection and scenario generation control as well as the speech recognition function. A DSP board is recommended to reduce the EWOBT computer processing requirements.

The system must have the capability of developing the interface requirements to recognize words or phrases and output the word to the PAM data program. The development kit should allow addition of new words because many of the words, such as ship or radar names used for the Navy operations, may not be in system dictionaries. The system should be capable of easily adding new words to the dictionary without an inordinate amount of time to train new words.

The input and noise rejection requirements come from the CIC environment which is very noisy from people and machines. The recognizer should have a capability of noise rejection or suppression without requiring customization of the speech recognizer. The input to the speech system should be from the intercom system or directly from the microphone. The microphones for intercom systems often have a small dynamic range of 300 to 5000 hertz. The speech recognition microphone dynamic range is 200 to 15,000 hertz. The recognition system should be able to accept input from an intercom system or a narrower dynamic range microphone. The system should have some inherent robustness in order to overcome or cancel the effects of different microphones.

This project has been designed to test the feasibility of an automatic speech recognition system to capture a verbal report for the PAM model. The requirements used to specify the requirements for the trade-off studies do not include the environmental and operational requirements of the production system. The trade-off requirements are to test the capability of a speech-recognition system to capture the verbal reports.

For this project, the requirements are divided into two categories, essential and necessary. The essential capabilities are those the system requires in order for the speech recognition system to provide the minimum function of recognizing a structured verbal report from EW operators. The essential requirements include the ability of the speech recognition system to operate with the EWOBT and input data to the PAM software. The essential requirements of the speech recognition system are: capability of speaker-independent operation; connected speech recognition of structured verbal reports; hosted or installed on a PC-based system; a development kit to produce the data acquisition application; and a capability of adding new words to the vocabulary for application specific terms or reports.

The necessary requirements should enhance the recognition capabilities. The necessary capabilities provide wider use of the system within the CIC without restricting the verbal reports to a structured format. Necessary requirements extend the dynamic range of the inputs to allow easier integration with the existing system components. Finally, the necessary requirements provide easier use and development or enhancement when using the system. The necessary requirements of the speech recognition system are: key-word spotting capability of at least 25 words; capability of accepting a wide range of acoustic inputs from different microphones and telephone systems; and noise suppression or robustness inherent in the system.

7.2 Product Review

A review of the literature revealed about twenty companies that advertised speech recognition systems. These companies provided information and brochures about their speech recognition systems. Appendix A contains a list of the companies that provided information, name of the products, contact individuals, telephone and fax numbers.

7.2.1 Product Review Results

No system was found that met all of the system requirements. Appendix B provides a description of each of the products and whether they met the criteria. Table 4 has the tabulated results of the review, a check mark (✓) indicates whether systems passed the requirement criteria. The legend at the bottom of the table identifies the criteria. Four companies had products that met the essential requirements; IBM, Lernout and Hauspie, Linkon and Speech Systems. These companies were contacted to obtain further information.

7.2.2 Lernout and Hauspie

Lernout and Hauspie's product literature suggested that their system would meet all the criteria at 96.7% recognition accuracy. Discussions with company representatives showed the system advertised was in the development stage. The capabilities were available on a workstation and the company was trying to implement the system on a PC. Operation with audio cards and DSP is in development for a PC, but the present capability on the PC is only for isolated speech recognition. The active word capability of the continuous speech is thought to be 25 to 30 words, but has not been fully tested. The company's explanation for the difference between the literature and the product is; "the speech recognition model is capable of meeting all the requirements, but has not been fully implemented". The Lernout and Hauspie system at the current state of development for a PC-based system was not recommended for this project.

Table 4. - Results of product reviews.

Company	CS	SI	DK	PC	Words	KW	Input	Noise
ASPI			✓	✓			✓	?
BBN Systems	✓	✓	✓		✓	✓	✓	✓
Creative Technologies			✓	✓	✓		audio	?
Covox				✓	✓		audio	?
Dragon Systems	Digits	✓	✓	✓	✓		audio	?
Entropic	Digits	✓	✓				?	?
Intellivoice Communications	Digits	✓	✓	✓	40	✓?	phone	
I B M	✓?	✓	✓	✓	✓	✓?	audio	✓?
Kurzweil Applied Intelligence		?		✓	✓		audio	?
Lernout & Hauspie Speech	✓	✓	✓	✓	✓	✓	audio	✓
Linkon Corp	✓	✓	✓?	✓	15-20	✓	phone	✓?
Scott Instruments			✓	✓				
Speech Systems Inc	✓	✓	✓	✓	✓	✓?	audio	✓?
SRI International	✓	✓	✓		✓	✓	✓	✓
Texas Instruments	?	?	?		?		?	?
Verbex Voice Systems	✓		✓	✓	✓	✓?	audio	
Voice Processing Corp	Digits	✓?	✓	✓	✓		phone	?

Legend: **CS**- continuous speech, **SI**- speaker-independent, **KW**- key-word spotting, **PC**-PC operating system, **Words**- active words recognized, **DK**- development kit, **Input**- audio and phone input, **Noise**- noise suppression or rejection capabilities.

7.2.3 Linkon Corporation

Linkon Corporation has a workstation version that uses the software and DSP hardware for continuous speech recognition. The PC version of the system was developed from the UNIX workstation and is configured for telephone information processing applications. The company uses the term continuous speech recognition to mean continuous digit recognition. Discussions with technical representatives suggested that the continuous digits could be changed to words and these words could form the basis for key-word spotting. The initial discussions suggested that 31 words could be spotted, however; this was reduced to 15 to 20 in follow-on discussions. The development tool kit TeraVox operates only on the UNIX system. Any development would have to be programmed in a lower level machine or assembly language and Linkon does not provide the documentation for this development. The Linkon system was not recommended for this project.

7.2.4 IBM

IBM Continuous Speech Series (ICSS) has a connected speech system with an independent speaker capability using a number of popular audio cards such as SoundBlaster 16 or Pro Audio Spectrum 16. The system runs under Windows and includes a development kit for adding words and developing applications. The connected speech is in syntax or context (as termed by IBM) format and must be delineated by silence. The ICSS has an active vocabulary of 1000 words and can use a number of microphones of different dynamic ranges. The microphone characteristics are: dynamic, cardioid or supercardioid with an impedance of 600 Ohms. IBM supplies only the software so an audio board, microphone and Windows software must be purchased separately. The system meets the essential requirements for the speech recognition system and is recommended for feasibility testing to identify its accuracy.

7.2.5 Speech Systems Incorporated

Speech Systems Incorporated, Phonetic Engine 400 (PE400) product, is a connected speech, independent speaker system running under Windows. The system is sensitive to the use of different microphones and changing microphones may require adaptation and rebuilding the speaker model. The size of the vocabulary is only limited by the amount of memory which is supplied with the DSP board (either 4 MB or 16 MB of random access memory). The PE400 product includes a development kit, DSP board, microphone, software, manuals, training course and technical support for 90 days. The PE400 meets the essential requirements for speech recognition systems to capture a structured verbal report and is recommended for further testing.

7.3 Product Recommendations

Verhaeghe in 1992 wrote in Byte magazine that " continuous large vocabulary speech recognition systems was five to ten years away". The products on the market indicate this is correct especially for PC systems. What is available is connected speech recognition for medium size vocabularies that are speaker-independent. These systems have the capability to recognize the EW operator's verbal report when used with the press-to-talk feature of the shipboard intercom system. Two systems were recommended for a trade-off study are; IBM Continuous Speech Series (ICSS) for Windows and Speech Systems Incorporated Phonetic Engine 400 (PE400). Both systems meet the essential requirements for the feasibility study.

7.4 Product Criteria

These products have the capability of recognizing the type of emitter identified in the verbal report but neither of the companies could or would identify the accuracy for this application. In order to test the accuracy, two systems were purchased to show how many of the verbal reports could be correctly recognized. The criteria for the accuracy in a quiet environment was to recognize 75% of the verbal report phrases. The rational for this criteria was derived as follows. Only 80% of the reports are structured and both systems require a structured report. Therefore, the two systems can recognize a maximum of 80% of the reports.

The PAM methodology has not matured enough to identify how many or what percentage of emitters must be recognized to measure proficiency. In the SHARP I study, only a limited number of emitters were reported. The number of emitters that the PAM model needs to provide statistically significant accuracy was arbitrarily chosen at 60%. So 60% of the emitters would provide enough power to the calculations to accurately identify the operator proficiency.

In order to achieve 60% of the emitters, when only 80% of the verbal reports can be recognized, requires an accuracy in the speech recognition system greater than 75%. This calculation is shown in Figure 11 for a total of one hundred emitters. If 60 emitters must be recognized out of one hundred and only 80 are formatted for recognition then 75% accuracy is necessary to capture the required number of emitters for PAM.

$$\frac{60 \text{ required number of emitters for PAM}}{80 \text{ maximum number recognizable by format}}$$
$$= 0.75 \text{ of the verbal reports must be recognized}$$
$$= 75\ \%$$

Figure 11. Accuracy criteria calculations

If one of the systems can meet the accuracy criteria, the functional allocation of the speech recognition function to a PC-based computer is considered sufficient to continue testing and for use in the laboratory studies. The system, when optimized to capture the verbal report, should provide better accuracy than this baseline measure. If the seventy-five percent accuracy is not achieved then the speech recognition for a PC-based system may not be feasible and another method would have to be devised such as, workstations or speaker-dependent systems which generally provide higher accuracy but are either more expensive or are intrusive to the operator and ship-borne operations.

VIII. PRODUCT TESTING

8.1 Introduction

The product testing measures the accuracy of the PC-based speech recognition systems to recognize the type of emitter in an EW Operators verbal report. If the PC-based speech recognition system meets the criteria of 75% accuracy then a PC system will be used for the automatic speech recognition function for the SHARP project. The testing provides a Go\No Go measure of whether a PC-based system can be used for the SHARP project.

The 75% accuracy criteria was chosen for the default settings of each product. The syntax was a single phrase for each utterance and the same for both systems. Appendix C contains the phrases used for the testing. The ambient noise level was about the equivalent of a large office at 70 dBA. This would be equivalent to the CIC during normal operations. Because the testing required the same effort as needed to capture the operational verbal reports, data collected in the testing provided valuable information for the production system requirements. One additional test was conducted during this feasibility or trade-off study to identify the accuracy of the system in the presence of very high verbal noise. This test did not have a criterion set because the actual noise level during the condition one attack is unknown.

The information that is gained from the purchase, training, set-up and testing is valuable whether or not the system meets the accuracy criteria. The information will be used for detailed specifications of speech recognition systems for the production system.

8.1.1 Equipment

Two speech recognition systems were purchased for testing; the IBM Continuous Speech Series for Windows (ICSS) and Speech Systems Incorporated (SSI) Phonetic Engine 400 (PE 400). The IBM system was set-up and configured in an EWOBT and the

PE400 in another 486 machine. Both systems use connected speech format and require installation of application dictionaries and compilation of the syntax.

IBM provides only the software and documentation when you purchase the ICSS package. A Sennheiser model 431 microphone and Pro Audio Spectrum 16 audio card, recommended by IBM, were purchased for the study. The microphone characteristics are cardioid, uni-directional, noise canceling with an input impedance of greater than 600 ohms and cost $410. The audio card was purchased for $190 and provided only the speech conversion function while the EWOBT computer processed the digital vectors to provide the speech recognition. The sampling rate for the ICSS is set at 11,000 or 22,000 samples per second. The default is set at the lower level which provides a lower accuracy but faster speech recognition.

The PE 400 product supplies microphone, DSP board and software for the recognition system. The microphone, an AKG C410/W, that comes with this system has very similar characteristics to the microphones bought for the IBM system, but the dynamic frequency range is slightly smaller. The difference is these microphones is the AKG is a small miniature headset microphone that sits very close to the mouth. Other microphones may not work with the SSI system as the speaker models were built for this specific microphone.

All of the speech recognition in the PE 400 is controlled on the DSP board with the host computer storing only the software program. On start-up, the speech recognition program is downloaded to the DSP board. The DSP board purchased has four megabits of memory but can be increased to sixteen.

8.1.2 Equipment Set-up

Both systems were set-up to measure the accuracy of each system. Each system has a test feature that can be used to measure the accuracy once the syntax and vocabulary have been installed and compiled. Forty-five non-classified phrases from SHARP I testing provided representative phrases. These phrases are shown in Appendix C. The setting was a large office environment with a noise level at about 65 to 70 dBA. The environment was in an office with computers and air conditioning equipment in operation. The level was measured with a Radio Shack Model 33-2050 sound level meter set at a fast setting. This setting is the approximate level of the CIC during normal operations. Actual shipboard noise level testing will be done on the ships as part of the SHARP II trials.

Both systems were to be tested in the same manner using ten subjects reading the 45 phrases and measure the accuracy of each system. The IBM system was set-up and tested for operation first then the PE400 system was set-up. During the set-up of the PE400 system, the EWOBT with the IBM system was required for other testing with the DX card. There was some difficulty with the DX card installation and the ICSS system would no longer work even when the DX board was removed. It is believed some of the interrupts in the EWOBT were changed by the DX installation and the system could not be reconfigured before the start of scheduled testing. The ICSS set-up was tested but no data was collected for the accuracy of the system before it became inoperative.

A decision was made to test only the PE400 system for accuracy. In addition, a noisy environment was set-up to obtain information about noise rejection and suppression. The subjects were available for two sets of data collection so the opportunity was used for additional testing. No pass fail criteria were set for this noisy environment testing.

The noisy environment was produced by digitally recording the verbal reports of speakers and merging the reports so a relatively constant source of human voices were audible. Some of the conversations overlapped and some of the conversation was the same as the verbal reports delineated by silence. The level of these voices (noise) was fixed at 90 dBA by placing two audio speakers eighteen inches apart and measuring the sound pressure levels with the sound meter. The participants kept the microphone between the two stereo speakers to provide the noisy environment. Ninety dBA was chosen because that is the approximate level of a shout at one foot and the noise level the Occupational Safety and Health Association has set for an eight hour work day.

The participants voice levels were kept relatively constant between the two environments by using earplugs with stereo headphones. The earplugs blocked some of the noise and music in the headphones masked the noise environment. The music level was constant in both the quiet and noisy environments so the voice level would remain relatively constant. The participants overall speaking volume was greater with the earplugs and headphones but the automatic gain control the PE400 system adjusted to the louder input level.

The testing was accomplished using the automatic collect function on the PE400 system which randomizes the output of the phrases. This function is a Windows application that shows a dialog box of the phrases which the participants read from the computer screen. The phrase appears on the screen and the participant holds the right hand mouse button while he reads the phrase. When the phrase was complete the button was released and the system would process the phrase. When the processing was complete the next phrase would appear on the display. The phrase and word accuracy were automatically collected. Each phrase and word could be broken down and analyzed from the data but the accuracy was the only measure of interest for this particular study.

8.1.3 Methodology

The purpose of this trade-off study is to determine the baseline accuracy of speech recognition and to see if it is feasible to use a PC-based system for automatic speech recognition. Ten speakers were used repeating forty-five verbal reports in both a quiet and noisy environment. The quiet environment was measured at 65 to 70 dBA, which is slightly above the noise level of a large office with electronic equipment and people in normal conversation.

The noisy environment level was set at 90 dBA which is a very noisy level similar to people shouting a foot away from the microphone. The accuracy is measured as a percentage of the correct number of phrases and words to the total number. The number of phrases in each environment was 45 and the number of words was 395. The accuracy was calculated for both phrases and words in the quiet and noisy environments. The standard deviation was calculated for each.

The criteria used for the quiet condition was for an average recognition accuracy of 75% correct phrases in order to allocate the automatic speech recognition function to the PC-based speech recognition system. If this criteria could be reached with a speaker-independent system at the default setting, it is felt that with optimization much better accuracy could be obtained.

8.2 Results

First, the accuracy of the system in a quiet environment. The accuracy was found to be 77.7 percent for the phrases and 77.3 percent for the words with standard deviations of 0.155 and 0.150 respectively. The results for individual participants are shown in Figure 12. This accuracy exceeds the criteria set for the functional allocation but the standard deviation is rather large for this test. The range for the phrase accuracy varied from 91 percent to 51 percent and for words from 90 to 51 percent for the individual participants. The large variation is troublesome, but the criteria is met and further testing is required to optimize the system for the study.

The results of the noise testing is presented to show the effect of a very high noise level of human shouting and conversation on the speech recognition. The speech recognition accuracy in the 90 dBA noise was 15.8 percent for phrases and 17.8 percent for words with a standard deviation of 0.074 and 0.071 respectively. This is poor recognition, but a number of interesting facts were discovered when setting up and doing these tests that were not identified as problems during the initial investigation for this project.

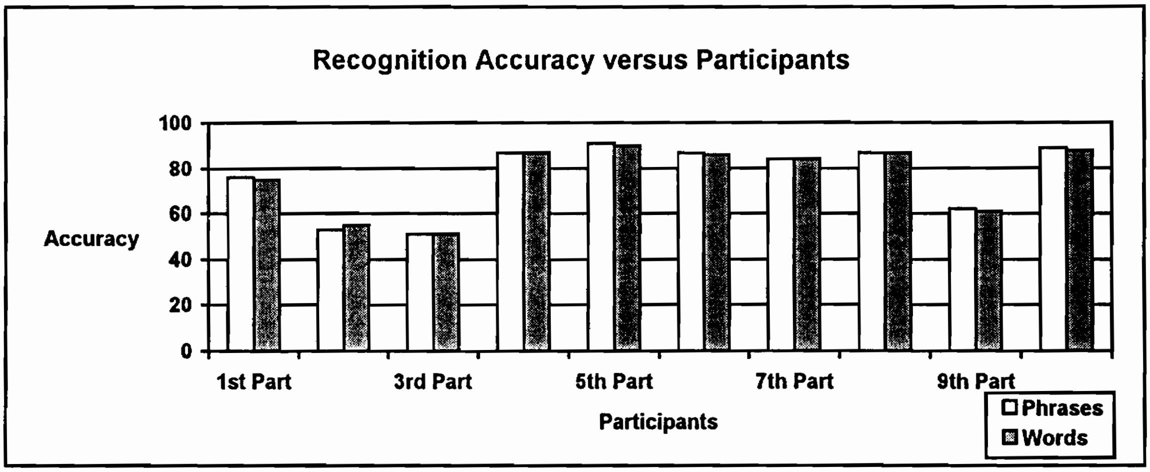

Figure 12. Recognition accuracy for individual participants.

8.3 Discussion

The system engineering process is a logical method of top down design so the lower level functions and design grow from common requirements. The process uses a structured approach identify problems early in the design process. The information from testing, feasibility studies, component changes, and functional allocation are fedback at the proper level to ensure the designers have information and specifications from a common level to accomplish the detailed design. The procurement, syntax integration and compilation, addition of new words to the dictionaries and testing identified many additional questions that need to be answered before this project is ready for design of an operational system.

The syntax used in this demonstration was very simple, using only a single phrase for each syntax. With the operational system the operators will not be reading a phrase, but will be uttering phrases with many combinational probabilities, making recognition more difficult. The speech recognizer should have no greater difficulty recognizing the words, but identifying what to include in the syntax for each of the emitter types may be very difficult to define for the users. The words or lexicon must be identified before the speech recognition system can be used for any data collection of the operational verbal reports. This includes any laboratory testing using operational scenarios. Once the operational reports are used the phrases and recordings will have a secret classification.

The problem with any speech-recognition system testing, human or machine, is variability in the speaker's voices even within a trial. Any accuracy measurements will vary with input from day-to-day and often from phrase-to-phrase. In order to overcome this problem, reports are often recorded and the recorded data used for testing. Communication system tests use this method to test quality and it provides good repeatability. Any testing of noise, bandwidth, microphone effects on speech recognition should use recorded data to ensure repeatable results.

In the computer speech recognition systems, the testing must be carefully planned. One of the plans to test the noise immunity of the speech recognition was to digitally mix noise with the recorded verbal reports. The problem with this method is the microphone and input circuitry often provide the noise suppression capabilities through directionality of the microphone or noise cancellation circuitry. The near field effects of headset microphones can not be accurately modeled in a digital filter to accurately shape and mix the noise with the digital voice recordings.

The effect of automatic gain controls (AGC) also affect the type of noise that the system will ignore. If the AGC is adjusted for each phrase, the ambient noise will have little effect as long as the operator speaks at a signal level that is greater than the acceptable signal-to-noise ratio for the gain control. Each system will have a different signal-to-noise ratio that will effect the overall accuracy of the signal in a noisy environment. The noise testing for the speech recognition system must be carefully analyzed before doing testing.

While data was not collected for the ICSS system a number of factors were observed when configuring the system and in the initial testing. The system recognized the phrases for the verbal report and appeared to have comparable accuracy to the PE400. The set-up and addition of new words to the dictionary was a little more difficult, but no large problems were encountered. The phrase recognition time appeared to be somewhat slower but no conclusive measure was made.

One area that should be tested is to see if the audio card and DSP board have different recognition times for the same accuracy. The CPU utilization should be monitored to see the effect of speech recognition on the CPU resources and if they will degrade the host computer functions.

IX. CONCLUSION

9.1 Recommendations

The project sets out the top level system engineering requirements for an automatic speech recognition system for the SHARP program. With the current system specifications, a PC-based speech system is feasible for an engineering evaluation during SHARP II studies. The Speech Systems PE400 product is recommended but the IBM system may work if the incompatibility between the speech recognition and the data extraction card is rectified. Testing with these systems will provide data that is useful in identifying the lower-level system requirements needed by the designers for the production system.

These systems were not purchased as production system equipment, but to test the feasibility of an independent speaker PC-based speech recognition system to accomplish the function of speech recognition in real time. Further testing of the system accuracy at capturing the operational verbal reports in both laboratory and ship-borne environments is essential in order to obtain accurate specifications for the production system.

The operational lexicon and the PAM accuracy requirements are seen as the two most critical requirements in the near term. It is recommended that personnel with the proper security clearance obtain the terminology and verbal report protocol to prepare the syntax for the laboratory and ship-borne studies in the SHARP II project. This will provide invaluable information for the system design.

9.2 Future Direction

Speech recognition appears at a crossroads with many of the workstation capabilities being transferred to PC-based systems. In a year or two, the systems may be capable of keyword spotting which may alleviate the need for formalized syntax and formatting of each verbal report. The Speech Systems PE400 has a limited capability for key-word spotting and this capability should be pursued.

The operational, support and logistic elements were not considered in the selection of the systems for the trade-off studies. These requirements for the production system are essential if the system is to work properly in the shipboard environment. When the PAM model is complete and the DX production models are complete they should be incorporated into the system requirements for the speech recognition system.

APPENDIX A

COMPANY INFORMATION

Company	Location	Product	Contact	Telephone/Fax
ASPI	Atlanta, GA	DSP card - ELF	Gladys de Moulpied	(404) 892-7265 (404) 892-2512
BBN Systems and Technology	Cambridge, MA	HARK	Graeme Smith	(617) 873-4636 (617) 821-4273
Creative Technologies	Milpitas, CA	Sound Blaster		(408) 428-6600 (408) 428-6699
Covox	Eugene, OR	Voice Master		(503) 342-1271
Dragon Systems	Newton, MA	Voice Tools	Brian Kinkade	(617) 965-5200 (617) 527-0372
Entropic	Washington, DC	HTK	Ken Nelson	(202) 547-1420
Intellivoice Communications	Atlanta, GA	DirectTalk	Jim Szyperski	(404) 875-3535 (404) 875-8036
IBM	Somers, NY	ICSS	Elton Sherwin	(800) 825-5263
Kurzweil Applied Intelligence	Wattham, MA	Voice series	Mark Flanagan	(617) 893-5151 (617) 893-6525
Lernout & Hauspie Speech	Woburn, MA	Speech recognition	Frank Carraba Jason Nicholson	(617) 932-4118 (617) 932-9209
Linkon Corp	New York, NY	LinkEngine KWS	Karen Garelick	(212) 753-2544 (212) 935-5635
Scott Instruments	Denton, TX	Speech recognition	Ray Cotten	(817) 387-9514
Speech Power, Inc	Iselin, NJ	Speech applications	Al Stomberg	(908) 225-0505 (908) 321-6562
Speech Systems Inc	Boulder, CO	Phonetic 400	Deborah Parsons	(303) 928-1110 (303) 938-1874
SRI International	Menlo Park, CA	DECIPHER	Ron Croen	(415) 859-3274 (415) 326-5512
Texas Instruments	Austin, TX	Telephone recognition	Peggy Hart	(512) 250-6437
Verbex Voice Systems	Edison, NJ	Listen for Windows	David Friel	(908) 225-5225 (908) 225-7764
Voice Processing Corp	Cambridge, MA	Telephone recognition	Desmond Pieri Canny Dual	(617) 494-0100 (617) 494-4970

APPENDIX B

COMPANIES

Atlanta Signal Processors Incorporated

Atlanta Signal Processing Incorporated (ASPI) is a hardware based company that has a limited speech recognition system. Their software is mainly for speech coding, compression algorithms and developing DSP applications. Lernout and Hauspie use ASPI's Elf board (to be discussed later) to provide continuous and speaker-independent applications. The speech recognition, for ASPI, is sixteen (16) discrete words, speaker-dependent operation. The system will accept input from audio or telephone RJ-11 inputs. The cost of the Elf application evaluation/development kit for the University is $1995 the cost of a DSP board is $1195.

Table 5- Atlanta Signal Processing Incorporated

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	F
Speaker-independent	Must recognize reports from different speakers	F
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Will operate on and 486 PC with the addition of an audio card or DSP board	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	F
Key-word spotting	Should be able to spot 25 active words	F
Input	Should accept both telephone and audio inputs	P
Noise rejection	Should be capable of operation in a "noisy" environment	?

BBN Systems and Technologies

BBN has a UNIX based speech recognition system with continuous speech, speaker-independent capability. The package includes a vocabulary and grammar development kit for application development. The system will interface with both telephone and microphone-based applications. The system has a vocabulary size of 1000-2000 words with phonetic based recognition and is suited for different operating environments. The cost of the system for University development is approximately $5000.

The BBN is only implemented on the UNIX based workstations.

Table 6 - BBN Systems and Technologies

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	P
Speaker-independent	Must recognize reports from different speakers	P
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	F
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	P
Input	Should accept both telephone and audio inputs	P
Noise rejection	Should be capable of operation in a "noisy" environment	P

Creative Labs Incorporated

Creative Labs are the manufactures of the popular Sound Blaster audio card. Their software VoiceAssist is an isolated word recognition system the can support up to 256 individual users. A microphone is included with the system and the application is mainly for command and control of Windows based applications. The vocabulary has 1,024 voice commands, however; only 32 are generic. There is a Developers Kit to add names or train commands within new applications. The cost of the development kit is $149 which also includes the VoiceAssist software and microphone.

Table 7 - Creative Labs Incorporated

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	F
Speaker-independent	Must recognize reports from different speakers	F
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	F
Input	Should accept both telephone and audio inputs	Audio
Noise rejection	Should be capable of operation in a "noisy" environment	?

Covox

Covox is a DOS and Windows based system used mainly for PC control applications. The system is called Voice Master and can have up to 1023 words per file. The voice recognition is a discrete, speaker-dependent system. An audio card, headset and cables are provided with the system.

Table 8 - Covox

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	F
Speaker-independent	Must recognize reports from different speakers	F
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	?
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	F
Input	Should accept both telephone and audio inputs	Audio only
Noise rejection	Should be capable of operation in a "noisy" environment	?

Dragon Systems

Dragons Systems is an Windows and DOS based system used for control and voice interfacing with program applications. The Voice Tools software allows development of voice-activated interfaces for other applications. The system is speaker-dependent (using voice files) or independent for a small vocabulary. It has the capability of recognizing continuous digits. The application interface is based on the C or C++ compilers. The system has an active vocabulary of up to 1000 words. The system includes a microphone but a IBM Audio Capture and Playback Adapter audio board is required.

Table 9 - Dragon Systems

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	Digits
Speaker-independent	Must recognize reports from different speakers	P
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	F
Input	Should accept both telephone and audio inputs	P
Noise rejection	Should be capable of operation in a "noisy" environment	?

Entropic

Entropic Research Laboratory Inc is a research lab for signal processing in acoustic, sonar and radar research. Their interest in speech recognition involves digital signal processing and algorithm development tool kits. These kits have been used to design and test speaker-independent connected-digit recognition systems on UNIX based workstations. This system uses pattern or whole word based pattern matching for recognition.

Table 10 - Entropic

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	Digits
Speaker-independent	Must recognize reports from different speakers	P
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	F
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	F
Key-word spotting	Should be able to spot 25 active words	F
Input	Should accept both telephone and audio inputs	?
Noise rejection	Should be capable of operation in a "noisy" environment	?

Intellivoice Communications Inc

Intellivoice supplies software and software development tools for telephone based speech recognition systems. The resource server uses on a 486/50 processor in conjunction with a DirectTalk 6000 card. The system is for recognition of voice inputs instead of entering data via the telephone touch pad. The vocabulary is approximately 50 words for speaker-independent operations. The system will accept up to 31 continuous digits and does do some key-word spotting to reject unrelated words and phrases. The system utilizes the Voice Processing Corp technology and allows for easy training of vocabulary in the field.

The system provides for discrete word recognition for vocabularies up to 1000 words. The continuous digit will not accept beeps or pauses in the speech. The system is for telephone based operations such as voice mail, voice input of data and numbers and telecommuting switching capabilities.

Table 11 - Intellivoice Communications Inc

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	Digits
Speaker-independent	Must recognize reports from different speakers	P
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	P?
Input	Should accept both telephone and audio inputs	Phone
Noise rejection	Should be capable of operation in a "noisy" environment	?

International Business Machines

IBM has a Continuous Speech Series (ICSS) for use on a PC running Windows 3.1 available in Dec 1993. The system uses many of the popular sound cards such as Sound Blaster 16 or Pro Audio Spectrum 16. The system is a speaker-independent and continuous speech system. The ICSS has an active vocabulary of 1000 words and over 20,000 words with phonetic representations in the dictionary. There is a development kit for application development of new requirements. The system is intended for application command /control, data entry/retrieval and forms filling. The cost of the system is $400.00 plus the cost of an audio board and microphone.

This system while advertised as continuous speech in all of the literature is a connected speech recognition system. Each of the words or phrases used by the system must be delineated by silence. IBM calls their syntax, "contexts". This implementation allows no key-word spotting.

Table 12 - International Business Machines

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	P?
Speaker-independent	Must recognize reports from different speakers	P
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	F
Input	Should accept both telephone and audio inputs	Audio
Noise rejection	Should be capable of operation in a "noisy" environment	?

Kurzweil AI

Kurzweil AI has a number if interesting products, but the company's main business is in dictation type systems and voice-activation of Windows and DOS applications. The system focuses directly on the medical community. Discrete words or phrases are used for the dictation. The system is speaker-dependent. The software is run on PCs and has a active vocabulary of 50,000 word from a 200,000 dictionary. The system is supplied with a ISA or EISA board and Headset microphone or handset.

Table 13 - Kurzweil AI

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	F
Speaker-independent	Must recognize reports from different speakers	F
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	?
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	F
Input	Should accept both telephone and audio inputs	Audio
Noise rejection	Should be capable of operation in a "noisy" environment	?

Lernout and Hauspie Speech Products

The Lernout and Hauspie Automatic Speech Recognition (ASR) system is a speaker-independent software package built for continuous speech recognition with regular language grammars. The continuous speech is based on Workstation platforms and not available at this time for audio boards. ASR can accommodate whole word models, but is based on phoneme models. The system is capable of key-word spotting through the application development software. The system uses Neural Network algorithms to provide a greater robustness for the word recognition in a noisy environment resulting from convolution or additive noise. The company has a MS-Window based Application Programming Interface system to aid in application development. The Windows based application development is for the audio cards only at this time. A Window based development system is expected for the ASPI Elf DSP by the end of March 94. The software allows for DSP hardware upgrading that will not affect the application programs. The cost of the software is $2000.00. The system requires an audio card and microphone. If the application development becomes available for the Elf card, the card itself is $1195.00.

Table 14 - Lernout and Hauspie Speech Products

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	P?
Speaker-independent	Must recognize reports from different speakers	P
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	P
Input	Should accept both telephone and audio inputs	Audio
Noise rejection	Should be capable of operation in a "noisy" environment	P

Linkon Corporation

Linkon Corporation supplies both software (SR-4 KWS) and hardware (FC-3000-4) for speech recognition. Linkon products are mainly UNIX based, however; the FC-3000 board will operate on a 386/486 computer. The input is telephone with four ports on the card. The key-word spotting software is speaker-independent based on phonetic models. Recognition accuracy is identified at 96%. This depends on size and content of the vocabulary and perplexity. The active vocabulary is 20 words. The system includes a continuous digit recognizer. The system has input from a telephone RJ-11 connector and would require modification for audio or microphone input.

The application development tool TeraVox only runs on UNIX operation systems. The application development for the PC-based systems must be written in C programming language using a Dialogic Corporation compatible driver interface. Linkon has not produced the documentation for the application development on PCs. Documentation for interfacing requires a Voice Software Reference Manual for MS-DOS from Dialogic Corporation. Words can be added to the dictionary using a program called LexTools. (This speech recognition system for the PC-based DSP appears to be an advanced development model not fully integrated into the product line.)

Table 15 - Linkon Corporation

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	P
Speaker-independent	Must recognize reports from different speakers	P
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	UNIX only
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	F
Key-word spotting	Should be able to spot 25 active words	P
Input	Should accept both telephone and audio inputs	Phone
Noise rejection	Should be capable of operation in a "noisy" environment	?

Speech Systems Incorporated

Speech Systems Incorporated is a PC-based system available with speaker-independent, large vocabulary, continuous speech recognition capability. The board contains two DSPs; one for acoustic analog conversion and the other for speech processing. The system includes an ISA board with 4 MB of memory, a dynamic range microphone and a System Development Kit (SDK). The SDK allows development of other applications. The speech input is connected speech using syntax defined for the application. The syntax must contain all of the words that would be spoken in the phrase in order for the system to correctly identify the format. This system works best with a push-to-talk button. Key-word spotting would have to be programmed into an application for the data acquisition requirements. Beta testing by the company (in response to a telephone call) indicates that key-word spotting could be accomplished, but phrases or reports provide better accuracy. This system allows adding words to the application dictionary by phonetic spelling. The cost of the system is $1195.00.

Table 16 - Speech Systems Incorporated

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	P?
Speaker-independent	Must recognize reports from different speakers	P
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	P?
Input	Should accept both telephone and audio inputs	Audio
Noise rejection	Should be capable of operation in a "noisy" environment	?

SRI International

SRI is a UNIX Workstation based product with plans in 1994 to include configurations of their Application Development Toolkit for PCs. The company is involved with software for large vocabulary continuous speech systems in development for language education and translation. The system called DECIPHER uses a hybrid system of Neural Network and Hidden Markov Modeling to provide the recognition capability. DECIPHER is a large vocabulary, speaker-independent, continuous speech recognition system. The system has been designed to maintain accuracy, despite variations in dialects, microphones and background noise. An Application Development Toolkit is in development at this time and should be available shortly. Application Programming Interface (API) is a set of subroutines that can be invoked by the application program for run-time control of the recognition system. The API allows preparation of a "grammar specification file" to indicate the word sequences expected for the application. The word dictionary has 160,000 words.

Table 17 - SRI International

Criteria	**Specification**	**Pass/Fail**
Continuous Speech	Must recognize continuous speech in verbal reports	P?
Speaker-independent	Must recognize reports from different speakers	P
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	F
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	P?
Input	Should accept both telephone and audio inputs	Audio
Noise rejection	Should be capable of operation in a "noisy" environment	P

Verbex

Verbex is a high end Windows command, navigation and data-entry system that uses continuous speech for inputs. The active vocabulary, depending on the memory of the system is from 300 to 2410 words. There is a development kit available for application development. The kit includes a headset microphone and board. The board contains two processors allowing operation on both fast and slow PCs. Input is via microphone or audio input jacks. The cost of the system is $695.00

Verbex uses word based recognition and requires training for each user. Thus, the system is not speaker-independent. Remote microphone input can be accomplished in a quiet environment.

Table 18 - Verbex

	Specification	**Pass/Fail**
Continuous Speech	Must recognize continuous speech in verbal reports	P
Speaker-independent	Must recognize reports from different speakers	F
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	P?
Input	Should accept both telephone and audio inputs	Audio
Noise rejection	Should be capable of operation in a "noisy" environment	F

Voice Processing Corporation

Voice Processing works with telephone based applications such as catalogue shopping, voice mail and credit card authorization. The system is PC-based and is speaker-independent. The application development kit is available for developing new applications. The software is used in Creative Lab's VoiceAssist and requires training for speaker adaptation. The system is only capable of continuous digit recognition. The cost of the Developer Kit is $9,995.

Table 19 - Voice Processing Corporation

Criteria	Specification	Pass/Fail
Continuous Speech	Must recognize continuous speech in verbal reports	Digits
Speaker-independent	Must recognize reports from different speakers	P?
Application Development kit	Must have a Application Development kit to create application interfaces for data collection	P
PC-based	Must operate on a 486 PC	P
Active vocabulary	Must have an active vocabulary capable of matching the verbal report	P
Key-word spotting	Should be able to spot 25 active words	F
Input	Should accept both telephone and audio inputs	Phone
Noise rejection	Should be capable of operation in a "noisy" environment	?

APPENDIX C

TEST PHRASES FOR DIFFERENT EMITTERS

MONITORING BRAVO SCAN BEARING ONE TWO FIVE GUIDANCE RADAR
GUIDANCE RADAR BEARING TWO SEVEN FOUR OUT OF LIMITS
NEW EMITTER BEARING TWO FOUR THREE PARAMETERS OUT OF LIMITS
BRAVO SCAN BEARING ONE SEVEN THREE GUIDANCE RADAR
GUIDANCE SIGNAL BEARING ONE ONE FIVE SEARCHING LIBRARY
SURFACE GUIDANCE SCAN ON BEARING TWO FOUR FIVE WILL SEND
GUIDANCE PARAMETERS BEARING ONE ONE FIVE WILL MONITOR
SURFACE GUIDANCE BEARING ONE THREE TWO INHIBITING ARC
GUIDANCE EMITTER BEARING ONE ONE THREE NOT AUTHORIZED
TRIBAL CLASS BEARING ZERO FOUR SIX SURFACE GUIDANCE RADAR
FRIENDLY GUIDANCE RADAR BEARING ZERO TWO TWO SCAN INHIBIT
SURFACE GUIDANCE RADAR BEARING ONE SEVEN EIGHT GONE OFF LINE
MODE CHANGE BEARING TWO TWO FIVE SURFACE GUIDANCE
GUIDANCE EMITTER BEARING ONE FIVE NINE JUST CAME UP
NEW EMITTER BEARING ZERO FOUR TWO SAME PLATFORM
WE HAVE A LOCK ON BEARING TWO THREE FOUR INBOUND MISSILE
SIGNAL GONE TO STEADY SCAN BEARING ZERO FOUR FIVE MISSILE
STEADY SCAN BEARING ONE SIX TWO MISSILE HOMER
DELTA SCAN BEARING ONE ONE FIVE MISSILE INBOUND
NEW EMITTER BEARING ZERO SIX THREE MISSILE
DELTA SCAN BEARING ZERO FIVE FOUR DELTA SCAN
LOCK ON BEARING THREE FOUR FIVE LOCKED ON MISSILE
SIGNAL GONE STEADY BEARING ZERO NINE ZERO DELTA SCAN
MISSILE LOCKED BEARING TWO ONE FIVE MISSILE LOCK ON
STEADY SCAN EMITTER BEARING TWO NINE TWO MISSILE HOMER
DELTA SCAN DELTA SCAN BEARING THREE TWO TWO MISSILE
MISSILE INBOUND BEARING TWO ONE NINE MISSILE HOMER
STEADY SCAN LOCK ON BEARING ONE SEVEN SIX MISSILE
NEW EMITTER STEADY SCAN BEARING TWO FOUR SEVEN MISSILE HOMER
DELTA SCAN INBOUND MISSILE BEARING ONE TWO SIX LOCK ON
WEATHER RADAR BEARING ONE FIVE SIX CHECKING PARAMETERS
NEW EMITTER BEARING ONE EIGHT FOUR NEUTRAL AIR
AIRLINER BEARING ONE THREE FIVE WILL CHECK TRACK
COMMERCIAL FLIGHT BEARING TWO ONE NINE WILL NOT OVER FLY
CIRCULAR SCAN BEARING TWO ONE SIX COMMERCIAL AIRCRAFT WEATHER
AIRBORNE WEATHER EMITTER BEARING ONE ONE EIGHT MODE CHANGE
MODE CHANGE BEARING THREE FOUR FIVE AIRBORNE MAPPING MODE
WEATHER RADAR BEARING ONE EIGHT EIGHT NEUTRAL AIR
NEUTRAL SEARCH MODE BEARING ZERO FOUR FOUR NO CHANGE
EMITTER CHANGE BEARING ZERO SIX THREE WEATHER MODE
CHECKED TRACK EMITTER BEARING ONE ONE FOUR NO OVER FLIGHT
HAVE NEW EMITTER BEARING ZERO THREE SIX AIR SEARCH MODE
AIRBORNE SEARCH BEARING ZERO TWO FOUR TURNED OFF
AIR NEUTRAL BEARING ONE FOUR SIX COMMERCIAL AIRLINER
AIR SEARCH ON BEARING ZERO FIVE FIVE CHECKING TRACK

BIBLIOGRAPHY

Acero, A. (1993). Acoustical and environmental robustness in automatic speech recognition. Boston, MA: Kluwer Academic.

Baber, C. (1991). Speech Technology in control room systems, a human factors perspective. New York, NY: Ellis Horwood.

Baranauskas, T. (1988). Anti-ship missile threat drives decoy development. Defense Electronics. March 1988 pg 109 - 120

Blanchard, B.S. (1992).Logistic Engineering and Management. Englewood Cliffs, NJ: Prentice Hall.

Blanchard, B.S. and Fabrycky W.J. (1990) System Engineering and Analysis. Englewood Cliffs, NJ: Prentice Hall.

Beaton, R.J. (1991). Human engineering evaluation of AN/SLQ-32(V) display device console: critique of existing system. Technical Report under contract No 230-11-110F-106-352485-1 August, 1991 (Draft V2.0).

Beaton, R.J. (1993). Shipboard assessment of required proficiency II. (SHARP II). Proposal for Vitro Corporation 16 September 93.

Beaton, R.J., Farley, W.W., McGee, M, Snow, M. (1994). Shipboard assessment of required proficiency: Development of a proficiency assessment methodology. Final Report (Draft). Virginia Polytechnic Institute and State University.

Committee on Armed Services (1987). United States Senate U.S. Government Printing Office Washington, D.C. Staff Investigation into the Iraqi attack on the USS Stark.

Committee on Armed Services (1988). United States Senate U.S. Government Printing Office Washington, D.C. Investigation into the downing of an Iranian airliner by the U.S.S. Vincennes.

Das, S., Bakis, R., Nadas, A., Nahamoo, D. and Picheny, M. (1993). Influence of background noise and microphone on the performance of the IBM Tangora speech recognition system. In 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing. (pp. II-71 -II-74).

Dyess, R.M. (1992). Operator Task Analysis of a Shipboard Electronic Warfare System. Masters Thesis, Virginia Polytechnic Institute and state University, Blacksburg, VA.

Gaster, N. (1993). Personnel discussion. 22 Sept 93

Green, C.A., and O'Shea, J.G. (1994). Modeling the AN/SLQ-32(V) operator functional recognition task. Report Number: DCL-1994-4 May 11, 1994 (Draft) Displays and Control Laboratory Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA

Hansen, J.H.L., (1993). Adaptive source generator compensation and enhancement for speech recognition in noisy stressful environments. In 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing. (pp. II-95 -II-98).

IBM Corporation. (1993). IBM Continuous speech series Version 1 Technical Reference. (SC22-0154-00). Gaithersburg, MD: Author.

Jaska, P.R. (1988). Aegis system still gets high marks. Defense Electronics October 1988. pp. 48-60.

Moscovic, S.A. (1992). Development of time and workload methodologies for micro SAINT models of visual display and control systems. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Rabiner, L. and Juang, B.H. (1993). Fundamentals of speech recognition. Englewood Cliffs, NJ: Prentice-Hall.

Sanders, M.S. and McCormick, E.J. (1993). Human factors in engineering and design. New York, NY: McGraw-Hill.

Schleher, D.C. (1986). Introduction to Electronic Warfare. Deerpark, NY: Artech House.

Systems Manual. Electronic Warfare On_Board Training Systems Manual

Walker, P.F. (1983). Smart Weapons in Naval Warfare. Scientific American. Vol 248, No 5, May 1983 pp. 53-61.

Waters, T. (1988). Stark naked. Discover. January 1988 pg 70-71

Woodson, W.E., Tillman, B. and Tillman, P. (1992). Human Factors Design Handbook. New York, NY: McGraw-Hill.