ARPA ORDER NO.: 189-1
3P10 Distributed Information Systems

R-1434-ARPA
June 1974

Military Applications of Speech Understanding Systems

R. Turn, A. Hoffman and T. Lippiatt

A Report prepared for

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Rand
SANTA MONICA, CA. 90406

The research described in this Report was sponsored by the Defense Advanced Research Projects Agency under contract No. DAHC15-73-C-0181. Reports of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

ARPA ORDER NO.: 189-1
3P10 Distributed Information Systems

R-1434-ARPA
June 1974

Military Applications of Speech Understanding Systems

R. Turn, A. Hoffman and T. Lippiatt

A Report prepared for

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Rand
SANTA MONICA, CA. 90406

PREFACE

This is one of a series of reports prepared for the Defense Advanced Research Projects Agency which present the findings of a study of voice data processing capabilities applied to defense requirements. The study was designed to augment research on speech understanding systems (SUS) currently being performed by other ARPA contractors. The present report focuses on the operationally attractive military applications of automatic speech recognition and understanding by computers. Other aspects of the study have been described in the following companion reports:

R-1356-ARPA, *The Role of Acoustic Processing in Speech Understanding Systems*, A. S. Hoffman, October 1973.

R-1377-ARPA, *Natural Language, Linguistic Processing, and Speech Understanding: Recent Research and Future Goals*, A. Klinger, December 1973.

R-1386-ARPA, *The Use of Speech for Man/Computer Communication*, R. Turn, November 1973.

The objective of this series of reports is to provide specific information on man/computer tasks in which the availability of a speech input capability would significantly enhance task performance. The findings are addressed primarily to the speech-understanding-research community and to the designers of man/computer interfaces. They should be particularly useful to the Information Processing Techniques branch of ARPA in its larger study of speech understanding by computers.

SUMMARY

This report identifies applications of speech understanding systems (SUS) in military man/computer systems which appear to provide operational benefits over current man/computer interfaces or which could lead to entirely new operational capabilities. It also provides an overview of the nontechnical factors in the military environment which are likely to affect the introduction of SUS capabilities in military systems. Among these factors are various political, manpower, and fiscal trends, such as the current pressure on military decisionmakers to consider cost savings as well as potential improvements in operational capabilities when judging the merits of new systems.

The military environment for SUS applications differs considerably from the "civilian" environment. For example:

- Man/computer tasks in military systems are often of long duration, are time-urgent, and may have critical consequences.
- High reliability is essential in many military systems.
- Military systems and their operators may be deployed in extreme climatic conditions or on mobile platforms, and they may be subjected to unusual stresses.
- Military users are accustomed to constraints and discipline in communications tasks.
- Communications security is essential in many military systems.

Consequently, the required characteristics of a military SUS interface can be expected to differ from the prototype SUS systems now in research laboratories. For example, recognition accuracy requirements will be more stringent, and military situations may not permit interactive dialogue for enhancing recognition.

Limited versions of many of the applications identified in this

report could be implemented using isolated-word speech recognition rather than continuous speech understanding. However, the continuous speech capability will be required for most of these applications if full operational benefits are to be obtained.

The potential SUS application areas discussed herein are divided into five categories. In each category, several applications are discussed in general terms and one or two are considered in detail. These categories and the applications considered in detail are:

1. Equipment and process control: The control of avionics equipment in a single-seat military aircraft. This is a typical "hands busy" situation where speech would provide the pilot with an additional communication channel.
2. Field data entry: An SUS interface for a field observer in the Army TACFIRE and TOS systems. A speech input capability could improve the observer's effectiveness and safety.
3. Cooperative man/computer tasks: A speech interface for the Tactical Coordinator (TACCO) on the Navy P-3C antisubmarine warfare patrol airplane. Speech input could significantly simplify and reduce the TACCO workload.
4. Data base management: An SUS interface for the Debarkation Control Officer on the Navy's LHA assault ships. Debarkation control requires rapid and frequent updating of a complex data base and thus could benefit from an SUS interface if near-real-time operation can be achieved.
5. Advanced applications: Applications that might be possible in the 1980s and 1990s, including automatic translation of foreign language speech, speech-operated typewriters, and computerized "staff officers."

We have not analyzed the costs of operational implementation of continuous speech understanding systems in various potential applications. Such analysis is not possible at present, as there is virtually no information available on the cost contributions of the SUS components in the present experimental systems, nor are there projections of what these costs might be four or five years from now when the first prototype systems can be produced. However, there are several factors that should lead to a general cost reduction, including the development of more efficient recognition algorithms and general advances in computer technology. The latter in particular promise large computer hardware cost reductions while increasing processing power and memory capacity. With decreasing costs and increasing operational needs for versatile man/computer interfaces, the cost-effectiveness of SUS can be expected to increase rapidly.

It is clear, however, that the transfer of speech understanding technology into operational military systems will be an evolutionary process, starting with limited applications of isolated-word speech recognition and gradually proceeding to implementation of continuous speech understanding and recognition systems as their operational suitability is demonstrated.

CONTENTS

I. INTRODUCTION

A speech understanding system (SUS) consists of hardware, software, and special man/computer interface equipment which enables speech to be used directly for computer input. The systems considered in this report may be designed to *recognize* isolated words or continuous speech (i.e., to produce a verbatim phonetic or written transcript) or to *understand* the spoken input (i.e., to deduce the correct meaning of the utterance without necessarily recognizing every word).

We are primarily concerned with SUS applications in military systems. In order for such applications to be attractive, they must satisfy at least the following criteria:

- The use of the speech interface must provide an operational capability not readily achievable by other means, or it must show a significant cost advantage over the alternatives.
- The speech interface must be natural for the given task and its operational environment.

The intrinsic characteristics of speech as a communication channel can provide significant operational advantages for an SUS. These characteristics, which are discussed in detail in a companion report [1], include the following:

- Speech is man's natural and primary communication channel.
- Speech is independent of vision and human voluntary motor activities (other than those required for speech generation); consequently, it can serve as a communication channel in situations of limited visibility, when an operator's hands are busy, or when he must move around.
- Speech contains information about the speaker--his

physiological characteristics (e.g., his vocal-tract structure), physical condition, and psychological state.

- Speech propagation is omnidirectional and requires neither a free line of sight nor physical contact with a transducer for conversion into computer-processable form.

However, these characteristics may also be sources of problems in the use of speech interfaces. For example:

- Speech production may be affected by mechanical forces on the speaker, the composition of the atmosphere, or the ambient climatic conditions.
- Speech production may be adversely affected by the physical and psychological condition of the speaker and by his ethnic and geographic background.
- Speech signals may encounter interference from other acoustic signals in the environment.

Analysis of the benefits of a proposed SUS application must include consideration of these advantages and problems, as well as the technical aspects of SUS implementation--acoustic signal processing, language and linguistic processing, and semantic analyses. Various aspects of these technical problems are addressed in two companion reports [2,3]. In particular, proposed SUS applications must be examined from the point of view of the SUS design characteristics outlined by Newell et al. [4].

An overall SUS applications analysis is far from simple, especially when the systems considered are only in the planning or R&D phases. Furthermore, while the introduction of a speech input capability in some applications may appear to produce only marginal operational benefits at a particular man/computer interface, it may provide considerable benefits in terms of the overall system. For example, replacing keyboard input devices with speech input capability could reduce operator training requirements and thereby alleviate a (hypothetical) shortage of skilled operators.

The implementation costs of an SUS, in particular, must be carefully considered from the point of view of the overall system's life cycle. There is no question that at present a speech interface requires more processing power and storage than an equivalent conventional input device. A cost comparison of the interface-equipment costs, therefore, is bound to lead to a general conclusion that the SUS equipment is much more costly than the conventional. However, if the introduction of an SUS can reduce manpower requirements (e.g., can eliminate the need for a copilot in tactical aircraft), the overall system's cost-benefit ratio may overwhelmingly favor the SUS.

Clearly, the problem of assessing the costs and benefits of SUS applications in military systems is complex and difficult. And still other factors will enter the cost-benefit analyses of these applications for operational use. For example, nontechnical factors, such as the national security policy, the political pressures on the military, and fiscal policies, must be considered. Section II discusses the current and projected mood in the military as it may affect the military applications of SUS. Section II also includes a brief overview of computer-based military systems presently in operation or in the R&D phase, and an assessment of the general implications of military applications for the design characteristics of SUS.

Other sections of this report consider specific application areas in more detail: equipment and process control (Sec. III); field data entry (Sec. IV); cooperative man/computer tasks (Sec. V); and data-base management systems (Sec VI). Section VII discusses additional advanced SUS applications that may be far in the future.

II. THE MILITARY ENVIRONMENT FOR SUS APPLICATIONS

The success of introducing speech understanding capabilities into military systems depends a great deal on the current general trends in the military environment--political, fiscal, operational, R&D--and on the nature of the military systems that evolve in response to environmental pressures. These dictate the general requirements for man/computer interfaces in military systems. Of course, the specific characteristics for military applications will be determined by the characteristics of the application areas, the tasks to be performed, and the operational environments.

We will briefly examine the general trends in the military environment and then consider in detail the military environment as it affects the SUS.

GENERAL TRENDS

The principal components of the military environment are the political, fiscal, human-resources, and operational trends, and the technological environment.

Political Trends

The present U.S. military posture is described as one of Realistic Deterrence, based on the concepts of the Nixon Doctrine and the Strategy for Peace [5]. This posture emphasizes the maintenance of *strategic sufficiency* in nuclear forces, modernization of general-purpose forces to deter nonnuclear threats, and, in cases involving low levels of conflicts and aggression in other countries, furnishing of military and economic assistance when requested and as appropriate, with the threatened nation expected to assume the responsibility of providing the manpower for its own defense.

This posture of sufficiency rather than superiority in weapon systems depends heavily on maintenance of the current margin of U.S. technological superiority over the Soviet Union to compensate for the "information lag" regarding Soviet weapon-systems development and to provide a hedge against technological surprise [5]. Of particular interest is computer technology, since the Soviet Union is now undertaking massive efforts to match the superior U.S. capabilities.

One of the implications of the Nixon Doctrine in the emerging multipolar world is the possible need for deployment of tactical forces and associated systems to a threatened country to provide support to indigenous forces. Computer-based tactical command-control systems, in particular, will need to be deployed to coordinate U.S. assistance with the operations of the native forces, as well as to provide the necessary direction, control, and surveillance and intelligence information processing. In such situations, the flexibility and effectiveness of man/computer interfaces and the interfaces for interaction with the indigenous forces will be very important and will have to be based on modern technology. Voice-operated data management systems, voice status reporting from the field, and limited-vocabulary translation could be important applications for SUS.

The strategic and theater nuclear-deterrence aspects of Realistic Deterrence have similar implications. Here the sufficiency posture requires rapid availability of information on emerging threats and an effective command-control system for processing, evaluation, and dissemination of that information. Again, automation of information processing and improvement in man/computer interfaces is a necessary prerequisite for effective crisis management in the emerging multipolar world.

In the domestic political arena, the military services are facing disenchantment regarding military programs on the part of the general public, the news media, and the Congress. Demands for total disengagement from Southeast Asia and the abolishment of the draft are examples of this disenchantment. The Congress has taken an increasingly critical view of military programs and systems and is demanding increased effectiveness and efficiency from both the systems that are procured and the personnel who man them. Automation and computer-aided operations seem the most likely means to provide these qualities for systems and their operators.

This, very roughly, is the national policy environment in which we must regard SUS applications in military systems. On the one hand, it is a favorable environment--the emphasis in maintaining the national strategic posture is on advanced technology. On the other hand, the environment is unfavorable in that the Congress is going to be very hesitant to spend large sums for automation efforts for the following reasons:

- o The SALT treaties and negotiations, along with recent improvement in relations with the Soviet Union and the Peoples Republic of China, have generated a more relaxed atmosphere.
- o A feeling exists that systems are already too sophisticated and should be simplified.
- o Many automation efforts are proposed for support areas where many feel we are "too fat" already.
- o Cost savings promised for automation have not materialized.
- o A feeling exists that increased automation actually reduces flexibility and reliability during dynamic crisis situations.

Fiscal Trends

The fiscal reality facing the military today is the staggering cost of military operations and systems. Inflation has radically reduced the purchasing power of appropriated dollars. The move to all-volunteer armed forces has led to large increases in personnel-related costs: For FY 1974, these costs comprise 56 percent of the total Defense Department budget [5]. Consequently, it is important for the military to use manpower as efficiently as possible and to capitalize on decreasing costs of computer hardware--to automate or provide semi-automated operations wherever possible. However, there is also a danger in the application of advanced technology in that the technologists and system designers are often inclined to use technology to improve performance rather than to reduce costs [6]. An increase in performance more often than not means development of new items to be added to the inventory, increases in training time for maintenance personnel, and so on. In the future, the military may be compelled to utilize technology to cut costs rather than to improve performance.

There are some serious considerations here for military systems design. One is the current tendency to develop weapon systems with general-purpose capabilities, i.e., systems that are all things to all people. But the associated complexity, the need for highly trained personnel for operation and maintenance, and the long development lead time of a general-purpose system often make it more expensive than several special-purpose systems.

In terms of system procurement and operation costs, there is an increasing tendency to apply stringent cost-benefit analyses to proposed new systems. Technological innovations in man/computer interfaces will likewise be subjected to such analyses. However, it is important to take into account the total system's life-cycle cost: development and procurement, training of operators and maintenance personnel, and all the costs associated with supporting the operations and maintenance structure, as well as the cost of disposal. In the case of innovations proposed for existing systems, such as providing the speech understanding capability at some existing man/computer interface, the cost comparisons must include the total system in its present state and with the proposed innovation. If the expected benefits due to the innovation can also be expressed in the same units of measurement (e.g., dollars), then the overall cost-benefit calculations can be made.

While SUS hardware is likely to be more costly than that for conventional input systems (such as keyboards), the SUS have considerable potential for reducing *overall* costs in systems where they are applied while at the same time improving performance. For example, in computerized military administrative systems, speech input capabilities could considerably reduce the present data-preparation and conversion activities, even though the computer interface hardware would probably be more expensive than current keypunch equipment, card readers, or optical character-recognition devices.

Human-Resources (Manpower) Trends

The military forces have nearly completed the transition to all-volunteer manpower; there are currently no draft calls issued. A profound implication of this is that the military is, in effect, in competition with the private sector for the individuals with the qualities and skills it needs. This has necessitated radical upward revision of military pay scales, resulting in large increases in manpower costs.

Another important consideration is the decline in the quality of the enlistees over the last five years; the percentage of those with above-average mental ability has decreased from 38 percent of the total new personnel in 1969 to 33 percent in the first half of 1973 [5]. At the same time, the proficiency required for the use of new military systems is increasing. Here, again, it may be necessary and would certainly be highly desirable to use advanced technology to provide simpler man/computer interfaces to reduce the training requirements. In other instances, total automation of previously manual or semiautomatic tasks may reduce the requirements for less-skilled personnel, thereby releasing funds for procurement and training of higher skills for tasks that are not totally amenable to automation. At the same time, there is a general tendency for today's youth to be more sophisticated technically, particularly in computer technology. Thus, there is a need for man/computer interfaces which will require less training time and be usable by less skilled individuals, and also a need for better utilization of the technically superior individuals who are available.

Of special interest from the SUS point of view is the drive to increase the proportion of women in the armed services. Although at present women comprise only 1 to 2 percent of the service force, in the early 1980s it is expected that 5 to 6 percent of the personnel in nearly all of the specialty fields of the services will be women. It is likely, however, that many of the women in the armed forces will gravitate toward those specialties that involve man/computer interfaces. Hence, the SUS developed for these systems will need to provide for recognition of both female and male speakers.

Operational Trends

Computer systems are used in almost all areas of military activity, from routine administrative tasks to real-time force control in strategic conflicts. They are deployed in all kinds of environments, from submarines to spacecraft, and in all climatic conditions. While not all of these systems contain interactive man/computer interfaces, many of them do. Systems that involve human interaction are rarely operated in isolation and, as a rule, their operators are subject to disturbances from other systems--acoustical noise, mechanical vibrations, or other disturbances that interfere with the operator's task or concentration. Often the computer systems operate on moving platforms. In other instances, of course, the military systems are very much like their commercial counterparts, housed in permanent buildings in well-regulated climatic environments with minimum interference from other systems.

The operations and tasks for which computer systems are used may include data input tasks in field situations or at the system facilities; control of equipment, weapons, vehicles, or processes (for example, in remotely piloted vehicles); performance of cooperative man/computer tasks; monitoring of automated activities; and routine tasks of many kinds. The performance requirements and workloads placed on the operators of these systems likewise include highly critical tasks with potential for high losses, time-urgent tasks with requirements for rapid response, vigilance tasks, and conventional, noncritical tasks.

Of special interest for the present study are those tasks in which application of speech understanding capabilities will lead to significant operational improvements. These are primarily tasks performed by an operator who must accomplish several simultaneous tasks and who could benefit greatly from the ability to use speech as the communication channel in his interaction with computerized systems. Some of the advantages (and disadvantages) of the speech input capability are examined in a companion report [1]; the potential operational advantages for several military applications are considered in detail in subsequent parts of this report.

Technological Environment

The application of speech understanding capabilities in military man/computer systems is, in large part, a technological problem, since the SUS is implemented in computer hardware. The amount and complexity of this hardware depends on the specific characteristics of the speech interface, e.g., vocabulary size, language structure, required processing speech, and the like. The corresponding software requirements are dependent on the algorithms used.

Fortunately for SUS applications, the present technological environment is characterized by steady advances in the large-scale integrated circuit (LSI) manufacturing technology; drastic reductions are being made both in the physical dimensions of the systems and in the cost of processors and memory units. For example, the All Applications Digital Computer (AADC) hardware development by the Naval Air Systems Command [7] is expected to produce systems equivalent to the IBM 360/195 and the CDC STAR-100 which will have very small volumes--1 cubic foot and 1.5 cubic feet, respectively--and whose cost will be less than 1 percent that of the present systems.

If such advances in hardware technology are realized, it is likely that the hardware costs of the future SUS can be reduced to levels of minor significance in cost-benefit comparisons with other man/computer interfaces; as a result, operational considerations will dominate such comparisons. Hopefully, these technology advances will also help to alleviate the problems of cost vs. technical sophistication which are a serious concern of numerous military managers. A representative view of this problem in relation to communication technology has been expressed by Brig. Gen. R. L. Edge [8]:

> There is a natural and legitimate desire to exploit technology. But when this desire takes the form of "we can do [a task in a] more sophisticated or more modern or more exotic way, therefore we *must*"--this is when the difficulty starts. Instead, we should insist on settling for the best cost effective mix which just *barely* does the job effectively. Why "just barely"? The simplest answer is that we shouldn't want to accept the penalties attached to over-design. These penalties can take many forms--increased weight, increased cube, decreased reliability with its increase in maintenance demand and increased number of

> specialized personnel, etc., all of which can result in diverging needed transportation to the tactical area; and, of course, the increased cost and delayed use which generally results from increased sophistication.

From the point of view of U.S. defense policy, the maintenance of U.S. technological superiority over the Soviet Union has become an important consideration. As stated by former Secretary of Defense Richardson: "We must conduct a vigorous research and development program to maintain force effectiveness and to retain a necessary margin of technological superiority" [5]. Not only is it important to maintain technological superiority in laboratory capability, this superiority must be exploitable in operational systems. The SUS research is one area in which the United States is likely to develop technological superiority, and therefore SUS developments can contribute to the overall technological superiority of the U.S. military systems.

MILITARY SYSTEMS

In the following, we present a brief overview of the types of military systems that presently exist and their man/computer interfaces. No comprehensive, consistent, nonoverlapping classification of military systems has yet been developed. However, such a classification is not necessary for the present purpose. We need only convey the "mood" of the various classes of military systems; then we concentrate on the different types of man/computer interfaces likely to be found in these systems.

One widely used classification scheme, and the one which we will follow below, categorizes military systems as strategic, tactical, intelligence, and support. Support systems include logistics, telecommunications, R&D, and administrative services. Implicit in all classes of systems, especially in the strategic and tactical, are *command-control systems* for managing the forces. These, in particular, involve information processing and man/computer interfaces.

Military system operations have customarily been categorized into the following phases (which are, however, becoming rather blurred):

- <u>Routine, peacetime activity</u>: Operation of systems for routine functions which may be either their main activity or maintenance activities to ensure preparedness for designated roles in other phases. Various aerospace surveillance systems represent the first type of function, while systems for launching ICBMs are examples of the second.
- <u>Crisis situation</u>: Activity in a time period when armed conflict is imminent or highly likely. This phase is characterized by high alert levels and preparations for possible conflict.
- <u>Conflict phase</u>: The actual war-fighting situation.
- <u>Postconflict phase</u>: A state of reconstitution, recovery, negotiations, and possible limited engagements.

As would be expected, the nature of the tasks performed by the systems and their human operators may be different in different phases; the criticality and time urgency of the tasks also vary, as do physiological and psychological stresses on the operators. These stresses, in turn, affect the operators' performance at the man/computer interface and the effectiveness of using the speech understanding capability. The implications of this variation on the man/computer interfaces include the following:

- During peacetime operation and exercises the interfaces are manned by low-level personnel (junior officers, senior enlisted men) who become proficient. During crisis or conflict phases, these interfaces are likely to be manned by higher-level personnel who tend to be unfamiliar with them and possibly unaware of what is available in the data base. The availability of a relatively sophisticated speech interface may help to bridge these operational difficulties.
- Regarding the criticality of time and the system effects, the interface may cause errors which can go un-

detected and have catastrophic effects (e.g., picking up the wrong tape, etc). A speech interface could help minimize these possibilities, since it provides the operator with additional feedback (i.e., the sound of his own utterance, and feedback from the speech recognition system).

In general, systems having radically different roles in the peacetime and crisis situations must be able to handle the changes in roles rapidly and without undue confusion.

The various roles of human operators in man/computer tasks and the nature of such tasks are discussed in a companion report [1]. We shall now consider various types of military systems.

Strategic Systems

It is customary to categorize a military system as a strategic system if its mission involves the delivery of nuclear weapons to enemy territory over intercontinental distances (strategic offensive forces) or the protection of the U.S. territory against enemy nuclear weapons (strategic defensive forces). Included in the strategic offensive forces are strategic bombers, intercontinental ballistic missiles (ICBM), and submarine-launched ballistic missiles (SLBM). Strategic defensive forces contain anti-ballistic missile systems (ABM), air defense aircraft, antisubmarine warfare (ASW) systems, and early warning, detection, and surveillance systems.

Computerized information systems are already in use in most of the strategic systems. The majority of these information systems are located in stationary, fixed facilities; others that must have a high degree of survivability (such as the command, control, and communication systems for the control of strategic offensive forces) are located in airborne command posts. Representative strategic systems include the following:

o Command-control systems for strategic offensive forces.
  -- Airborne Command Post (ACP): A system that provides

the National Command Authority (NCA) and the Strategic Air Command (SAC) with a command and control system for the strategic offensive forces. The ACP is installed in a large aircraft that will be operable during the conflict and postattack phases of a general war. The system contains computers, displays, data bases, and the airborne parts of the Command Data Buffer (CDB) and Airborne Launch Control System (ALCS).

-- Strategic Air Command Automated Control Systems (SACCS): A system that transmits, collects, processes, and displays data to assist the SAC Commander in Chief in commanding and controlling his forces.

-- World Wide Military Command and Control System (WWMCCS): A system of computers, communication links, displays, etc., to provide for worldwide control of U.S. forces.

- Command Data Buffer (CDB). A system that provides rapid, flexible remote retargeting of the Minuteman ICBM.
- Specific control systems for targeting and launching the Minuteman missiles, the SLBMs, and the strategic bombers.
- Early warning and surveillance systems to detect enemy ballistic missile and airborne attacks.

  -- Ballistic Missile Early Warning System (BMEWS): Radar and associated data processing systems to detect mass missile attacks on the United States and Canada.

  -- Semiautomatic Ground Environment (SAGE) system: A semiautomatic air-weapons control and warning system for detecting, identifying, tracking, and interceptor control of airborne weapons attacking the United States and Canada. The earliest system of its kind, SAGE involves a large amount of man/computer interaction.

- -- Airborne Warning and Control System (AWACS): A survivable airborne system providing surveillance capability and command, control, and communications functions. Its distinguishing technical feature is the capability to detect and track aircraft operating at high and low altitudes over both land and water.
- -- Various systems, both airborne and shipborne, to provide detection and neutralization of enemy submarines.
- -- Supporting data-base systems for targeting and retargeting of strategic forces, force reconstitution after a conflict, damage assessment, and the like.

The principal criteria that the man/computer interfaces in the strategic systems must meet are reliability and controllability to prevent accidental or unauthorized initiation of strategic war. The identification and authentication of individuals through speech characteristics could provide a reliable controllability and security feature.

Tactical Systems

Tactical systems are designed for use in a wide range of possible conflicts, from small subtheater and localized conventional warfare to theaterwide nuclear operations. The tactical information systems support the planning, coordination, and control of such operations. They include the tactical command-control systems, surveillance and reconnaissance systems, situation status display systems, and weapon control systems (e.g., aircraft avionics systems). They must be capable of operational flexibility, rapid deployment into a variety of climates, and interaction with other services or the armed forces of other nations, and they must be operable in inhospitable environments.

The man/computer interfaces in tactical systems are subject to a great deal of interference, and the tasks performed are likely to vary with the changing operational situation. Operating personnel, likewise, are likely to change more often than in the strategic systems. In many situations, the operator may be in personal danger or under psychological and physical stresses.

The primary functions of tactical information systems are the following [9]:

- Coordination of data collection from on-site sources and from external sources and systems.
- Coordination of data to obtain a clear picture of the tactical situation.
- Processing of data required for the decisionmaking function.
- Communication of decisions and actions to weapons, other users, or other systems.

Man/computer interfaces in these systems should be designed to relieve the operator as much as possible from tiring and repetitive operations in order to allow him to concentrate on tasks requiring judgment and experience. Typically, these tasks must be performed in real time and they require constant attention. The operators are mostly enlisted personnel with various educational backgrounds. In exercises or in battle action, they are likely to man their duty stations for long periods and are thus subject to fatigue and physical discomfort which can lead to reduction of vigilance levels and increased error rates. The extreme climatic conditions that tactical systems are likely to meet further affect operator performance. These factors are recognized by the designers of tactical command-control systems, and they attempt to simplify the man/computer interface as much as possible. Implementation of speech input capabilities may provide further simplification, with consequent improvement of operator performance.

Some of the major tactical command-control systems used by the military services are the following:

- o Army Tactical Data System (ARDATS). A collection of information processing and communication systems designed to provide the commanders with accurate, secure, and timely information, and to automate the functions of battlefield fire control. The major components of this system are [10]
  - -- TACFIRE: Tactical fire-direction system for artillery command and control. The target information is provided by forward observers through DMED digital message entry devices.
  - -- TOS: Tactical operations system for selected functions in intelligence, operations, and fire-support systems.
  - -- Missile Minder: An air defense control and coordination system.
  - -- ATMAC: Air traffic management automated centers.
  - -- CSSS: The combat services support system for providing weapons and supply status information, etc.
- o Army Automated Battlefield (IBCS). Integrated systems to provide automated battlefield environment data using sensor systems, dedicated computer systems for targeting, logistics support, fire control, and communications. Included in IBCS will be the various elements of ARDATS as well as new components.
- o Air Force Tactical Air Control System (TACS, 407L). A field-deployable tactical air control system designed to provide aircraft status information, air traffic control, tactical airstrike-request processing and strike planning, and many other related functions [11, 12]. Among its components are
  - -- TACC: Tactical air control center.
  - -- CRC: Control and reporting center.
  - -- CRP: Control and reporting post.

-- DASC: Direct air support center.

-- ALCE: Airlift control element.

- Air Force Tactical Information Processing and Interpretation System (TIPI). A modularized family of equipments designed to satisfy the complete spectrum of tactical intelligence requirements for the Air Force and also for the Marine Corps general-purpose forces.
- Navy Tactical Data System (NTDS). A system for shipboard command and control of tactical aircraft, surface ships, and submarines that furnishes the ship commanders with automatic, real-time combat situation information.
- Navy Integrated Tactical Air Control System (ITACS). A system that provides air-ground communications with discrete address to specific aircraft users, navigation collision avoidance, command and control, and air traffic control.
- Marine Tactical Data Systems (MTDS). A system that provides tactical control of Marine Corps air elements. MTDS subsystems include

  -- TACC: Tactical air control center.

  -- TAOC: Tactical air operations center.

  -- TDCC: Tactical data communications center.
- Marine Tactical Command and Control System (TACCS). A system that provides the Marine Corps with integrated tactical command and control encompassing areas from air operations to logistics. Among its elements are

  -- TCO: Tactical command operations element.

  -- TAO: Tactical air operations element.

  -- MIFASS: Marine integrated fire and air support system.

  -- MIPLOG: Marine integrated personnel and logistics system.

  -- MAGIA: Marine air-ground intelligence system.

Among the nonmilitary systems with operational requirements similar to those of the military tactical command-control systems are

- ARTS-3. Automated Radar Terminal System for air traffic control around large airports, operated by the FAA.
- The Manned Space Flight Control and Communication system, operated by NASA.
- National, state, regional, and local law-enforcement command-control and intelligence information systems [13].

To summarize, tactical command-control systems can be characterized as dealing with the control and guidance of dynamic processes which are often influenced by unpredictable and noncontrollable external factors. These systems must continuously acquire information on the status of the controlled processes and the external factors in order to direct the processes toward desired goals (which sometimes are poorly defined and dynamic).

Avionics and Equipment Control

Several types of man/machine interfaces (and man/computer interfaces) are required for controlling military avionics systems and equipment. The pilot, especially of a single-seat fighter aircraft, must control, operate, or monitor equipment for communication, navigation, fire control for missiles and guns, aircraft flight control, electronic countermeasures (ECM), target acquisition, and the like. In doing this he interfaces with a variety of controls, visual displays, and acoustic channels. The potential for a speech interface in these tasks is discussed in detail in Sec. III.

Specific avionics development programs include the following:

- DAIS. The Digital Avionics Information System developed by the Air Force Avionics Laboratory [14].
- Integrated avionics systems for the Air Force's FB-111 and B-1 aircraft.
- Helicopter avionics systems for all services.

- A-NEW. The advanced antisubmarine warfare (ASW) avionics system for the Navy P3-C aircraft, which consists of sensors and data processing and display equipment.
- SEEK BUS. An Air Force program to develop and demonstrate a time-ordered, secure, jam-resistant, digital air-to-ground and air-to-air communication system. It provides for automatic aircraft position reporting as well as for transmission of pilot-generated messages.

Supporting Systems

A number of military systems are categorized as supporting systems. These deal with logistics, maintenance, administration, R&D, medical care, security, test and evaluation, training and instruction, general communications, and so forth. Among the systems in this category presently in operation or under development by the military services are:

- ALS. The Advanced Logistics System for the Air Force.
- ADSAF. The Automatic Data System for the Army in the field.
- Base-level ADP systems for all services--personnel, finance, supply.
- VAST. The Versatile Avionics Shop Test system for the Navy.
- LCSS. The Land Combat Support System for the Army, to be used for testing and maintenance of Army missile systems and their electronic components.

The man/computer tasks and interfaces in the supporting systems include data-base inquiry and maintenance via remote terminals, data entry terminals (as in warehouse inventory-taking), and on-line programming via remote terminals. Characteristically, the tasks performed are not time-urgent, the systems are mostly in permanent fixed locations (although some are mobile), and the operators are free from environmental or climatic inconveniences.

SPEECH UNDERSTANDING SYSTEMS IN THE MILITARY ENVIRONMENT

Newell et al. [4] have characterized the SUS in terms of nineteen "problem areas" which also correspond to the principal design parameters of these systems. In the following, we shall discuss these parameters from the general viewpoint of potential military applications. This discussion is intended to help clarify the implications of the military environment on the design and use of SUS as man/computer interfaces in military systems. As specific applications are developed later in this report, these same problem areas will again appear. Each application will have its own tradeoffs in terms of these areas.

Continuous Speech

An important design decision in military applications of speech interfaces deals with the question of whether a *continuous* speech understanding capability is required or whether *isolated-word* speech recognition is adequate. Existing isolated-word speech input systems require a pause (from .1 to .25 seconds) between words or phrases that are treated and recognized as independent entities. This restricts the speaking rate and, being somewhat unnatural, may accelerate the onset of fatigue. In addition, isolated-word speaking may require more concentration on the part of operators. While the system users and operators can be trained to adjust to these requirements, the replacement of a trained operator with someone not so trained may lead to reduced system effectiveness in emergency conditions. Continuous speech input capability, especially when coupled with looser syntactic and semantic constraints, is certainly preferable from an operator's point of view, but limited versions of many of the potential military applications could be handled with the isolated-word speech recognition capability.

Multiple Speakers

System design must take into account the number of different speakers who may concurrently (but not necessarily simultaneously) use the system. To handle numerous speakers, the system must either contain speaker-independent recognition/understanding algorithms or store individualized profiles of the voice characteristics and speaking habits of each speaker. If the speech interface is associated with equipment that is operated by only one person at a time (such as voice-controlled avionics in an aircraft), each speaker could carry his speech-characteristics information on some portable storage device, such as a card or tape cassette, which is loaded into the system when he assumes control.

The man/computer interfaces in large, on-line data-base management systems found in command-control and supporting systems are operated by many operators simultaneously. Such systems may need to operate around the clock with several shifts of operators. In crisis or conflict situations, the user population of some systems may be volatile, making speaker-independence of the system a definite requirement.

Speaker Dialect

Each speaker will have specific voice characteristics (male or female), and speaking and pronunciation habits (accent, age, background). Military personnel, characteristically, have heteorogeneous backgrounds, and efforts are under way to further this trend--to increase the enlistment of females and to increase the integration of ethnic minorities. It may thus become increasingly difficult to justify the selection of operators for SUS applications on the basis of their speaking habits (e.g., to require that they be males who speak the "general American dialect").

Further, the operators in some of the tactical systems may be personnel from the military services of other countries who are likely to speak with strong accents. Nevertheless, the initial application of speech understanding capabilities in military systems can be expected to be experimental and will require imposition of restrictions on dialects. The dialect problem will have to be dealt with, however, before wide-scale operational use can be implemented.

Environmental Noise

Many military computer systems are operated in environments where the ambient noise can be controlled or has well-known stable characteristics. Other systems and their input devices are airborne or in transportable shelters where the ambient noise due to auxiliary equipment, other operators, etc., cannot be effectively abated. Terminals for data collection may even operate under battlefield conditions. Special noise-canceling microphones may alleviate a part of this problem in SUS applications.

Another problem in SUS applications in fighter aircraft or manned space vehicles results from the use of oxygen masks, which cause noisy breathing, and noise interference from oxygen metering valves. One solution to this problem involves the application of special acoustic signal processing techniques. Finally, the helium-oxygen atmosphere in undersea vessels causes changes in speech resonant frequencies and leads to loss of intelligibility in voice communications. Special filtering equipment to compensate for this phenomenon is being developed [15].

The Transducer

The transducer (microphone) and its associated communication channels can both introduce distortion and various forms of electrical noise. The ability to use the ordinary telephone, as well as radio transmissions, for SUS is quite important in many military applications. For example, the use of speech for input of fire control information in the Army's TACFIRE system would involve radio or field telephone communications. Further, in the TACFIRE application, it is also necessary to digitize the speech data at the input terminal for communication in short bursts over a narrow bandwidth channel in a store-and-forward mode.

One of the basic considerations in any SUS application at remote input terminals is the amount of speech signal processing which must or can be done at the remote terminal prior to transmission to the central facility for storage and use. There are several factors to be considered:

- <u>Bandwidth required for the transmission channel</u>. Preprocessed speech may be sent in digital form which uses less bandwidth. This may be important in tactical systems, but it is less important in administrative systems.
- <u>Transmission time</u>. Transmission of a spoken message takes much longer than transmission of a digitized message. This is important in applications such as TACFIRE, where the operator's safety may depend upon limiting transmissions to bursts of a few seconds.
- <u>Communications security</u>. Security transformations applied to voice communications tend to distort severely the decoded form of the speech signal. This does not occur with digital representation. In fact, in many security transformations, the speech signal is digitized, transformed in this form, and reconstituted into acoustic form at the receiver. This procedure shares much of the technology for the SUS front end and could be quite efficient.
- <u>Error control</u>. Voice transmissions over telephone lines or radio are subject to noise and distortion. In the digitized form, however, error detection/correction transformations can be applied to reduce these problems.
- <u>Processing requirements and costs</u>. Preprocessing equipment is required at the terminal, and a processor at the central facility. If the system must handle numerous terminals at the central processing facility in a time-shared manner, quite powerful processors may be needed.

Another consideration in the use of the speech interface at a remote terminal is the nature of the existing communication system; it may be necessary to use this system and its specific way of handling voice communications. Also, the communication system may utilize some type of vocoder or other speech compression techniques, and it may be necessary to build the speech input capability around these systems.

System Tuning

A speaker-dependent SUS must be supplied with the voice and speech characteristics of its users. This is normally accomplished by having each speaker (or a representative subset of speakers) read into the system the entire vocabulary (or certain subsets of it) one or more times.

In military SUS applications, several considerations affect the amount of tuning required or desired. In some systems where there is a large and dynamic user population, it may be desirable to minimize the tuning requirements. On the other hand, if the same system has access-control requirements, a high degree of speaker-dependence would permit identification of the individual users. In tactical systems deployed in the field or in extreme climatic conditions, health problems which affect voice characteristics but not the general ability to perform assigned duties may require frequent retuning of the SUS interfaces and thus may impair the system's effectiveness.

User Training

In most of the envisioned SUS applications, it is also necessary to train users to cooperate with the SUS in order to improve its performance or permit design simplifications. That is, the user must learn to communicate in a constrained language as well as to avoid certain speech habits. Newell et al. [4] have pointed out that humans can adapt to the use of new words or syntactical rules rather easily, but they cannot so easily alter the speech generation processes of their native languages or dialects. The latter are likely to cause a problem also in military SUS applications, since a variety of dialects must be expected among the operators.

Military communications are characterized by the use of jargon and code names, and the military training process includes teaching the "military language" and communications practices. Hence, military personnel are accustomed to training and can be expected to adapt to the SUS language requirements more readily than will civilian users. Thus, the user training problem should not be a serious consideration in the design of SUS input languages. Indeed, it may be possible to capitalize on the trainability of military users in designing constrained SUS vocabularies and syntactic structures which help to relax other design parameters.

Vocabulary

The size of the vocabulary allowed in SUS applications strongly affects the versatility and flexibility of the system, as well as the operator's ability to perform his tasks effectively. For example, a small vocabulary constrains the expressive power of the language but is easier to learn--it is certainly easier to remember 100 acceptable words than it is to recall which of 500 words may be used. For the SUS, large vocabularies mean large processing and storage requirements. One accepted approach to providing larger vocabularies while keeping processing within bounds is the use of syntactic constraints to organize the vocabulary so that only a relatively small subset needs to be searched at any one time.

The military has always strived for high intelligibility and precision in voice communications. To this end, special vocabularies have been selected and made mandatory for critical communications [16]. This experience can also be used in the selection of vocabularies for SUS applications. Here it is important to distinguish between isolated-word and continuous speech systems. For the latter, the word-to-word transitions are important and the vocabulary should be constructed to minimize transition ambiguities.

Various other difficulties in speech processing can also be alleviated by vocabulary selection. For example, if the recognition of "stop consonants" (e.g., p, t, k) at the beginning of words causes problems, such words could be eliminated from the vocabulary.

The military also has a long-standing practice of designing "dynamic vocabularies" of code words to be assigned to units or activities. Usually a code consists of a pair of natural language words not normally in the vocabulary, and its operational use varies from a few hours to weeks. Examples of code words are Snowflake, Sly Fox, Red Prince, and Chowder Hound Five. In the SUS environment, code words should be selected to enhance recognition. A special code-word dictionary could be compiled for this purpose.

In Secs. III through VII of this report, several potential SUS applications in military systems are described and specific examples of vocabularies are given. A few general comments can also be made about the vocabulary requirements for various classes of applications:

- o <u>Voice control of equipment and processes</u>. A useful vocabulary size for control of avionics, communication systems, and displays in single-seat aircraft is 100 to 150 words.
- o <u>Status reporting and field data entry</u>. In these applications the vocabulary size depends on the nature of the data. In one study of unrestricted voice communications between pilots and ground control stations [17], a 1200-word vocabulary was identified. Field data entry for fire control or battlefield intelligence purposes may require a large vocabulary (to describe enemy forces, their location, landmarks, etc.). However, syntactic rules can be used for selecting smaller subvocabularies.
- o <u>Cooperative man/computer tasks</u>. These applications involve tactical command-control, air traffic control, general problem solving, equipment checkout, computer-aided instruction, and the like. The vocabularies involved depend on the specific applications. Usually they include 50 to 100 commands to the computer, about 128 alphanumeric characters and punctuation marks for spelling and numerical data, and a set of names for data sets and computational variables.
- o <u>Data-base management</u>. This area may involve command and control systems, administrative data bases, intelligence, logistics, etc. In these SUS applications, the vocabularies should permit information retrieval by using key words and phrases (the vocabulary size can easily exceed 1000 words/phrases here). For speech input and update of the data base (e.g., personnel

> records), the vocabulary may need to be entirely open-ended in order to input names, place names, educational and employment histories, and the like. However, following the usual voice communication practices, the names could be spelled verbally or entered by keyboard.

One of the main advantages of military systems for SUS application is that constrained vocabularies are already in use and the military users are not likely to feel unduly restricted. Further, the present military vocabularies contain many words not mnemonically related to their meanings, which allows the construction of vocabularies to enhance speech recognition.

Syntactic Support

Syntactic support refers to the structuring of the commands, requests, and statements presented to the SUS. Are the positions of different word categories rigidly specified? Are alternative structures allowed? Are free, natural language expressions allowed?

The objective in imposing syntactical constraints is to provide to the SUS additional information for resolving recognition ambiguities and to identify relevant subvocabularies. In the SUS (as contrasted to recognition systems), syntax and semantic context are also used to deduce the meaning of an utterance even without complete recognition of the acoustic signals.

As in the case of vocabularies, the military communications have a tradition of rigid syntactical restrictions. For example, reporting of hostile air contacts must follow the sequence, What? Where? Whither? When? Likewise, nearly all other types of military messages are formatted and standardized, although there may exist many different reporting formats. For example, the message catalog for the SEEK BUS digital air-to-ground communication and status reporting system includes over 50 formats. A typical SEEK BUS system message has 10 to 15 sequential word positions. More specific examples of the syntax used in potentially attractive SUS applications are discussed later in this report.

Semantic Support

Semantics has to do with meanings of sentences and messages. In SUS applications, semantic processing would be used to help resolve ambiguity in homophones (words with identical pronunciation but different spelling) and homonyms (different meanings of the same word) when syntax and grammar cannot do the task.

The semantic aspects of SUS applications in the military are related to the specifics of the task performed. In a multitask application, the interpretation of, or the action taken upon receiving, a given word or phrase may depend on the task.

The semantic context may be specified by the user in advance or may be determined dynamically from the vocabulary and syntax being used. Therefore, it is not likely that many military applications will need sophisticated semantic processing; military tasks tend to be well formulated, and the context of the input statements tends to be clear.

The User Model

The term "user model" refers to the information stored in an SUS about a user (or class of users) regarding his language habits, word usage, speaking idiosyncracies, current knowledge regarding the task, relevant experience, psychological "hangups," and the like. The purpose is to enhance the semantic processing performed by the SUS and, if possible, to adapt the interaction process to accommodate the user's preferences. Although the generation and use of such models is a difficult and poorly understood process, their development promises to take the man/computer system closer to real symbiosis.

In situations that place severe stresses on human operators, considerable changes in the operators' behavior may occur despite their training in the use of the system. For example, an operator may abandon the vocabulary and syntax prescribed for SUS use and revert to using expressions more familiar and natural to him. If this can be anticipated and included by his psychological model in the SUS, the loss of SUS effectiveness may be avoided.

High-level decisionmakers in the military, if they prefer to interact directly with the information system through a speech interface, will probably be quite reluctant to change their speaking habits and vocabularies to suit the needs of the SUS. Models of their speaking habits can be of great help in increasing SUS recognition accuracy, thereby making the system useful for these applications.

System-User Interaction

Newell points out that the total success of an SUS application is determined mainly by how skillfully the system handles interactions with its users [4]. There are two facets to this interaction:

- Task-oriented interaction. Feedback regarding the reception of commands to perform a task, task accomplishment, errors, inconsistencies, etc., which all are somewhat independent of the interaction medium.
- Speech-processing-related interaction. Feedback on recognized/understood utterances, requests for clarification or rewording of an utterance, error messages, and other interaction designed to aid speech processing.

The ease and naturalness of using a man/computer interface are important factors in determining user acceptance of any such interface. Ease of use is particularly important in systems where the operator must perform several tasks concurrently, control real-time activities, operate in physically uncomfortable environments, or be subjected to psychological stresses. Numerous military systems are characterized by one or more of the above.

It appears generally desirable, but particularly so in military SUS applications (e.g., in tactical systems and in real-time equipment and process control), to reduce the amount of speech-processing-related interaction to a minimum. This implies higher recognition accuracy, more constrained vocabulary and syntax, selection of the vocabulary to enhance recognition, and more extensive user training and system tuning.

Whatever interaction will still be required should convey only the most essential feedback information and use the simplest format.

The selection and design of the computer-to-man communication link for this purpose is equally important: It affects the equipment required at the terminal, the communication bandwidth, and the effectiveness of the interaction process.

There are two choices for the interaction medium, speech and visual display. *Synthetic speech* output from the SUS processor has the advantages that (1) the entire interaction would be in the same medium, i.e., speech; and (2) the user's visual channel would not be interfered with by the feedback messages and the need to shift his attention. The disadvantages are that (1) the feedback message is volatile and requires a specific request for its repetition; and (2) speech messages take a longer time than visual messages. Finally, the user's speech channel may already be saturated with other speech communications.

The use of a *visual display* for feedback may require additional equipment, may compete for display surface space, and may require shifting of the user's attention (possibly upon an audible warning signal), but it provides a message which is stationary and can be read rapidly and repeatedly.

Finally, feedback in any form, while desirable for assisting the SUS and increasing reliability, also slows down the task-oriented interaction, thereby reducing the intrinsic speed advantage of the speech interface.

Reliability

Reliability of a speech recognition/understanding system may be measured in terms of the percentage of correct task accomplishment. For speech *recognition* (as in isolated-word speech systems), this would be the percentage of correctly identified words and phrases; for speech *understanding* (as in continuous speech systems), it would be the percentage of correctness in the final semantic interpretation of the utterances. However, certain numerical input information must be recognized accurately by both.

In noncritical SUS applications, poor reliability results in wasted time due to the need to repeat the input utterances, and this leads to poor user satisfaction. In many military applications, however, there is no time to waste, and possible misinterpretation of an input command or statement may have serious detrimental effects. In these applications, high SUS reliability is essential and must be provided.

The conventional military voice communications protocols and military manner of issuing commands can be applied to the speech interaction protocols to provide certain safeguards against misinterpretation. For example, a military action is customarily ordered by a "Stand by to do X" command. This is acknowledged by a "Standing by to do X" statement. The action is then initiated by an "Execute" or "Execute X" command. However, in various emergency situations, it may be essential to have very high reliability in the recognition of the first utterance.

The potential problem of utterances not meant as SUS inputs being so interpreted in situations where a SUS is on-line and monitoring all speech inputs can also be handled by following standard military communications practices. For example, the SUS can be assigned a code name, which is easy to differentiate from the normal conversational vocabulary (e.g., "Lola"). The SUS is then addressed by that name (e.g., "Standby Lola," "Lola out").

Another reliability consideration deals with providing for "fail-soft" design features and for independent backup systems. Situations may occur in which the operator's speech is suddenly affected, sometimes so radically that the speech interface becomes inoperative: smoke or other fumes in the facility, coughing spells, laryngitis, and other temporary voice afflictions, or unexpected acoustic noise levels in the environment. In a milder case, fail-soft design features could be used. For example, the SUS could be changed from a continuous speech understanding mode to an isolated-word recognition mode, or a very restricted emergency vocabulary/syntax could be prescribed. For SUS backup, a mechanically operated tone generator or a very simple keyboard could be provided. In the emergency mode, the operator should be able to provide the most essential task-oriented inputs, possibly at reduced rates; as a minimum, he should be able to inform the system of his condition.

Typical error rates for several experimental speech understanding and recognition systems used under laboratory conditions are listed in Table 1 [18].

Response Time

The problem of response time, also known as the "real time" problem, deals with the time required for interpreting input utterances. The amount of time to process 1 second of speech can be used as a measure measure. There are two components of the response time:

- o Task-determined. How quickly must the input utterance be correctly interpreted in order to permit efficient performance of the task (including the presentation of feedback to the speaker)?
- o Speech processing related. How quickly should the operator receive response of his utterance in order to effectively use the speech input channel and perform his specific task?

In general, in an SUS that processes continuous speech the feedback cannot be generated at the same rate and simultaneously with the input utterance. This is due to the need to examine the entire utterance to determine the semantic context before its correct meaning can be deduced and feedback provided. In simple isolated-word recognition systems, however, the feedback for each word can be produced as soon as it is recognized.

In systems where feedback is provided by synthesized speech, it may be desirable to withhold feedback until the entire sentence is completed. Otherwise the speaker may become confused as he tries to talk and listen at the same time. But the time spent on an utterance of N seconds is now at least 2N seconds. Feedback through visual channels may alleviate this confusion significantly and shorten the feedback presentation time, but this mode may require extra equipment and/or display space, or it may divert the users' attention.

In tasks requiring continuous input task performance, such as in various source data input systems, the speech processing rate determines

Table 1

PERFORMANCE OF SPEECH PROCESSING SYSTEMS

Facility and Investigator	System Capabilities	Percent Correct Recognition
BBN, Bobrow [19] (1969)	109 isolated words, single speakers	91-94
SRI, Vicens [20] (1969)	54 isolated words, single speakers	98-100
	54 isolated words, 10 speakers, pooled data, arbitrary training order	79.4
	561 isolated words	91.4
Calgary University, Hill [21] (1969)	16 isolated words, 12 unknown speakers (system trained on different speakers)	78
IBM, Dixon and Tappert [22] (1971)	250-word vocabulary, continuous speech, several speakers	75
Threshold Technology, Inc. Martin [23] (1971)	10 digits, pairs and triples, 170 male speakers (including 77-dB background noise, light labor for talkers), no adjustment from initial setting	90
Threshold Technology, Inc. Herscher and Cox [24] (1972)	10 isolated digits, male and female speakers	99
Univac, Medress [25] (1972)	100 words, 5 speakers (one used for training)	94
Texas Instruments, Doddington (1973)[a]	10 digits, continuous speech	99

[a]Private communication, July 1973.

the input data rate. Any need for feedback at all is bound to slow down the input rate and reduce the speed advantages of the speech interface.

Whether or not the response time is a critical problem in military applications depends on the specifics of the application. In general, those situations where real-time recognition is required (such as emergencies, or real-time control of equipment or processes) also use short input commands (and limited vocabularies) which are fast to process. Longer, continuous speech input utterances for source data input at low rates, and for data-base management, are not likely to demand real-time processing.

Processing Power

The processing power of a computer system is usually measured in terms of MIPS (millions of instructions per second). It has been suggested that for SUS applications the measure might be MIPS per second of speech processed [4]. We will designate this here by MIPS/S. The processing-power requirement clearly depends on most of the SUS problem areas discussed above, but no clear-cut quantitative functional relationships are available relating these to power in terms of MIPS/S. However, it is clear that any relaxing of the constraints on vocabulary size, syntax, speakers, etc., while maintaining the required response-time constant, is bound to increase the processing-power requirements.

In military systems, processing power is only one of several requirements. Also important in airborne and other mobile applications are the system's size and weight, power consumption, cooling requirements, environmental ruggedness, and the like. Military applications can involve installation of the processor in an aircraft, jeep, truck, or ship. Environmental conditions may involve vibrations and shocks, wide variations in ambient temperature, and exposure to g-loads.

However, current advances in computer technology promise the availability of large amounts of reliable, rugged, miniaturized processing power. For example, as already discussed, the Navy's AADC promises processor modules and special signal processor units which are manufactured on single 3-inch-diameter LSI wafers and, thus, radical improvements can be expected in all critical design parameters.

Other developments in computer hardware technology--solid-state mass memory units built with LSI techniques and advances in display devices--reinforce the expectations that adequate processing power will become available for sophisticated SUS applications.

The requirement for increased processing power may, however, slow the introduction of speech interfaces into military systems, since large investments have already been made in existing hardware. Military managers are likely to be reluctant to retrofit their systems to accommodate speech interfaces, especially in administrative applications, and are likely to wait until a general upgrade of their hardware is scheduled.

Memory Capacity

The present approaches to speech recognition/understanding require the storing of considerable amounts of information for each vocabulary item and for each speaker; 2000 bits per word is not unusual for high-accuracy recognition. For large vocabularies and many simultaneous speakers, the storage requirement may amount to millions of words of high-speed storage. However, memory technology is undergoing the same rapid development as processor technology, and sufficient miniaturized high-speed memory is expected to be available to handle almost all military requirements for SUS applications.

System Organization

The hardware and software organization of an SUS application as well as that of a system utilizing an SUS as an interface can greatly influence the SUS effectiveness and success. For example, to effectively provide on-line and near-real-time speech channel support for many users, as would be needed in a tactical command-control system, all possible advantages may have to be realized from advances in hardware architecture, data-base management, and special-purpose techniques for signal and natural language processing.

Complex military systems ensue from the complex nature of the tasks that they support. For example, multiprocessor architectures are being introduced for achieving computational power through concurrent processing, as well as for providing reliability and graceful degradation in the case of system malfunctions. Addition of speech interfaces is likely to require additional processors and increase the need for multiprocessing hardware architectures. Other system-oriented features which may need to be implemented include speech communications security and data security within the system (in the recognition/understanding part, in particular).

To summarize, the system organization problem tends to be more acute in the military applications of the SUS than in comparable civilian applications.

Cost

Cost has always been a major problem in developing and procuring military systems. The recent statements of high-level military managers reflect the current trends and emphases in procurement costs, i.e., technology should be used to cut costs rather than for achieving incremental (and sometimes marginal) increases in operational performance [8].

The major cost elements in an operational speech understanding/recognition system are the additional computing power and memory capacity which must be provided (either as a dedicated system interfacing with the task-processing computer system, or as a part of the latter); the recognition/understanding software; the software/hardware for interfacing with the task processor and generating feedback; and the user's input and feedback devices. In addition, various hidden costs arise from the speech interface equipment (such as space requirements in aircraft).

Another cost problem is associated with backup equipment for the SUS. Unless the reliability of the SUS hardware and software is extremely high, the conventional man/machine interface equipment would need to be kept available to assure operational continuity in cases of SUS failure. Hence, for applications where the SUS would replace the conventional interfaces, the SUS cost could turn out to be added to that of conventional interfaces.

Even in applications where the SUS equipment would replace the majority of existing conventional manual input equipment (keyboards, thumbwheels, dials, cursors, and such), the initial cost comparisons may be quite dramatically against the SUS, since manual input devices are among the lowest-cost items in any computer system, and the supporting processing and storage requirements are small. Therefore, the justification for a speech interface must come from the operational benefits derived in task performance, or from the cost aspects of the overall system.

The *operational benefits* arise from various intrinsic capabilities of speech as an input medium, such as releasing the operator's hands for other tasks, allowing mobility, and permitting high data rates.

The SUS benefits which could reduce the *overall system's operating costs* include those which can reduce manpower needs--one of the largest cost items in the military. For example, implementation of direct source data input through speech interfaces may eliminate several processing steps in present data gathering operations: longhand transcription, typing, keying and verifying operations, and intervening physical transportation of the data. Even greater manpower savings could be achieved in avionic control applications if SUS could eliminate the need for a copilot in tactical aircraft.

Operational Availability

The research, development, engineering, and testing (RDT&E) process of a military system spans several years (typically, 6 years) from concept formulation to operational use. This process includes the formal steps of exploratory development, advanced development, engineering development, operational testing, and operational use. The present estimate for the availability of a continuous speech understanding system satisfying the ARPA speech understanding research goals in the laboratory environment is 1976 [4]. Hence, the operational use of such a system in the military cannot be expected before 1980.

Speech recognition systems or more limited capabilities, such as isolated-word speech recognition systems, can be expected to be in operational use earlier. At present, the military services are supporting evaluations of operational benefits and costs of using isolated-word speech recognition systems:

- The Air Force Avionics Laboratory will evaluate a 144-word system for equipment control in a simulator early in 1974.
- The Navy Air Systems Development Center supports studies of avionics use of isolated-word speech input systems and is exploring the use of speech in man/computer tasks in ASW aircraft in a simulator.
- The Defense Supply Agency had an early but abortive experience with voice-controlled sorting of parcels in one of its depots but will perform additional experimental evaluations.

However, even if a speech input unit is perfected and made ready for operational use in an off-the-shelf manner, its incorporation into an operational system may still require several years. It is necessary to convince the system's designers and users of its operational benefits, to analyze and favorably resolve the problems associated with its incorporation into the system, and to convince the funding agencies of the necessity of its deployment.

FURTHER MILITARY-ORIENTED DESIGN REQUIREMENTS

Security

Communications and data processing security is an important design requirement in most of the military systems where SUS applications appear attractive. The use of speech as an input medium extends the protection problem from the electromagnetic domain into the acoustic domain, where eavesdropping technology is highly developed. It is important to take appropriate measures to assure that input messages which have been secured in the electromagnetic communication links are not compromised through acoustic "bugging" prior to and during transmission.

Adaptability

Military operations in the modern, multipolar world are characterized by constantly changing strategic and tactical requirements. In order to be responsive to changing tasks, the involved military systems and their man/computer interfaces must be able to adapt quickly to new requirements. In the SUS context, this implies the ability to adapt the vocabulary, syntax, user and system training, and the nature of the feedback presentation to handle different tasks or different users. In particular, the production and maintenance of recognition/understanding software may present problems in the military environment due to the complexity of such software.

Modularity

Closely coupled with the adaptability requirement is the need for modularity in SUS-oriented hardware and software; that is, the ability to assemble an SUS system to suit a particular application from sets of standard hardware and software building blocks. Modularity in the system permits structuring of different systems from the basic elements or modifying the existing systems by addition, removal, or rearrangement of the building blocks.

A prerequisite of modular design is the identification of the general structure of the system. The SUS model proposed by Reddy et al. [26] is a significant step in this direction. Indeed, the flexibility to move from the continuous speech understanding mode to the isolated-word recognition mode, as well as to allow concurrent operation in the word spotting mode, would provide a great deal of the capability required for reliability and graceful degradation under conditions of system malfunctioning, unexpected interference from the environment, or problems with the speakers.

Transferability, Standardization, and Interoperability

In the military, systems that are custom-made for a particular user, a particular computer system, or an esoteric version of a general task must usually be avoided. While "standardization" is generally understood to mean designing a class of equipment to satisfy the same specifications, "transferability" and "interoperability" need further elaboration:

- Transferability. The potential of equipment and computer software developed for a particular application by a particular user to be used for the same application by other users at their facilities without extensive modification. In the case of computer software, this also includes the ability to use the same software on different computers (which have roughly the same computational and storage capabilities).
- Interoperability. The ability of two separately designed and operated systems (e.g., the command-control systems of two military services) to exchange information readily or share the load, or the ability of one of the systems to assume the essential tasks of the other in case of loss of one of the systems.

These design requirements (or, rather, design goals) may tend to be second-order considerations in experimental speech understanding/recognition systems, but they become more important in operational systems.

TECHNOLOGY TRANSFER

New technology is introduced into the military systems in two major ways. The first is through commercial industry, which uses its own funds to apply the results of basic research to a specific product or system intended for sale to the military. Typically, this involves the development of a demonstratable prototype. The second way is for the military to pursue applications of basic research in their own laboratories or through contractors. The approach taken often depends on the nature of the military system involved.

The developers of military administrative, management, and process control systems tend to rely heavily on the commercial markets for hardware and software. Hence, they are reluctant to invest in advancing the technological state of the art. Therefore, new technologies, such as SUS, can be expected to be available in the commercial market well before they are introduced in military administrative and management systems.

The transfer of technology into weapon systems and command-control systems typically follows the second approach. Here the military system designers tend to pursue actively the introduction of new technology and the development of prototype systems; they tend to be leaders rather than followers of the commercial markets. Therefore, applications of results of speech understanding research are likely to take place earlier in tactical systems (e.g., aircraft, tactical command-control systems) than in administrative and management systems. Indeed, this seems to be the case at the present time: All current SUS-oriented development efforts in the military laboratories are related to applications in avionics control or man/computer interaction in tactical systems.

SUMMARY

In this section we have attempted to present the "mood" of the military environment as it affects the application of speech understanding and recognition capabilities in military man/machine interfaces. The major points of the discussion were the following:

- There are many kinds of military systems, applications, and operational environments in which man/machine interfaces are found. Uses of speech recognition and understanding systems at these interfaces promise considerable operational advantages in several types of military systems.
- Costs of equipment and manpower are a dominant problem area in the military. Senior military managers have suggested that new technology should be applied to reducing costs rather than gaining additional operational advantages.
- A potential SUS application must be judged from the total-system point of view, not merely by comparing

the SUS equipment cost with the costs of alternative input devices. For example, if the SUS interface in avionics control contributes to lessening the need for a copilot in a tactical aircraft, the system cost reduction may be dramatic.

- Limited versions of many of the potential military applications of SUS could be implemented with isolated-word speech recognition. Continuous speech capability is necessary for gaining the full advantage of speech interfaces.
- The user population in military SUS applications is likely to be very heterogeneous, and selection of users on the basis of dialect and speech habits is not always possible or desirable. On the other hand, training is an integral part of military life, so users could be easily trained to use a constrained, perhaps unnatural vocabulary.
- The traditional military command and voice communication practices allow considerable flexibility in construction of SUS vocabularies and in specification of syntactic structures. This flexibility can be used to reduce technical problems in speech understanding and recognition.
- Reliability is an extremely important requirement for most of the potential military SUS applications. In some of these only a limited form of feedback can be provided or is desirable; therefore, these applications require a high level of recognition/understanding accuracy.
- Some military SUS applications may be in systems where the operators are subjected to more physical discomfort and psychological stress than would normally occur in civilian or administrative systems. Simultaneous performance of multiple tasks is the rule rather than the exception in numerous potential military SUS applications.

- Miniature computers with high processing speeds and large internal memory are likely to be available in a few years. Hence, the processing and memory capacity required for continuous speech SUS applications is likely to be available in four or five years.
- Computer and communications security is an important operational requirement in many military systems and potential SUS applications.
- Transfer of speech technology to the military will tend to be spearheaded by the research laboratories and design centers of the military services in which weapon systems and command-control systems are developed. Applications in administrative systems are more likely to wait for commercial availability.

Specific SUS applications in military systems are described in the following sections.

III. APPLICATIONS IN EQUIPMENT AND PROCESS CONTROL

There are many types of man/machine systems in both military and commercial use where an operator either directly or indirectly (through a computer) controls the operation of equipment or processes. Some of these systems may need continuous attention and control, others can operate unattended or remain stable in a state specified by the human controller. Continuous control is required for equipment or processes that operate in unpredictable environments (e.g., driving an automobile in heavy city traffic) or where the controlling force must be continuously applied by the human operator (e.g., manually flying an airplane). Discrete control is used to change the operating state of equipment which can operate stably in the desired state (e.g., changing the transmission frequency of radio equipment or the display of specific information by a computer system) or which contains its own automatic control system (e.g., autopilot control of an aircraft).

The use of speech for controlling equipment is essentially a discrete control process--a word or phrase must be uttered, processed, and *correctly* understood or recognized before the desired control action can be initiated. This requires a discrete amount of time. If control must be applied in less than this amount of time, speech cannot be used. There are exceptions, of course; for example, changing the state of a continuous process under emergency conditions (e.g., shouting "stop" to terminate a process, or uttering "start" to initiate an activity).

The behavior of the discretely controlled system is not limited to stable operation in the selected state. Indeed, the requested behavior may be quite complex, requiring complex control instructions. For example, the Stanford Research Institute's simulated speech-controlled robot [27] can be requested to "pick up the green box and put it on top of the large black box," and the chess program [28] at Carnegie Mellon University can be instructed to make a move on the chessboard (in its internal computer representation). The other alternatives for controlling the robot would be to guide its actions through a continuous control system or by a sequence of discrete control actions.

In the following, we shall discuss three specific applications. Two of these, sorting processes and teleoperator control, will be described briefly. The third, speech control of aircraft avionic equipment, will be examined in more detail.

SORTING PROCESSES

One of the earliest applications of isolated-word speech recognition to be tried in practice is for sorting tasks in mail, parcel, and baggage handling [29-31].

These experimental systems are all built around a chute and conveyor belt which channel the material to be sorted into desired bins. An SUS was briefly tried for mail sorting in a post office in Philadelphia; in this application, sorting was based on ZIP codes. The effort was abandoned because of the unacceptable error rates--6 to 8 percent per digit. However, the environment was noisy, and untrained speakers with various dialects were used. Moreover, there was no tuning of the system. Nevertheless, the Post Office Department is still interested in this application, although it would like to have continuous speech capability.

A parcel-sorting SUS [30] was installed in 1971 by the Defense Supply Agency at the Memphis Depot. This activity was also discontinued after a six-month trial period, again because of an unacceptable error rate. In this case too, the environment was noisy and the operators, although selected, were poorly trained. The vocabulary used consisted of ten numerals and five additional command words: Billy, Jesse, Mistake, Preset, and Do-It-Again. The digits were spoken as two-word groups, with a 250-millisecond pause between digits. Feedback was provided on a visual display unit. Although this system was not wholly successful, the application of speech control for parcel sorting is still considered attractive, and DSA will conduct another trial with more accurate equipment.

A baggage-sorting experiment is presently being conducted by United Airlines at Chicago's O'Hare Airport, and by TWA at New York's

Kennedy Airport [29,31]. Both use limited vocabularies consisting of 25 words and digits and the names of the more common airports. Each speaker tunes the system by repeating the entire vocabulary 10 times. Feedback is provided by a display unit. The objective of this system is to achieve 33.3 sorting operations per minute, allowing 6 stations to keep a 200-bags/minute conveyor belt fully loaded.

These mail- and baggage-sorting applications represent very simple equipment-control operations. In all cases the speech interface promises performance improvement and cost savings over the previous keyboard control of the same tasks. These tasks are performed in the "hands busy" situation where speech provides the needed additional communication channel. Continuous speech recognition is not essential but becomes more and more desirable as digit groups get longer. High reliability (or adequate backup) is needed, as essentially continuous processes are controlled. Accuracy must also be high: Correction of recognized errors causes delays, while unrecognized errors cause misrouting of items and customer dissatisfaction.

The outlook for SUS applications is very good in sorting operations in warehouses, assembly-line operations, post offices, and the like.

CONTROL OF TELEOPERATORS AND ROBOTS

A *teleoperator* system is any remotely controlled system. In these systems, man is an essential element and performs all of the control functions. He remains in a safe, comfortable environment and uses a two-way communication link to control the actions of the remote equipment [32,33]. A *robot* is a remote system with a greater degree of autonomy than a teleoperator system [33,34]. Robots are equipped with more sophisticated sensors and internally programmed behavior rules. Control by man is on a grosser level, consisting of orders to perform complex tasks, including autonomously controlled motion of the robot to the task site and searching for the objects involved in task performance.

Teleoperators

A teleoperator system usually contains the following components: manipulators and end effectors, sensors, a mobility subsystem, a communications receiver and an information processor, an information display, a man in the control loop at various levels of sophistication, a set of controls, and a transmitter [33]. The processor may be at the remote control facility, at the manipulator, or at both locations. The system may also be used to perform sets of well-defined activities which could be controlled locally at the manipulator site. Some typical teleoperator systems are the devices developed by NASA for unmanned exploration of space and for operation in space stations, on planetary surfaces [35], and under water. In the military, space applications include space station and satellite operations, inspection, and repair. The principal nonspace applications involve various types of remotely piloted vehicles (RPV) and remotely manned vehicles (RMV) [36].

The feasibility of using speech input for controlling teleoperator systems depends on the nature of the control loop required. If a continuous (analog) control must be applied, a joystick or similar control device is used. If the control can be applied as a sequence of discrete steps, then speech commands can be employed (e.g., the controller can order the manipulator to move left by ordering "left, N feet"). Here the system provides continuous feedback allowing the operator to determine the location and behavior of the teleoperator.

To save communication-channel bandwidth in these SUS applications, speech recognition or understanding would be done at the control site. Digitally coded commands would be transmitted to the teleoperator. Isolated-word speech recognition capabilities would be sufficient. A vocabulary of 50 to 100 words could provide a great deal of flexibility in the teleoperator control.

The principal application of RPVs in the military is for reconnaissance, although their use as remotely controlled strike aircraft is being actively explored. The important considerations in RPV control system design are economy in RPV cost, survivability until targets are reached, effective controllability, and secure, nonjammable communications links [37]. An important cost and operations consideration is an operator's ability to monitor and control several RPVs simultaneously, i.e., keep the vehicles on prescribed flight paths in the enroute part and acquire the target and direct the final descent for weapon release. The use of voice commands, such as "RPV-X, N degrees right," may be feasible and may permit more rapid application of control than the use of a joystick control device.

The entire topic of man/computer interface design for RPV control is still very much unresolved. The speech interface may provide an operationally attractive option.

Robots

Robots, such as the proposed remotely controlled planetary rovers [38], are equipped with sensors for providing environmental information to the internal control mechanism (as well as remote controllers) which permit the system to operate in an adaptive manner (e.g., navigate in rough terrain, spot and acquire items to be manipulated, etc.). Such devices can be commanded to perform an entire sequence of actions autonomously. For example, a robot may be ordered to search for a specific object, transport it to a specified location, and then perform some operation on it.

The use of voice commands to specify a complex action is operationally attractive in robot control activities. Such actions can be described in (constrained) natural language, and the robot can be left alone to perform the action. Here, continuous speech recognition appears more attractive than isolated-word speech. Work at SRI with a simulated voice-controlled robot [27] has contributed considerably to the development of the vocabulary and linguistic aspects of speech control of robots.

AVIONICS SYSTEMS

Speech control of the avionics equipment in military aircraft promises several potential operational benefits. Applications in single-seat fighters, in particular, are of considerable interest to the Air Force and Navy, since single-seat aircraft represent the classical "hands busy" situation. Therefore, we will examine in detail the possibility of a speech-operated "computer copilot" and its speech interface with the pilot. As stated earlier, the elimination of the need for a copilot in tactical aircraft would be a significant benefit of the SUS.

Avionics Equipment

A modern military tactical aircraft is a highly complex machine. It contains electronic equipment, processors, and displays for the following major functions [39]:

- Flight control of the aircraft (e.g., autopilot, instrument landing system (ILS)).
- Navigation (e.g., inertial system, doppler radar, Loran).
- Fire control (fire control radar, forward looking infrared (FLIR) sensors, target illuminator).
- Electronic countermeasures (ECM).
- Communications (UHF, satellite communications).
- Test and fault-location equipment.
- Weapon selection and delivery systems.

In the present generation of operational tactical aircraft, these systems tend to be autonomous and to possess their own controls, processors, and displays. In the future, the Air Force's Digital Avionic Information System (DAIS) [14] will integrate many of these systems into a single system. Even then, however, the pilot will have to operate numerous selector switches and controls to select information displays, specify the information to be presented, select communications channels, select ECM and IFF (identification friend or foe) channels, select weapons and specify their control parameters, and at the same time scan the air environment for enemy aircraft, surface-to-air missiles (SAM), and friendly aircraft. Concurrently, the pilot may be preparing to execute a close-air-support action of friendly ground troops. All these activities tend to saturate the pilot's capabilities. Turning his attention from aiming a weapon to operating manual controls, even for a moment, is sufficient to break the visual contact he needs to maintain.

Larger aircraft have more crew members but they also have more complex equipment to support the variety of missions they are designed to handle. For example, the B-1 avionics system involves 22 minicomputers. Helicopters, VSTOL aircraft, and carrier-based Navy aircraft are of similar complexity.

Avionics Control

Pilots experience considerable inconvenience with manual operation of the numerous avionics controls (e.g., the Navy A-7 aircraft contains 6 control boxes requiring frequency/channel selection; for this there are 15 separate controls; in addition there are 25 other multiposition rotary switches, 7 variable controls, and a couple dozen toggle switches [17]). Numerous pilots have expressed hope that voice control of the avionics equipment will eliminate this inconvenience.

At present, both the Air Force and the Navy are continuing development of voice-operated cockpit control equipment. The Air Force Avionics Laboratory at Wright-Patterson Air Force Base is going to evaluate the Threshold Technology, Inc., isolated-word recognition system in an aircraft cockpit simulator. The Navy Air Development Center at Warrington, Pennsylvania, is working with Scope, Inc., in developing a voice control system for Navy aircraft. NASA has also been exploring voice-operated avionics for VSTOL aircraft.

SUS CHARACTERISTICS

We will now examine the application of SUS for avionics control in terms of the SUS characteristics discussed in Sec. II.

Continuous Speech. Continuous speech is desirable in avionics control applications, but not essential. Present developments use the isolated-word recognition approach. However, control phrases to be recognized may consist of 4 or 5 words which would be more naturally uttered as a continuous speech sentence.

Multiple Speakers. Only one speaker, the pilot, will use the system at a time. However, even during the same day a given aircraft may be flown by different pilots. Each pilot could tune the system during the preflight checkout process, or he could insert a tape cassette or some other storage medium containing his prerecorded speech characteristics.

Dialect. Pilots may come from different geographical and ethnic backgrounds. However, they all attend a military academy or officer training school, as well as passing through extensive flight training where communications intelligibility, among other skills, is emphasized. Hence, it is likely that their departures from the standard American dialect will be slight and can be handled as part of the system tuning.

Environmental Noise. While the aircraft's engine contributes to the noise environment in the cockpit, the principal noise source is the pilot's oxygen mask: Noise is produced by breathing and the action of the oxygen-metering valves [40]. When the pilot is subjected to high g-loads the problem is even worse. The valves produce a uniform noise spectrum which is independent of the speaker and altitude and which obscures accurate word boundary detection in isolated-word recognition systems [41]. Approaches to overcome this problem include the following:

- Detection of valve action to generate a gating signal for turning off acoustic input. (This approach has several drawbacks, including the need to modify standard equipment.)
- Modification of the oxygen-mask microphone (which may also be undesirable).
- Use of additional processing of the acoustic signal. An effective system to implement this approach has been developed [40].

Transducer. The normal transducer in a military aircraft is a microphone built into the oxygen mask of the pilot. Recent research on the breathing-noise problem has shown that masks modified to use specially designed microphones with -3 dB frequency-response knees at 300 to 400 and 3000 to 3200 Hz greatly reduce the breathing noise, thus making SUS feasible [41].

Tunability. Tuning the SUS for use by a particular pilot will pose no particular problems. If the processor has sufficient storage capacity, the speech characteristics of a few pilots who normally fly the aircraft could be stored internally. Otherwise, the system should permit loading of the speech characteristics from a tape cassette. On-line tuning of the system each time a new pilot takes over would be unacceptable in tactical situations where many missions are flown, rapid turnaround is desired, several shifts of pilots are used, and the pilots tend to be fatigued.

User Training. Pilots are already highly trained individuals. If the speech input system provides operational advantages, they can be expected to learn the vocabulary and syntactic rules easily. However, pilots undergo considerable stresses while flying combat missions, and all efforts should be made to design the interaction language to accommodate their speaking habits.

Vocabulary. The avionics control tasks can be handled with vocabularies of 100 to 150 words. A vocabulary can be arranged into subvocabularies that are associated with particular equipment being controlled or particular control tasks. For example, for voice selection of radio channels, the following vocabulary was proposed [42]. Each of the 30 channels was assigned a name from the word list:

Tiger	Stork	Bravo	Rally
Gypsy	Hero	Igloo	Lucky
Shark	Pigmy	Cupid	Razor
Zebra	Wolf	Fox	Taxi
Polka	Echo	Wasp	Yogi
Oasis	Chief	Jumbo	Tango
Angel	Eagle	Star	
Decoy	Topaz	Lasso	

A more complete avionics control vocabulary, subvocabularies, and the associated syntactic structure are shown in Fig. 1. This vocabulary was designed for the Scope Electronics, Inc., VICCI (Voice Initiated Cockpit Control & Interrogation) system [43].

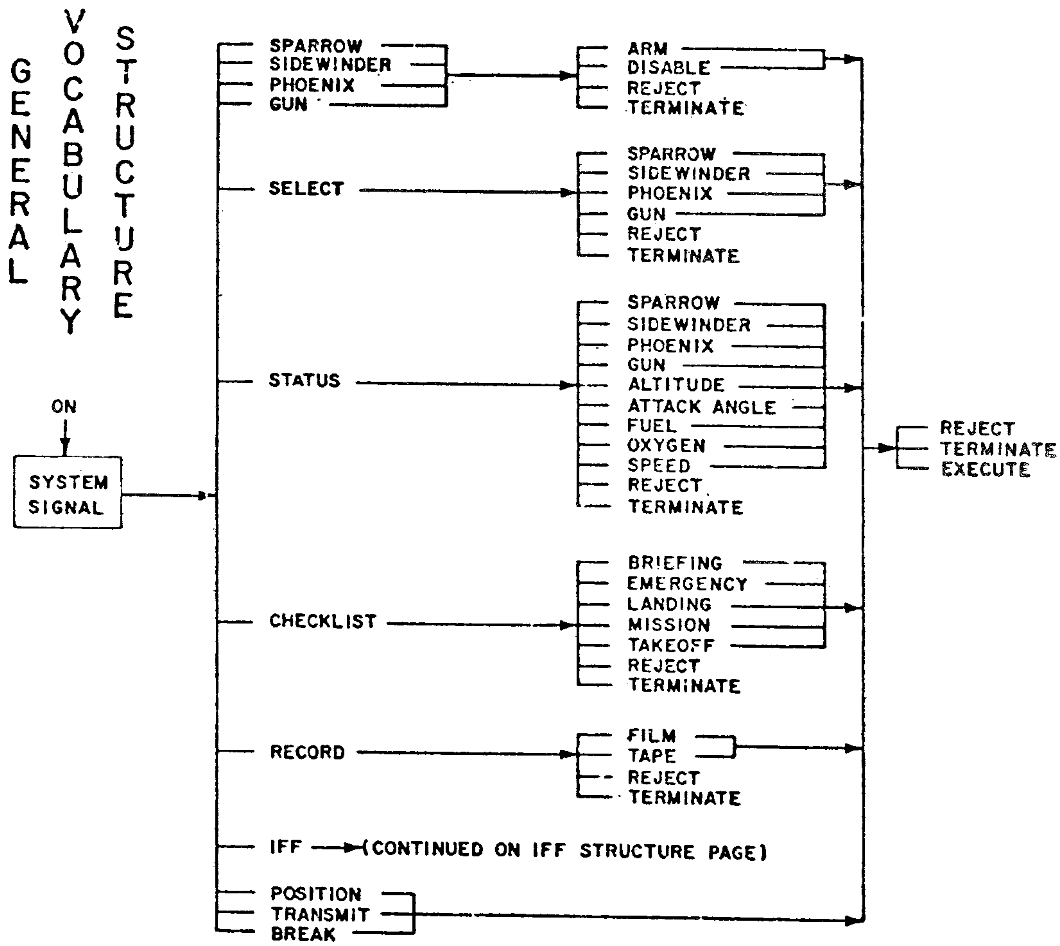

Fig. 1 — A proposed vocabulary and syntax for avionics application

IFF VOCABULARY STRUCTURE

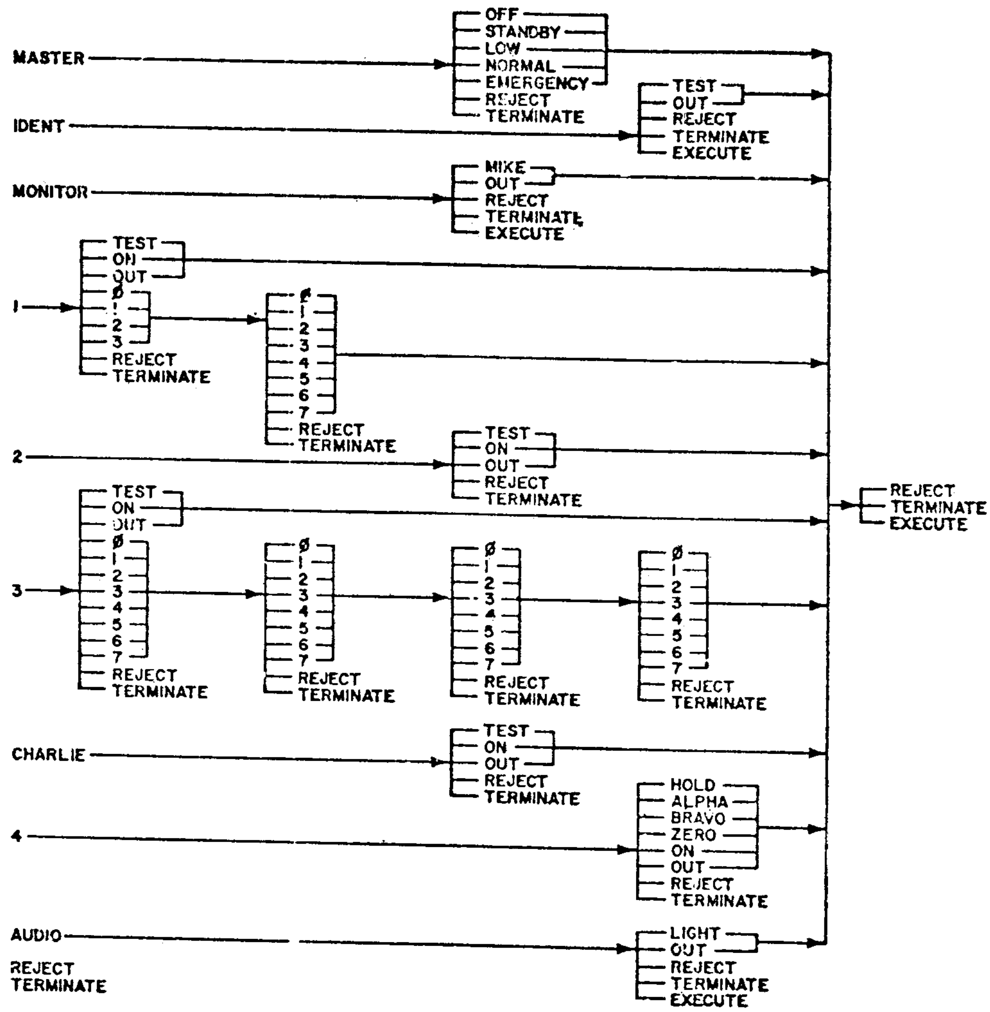

Fig. 1 (continued)

Syntactic Support. A large amount of the SUS syntactic support can be provided by rigid command structures. This allows specifying relatively small subvocabularies for various avionics control commands, as shown in Fig. 1. Pilots are accustomed to issuing stylized commands and statements.

Semantic Support. The present avionics applications are highly constrained. No ambiguity should arise if the vocabulary is properly selected. Hence, only minimal semantic support seems to be required.

User Model. The task is very simple and it is sufficient to provide only the pilot's voice and speech descriptions. However, if it is desired to equip the system with a capability to monitor the pilot's physical and psychological condition, information must be provided on each pilot's "normal" condition and normal reactions to stress. That is, a model of the pilot's physical and psychological condition must be provided.

Interaction. Feedback on the system's recognition of the pilot's commands must be provided either visually or by using synthesized speech. However, if the pilot maintains voice communication with other aircraft and ground stations (as is highly likely), speech feedback may interfere with these communciations, and vice versa. Visual feedback, on the other hand, requires a specific display or an overlay on some existing display. The latter may be preferable if the pilot can see the display without having to interrupt his other activities (i.e., if a "heads up" display is used).

Reliability. Proper control of the avionics equipment is vital to the aircraft's safety and mission success. The pilot is performing real-time tasks and hence cannot engage in lengthy interactions with an SUS to get his command properly recognized. It would appear that one request to repeat the command is all that a pilot could tolerate, so recognition must be accurate even in the cockpit noise environment, when the pilot is wearing an oxygen mask, is subjected to vibrations and g-loads, and is under psychological stresses. This implies that the recognition algorithms must be insensitive to a considerable range of changes in the pilot's voice characteristics and environmental noise levels. Experiments have shown [41] that each of these factors may in-

dependently affect the recognition accuracy by as much as 10 percent.*
For various combinations, however, the accuracy loss tends to be less than linear. Indeed, for some combinations there may even be gains in recognition accuracy.

In isolated-word recognition, where the recognition probability of every word is independent of others and has the same value, p_w, the probability of accurately recognizing an N-word utterance is $P = p_w^N$. For $P = .95$ and $N = 3$, we have $p_w = .983$. If this accuracy is to be maintained under all environmental conditions described above, the nominal recognition accuracy must be nearly 100 percent.

Clearly, a backup capability must also be provided. The simplest approach would be to keep the existing manual controls. These would have to be operated by the SUS automatically to reflect the status of the equipment.

Provisions must also be made to prevent inadvertent actuation of controls by other verbal communications activities. Either the SUS must be specifically addressed, or a push-to-talk switch must be used.

Response Time. When the pilot makes a (spoken) request for some control action, he wants it done immediately. However, a 1- to 2-second delay may be acceptable for some of the control functions.

Processing, Storage, and System Organization. The conventional SUS hardware includes an A/D converter, a special-purpose digital FFT subsystem, and an airborne general-purpose computer. If the Navy AADC airborne computer development proceeds as expected, it will provide more than sufficient processing speed and storage capacity and can be used to integrate all the avionics functions, including the SUS processing. Other available airborne computers may also be adequate, but they may have to be dedicated for the use of the SUS.

Cost and Operational Availability. Naturally, the equipment cost for a speech interface for avionics control will exceed that for the present manual controls (especially if these are to be retained for backup). However, both the Air Force and the Navy are actively

---

*Personal communciation from Cmdr. R. Wherry, Naval Air Development Center, Warrington, Pennsylvania.

investigating integration of their avionics systems. Introduction of speech control in the initial design phases of these integrated systems may cost far less than it would to retrofit them into the system later, and it could improve their effectiveness. There is no question that a reliable speech interface for controlling avionics functions would decrease the pilots' workload.

Finally, a substantial cost saving could be achieved if the co-pilot could be eliminated from present two-seater tactical aircraft (e.g., the F-4, the A-6) or their future replacements.

SUMMARY

This brief analysis of potential SUS applications in equipment control, especially for avionics control in tactical aircraft, has indicated that such applications, indeed, appear operationally advantageous. It appears that a continuous speech understanding capability is desirable, but not necessary, for the SUS; reliability is important; and required processing power and storage capacity can be expected to be available.

IV. APPLICATIONS IN FIELD DATA ENTRY

Field data entry as used here means essentially one-way communication to systems with little or no interaction or feedback. These applications can be likened to source data automation activities where the operator is remote from the system, is usually mobile, and has minimal equipment. Since the primary purpose in field data entry is to acquire and enter data into the system, the most important requirement for an SUS application here is high recognition accuracy. The opportunities for two-way interactive communication in field data entry tasks will be discussed in Sec. VI along with other SUS applications in data management systems.

Potential SUS applications in field data entry range widely within the military from speech input to the SEEK BUS digital communication system in the cockpit of a fighter aircraft to warehouse inventories taken with tape recorders. Most of these represent "hands busy" situations.

This section will address two specific field data entry SUS applications, one in source data automation and the other in field tactical communications.

SOURCE DATA AUTOMATION

There are numerous SUS application areas in source data automation. All are characterized by requirements for high mobility by the operator, the desire to carry only the minimal equipment, and the need to keep the hands free for performing other tasks. One implication of these requirements is a constraint on the amount of feedback and interaction that can be provided for checking on recognition correctness. The use of synthetic speech may provide one means for providing the feedback.

A typical SUS application in field data entry is currently being investigated by the Electronic Systems Division (ESD) of the Air Force. The Air Force Military Airlift Command (MAC) has a computerized cargo control system which requires that as cargo arrives at a MAC warehouse certain information from the bill of lading or invoice must be entered into the system. The system then compares the actual arrival time with the planned arrival so that the cargo can be properly scheduled for further shipment. Currently, the warehouse personnel transcribe the required information from the lading bill on a coding form which is then keypunched and read into the computer. This procedure is subject to high error rates and has become a major bottleneck in the system's information flow. As a result, shipping schedules frequently cannot be met and have to be revised.

Use of a speech recognition system in this process would permit warehouse personnel equipped with small radio transceivers to interact directly with the system and rapidly enter the required information. The net result would be not only improved information flow and reliability, but also reductions in manpower needs.

FIELD DATA ENTRY IN TACTICAL COMMUNICATIONS

There are currently numerous uses for field data entry devices in the Army and the Marine Corps. Until recently, communication was handled man to man via voice or teletype. Now both services have procured, and are in the process of procuring more, field-deployable tactical computer systems. These systems are used for many different purposes and require different kinds of data input/output devices. The input devices currently being used for input from field units go by a number of names, including DMED (Digital Message Entry Device), FFMED (Fixed Format Message Entry Device), and MID (Message Input Device). Although the devices vary slightly in their construction, they all have essentially the same function, that is, they allow an operator to dial in and transmit a 25- to 30-character message. All characters are numeric codes which vary from 0 to 9 or 0 to 15, depending on the device and the application. The primary current and planned use of these devices is to transmit preformatted digital messages over existing voice channels, either radio or direct voice, into the tactical computer systems. In some cases, they will supplement voice communications by providing a small lightweight "teletype" capability to the smaller forward units in the field. In other cases, the devices will be used as a direct interface between forward observers (artillery spotters) and the fire control computers. These devices will be used mostly to transmit data concerning

- o Fire control
- o Reconnaissance
- o Unit status
- o Requests for support

This method of data transmission offers a number of advantages over the normal voice communication. The message is stored as it is being input and then is transmitted in burst mode, which takes between 1.3 and 1.5 seconds. This increase in data transmission speed results in reduced bandwidth requirements and reduced detectability by enemy forces. It also provides a direct interface to the field computer systems. In the case of the Army's TACFIRE system [10,44,45], the forward observer is tied to the artillery fire control system; and in the cases of the Army's Tactical Operations System (TOS) and the Marines' Data Transmission and Switching System (DTAS), the remote units are tied directly to a message switching storage and retrieval system. The advantages are obvious: There is no need to transcribe and route messages into those systems, thus there are no transcription errors and message handling and request approval coordination are expedited.

These input devices are not without problems, however. In recent field tests, operator input error rates were high, and the Army is very concerned about the man-machine interface. The devices are typically used by the lowest echelons under adverse conditions. These include company-level and forward observer personnel who are probably under the most stress. To stop what they are doing and dial in messages with thumbwheels in a fixed format is difficult, especially at night. During daytime they probably are using binoculars to scan the battlefield, and they must put them down to operate the equipment. A hands-off device such as an SUS could eliminate some of these problems while maintaining many of the advantages of digital communications.

An SUS could also provide the following additional advantage: If a voice authentication capability is included in the SUS, it may solve a security problem that may arise if the present devices fall into the wrong hands. With a little knowledge of authentication procedures, the enemy could penetrate the system by entering phony and misleading data.

The SUS also has potential disadvantages. The first is that in these field applications, communication is essentially one-way, and there is little interaction with the system other than observing an acknowledge light on the device. It would be difficult for the operator to know if his message was "understood." It may be possible, however, to circumvent these problems by establishing some strict "reject" criteria in the SUS which could be tied to the acknowledge response. The second, and probably most important, problem is that of speech data compression prior to transmission. For routine company-level communication this is probably not necessary, but for forward observers the current short burst mode of transmitting digital data provides the observers a measure of protection not provided by voice. Therefore, where security (both physical and communications) is important, the SUS should ultimately be compatible with voice compression and vocoder techniques.

A TACTICAL OPERATIONS SYSTEM SCENARIO FOR SUS

As described previously, TOS is essentially a message switching, storage, and retrieval system which connects all levels of command in the field for the Army. Messages can be entered and retrieved from interactive terminals at the battalion level and above. From the company level, however, messages are entered in fixed format with one-way input devices. To illustrate the use and advantages of an SUS in this environment, we shall present a simple scenario, first as it would be with the existing methods and then as it would be with a sophisticated SUS.

A company is told to take and hold the territory on the far side of a river. According to the latest intelligence reports, there is a bridge crossing the river which the company is to use for its mission. However, when the company reaches the bridge they find that it has been destroyed by the enemy. Using the TOS as presently conceived, the company commander would probably have to send the following types of messages with his input device:

- Reconnaissance: The bridge has been destroyed by enemy action.
- Mission status: Mission delayed.
- Request: Send portable bridge.

If he was really in a hurry, he probably would back this up with a voice transmission to his battalion, saying, "The bridge is blown, I need a portable one immediately, when can I get it?" Although redundant, the entry of the messages into the TOS is a requirement in order to maintain a current data base for the higher levels of command. Theoretically, the TOS is supposed to aid in expediting requests such as this, but the Army personnel interviewed indicated that the TOS request would be "backed up" by voice because of the urgency of the situation.

Now, with a relatively sophisticated SUS integrated with the TOS the sequence could be as follows. The company commander gets on his radio and transmits the following: "This is Company Bravo. Mission delayed because bridge XYZ has been blown by enemy. Send portable." The SUS would first identify the speaker based on his voiceprints, then it would proceed to generate the necessary update messages to the TOS data base on the reconnaissance information concerning the bridge, the mission status, and the request for the portable bridge. In addition, since the original message was transmitted by voice, the battalion personnel, having monitored the transmission, could immediately acknowledge and begin processing the request for the portable bridge. Instead of sending four messages, the company commander, using SUS, only sends one. Granted, the SUS described here is a relatively sophisticated one, but even if the operator were constrained to "loose" formats, this method would be superior to the fixed-format entry device.

SUS CHARACTERISTICS

We shall now consider SUS characteristics in field data entry applications such as TACFIRE and TOS.

Continuous Speech. Continuous speech is certainly desirable, especially for some of the TOS applications. Single-word or single-phrase recognition, however, could be used quite readily with highly formatted messages.

Multiple Speakers. The system would definitely have to handle multiple speakers. Both TACFIRE and TOS would require the capability to handle as many as 10 speakers at one time. Probably at least 3 times that number would have to be able to use the system during different periods.

Dialect. Typically, the users of an SUS in the environment described will come from a variety of backgrounds, although they probably will all be male. This implies differences in ethnic background, geographical locale, and level of education, and a corresponding variation in dialects.

Environmental Noise. In battlefield applications, environmental noise may present the biggest problem. The background noise may not only be loud, it may also vary greatly in its characteristics, from explosions to vehicle noise. Another form of noise that may present problems is that generated by the speaker who whispers, and whispering would probably be a requirement for forward observers who are operating covertly.

Transducer. To counteract the possible background noise problem, noise-canceling microphones will probably be a requirement. For current applications, these would be coupled with the normal radio and telephone communication nets. In the future, the military plans to use all-digital communications using speech compression techniques to reduce bandwidth requirements and provide more security. Therefore, potential SUS systems should be designed to be compatible with speech compression and vocoder techniques.

Tunability. As mentioned previously there may be a large number of potential users (as many as 30), although probably no more than 10 would be using the system at one time. This may present some storage problems, since the vocabulary for some applications (especially TOS) could get quite large, depending on system design and the degree of flexibility desired.

User Training. In these applications the amount of training required will depend heavily on how rigidly the system is designed. If the system is relatively format-free, then training will be minimal. If the system requires highly formatted inputs with special vocabulary in order to reduce processing requirements and increase reliability, then more training will be required. This should, however, present less of a problem in the military community than in the civilian community, since military personnel typically are used to special training and communications discipline.

Vocabulary. The vocabulary will vary greatly depending on the particular application. A forward observer involved only in fire control will require a much smaller vocabulary than a company commander interfacing with the TOS. Again, the vocabulary will vary with the flexibility desired, and the degree that inputs are formatted. A typical FFMED format [46] is shown in Fig. 2.

Syntactic Support. Syntactic support will also vary with the flexibility desired of the system. It is likely, however, that input will be formatted into some specific input sequence which is relatively natural for military personnel. Therefore, there would be strong syntactic support for the SUS.

Semantic Support. Military jargon and vocabularies which have evolved with military communications already tend to be clear and unambiguous. Therefore, for TACFIRE and TOS applications it can be expected that in most cases the meaning of an utterance can be derived from vocabulary and syntax.

User Model. For these applications it is probably sufficient to provide only voice characteristics adequate for the vocabulary being used and, when necessary, to provide speaker identification. It would also be desirable to provide, in the case of formatted messages, the capability to handle words uttered in an incorrect sequence.

Interaction. As described previously, the current systems make use of an acknowledge light on the equipment. At a minimum, the SUS

RECONNAISSANCE REQUEST

1 2	3	4	5	6 7 8	9 10 11	12 13 14 15	16 17	18	19	20	21	22
AUTH	FORMAT ID	MSG NR	TARGET LOCATION			LATEST TIME	DATE	TARGET	RECON TYPE	SCALE X 1000 SEE INST	ANALYSIS TYPE	NO OF PRINTS
			CHART	RIGHT	UP							
	P	1	1 AN					1 BASE CAMP	1 VISUAL	1 1	1 STRIKE	
	O	2	2 WT					2 VEHICLES	2 V PIN PT	2 5	2. NR/ TYPE	
	S											
	I	3	3 XS					3 SUPPLY POINT	3. VERT STRIP	3 8	3 TGT CMPLX	
	T											
	I	4	4 XT					4 TRPS IN CONTACT	4 FWD OBL	4 10	4 MTL/ EQUIP	
00 THRU 99	O			000 THRU 999	000 THRU 999	0000 THRU 2400	01 THRU 31					1 TO 10
	N	5	5 XU					5 TRPS IN OPEN	5 SIDE OBL	5 12	5. TRAFFIC	
	0	6	6 YS					6 BUNKERS	6 SONI STRIP	6 15	6. MVE/ ACT	
		7	7 YT					7 SAMPAN	7 VERT PAN	7. 18	7 CONST	
		8	8 YU					8 LZ PREP	8 FWD PAN	8 20		
		9	9 ZS					9 PRESTRIKE	9. SLAR	9 30		
		0	0 ZT					0 VOICE COMM	0 IR	0 NONE		

Fig. 2—A FFMED message format

should provide internally some calculated measure of "understanding reliability" and then acknowledge only if these inputs yield a measure above some threshold. This threshold would probably be some function of the degree of reliability required for a particular application. For example, fire control messages would require a higher threshold than routine status messages. If the operator failed to receive an acknowledgment, then he could indicate a retransmission and retransmit the entire message. As stated earlier, this section deals with applications which are essentially one-way input with little or no feedback. Synthetic speech feedback on the communication net is a possibility, but this uses up already scarce net time and bandwidth--another problem the Army is attempting to solve.

Reliability. For these applications the primary objective is to input information which will be processed by humans. The partial exception to this is, of course, fire control, where inputs are processed by the machine but the output is monitored by humans. The other exception is message header information such as message type and sender identification which are used to route and store messages. In TACFIRE and TOS applications the emphasis is more on recognition of message content than on understanding. Much of the data are numerical in nature (e.g., times, positions, etc.), and accuracy can be of the utmost importance. For example, the coordinates transmitted when calling for artillery fire must be essentially 100 percent accurate. The reliability can be enhanced through limited feedback techniques as described in the preceding paragraph on interaction, but having to repeat messages frequently would be undesirable.

Response Time. In no case is response time critical in terms of the mission. The system should be fast enough, however, to comply with good human-factors standards for man/machine interfaces. Response times of from 1 to 5 seconds should be adequate.

Processing, Storage, and System Organization. The major constraint here is that the system be field deployable, which implies limited space and weight and a severe environment. Therefore, careful consideration would have to be given to the design of a field-deployable SUS, since this type of hardware is considerably more expensive than hardware used in civilian environments.

SUMMARY

It appears that there are several good applications for an SUS in field data entry. The potential benefits are particularly good in the tactical environment, where advantage can be taken of some of the intrinsic benefits of speech input, such as freeing of hands for other tasks, speed, reduction in communication redundancy, and voice identification. The tactical environment also presents problems in applying SUS, however, especially environmental noise and reliability requirements.

V. APPLICATIONS IN COOPERATIVE MAN/COMPUTER TASKS

We have categorized as "cooperative man/computer tasks" those tasks in which the human operator and the computer both contribute to solving a problem or performing a task. Typically, the operator selects processes to be performed by computer and the data to be processed. The computer, in addition to responding to the operator's requests, also handles communications and sensor data inputs and generates the appropriate output messages and control signals. The required man/computer interaction proceeds through an interface which includes graphic or digital displays, mechanical input devices, and, as we are exploring in this report, spoken communications.

Several representative cooperative man/computer tasks are listed below. We have deliberately excluded tasks associated with data-base management, as these will be discussed separately in Sec. VI.

- Computer-aided checkout, diagnosis, and instruction.
- Computer-aided situation monitoring and control (e.g., air traffic control tasks).
- Target search, acquisition, and weapon system control (e.g., the Tactical Coordinator task on the Navy's P-3C antisubmarine warfare (ASW) aircraft, which is similar to a number of tactical command-control systems).
- Computer programming and interactive problem solving (e.g., computer-aided on-line simulation, design, or analysis).

We will briefly discuss the potential applications of speech interfaces in these tasks.

CHECKOUT, DIAGNOSIS, AND INSTRUCTION

Applications in this category tend to involve the use of preprogrammed questionnaires and instructions to guide the user's actions or responses. For example, in checking out the operational condition of some equipment (such as space vehicles, missile systems, and aircraft), the computer program presents instructions to the operator to make certain measurements or observations and input the results. The computer then evaluates these results against prestored criteria and chooses the appropriate course of action in the checkout process. Similar procedures are followed in computer-aided diagnosis of malfunctions in equipment, or diagnosis of health problems in human patients. Computer-aided instruction, likewise, is based on presentation of instructional material as a function of the trainee's responses.

Equipment checkout and malfunction diagnosis are likely to be "hands busy" situations, since the user has to operate the equipment, make measurements, choose test sequences, and the like. A speech input capability for responding to computer instructions appears attractive, particularly when operator mobility is required. At present, an isolated-word speech recognition system is being developed for NASA.* The system will have a 100-word vocabulary and will be tested at Cape Canaveral.

In the military there are numerous types of complex systems (e.g., aircraft, missile systems) which require lengthy and involved operational checkout procedures. Fault diagnosis and isolation in these systems is, likewise, difficult. Several computer-aided systems have been developed by the military services for this task, including the Versatile Avionic Shop Test (VAST) system which is being installed aboard Navy carriers [47]. This system presently provides for operator interaction through a keyboard. A speech interface could provide operational improvements here, and an experimental program for simulating a voice-operated military maintenance system has been implemented at the System Development Corporation.†

The following scenario (suggested in Ref. 48) illustrates SUS applications in shipboard electronic equipment checkout. It is assumed

---

*Personal communication from Ron Rungie, McDonnell-Douglas Corporation, Huntington Beach, California, May 1973.

†"Military Maintenance Model," unpublished report, System Development Corporation, Santa Monica, California, 1972.

that the computerized checkout equipment provides for both speech input and synthesized speech output. A relatively inexperienced electronic technician detects a malfunctioning piece of electronic hardware. Using a small portable transceiver, he tells the computer in his own dialect and jargon which piece of equipment is malfunctioning. With its speech synthesizer the computer asks which, if any, fault lights are on. Based on this and ensuing dialogue, the computer guides the technician through a fault-isolation procedure until the malfunction is identified.

If the system is designed to present questions such as, What is the voltage between probe points A and B?, the technician's answers can be constrained to be short and to consist of a small, task-oriented vocabulary. After the malfunction is located, the computer searches its parts inventory data base and tells the technician the part numbers and locations of appropriate replacement parts. The system then guides the technician through the repair process and final testing. Throughout the activity, the technician's hands are free to use his equipment. He can move around and can get computer assistance anywhere in the ship. The technician can be relatively untrained, as he acts mainly as the eyes and hands of the computerized maintenance system. In view of the vast amounts of electronic equipment on present and future ships, such a system could provide significant operational benefits by reducing the requirements for highly trained electronics specialists.

Computer-aided instruction (CAI) systems, likewise, are finding increasing interst in the military services. Technical training of servicemen is a large operation: Tens of thousands of servicemen are trained yearly. A basic requirement for a CAI student terminal is that it provide for effortless communication with the computer system; the manipulation of the terminal should not distract the student from the course material. A speech interface may be able to provide the desired naturalness of man/computer interaction.

MONITORING AND CONTROL: AIR TRAFFIC CONTROL

In air traffic control operations, both military and civilian, human air traffic controllers interact with the aircraft pilots and with air traffic control computer systems which display aircraft location and flight plan data. At present, the controller must not only perform monitoring, managing, and decisionmaking tasks, he must also do data processing, manipulation, and recording. In addition, he is a data transmission device and organizer of data flow. The controller is in continuous communication with a network of humans (other controllers, pilots, supervisors, coordinators, etc.) over a variety of channels (radio, telephone, etc.), and he is subjected to high levels of stress due to many conflicting demands on his attention and time [49].

In civilian air traffic control, the ARTS system now under development by the FAA and already in limited use will provide considerable automation of enroute flight control data. At commercial airports, a lesser degree of automation will be provided [50]. In the military, the tactical command-control systems of all four services have computerized air traffic control elements and provide man/computer interfaces for the controllers.

In these computerized air traffic control systems, which provide for automatic monitoring of aircraft adherence to assigned flight patterns in terminal areas, the air controllers enter into the computer the instructions they transmit to the pilots over voice communication links. At present this is done by using keyboard terminals, but a speech interface would permit both tasks to be performed simultaneously, thereby considerably reducing the controller's workload. Further, a speech interface would enable voice commands to be used for requesting flight information from the data base.

The SUS characteristics for this purpose are similar to those previously discussed for other application areas in Secs. III and IV. However, some specific points need to be made.

Continuous Speech Capability. Isolated-word speech recognition would be adequate for air traffic control applications, but a continuous speech capability is preferable. This is especially so in terminal-area air traffic control tasks where the controllers are under high stress and may find the deliberate pacing of their utterances annoying. However, an existing terminal-area air traffic control simulation being generated in the laboratory environment shows that an isolated-word speech interface does not require excessively unnatural speaking.*

System Tuning and User Training. Civilian air traffic control centers are fixed facilities staffed with highly trained, permanent personnel. The user dialect and training aspects of the SUS could be handled during the normal air traffic controller training period. Considerable tuning of the system may be acceptable, and sufficient internal storage should be available for the speech characteristics and models of each controller.

Environmental Noise. The principal acoustical noise sources in the control rooms are other controllers performing their tasks and their equipment. This noise can be expected to remain rather stable, and appropriate microphones or preprocessing steps should be adequate to reduce the interference below critical levels.

Vocabulary, Syntax, and Semantic Support. The vocabulary used in air traffic control situations has been studied extensively [51,52]. As can be expected, the vocabularies differ at different locations. However, they all tend to contain airline identifications (American, United, etc.), aircraft type descriptions (DC-8, 747, etc.), numerals and letters of the alphabet, instructions to the point (climb to, descend to, etc.), and directional terms (north, southwest, etc.). Parts of a 136-word vocabulary used by the Texas Instruments simulation* are listed in Table 2. The syntax is quite rigid and messages are highly formatted even in the present manual operation. There should be no difficulties in maintaining syntactical constraints for an SUS. The syntactical structure used in another (human) simulation of voice-operated, computer-based air traffic control which involves voice inputs by pilots [53] is presented in Table 3. Vocabularies for SUS applications in military air traffic control systems can be expected to be somewhat different. For example, control of the landings of carrier-based Navy aircraft differs considerably from control of close support operations of Air Force tactical aircraft.

---

*Personal communication and unpublished notes from George Doddington, Texas Instruments Corporation, Dallas, Texas, June 1973.

Table 2

A SAMPLE VOCABULARY FOR AIR TRAFFIC CONTROL

Air Force	Track	Speed
Navy	Offset	Knots
American	Start hold	Altitude
Braniff	Reports	Feet
Continental	Enter	Heading
Delta	West	Degrees
Eastern	South	Radial
Frontier	Southeast	Miles
National	East	Aerobat
Ozark	North	Agwagon
Pan Am	Navajo	Cardinal
Texas International	Beechcraft	Centurion
Piedmont	Cessna	Skyhawk
United	Piper	Skylane
Northeast	Baron	Skymaster
Northwest	Bonanza	Skywagon
Southwest	Duke	Stationaire
Boeing	King Air	Turbo Skywagon
DC	Musketeer	Aztec
AT	Queen Air	Backup
Climb to	Turbo Barron	Clear
Descend to	Sector	Delete
Handoff	Beacon	

Table 3

AN EXAMPLE OF SYNTAX FOR AIR TRAFFIC CONTROL

1. (Type of Aircraft) IN (Location) (Altitude) (Distance-Direction)
   OUT

   Example: "AA four five four IN two zero south altitude three five altitude two zero," interpreted as, "AA 454 is inbound, presently 20 miles south of the airport at 3500 feet, descending to 2000 feet."

2. (Type of Aircraft) IN RUNWAY (Number) SEQUENCE (Number)
   OUT

   Example: "TWA two nine IN RUNWAY three four SEQUENCE three," interpreted as, "TWA 29 has been assigned runway 34 for landing as number 3 in sequence."

3. Q(Type of Aircraft) STATUS
   TIME
   RUNWAY SEQUENCE

   Example: "Q TWA two five STATUS." This is interpreted as a query: "What is the location of TWA 25?"

4. (Specification) IS (Type of Aircraft) (...) END

   Example: "Sequence IN one IS two four eight sequence IN two IS three four seven LG sequences IN three IS blank," interpreted as, "Sequence number one inbound is held by TWA 48; sequence number two inbound is held by aircraft 347 LG; sequence number three inbound is blank."

Reliability, Interaction, and Response Time. Control of aircraft operations is a critical task, and high reliability is required of the equipment, the man/computer interfaces, and the controllers themselves. The controllers' tasks are time-urgent and lengthy dialogues for correct recognition of utterances cannot be tolerated. Thus the recognition system must require no more than one repetition of an utterance to achieve 100-percent accuracy. As discussed previously, considerable amounts of system tuning, user training and models, and syntactic constraints may be applied to achieve this accuracy. Adequate backup provisions must be provided for both the computer system and the speech interface. Since the system is used to transmit instructions to pilots (or their autopilot and navigation equipment), a slight response-time delay may be acceptable for all but emergency requests and instructions. However, after passing on the instructions to one aircraft, the controller is likely to turn his attention to another, and he could not tolerate excessive delays in receiving feedback to his previous action.

Processing Power, Memory Capacity, and System Organization. An air traffic control center near a major airport is likely to have several controllers working simultaneously. In this case, each must be provided with a dedicated speech terminal processor, or a sufficiently powerful processor must be provided to handle all the controllers in a time-shared manner. Since time-shared operation introduces computational requirements of its own, the processing power required to handle N controllers simultaneously, each requiring processing power of X MIPs, requires (N + k)X MIPs, where the value of k depends on the computer system organization and the nature of its operating system. In military tactical air control systems, the processor must also satisfy size, weight, power consumption, and ruggedness constraints. However, if the Navy's AADC development succeeds, adequate processing power and memory capacity will be available.

Security. In tactical air control systems, provisions must be made for communications security and protection against enemy jamming. This implies digitizing the controller/pilot communications for subsequent application of transforms. Since the SUS also requires voice digitizing, it may be possible to combine this with security and antijamming processing.

Controller Mobility. An intrinsic advantage of the speech interface over other types of implementation is the user's freedom of movement. In air traffic control systems, this may permit more effective layout of the displays and improvements of controllers' performance.

The above discussion of potential benefits of SUS applications in air traffic control tasks represents only a preliminary examination. More detailed analyses are required before conclusions can be drawn about the operational benefits that could be achieved.

TARGET SEARCH, ACQUISITION, AND WEAPON CONTROL: THE TACTICAL COORDINATOR TASK IN THE P-3C

The Navy P-3C is a four-turboprop-engine aircraft designed for patrol and ASW [54]. The distinguishing features of the aircraft include advanced submarine detection gear which interfaces an on-board computer, the ordnance system, and the armament system. The mission of the P-3C is to search, locate, and kill submerged targets.

The determination of a target's existence depends upon the operation of various sensing devices, both on board the aircraft and dropped into the water. Their action, as well as the navigation, communication, and data processing functions for their support, are coordinated by the ASQ 114(V) computer. The overall monitoring of the search operation is done from the Tactical Coordinator (TACCO) station, and to a lesser degree, by the pilot station. Other stations are provided for the navigator/communications operator and for sensor system operators.

During an ASW patrol mission the TACCO Officer has many roles and duties. Principal among these are [55]:

- Tactician
  - Coordination of the ASW search, surveillance, and detection.
  - Coordination of ASW localization and attack systems.
  - Coordination of the intelligence collection and dissemination system.
- Communicator/coordinator
  - Communication using voice (selection of communication

channels; encoding/decoding operations; communication with controlling agencies and other ASW units).

-- Communication using data link and teletype.

-- Communication with the aircraft crew over the intercommunication system (ICS).

-- Communication using pilot display/command signals (providing pilot information for navigating and station-keeping the aircraft while on search or attack missions).

- <u>Navigator</u>

  -- Navigation using inertial/doppler systems.

  -- Navigation using TACAN, VOR, Loran, or celestial systems.

  -- Navigation using radar or visual/DR systems.

- <u>Sensor manager</u>

  -- Management of radar, TV, and MAD (Magnetic Anomaly Detection).

  -- Management of acoustic sensors and sonobuoys.

  -- Management of visual search scan.

  -- Management of electronic countermeasures (ECM).

- <u>Weapons manager</u>

  -- Management of ordnance systems (sonobuoys).

  -- Management of air-to-ground weapons systems (mines, depth bombs, torpedoes, rockets).

- <u>Assessor of systems</u>

  -- Preparation and inspection of systems.

  -- Assessment of systems' status.

The TACCO Officer operates in his various roles and performs his duties at the Tactical Data Display System (AN/ASA-70) console. The system includes the following display and control elements:

- A multipurpose data display (IP-917/ASA-70) which is a CRT system that allows displaying of scan-converted radar, low-light-level TV (LLLTV), and tactical digital data (alphanumerics, symbols, and graphics).

- o Tracking ball (for moving the cursor on display).
- o Keyboard.
- o Matrices of pushbutton indicators (for selecting and activating controls and displaying their status).
- o Control dials.

In his roles as sensor manager, weapons manager, and tactician, the TACCO is very busy and must continuously operate his display controls (the cursor tracking ball, keyboard, and various pushbuttons). Providing the TACCO with speech input capability could considerably alleviate his workload and increase his task performance effectiveness. The Naval Air Development Center in Warrington, Pennsylvania, is actively engaged in exploring this question and is planning to set up an experimental SUS capability in a TACCO station simulator.*

A specific TACCO task which requires cooperative problem solving with the computer is the establishment of points on the submarine track, using triangulation techniques. Here the TACCO Officer chooses the information from those sonobuoys which seem to offer the best triangulation data for accurate track determination. This involves operating of the cursor and keyboard, in addition to requesting other displays and communicating with the pilot or navigator.

We will now examine various SUS characteristics in the light of the TACCO application.

Continuous Speech Capability. As in most of the applications we have identified, isolated-word speech recognition capability will provide considerable operational benefits, but continuous speech understanding/recognition capability would eventually be needed for achieving the envisioned operational advantages. The TACCO Officer regularly uses voice communications with crew members and controlling agencies; he may find it annoying to keep changing his speaking habits when he is interspersing the speech interface with these communications.

System Tuning and User Training. During a given mission, there are at most two or three users of the TACCO speech interface. It would

---

*Personal communication from Cmdr. Robert H. Wherry, Naval Air Development Center, Warrington, Pennsylvania, April 1973.

appear that considerable system tuning can be provided for without affecting the operational benefits. Information on the users--the user model--should be provided to accommodate speaking fatigue and effects on voice characteristics due to long station-keeping missions. and excitement (e.g., when engaged in weapons-dropping operations). The TACCO task is complicated and requires considerable training. User training for the speech interface could be incorporated in the regular training program. In the foreseeable future, the operators can be expected to continue to be male officers.

Environmental Factors. The ambient acoustic noise is mainly that of the P-3C turboprop engines. This, however, is rather stable and predictable and could be taken into account in the speech-interface design. The aircraft is pressurized and air-conditioned. Oxygen masks are used only in emergencies. Most of the P-3C ASW operations are in an orbiting flight pattern and in relatively slow flight; there are no high-level acceleration forces involved. However, ASW missions are conducted in all weather conditions and, in addition to the regular vibration due to the engines, there may be considerable shaking and buffeting in adverse weather.

Vocabulary. In the present TACCO station implementation, the operator is provided sequences of cue messages to guide his tasks [54]. For example, at each decision point a menu of allowed actions is presented on the CRT display, and a specific format is displayed whenever the TACCO has to provide a numerical value. For the speech interface, a basic vocabulary of 100 to 150 words would be needed. The utterances are likely to be sentences composed of 2 to 5 words. Table 4 presents a list of words likely to be in the vocabulary, derived from those in present use [54].

Syntax and Semantic Support. The TACCO's interaction with the computer is currently mediated by the computer programs and thus is highly formatted. The SUS application could, at least initially, capitalize on this and impose considerable syntactical constraints. These should not significantly affect the TACCO performance. An example of the syntax that might be used is presented in Tables 5a and 5b.*

*See previous footnote.

Table 4

A PARTIAL SAMPLE VOCABULARY FOR TACCO TASKS

Change	Depth	Bearing
Display	Shallow	Contact
Start	Deep	Assign
Executive	Optional	EOM
Affirmative	Time	Radius
Negative	Verify	Range
Unknown	Correct	Circle
Understand	Interval	Horizon
Repeat	Seconds	Scales
Stop	Minutes	Position
Cease	Velocity	Expand
Erase	Bias	Category
Report	Mark	Sub
Record	Preset	Ship
Data	Release	Aircraft
Restart	Charge	Sensor
Recover	Search	Feet
LOFAR	Load	Knots
Number	Mine	RO
Wind direction	Designate	ECM
Wind speed	Unload	TV
Scale	Inventory	Visual
Amplify	MAD	Link
Select	PT	Tape
Track	DIFAR	HSP
Hook	Symbol	Index
Insert	Slew	Auto-track
Buoy	Normal	Fix
Torpedo	Expendable	Reference
Latitude	Fly-to-point	Predict
Longitude	Modify	Orbit
Reject	Accept	Interval
Impact	Arm	Hydro
Preset	Open	Close

Table 5a

A POSSIBLE SYNTAX FOR TACCO APPLICATIONS

Function	Syntax
1. Establish communication link	"*listener* (THIS IS *sender*)"
2. Alter the state of something	"CHANGE *control/display/etc.* TO *desired position*"
3a. Present selected information to the sender visually	"DISPLAY *selected information* (ON *display position*)"
3b. Present selected information to sender auditorially	"REPORT (TO *operator*) *selected information* (EVERY *no. of seconds*)"
4. Record selected information	"RECORD *selected information* (ON FILE *file descriptor*)"
5. Initiate a procedure	"START procedure"
6. Ask a yes-no question	"IS (IT *true or false* THAT) *conditional statement*"
7. Respond to yes-no question	"AFFIRMATIVE" "NEGATIVE" "UNKNOWN"
8. Confirm what has been said	"UNDERSTAND *message*"
9. Request a repeat	"SAY AGAIN (ALL AFTER *message portion*)"
10. Remove previous requests	"QUIET" "CREATE REPORT OF *selected information*" "ERASE DISPLAY OF *selected information*"

Table 5b

CONDITIONAL QUALIFIERS FOR TACCO APPLICATION SYNTAX

Conditional Qualifier Phrase	Meaning
IF	"If the 'conditional phrase' *is now* true."
{ WHENEVER EVERYTIME EACHTIME	"If the 'conditional phrase' *is now* true and *everytime* it is found to be true in the future."
{ UNTIL AS LONG AS WHILE	"If the 'conditional phrase' *continues to be* true."
{ AFTER ONCE AS SOON AS WHEN	"If the 'conditional phrase' *becomes* true."
EVERY	Time scale (EVERY 5 SECONDS)
EACH	Number scale (EVERY 500 FEET)

Work is also being sponsored by the Naval Air Development Center on computer simulation of the human operator actions that could also be used to assess the effectiveness of speech interfaces for various tasks [56,57]. Regarding semantic support, the TACCO does have several operational roles (listed above), and it may be necessary for the SUS to deduce which of these he is involved in for a given utterance. However, it would appear that only a small amount of semantic processing would be needed in the initial, limited version of this application.

Reliability, Interaction, and Response Time. While a high degree of reliability is desired in all aspects of the TACCO task, it is essential when the TACCO Officer is in the role of weapons manager, where he controls the release of ordnance and air-to-surface weapons. In other TACCO roles, such as navigator, tactician, or sensor manager, the Officer has more time available for repetition of an utterance if it is improperly interpreted by the SUS. In all cases, interaction with the SUS can be provided on the TACCO CRT display units as is done in the current system. The display could, likewise, be used to present the operator the syntactical constraints and the allowable subvocabularies. Fast response is necessary for some of the tasks, such as actuating equipment or commencing activities at a specific point in time (e.g., when the aircraft is precisely above an identified enemy submarine).

Processing Power, Memory Capacity, and System Organization. The processor presently used in the P-3C is the ASQ-114, a miniaturized, general-purpose, digital computer. It performs the functions of navigation, flight control, armament system support, and sensor data processing and display. Its maximum speed in performing additions with instruction fetch overlap is .5 MIPS. The storage capacity is 65K 30-bit words. For SUS application, either a more capable general-purpose processor or a dedicated special-purpose processor must be provided. The P-3C is a relatively large aircraft and should be able to handle the additional space, weight, and power requirements. As discussed before, powerful miniature computers for SUS applications can be expected to become available in the late 1970s.

Security. All interactions with the SUS will take place within the P-3C aircraft. Consequently, security should not be a problem at the TACCO speech interface.

The cost and operational availability of the SUS for the TACCO application are quite similar to those already discussed for other potential applications; the SUS equipment is likely to cost more than the present manual input devices, but it could improve the overall system's cost and effectiveness. The operational availability for isolated-word speech can be achieved in a few years, and several more years will be required for continuous speech.

In general, the TACCO task is representative of other man/computer cooperative problem-solving tasks in tactical command and control systems, computer-aided engineering design of equipment and systems, and computer-aided training. Further, detailed study of this SUS application is clearly warranted.

COMPUTER PROGRAMMING AND INTERACTIVE PROBLEM SOLVING

Computer programs are presently generated either by printing on a coding sheet for subsequent keying into punched cards or onto magnetic tape, or by using the keyboard of an on-line terminal for direct generation of the program in computer files. The former tends to be inefficient, the latter may require expensive on-line terminals. A different method was recently tested at the University of Pennsylvania: Programmers dictated their programs into a central audio recording/playback system [58]. The result was an immediate saving in coding time of 16 percent and an estimated 42-percent saving for programmers experienced in dictation of programs. The languages used were COBOL and PL, which are not really designed to be easily speakable.

Coding time might be reduced still further by the use of a programming language designed for audio recording. However, human operators will still be needed to transcribe the recording into computer-readable form. This may present an interesting SUS application: An SUS could be used directly with individual programmers at on-line stations, perhaps with voice answer back. Alternatively, low-cost television terminals might provide better interaction and more effective programming.

One of the problems with present programming languages is the lavish use of punctuation marks which interrupt natural speaking. However, a recent article has suggested a natural language programming system which tends to avoid extensive punctuation marks and therefore is also readily speakable [59]. Table 6 lists the initial vocabulary suggested for this language. The syntax of a speakable programming language may still have to be rather constrained, but it could be made acceptable to programmers if a few synonyms are allowed and the use of

Table 6

A SAMPLE VOCABULARY FOR SPOKEN PROGRAMMING

add	from	period
all	get	place (verb)
answer	go	point
argument	greater-than	print
box	half	product
by	halt	put
calculate	identify-as	read
call	if	record (noun)
cancel	in	repeat
character	integer	result
colon	is	right-bracket
column	it	right-paren
comma	item	row
cosine	left-bracket	semicolon
cotangent	left-paren	set
delete	let	sine
demand	line	size
deposit	log	square
digit	matrix	square root
divide	maximum	star
divided-by	minimum	step
divided-into	move	store
do	multiply	string
done	name (verb)	subtract
enter	new	subtracted-from
equals	next	sum (verb)
exponent	not-equals	table
figure	number	take
file	obtain	tangent
find	of	that
for	or	the
form	page	then
fraction	part	this
times	to	total (noun)

punctuation marks can be reduced. Continuous speech seems highly desirable for SUS applications for programming.

SUS applications in interactive man/computer problem-solving situations and in question-answering programs require flexibility in the use of natural language, as well as continuous speech understanding capability. A great deal of research has been done in computer linguistics [60]. Much of the current speech understanding research is, likewise, performed in the context of such systems [3,61,62]. Consequently, this application area is receiving considerable attention and we will not attempt to explore it further.

SUMMARY

The potential SUS applications in air traffic control and in TACCO tasks promise considerable operational advantages and warrant further, detailed study. Applications in computer-aided equipment checkout and malfunction diagnosis could alleviate the present "hands busy" situation and thereby improve the performance of these tasks. Finally, the benefits of SUS applications for on-line problem solving and programming may also be considerable.

VI. APPLICATIONS IN DATA MANAGEMENT

There are numerous potential applications in the military for voice-operated data management systems, and in most cases these applications are quite similar in concept and operation to the experimental system described by Newell, et al. [4]. Data management, as used in this report, includes spoken queries of the data base as well as data entry for the data-base update (e.g., voice-operated keypunch). The SUS applications in data management systems differ from those in field data entry (discussed in Sec. IV) in that they will support strong feedback to the user through some device, such as a CRT display or synthesized speech.

In this section, we shall divide the potential data management applications into two categories, administrative and tactical, which differ considerably in the demands they place on the system and their operating environments. Administrative systems typically operate in permanent installations in a relatively quiet environment; in this respect they are similar to the ARPA experimental systems. Tactical systems, on the other hand, must be deployable, are likely to operate in noisy, hostile environments, and may have severe response time and reliability requirements. Strategic data management systems, although not explicitly treated here, can be expected to have reliability and response-time requirements similar to those of tactical systems, but they will operate in environments similar to that of the administrative systems.

ADMINISTRATIVE SYSTEMS

There are several administrative data management systems in the Air Force which already use or are proposing the use of interactive terminals for data-base update and retrieval, making them near-term candidates for voice applications; these include the following:

- Air Force Advanced Logistics System (ALS)
- Air Force Personnel System
- Air Force Base Communication System - 1985

The other services have similar applications. These will not be discussed in depth because, as mentioned previously, the majority of potential SUS applications in administrative systems are fundamentally the same as the SDC and ARPA experimental systems [4,63,64]. We shall discuss the Air Force Base Communications system, however, because the Air Force in its study [65] specifically suggests future uses for speech interfaces that go beyond the usual administrative data management systems.

The base communications mission analysis described in Ref. 65 was an effort by the Air Force to identify, investigate, and propose conceptual solutions to base communications and information transfer problems in the 1980s. Only intrabase communications and information transfer needs were considered in the analysis. Although not explicitly stated, the apparent objective of the study was to establish a design for an essentially "paperless" administrative system.

The study recommended a totally integrated, broadband frequency-division multiplex system which would furnish all information transfer services, including analog voice, digital data, and analog pictorial data. The system would integrate the telephones, data terminals, video systems, communications processing, and the computer facility into one communications system. Several future uses for speech input were recommended:

- Voice identification to limit access to data base.
- Speech-operated input/output devices.
- Speech-operated data management system.
- Speech-operated typewriter functions.

The study pointed out, however, that the suggested applications of speech are beyond the capabilities of the current technology and that it would be unlikely that such applications would be operationally available in the 1985 time period.

The estimated personnel savings that would result from the integrated base communications system (without any SUS applications) were not dramatic--a reduction of 3 to 10 percent in manpower slots per base. The estimated investment per base was in the $80 million to $100 million range. However, the dollar savings may be even less than the estimated manpower reduction would suggest, since some slots for less skilled personnel may have to be traded for slots for highly trained technicians to maintain the system. Even with this relatively low yield, the Air Force is likely to proceed with the program, especially since much of its current intrabase communications equipment is antiquated and needs to be replaced.

This example brings up an important point regarding SUS applications to administrative systems in the military. Simply stated, the benefits (mainly in terms of overall cost savings) to be expected from improving administrative systems through investments in more advanced automation techniques may not seem clear to the military managers. Interactive data management systems, for example, have been available commercially for some time, but very few have been installed by the military in support of administrative functions.

Therefore, unless the benefits become greater and more apparent, or the required investment is significantly reduced, the military will be reluctant to make large investments to retrofit their administrative systems to provide more automation. With this in mind, it is likely that the *near-term* SUS applications will have to concentrate on providing operational benefits in tactical and strategic systems, rather than on providing cost savings for administrative systems.

TACTICAL SYSTEMS

Some examples of tactical systems where SUS could be applied to data management functions are:

- Army Tactical Operating Systems (TOS).
- Air Force Automated Tactical Air Control Center (485L-TACC).
- Navy Management Information System (MIS) for the Landing Helicopter Assault (LHA) ship.
- Marine Digital Switching and Transmission System (DTAS).

Two of these systems--the Army's TOS and the Navy's MIS--will be discussed in more detail because the application of an SUS seems to offer definite operational benefits for them.

ARMY TOS DATA-BASE QUERY AND UPDATE

The TOS data management functions are carried out at the battalion level and above, using a message input/output device called a MIOD, which is essentially a CRT with a keyboard. Using this device, personnel at battalion, brigade, division, and corps levels can enter and retrieve preformatted messages from the data base. The TOS has been initially designed to maintain Army Tactical Information in five areas:

- Friendly unit information (requests, status, etc.).
- Enemy situation.
- Enemy order of battle.
- Nuclear fire support.
- Effect of enemy nuclear strikes.

As pointed out above, both queries and inputs are accomplished within the constraints of preformatted messages. Some examples of the formats used in the prototype TOS (DEVTOS) [66] are shown in Figs. 3 and 4. They illustrate the type of vocabulary and syntax used.

To accomplish an update or query, an action officer first makes up an input worksheet for the message type that he desires. This is given to the MIOD operator, who first types a three-letter code requesting the required format, which then appears on the screen. He then proceeds to "fill in" the data as prescribed by the displayed format and transmits the message. A hard copy of the message entered will, if requested, be typed out at the operator's line printer. For purposes of transition to manual backup should TOS fail, the message worksheets are saved and filed.

As this brief description of the TOS operation indicates, there is a great deal of redundancy in the effort required to generate an input or query message. Without changing the basic operation of the TOS, and by adding an SUS capability, the action officer could interact

U A 4 — FRIENDLY UNIT INFORMATION

PRECEDENCE ! HARDCOPY !

(TASK ORGANIZATION / TASK FORCE QUERY)

ORIGIN ! / ; SCTY ! / ; FORMAT / ;

/UNIT-ID OR SWBD-DSGTR;
UNIT / ;

/NAME OF TASK FORCE;
TF-NAME / ;

/R-O/ECHELON;
ECHELON / / ;

TYPE / ;

BRANCH / ;

CATEGORY / ;

NATION / ;

/ASGD OR ATCHD/ARTY MISSION;
SUBOR-TYPE / / ;

SUBOR-TO / ;

TIME-FRAME / FROM / / TO / ;

/DATA MESSAGE ORIGINATOR; /R-O/SCTY-LEVEL;
ENTERED-BY / ; CLASSIFIED / / ;

U J 4 — FRIENDLY UNIT INFORMATION

! !

(UNIT STATUS QUERY)

ORIGIN ! / ; SCTY ! / ;

PERS / / ; PERS-PCT / / ;

TANKS / / ; TANKS-PCT / / ;

WHEEL-VEH / / ; WH-VEH-PCT / / ;

TRACK-VEH / / ; TR-VEH-PCT / / ;

ARTY / / ; ARTY-PCT / / ;

MISSILES / / ; MSL-PCT / / ;

ACFT / / ; ACFT-PCT / / ;

CBT-EFFECT / / ;

UNIT / ;

ECHELON / / ; TYPE / ;

BRANCH / ; CATEGORY / ; NATION / ;

SUBOR-TYPE / / ;

SUBOR-TO / ;

TIME-FRAME / FROM / / TO / ;

/DATA MSG ORIGINATOR;
ENTERED-BY / ; CLASSIFIED / / ;

Fig. 3—TOS query message formats UA4 and UJ4

ECHELON	Entry
Army	ARMY
Command	CMD
Army support command	ARMSUPCOM
Corps	CORPS
Corps artillery	CORARTY
Division	DIV
Division artillery	DIVARTY
Division supply command	DISCOM
Brigade	BDE
Group	GP
Regiment	REGT
Squadron	SQDN
Battalion	BN
Battery	BTRY
Company	CO

TYPE	Entry
105 mm	105MM
155 mm howitzer	155HOW
155 mm howitzer, self-propelled	155SP
175 mm gun	175MM
8 inch howitzer	8INHOW
Air cavalry	AIRCAV
Air mobile	AIRMBL
Airborne	ABN
Airborne cavalry	ABNCAV
Airborne helicopter	ABNHEL
Armored cavalry	ARMCAV
Aviation	AVN
Bridge	BRG
Bridge building	BRGBLD
Combat	CBT
Floating bridge	FLTBRG
Hawk missile	HAWK
Heavy equipment	HVEQUIP
Mechanized	MECH
Honest John missile	HJ
Light equipment	LTEQUIP
Missile	MSL
Nike-Hercules	NH
Panel bridge	PNLBRG
Pershing missile	PERSH
Sergeant missile	SGT

RELATIONAL-OPERATOR (R-O)	Entry
Equal to	EQUAL
Equal to	*(blank)*
Less than	LESS
More than	MORE
No more than	NOMORE
No less than	NOLESS

CATEGORY	Entry
Air defense	AD
Ground combat units or combat units	CBT
Fire support units	FS
Combat support units	CBTSPT
Combat service support	CSS

NATION	Entry
United States of America	US
Federal Republic of Germany	GY

CLASSIFIED	Entry
Secret	SECRET
Confidential	CONF
Unclassified	UNCLAS

BRANCH	Entry
Air defense	AD
Armored	ARMD
Artillery	ARTY
Aviation	AVN
Engineer	ENGR
Infantry	INF
Maintenance	MAINT
Medical	MED
Military intelligence	MI
Military police	MP
Ordnance	ORD
Quartermaster	QM
Signal	SIG
Transportation	TRANS

SUBOR-TYPE	Entry 1
Assigned	ASGD
Attached	ATCHD
Arty-Mission	*Entry 2*
General support	GS
General support reinforcing	GS-REINF
Direct support	DS
Direct support reinforcing	DS-REINF

Fig. 4—Vocabulary for TOS query messages UA4 and UJ4

directly with the system. He could call up the appropriate format using some plain language identifier and proceed to "fill in" the format using voice, getting immediate feedback on the CRT. When satisfied with the message content, and adding any text or remarks using the keyboard, he would then transmit the message. When the hard copy is returned, he can file a copy of it for manual backup purposes. With such an SUS capability, not only would inputs and queries be expedited, but potential savings in personnel could be realized by eliminating the requirement for terminal operators.

CONTROL OF DISPLAYS IN TOS*

As part of the TOS, the Army intends to develop a large-screen display for the top level of command--the division and the corps commanders. This display would be a substitute for the situation map that is normally used to keep track of enemy and friendly positions. This map is kept current manually by personnel who move map pins as the position and status of units change. The proposed large-screen display would be updated from a display control console by an operator who obtains information from the TOS data base, using a standard TOS input/output CRT. The large-screen display will not contain any detailed information about individual units. If the commander or his chief of staff desire more information, they typically will ask a staff member, who in turn will go to the TOS console and enter a query message. When the response to the query is complete, the staff member will then brief the commander or give him a hard copy of the response. Tests with TOS have shown that commanders don't like to deal directly with the "system" if they have to sit down and type in their own query messages.

With a modified TOS having an SUS capability, the system could work quite differently. First, assume that the large-screen display is tied directly to the TOS in such a way that as unit status messages come in, the display is updated automatically, causing the unit

---

*The TOS application was only chosen as an example. The conclusions of this section would apply equally well to the NORAD, SAC, or National Command Authority systems.

(company, battalion, etc.) symbol on the display to blink for some specified period of time. Also, assume that the commander has on his desk a CRT for displaying detailed information on a particular unit. With the SUS he could control the information presented to him just by "asking" the system rather than having to ask his staff and wait for them to interact with the system. For example, suppose a unit status message came in on Company Bravo. The large-screen display would be automatically updated and Bravo's symbol would blink. The commander then could simply say, "Display status for Bravo," and he would get the status display at his desktop CRT without having to go through his staff. As another example, assume that he is interested in the estimated strengths of enemy units. In this case he might say, "Blink enemy units where strength is greater than one hundred." The appropriate symbols on the large-screen display would then blink, giving him a better view of the "big picture" than he could get by having his staff point out the locations on the display. The SUS capability appears to not only be quicker and more efficient, but may provide an interface with commanders that they may be more willing to use.

As we have stated, speech may well provide the best man/machine interface for commanders. Interfaces such as teletype or a CRT with a keyboard have in the past not been successful in providing a coupling between high-level decisionmakers and information systems. This has been true both in the military and in the civilian sectors. The reason for this is unclear. Some analysts have speculated that direct interaction with a computer is below a commander's dignity; others say that the interfaces are too complex, rigid, or unnatural, and their use too hard to learn and remember.

Still another view is that a decisionmaker's short-term memory becomes overloaded and he loses his train of thought when he has to type a series of highly formatted commands into the system. This explanation may be especially valid when the decisionmaker is using a large-screen display where he must perform pattern-recognition tasks rather than sequential processing. A speech interface may be the best interface for direct interaction between commanders and information systems--it is natural to use and causes very little interference with other activities.

However, such an SUS application could present many design problems. Commanders are unlikely to conform to rigid vocabularies and syntax, thus placing a considerable "stress" on the SUS. They are also likely to be less tolerant of system errors and incorrect responses. Therefore, if the system is to be effective for a commander, its reliability must be high or he will again relegate its use to his staff, and much of the potential savings may be lost.

NAVY MANAGEMENT INFORMATION SYSTEM

The Navy MIS for the LHA amphibious assault ship [67] is designed to support the following shipboard functions during an amphibious assault:

- o Supporting arms coordination.
- o Force logistics control.
- o Intelligence, data-base maintenance and query.
- o Debarkation control.
- o Helicopter direction.

An amphibious assault is a highly orchestrated operation in which many events are timed down to the minute for a very large force of troops, ships, and airplanes. The MIS is designed to aid in tracking these events and their status throughout the entire operation. In the present system, each of the functions listed above has its own CRT terminal which is used to update and query its files during the actual assault. The application of SUS to the MIS is probably advantageous for all these functions because of the severe time constraints placed on all the operators. The debarkation control function will be discussed in more detail below, however, because its time constraints are probably the worst.

It has been estimated that during an assault the debarkation control console operator could, on the average, have to make an update or a query to his data base as often as every 30 seconds. The purpose of debarkation control is to ensure that each of the landing-craft assault boats leaving the LHA ship are loaded with the proper items (personnel and material) in the proper order and that they depart at the proper time. This offload is supposed to take place according to a plan pre-stored in the MIS data base. During the assault, anomalies develop which must be reflected in the data base. For example, certain items may not get loaded on the proper boat and may have to be scheduled for another boat, or a boat may not leave on time. In other cases, the offload sequence must be changed because of changes in requirements on the beach. These anomalies from the original plan call for numerous updates and queries which must take place very rapidly if the debarkation control officer is to make timely and accurate decisions.

An SUS coupled to the MIS would certainly enhance its operation because of the speed of the speech interface. This assumes, of course, that the SUS can operate in near real time and does not require long processing times. Like the Voice Data Management system described by Newell, et al. [4], the MIS has a highly constrained update and query language. The fixed-function word vocabulary is shown in Table 7. The MIS also uses a fixed set of files that require a fixed vocabulary for access. This would provide an SUS with strong semantic and syntactic support, which should aid in reducing the processing time required.

SUS CHARACTERISTICS

This discussion of SUS characteristics will be aimed at tactical data management applications such as those described for the TOS and the MIS. Since most of the SUS characteristics for these applications are the same as those discussed in Sec. IV, we will present only those topics which differ significantly.

Vocabulary. For the TOS and MIS applications, the vocabularies could vary considerably depending on design, but 1000 to 2000 words would probably be appropriate. Table 7 gives some indication of the type of vocabulary that might be required.

Interaction. As mentioned previously, the systems would have CRT interfaces which could provide for a strong interaction with the SUS.

Response Time. Response times for the MIS may be far more crucial than for the TOS, not so much in terms of the time criticality of the

Table 7

MIS QUERY LANGUAGE FUNCTIONAL VOCABULARY

ABS
AND
ANY
APPEND
BACK
BEGIN
BT (break)
BY
CALL
CAT (category)
CHANGE
CHAR (character)
CLOSE
COLS (columns)
COUNT
CRTT (cathode ray tube)
DECS (decimal)
DELETE
DISPLAY
DO
EACH
END
EQ (equal)
FILE
FINAL
FOOT
FOR
FROM
GE (greater than or equal to)
GOTO
GRID
GROUP
GT (greater than)
HEAD
HEADED
HOUR
HSP (high speed printing)
IF
IN
INPUT
INSERT
INTO
IS
LE (less than or equal to)
LENGTH
LET
LIST
LT (less than)
MAX (maximum)
METERS
MILES
MIN (minimum)
NE (not equal)
NO
NONE
NOT
NUM (number)
OF
OLD
ON
OPEN
OR
OUTPUT
PAGE
PASS
PCH (punch)
PCT
PROC (procedure)
READ
REPORT
REWIND
REWOUND
REWRITE
ROWS
SAVE
SELECT
SET
SKIP
SORT
SPACE
SUB (subtract)
SUM
TAB
THEN
TITLE
TO
TRAIL
UNWIND
UNWOUND
UPDATE
USING
VALUE
WHERE
WITH
WRITE
YARDS

response, but rather in terms of workload on the user. Therefore, the total response time of the SUS and the MIS combined should probably be no more than 5 seconds.

SUMMARY

Decisions to use data management and voice-keypunch SUS for administrative information processing systems will, in most instances, depend on demonstrated cost savings and commercial availability. In this area, source data automation appears to be the most likely application. Applications of SUS in tactical systems appear more promising, as they would provide both improved operations and reduction in uniformed manpower. From the technical point of view, however, they present serious design problems because of the possible high noise levels in operating environments and the need for very high recognition accuracy. Nevertheless, speech may be the most effective interface for commanders and other high-level decisionmakers with their information systems.

VII. ADVANCED APPLICATIONS

Stepping back from practical concerns with segmentation, parsing, signal-to-noise ratios, and all the other necessary parameters of current SUS research, we should like to suggest some of the possibilities that may lie ahead in the 1980s and 1990s--after years of artificial-intelligence research and years of experience with various aspects of speech as a man/computer interface, and after the demonstration and operational use of flexible, continuous speech understanding systems.

TRANSLATION OF SPOKEN NATURAL LANGUAGE

The capability of speech understanding for task accomplishment may be extendable into reinterpretation (paraphrasing) of original spoken messages into other words which convey the same message (i.e., leading to the same task accomplishment). A step beyond such a reinterpretation capability is the translation of messages into another language (natural or artificial). This capability would be of great value at international gatherings, in international organizations, and, in the military, in interacting with the military forces and civilian populations of other countries. Automatic translation of spoken utterances would also be valuable for cooperative joint space exploration with astronauts of other countries (e.g., the planned joint U.S./ Soviet space station).

Of course, much work has been and is being done in language translation by computer and, contrary to some pessimistic assessments, progress is being made in the United States and other countries [68]. For example, a research group at Kyoto University in Japan has been trying to coordinate a project in mechanical language translation with work in speech recognition and synthesis [69,70]. They have built a computer-based system for translation from Japanese to English and vice versa which is now being used with an 8000-word dictionary, 400 idioms, and 900 syntactic rules. The major problems now are with semantic requirements. One of the experimental applications of this system is in the translation of sentences about elementary geometry.

Here the terms have definite meanings, and syntactic ambiguities can be resolved by conferring their semantics. In the semantic table, terms are connected with each other into logical structures. Simple sentences are easily handled, but compound sentences are still a problem.

However, further research will eventually resolve the syntactic/semantic problems and permit real-time language translation. At first, the subject areas of conversation have to be constrained to specified world models. The reinterpretation capability within a given language can be used to preprocess the input utterances into stricter syntactical forms prior to application of translation rules, and it may also be applied to resolve translation ambiguities. Ultimately, as the ability to extract and interpret prosodic information grows, a wealth of information about the utterance and the speaker's feelings about it could also be conveyed by appropriate synonyms or with additional signals.

SPEECH-OPERATED WRITING MACHINES

Given the capability of reinterpretation and paraphrasing of input utterances, further research in acoustic and semantic processing can be expected to lead to the ability to produce an accurate phonetic representation of the input utterance--a representation that could be used to drive a speech synthesizer to precisely repeat the input utterance. Further semantic processing and context information could now permit producing the corresponding orthographic representation, the written form of the input utterance.

A speech-operated writing machine--a speech typewriter--is an age-old dream of inventors and speech researchers [71]. No one questions the value of such a device if it could be made available at a reasonable cost. For example, the required processing could be made available on a time-shared basis from a central processor and both on-line real-time and "batch-process" service could be provided. The latter would process dictation from previously recorded tapes, very much like the programming application discussed in Sec. V. To simplify the processing, a choice of vocabularies could be offered (e.g., for business letters or for reports in various subject areas), essentially grammatically correct language could be insisted on, and considerable use could be made of user models.

The hope for such a capability is illustrated in the report of the recent Air Force Mission Analysis on Base Communications - 1985 [63], where a scenario of the use of NEWCOMM, the future base communication system, suggests:

> When he arrives at his duty station in the Base Aircraft Maintenance Office, Captain Case is surprised to learn that there is no secretary assigned directly to his office.... However, he soon learns that his communication system and terminal provide all the secretarial support he needs in his job. For example, by merely pressing the "Dictation" button on his terminal and dictating into it, he can edit the text as it is displayed on his terminal and receives a smooth copy of the dictated letter (or report) for his signature as it is printed out by the terminal.

SUS-BASED COMMAND-CONTROL SYSTEMS

As discussed in previous sections, important by-products of the use of speech interfaces are the capabilities of speaker identification and verification, and the potential for monitoring his physical and psychological conditions [72]. These capabilities have important implications in command-control systems applications. The former can be used to provide for effective security controls, since certain speech characteristics are nearly impossible to mimic; and the latter offer potential for periodic tests of the operator's fatigue and vigilance levels. For example, operators can be asked periodically to repeat a sentence which then can be analyzed to detect significant changes. Likewise, it may be possible to detect emotional conditions from the speech characteristics which could reveal, for example, whether or not a person is capable of continuing his tasks.

Since the speech signals must be digitized for SUS processing, this could be done at the users' input devices, and transformation techniques could be applied to the digitized speech to provide communications security. These potential capabilities--physical and psychological monitoring, and communications security--have important operational implications in military command-control systems. Indeed, for these benefits alone it may be desirable to implement all-SUS computer interfaces in these systems. However, with the addition of a speech-operated, computerized dictation system and a language translation capability, the attractiveness of all-SUS command-control communications becomes overwhelming.

BIOMEDICAL MONITORING

Everyone is familiar with how the voice is affected by nasal congestion, sore throat, fatigue, and a host of other physical conditions. A diagnostic physician can use the voice as an aid in confirming the existence of certain pathologies (e.g., laryngal cancer). Neurologists are well aware of how certain central nervous system pathologies and dysfunctions may affect the speech perception, processing, production, and generation mechanisms. Moreover, the maturation cycle of an individual is reflected in his speech, both at a neurological level and at a physiological level. Finally, the evolution of species is reflected in their audible communication patterns as well as in the physiological development of their audio apparatus.

All of this suggests that the speech production process as well as the speech understanding process can play an important role in some or all of the following areas:

- As an identification device for individuals (discussed above).
- As a diagnostic aid in confirming or obtaining early warning of certain pathologies--certainly in the speech production system, but possibly in the respiratory system--and of certain diseases.
- As an aid to diagnosing and evaluating treatment of a number of central nervous system conditions.
- As a means for evaluating and enhancing the learning-cycle process in individuals in terms of both muscle control and higher levels of learning.
- As a means of studying the aging process and possibly arresting the decline of certain neurological processes.

- As a means of studying the perception, memory, and vocal expression processes, to gain better understanding of these and to enable utilization of some of the human processing techniques for the development of useful digital algorithms.
- As a tool in the study of evolution and of cultural anthropology.

Speech understanding research would help enhance some of the above capabilities by furnishing clues to human perception and analysis processes. The development of other capabilities would need close cooperation between those involved in acoustic signal processing and the clinical community.

COMPUTERIZED "STAFF OFFICERS"

One attractive property of speech as a man/computer communication medium is that it can be used for simultaneous communication with both computers and humans. Capitalizing on this, and on the expected future developments in natural language processing, speech understanding systems, decision theory, and other related research areas in artificial intelligence, we can hypothesize for the 1990s the following intriguing applications, which can be called "computer-assisted responsibility."

A computer system, most likely in a secure central facility, monitors the verbal deliberations of a decisionmaker, his councils, and his staff (e.g., the President, military commanders, international negotiators, legislators, etc.), who are connected to the system via secure communication systems. In real time, the system performs one or more of the following tasks:

- It constructs a model of the planned action and analyzes it for logical consistency, practical aspects, conflicts with other plans or actions, potential reactions by those affected, and the like. For this the computer contains a vast, efficiently organized data base. It outputs its findings and warnings as the deliberations proceed.

- It monitors the planning, deliberation, or negotiation process itself for logic, facts, and, if desired, whatever attitudes or intentions of the participants can be deduced by linguistic analysis of their statements and acoustic analysis of their utterances.
- It offers facts associated with the deliberations and raises relevant points that are apparently being overlooked.
- It responds to specific questions posed.

There appears to be no decisionmaking situation which would not benefit from such a nearly-omniscient "staff officer." The U.S. delegates in complex international negotiations, such as SALT, could have definite advantages if such a system were at their disposal (how to get the other participants to agree to its use is a different question). Military commanders and planners at all levels could make better decisions with such assistance.

For example, consider the following decisionmaking situation in the military: A division commander is planning an attack to capture an enemy stronghold and is discussing his attack plan with his staff. A CSO (Computerized Staff Officer) terminal monitors the planning. The terminal is connected to the central facility over secure satellite communication links, as is the commander's local tactical command-control system data base. The latter contains intelligence, logistics, geographical, planning, and other information relevant to the commander's battle sector. The commander formulates an attack plan for the next day requiring an armored column containing elements of Battalions A and B to proceed from point X to point Y, crossing a bridge on river Q. The CSO terminal beeps and flashes a message: "Last night bridge Q was shelled and damaged. Heavy tanks of Battalion A cannot cross safely. Engineers estimate 3 days for repair. Bridge still in the range of enemy guns." The commander modifies the battle plan.

Besides the obvious technical questions, and the present deficiencies, there are other, mainly political questions which may adversely affect the design (or even research) of the CSO systems, especially for applications in support of national or international decisions. The fear of control by computers, the potential power yielded by the designers and programmers of such systems, the incompleteness of designs, and the security questions are but a few of the concerns that are likely to arise. Further analysis of these, however, is beyond the scope of this study.

Some of the potential capabilities of CSO systems can be obtained through the use of standard management information systems using conventional input terminals. Indeed, every such system has some benefit to the decisionmaker as its *raison d'être*. However, really significant support to the decisionmakers requires the use of speech interfaces and on-line monitoring by the CSO computers.

VIII. CONCLUDING REMARKS

We have identified a number of SUS applications in military man/computer systems which, on the basis of preliminary analyses, appear to provide operational benefits. Several of these applications are already being investigated in military research laboratories (e.g., applications in avionics control and source data automation), although, to our knowledge, none are in operational use. Other, simpler applications (e.g., in sorting tasks) are being developed and tested in industry. Various aspects of SUS application in general data management systems (such as the Voice-DM, Voice-KP, and the like) are being used in ARPA-sponsored speech understanding research projects as research vehicles.

It is clear that even though each of the identified military SUS applications promises some element of operational advantage (enhanced use of the interface, increased user mobility, expedited tasks, reduction of operator's workload, increased safety, or potential for manpower reduction), not all are equally important. Hence, it would be useful to rank order the proposed applications on the basis of some preference measure. While cost-effectiveness is a natural choice for such a measure, the lack of adequate SUS cost data precludes its use. Instead, we shall use two qualitative factors to establish rough priorities:

1. Potential operational payoff of the SUS application.
2. Technical feasibility of implementing a continuous speech SUS for this application in the 1975-1980 time period; the prime considerations here are linguistic--the size of the vocabulary and the syntactic/semantic freedom that must be provided in order to realize the expected operational benefits.

We have estimated both factors for each of the main application areas discussed in the report. The results of this exercise are presented in Table 8. We emphasize that these findings are subjective and largely intuitive, and they reflect the various biases of those participating in the evaluations. Depending on the objective, two different rankings can be derived from Table 8:

1. From the point of view of near-term transfer of SUS technology into military systems, the technical-feasibility factor dominates. The military is likely to concentrate on applications which can be implemented with SUS technology that is now becoming available (e.g., isolated-word speech recognition and continuous speech understanding/recognition with highly constrained languages).

Table 8

POTENTIAL PAYOFF AND FEASIBILITY IN 1975-1980
OF SUS APPLICATION AREAS

Application	Potential Payoff	Technical Feasibility
1. Control of robots and teleoperators	Medium	High
2. Avionics control	Superior	High
3. Field data entry in tactical systems	Superior	Medium
4. Field data entry in noncombat systems	High	High
5. System checkout and diagnosis	Superior	High
6. Computer-aided instruction	Medium	Medium
7. Air traffic control	Medium	Medium
8. Man/computer tasks in tactical systems (e.g., TACCO)	Superior	High
9. Computer programming, problem solving	Medium	Medium
10. Administrative data management systems	High	High
11. Tactical data management systems (e.g., TOS)	High	Medium
12. Commander's interface with computer	High	Low
13. Spoken language translation	High	Low
14. Computer-enhanced conferencing	Medium	Medium
15. Speech-operated writing machine	Superior	Low

2. From the long-term point of view, the potential payoff is the overriding consideration. Here, the technical-feasibility factor indicates the need for further research.

The near-term SUS applications can be ranked into three groups as follows:

1. First priority
   - -- Avionics control (superior feasibility, high potential payoff)
   - -- System checkout and diagnosis (superior feasibility, high potential payoff)
   - -- Man/computer tasks in tactical data systems (superior feasibility, high potential payoff)
2. Second priority
   - -- Field data entry in noncombat systems (high feasibility, high potential payoff)
   - -- Administrative data management systems (high feasibility, high potential payoff)
3. Third priority
   - -- Control of robots and teleoperators (medium feasibility, high potential payoff)

The long-term SUS applications can also be ranked into three groups, using potential payoff as the primary parameter. The emphasis is on research effort required to achieve SUS capabilities for implementing high-payoff applications which appear not to be technically feasible in the 1975-1980 time period.

1. First priority
   - -- Field data entry in tactical systems (superior potential payoff, medium feasibility)
   - -- Air traffic control (high potential payoff, medium feasibility)
   - -- Tactical data management systems (high potential payoff, medium feasibility)

2. Second priority
   - -- Speech-operated writing machine (superior potential payoff, low feasibility)
   - -- Commander's interface (high potential payoff, low feasibility)
   - -- Spoken language translation (high potential payoff, low feasibility)
3. Third priority
   - -- Computer-aided instruction (medium potential payoff, medium feasibility)
   - -- Computer programming, problem solving (medium potential payoff, low feasibility)
   - -- Computer-enhanced conferencing (medium potential payoff, low feasibility)

We would like to emphasize several significant observations which were made in Sec. II regarding the transfer of SUS technology to military systems:

- o There is a great deal of cost-consciousness in the military at the present time. The reduction of operational costs, rather than the improvement of operational effectiveness, is the preferred rationale for introducing new systems. An SUS application which could achieve both would certainly find enthusiastic support.
- o In evaluating the cost of an SUS application, it is important to consider the *total system* cost, rather than that of the speech interface alone.
- o Limited versions of most of the identified SUS applications could be implemented with isolated-word speech recognition systems. However, further operational advantages could be achieved by implementing continuous speech understanding interfaces.
- o The long-standing military communications practice of improving communications intelligibility through

specially chosen vocabularies and syntactical constraints will facilitate the implementation of SUS applications. This practice leads to higher recognition/understanding accuracy and higher interface reliability--a definite requirement in most of the potential military applications.

- Several attractive military SUS applications (such as in field data input) require very high recognition accuracy the first time around; there are no opportunities for dialogue.
- A great deal of environmental noise may be present in several applications. Moreover, adverse climatic conditions, acceleration forces, vibration, long-duration missions, fast-response tasks, concerns over physical safety, and the like place unusual stresses on the military systems operators.

But these are only the general considerations in the speech technology transfer process. Before a user agency can justify the funding of operational development of speech interfaces for its systems, in particular when the replacement of a more conventional interface is involved, it must be able to specify the required speech interface and the associated subsystems; perform the (inevitable) cost-benefit analyses; consider the effects of relevant environmental constraints (such as limits on physical characteristics of the interface equipment, environmental noise, and operating conditions); postulate interaction protocols; analyze the reliability, availability, and maintainability aspects of the speech interface; and investigate techniques for alleviating critical environmental constraints.

While the military R&D agencies are quite capable of performing system analyses for conventional man/computer interfaces, the speech interface is sufficiently novel and controversial to require the development of a special technology transfer and applications analysis methodology:

1. Suitable *models* must be made of SUS performance, reliability, and error modes, and implementation in complex systems (time-shared, multiuser, real-time operation; computer security requirements; complex architectures; federated or networked systems).
2. Techniques must be developed for assessing performance, benefits, and costs. Tradeoff functions are needed for data processing capabilities, task requirements, complexity of world and user models, acoustic processing, semantic processing, and the like. Techniques must be developed for identifying and underscoring the uncertainties involved.
3. Techniques are needed for analyzing the human and environmental aspects of the speech interface with computers, for cataloging of the available techniques for "environment enhancement," and for analyzing their cost effectiveness. Techniques are also needed for handling uncertainties in environmental conditions and in human performance.
4. Suitable technological forecasting methods must be selected for projection of computer technology, developments in alternative interface implementations, and computing costs for future time periods. These can be adapted from general forecasting methodology to special cases involving projected introduction of SUS.
5. Performance and cost data from current experimental speech understanding research projects must be collected into a data base for use with the models being developed. The uncertainties in these data, and the consequent uncertainties in analyses using them, must be carefully identified and underscored.
6. The effects of operating protocols on performance and cost-effectiveness must be analyzed for a few selected applications.

Speech as a man/machine communication medium can offer benefits not provided by other, conventional means for human interaction with computers. Research in SUS technology is rapidly approaching the prototype development phase. The stage is set for transfer of the results of this research into operational systems. With this report, we have attempted to contribute to the first steps of the SUS technology transfer by identifying attractive application areas in military systems and making specific suggestions for further development of an SUS technology transfer methodology.

REFERENCES

1. Turn, R., *The Use of Speech for Man/Computer Communication*, The Rand Corporation, R-1386-ARPA, November 1973.

2. Hoffman, A. S., *The Role of Acoustic Processing in Speech Understanding Systems*, The Rand Corporation, R-1356-ARPA, October 1973.

3. Klinger, A., *Natural Language, Linguistic Processing, and Speech Understanding: Recent Research and Future Goals*, The Rand Corporation, R-1377-ARPA, December 1973.

4. Newell, A., et al., *Speech Understanding Systems*, Carnegie-Mellon University, Pittsburgh, Pennsylvania, May 1971.

5. Statement by the Secretary of Defense, Elliot L. Richardson, before the Committee on Armed Services, United States Senate, 93d Cong., 1st sess., April 1973.

6. Statement by the Director of Defense Research and Engineering, John S. Foster, Jr., before the Committee on Armed Services, United States Senate, 93d Cong., 1st sess., April 1973.

7. *Proceedings of the 1973 AADC Symposium, Orlando, Florida*, Naval Air Systems Command, Washington, D.C.

8. Edge, R. L., "Future Tactical Communications Requirements for the USAF," *Signal*, September 1972, pp. 13-16.

9. Chapin, G. G., "What Is Different About Tactical Military Operational Programs," *AFIPS Conference Proceedings, Vol. 42, 1973 National Computer Conference*, AFIPS Press, Montvale, New Jersey, 1972, pp. 787-795.

10. Crawford, A. B., "Army Tactical Data Systems," *Signal*, September 1972, pp. 17-19.

11. "The Electronic Air Force," *Air Force Magazine*, July 1971, pp. 32-55.

12. "The Electronic Air Force," *Air Force Magazine*, July 1973, pp. 38-62.

13. Whisenand, P. M., and T. T. Tamaru, *Automated Police Information Systems*, John Wiley & Sons, Inc., New York, 1970.

14. Klass, P. J., "USAF Pushes Digital Avionics for Aircraft," *Aviation Week and Space Technology*, June 11, 1973, pp. 60-63.

15. *Electronics,* November 6, 1972, p. 55.

16. Webster, J. C., and C. R. Allen, *Speech Intelligibility in Naval Aircraft Radios,* Naval Electronics Laboratory Center, San Diego, California, 2 August 1972.

17. Hitchcock, M., A. Krivitzky, and H. Connor, *Speech Bandwidth-Compression Study,* Final Report, Avala Air Systems Command Contract No. N0019-69-C-0184, Scope Electronics, Inc., Reston, Virginia, 31 March 1970.

18. Hill, D. R., "An Abbreviated Guide to Planning for Speech Interaction with Machines: the State of the Art," *International Journal for Man-Machine Studies,* Vol. 4, 1972, pp. 383-410.

19. Bobrow, D. G., A. K. Hartley, and D. H. Klatt, *A Limited Speech Recognition System II,* Report No. 1819, Bolt, Beranek and Newman, Inc., Cambridge, Massachusetts, 1969.

20. Vicens, P., *Aspects of Speech Recognition by Computer,* Technical Report CS127, Stanford University, Stanford, California, 1969.

21. Hill, D. R., and E. B. Wacker, "ESOTerIC II--An Approach to Practical Voice Control: Progress Report," *Machine Intelligence,* Vol. 5, Edinburgh University Press, Edinburgh, Scotland, 1969, pp. 463-493.

22. Dixon, N. R., and C. C. Tappert, "A Multi-Stage, Sequential Strategy for Automatic Recognition of Continuous Speech," *Workshop on Automatic Pattern Recognition Problems in Speech,* Rome Air Development Center, New York, 1971.

23. Martin, T. B., and E. F. Gunza, "Recognition of Connected Spoken Digits for a Large Number of Talkers," *Workshop on Automatic Pattern Recognition Problems in Speech,* Rome Air Development Center, New York, 1971.

24. Herscher, M. B., and R. B. Cox, "An Adaptive Isolated-Word Speech Recognition System," *Proceedings, 1972 Conference on Speech Communication and Processing,* Air Force Cambridge Research Laboratories, Bedford, Massachusetts, 1972, pp. 89-92.

25. Meddress, M., "A Procedure for Machine Recognition of Speech," *Proceedings, 1972 Conference on Speech Communication and Processing,* Air Force Cambridge Research Laboratories, Bedford, Massachusetts, 1972, pp. 113-116.

26. Reddy, R., L. Erman, R. Neely, et al., *Working Papers in Speech Recognition, I,* Department of Computer Sciences, Carnegie-Mellon University, Pittsburgh, Pennsylvania, April 21, 1972.

27. McCarthy, J., L. D. Earnest, D. R. Reddy, and P. J. Vicens, "A Computer and Ears," *AFIPS Conference Proceedings*, Vol. 33, Part 1, 1968 Fall Joint Computer Conference, AFIPS Press, Montvale, New Jersey, pp. 329-338.

28. Reddy, D. R., L. D. Erman, and R. B. Neely, "A Model and a System for Machine Recognition of Speech," *IEEE Transactions on Audio and Electroacoustics*, June 1973, pp. 229-238.

29. "Spoken Words Drive A Computer," *Business Week*, December 2, 1972.

30. "Look! This Conveyor System Obeys a Spoken Command," *Modern Materials Handling*, November 1971.

31. Feidelman, L. A., "A New Voice in Data Entry," *Modern Data*, April 1973.

32. Deutsch, S., and E. Heer, "Manipulator Systems Extend Man's Capabilities in Space," *Astronautics and Aeronautics*, June 1972, pp. 30-41.

33. Greene, T. E., "Remotely Manned Systems--An Overview," *Astronautics and Aeronautics*, April 1972, p. 44.

34. Rosenblatt, A., "Robots Handling More Jobs on Industrial Assembly Line," *Electronics*, July 19, 1973, pp. 93-104.

35. Interian, A., and D. Kugath, "Remote Manipulators in Space," *Astronautics and Aeronautics*, May 1969, p. 40.

36. Miller, B., "RPVs Provide U.S. New Weapon Options," *Aviation Week and Space Technology*, January 22, 1973, pp. 38-43.

37. Stein, K. J., "Man-Machine Interface Poses Problems," *Aviation Week and Space Technology*, January 22, 1973, pp. 62-66.

38. Moore, J. W., "Toward Remotely Controlled Planetary Rovers," *Astronautics and Aeronautics*, June 1972, pp. 42-48.

39. List, B., "DAIS: A Major Crossroad in the Development of Avionic Systems," *Astronautics and Aeronautics*, January 1973, pp. 55-61.

40. Kleiner, N., and K. H. Miller, *Voice Activated Cockpit Control*, Technical Report AFAL-TR-72-290, Air Force Avionics Laboratory, Wright-Patterson Air Force Base, Ohio, October 1972 (by Scope Electronics, Inc.).

41. Glenn, J. W., et al., *Voice Initiated Cockpit Control and Interrogation (VICCI) System Test for Environmental Factors*, Final Report, Naval Air Development Center Contract No. N00019-70-C-042, Scope Electronics, Inc., Reston, Virginia, 30 April 1971.

42. Scott, P. B., and J. R. Richards, *Speech Controlled Radio Channel Selector*, Technical Report AFAL-TR-71-266, Air Force Avionics Laboratory, Wright-Patterson Air Force Base, Ohio, October 1971 (by RCA).

43. *Simulation Voice Initiated Cockpit Control and Interrogation*, Scope Electronics, Inc., Reston, Virginia, 31 January 1970.

44. Miller, G. E., "TACFIRE, An Innovation in Artillery," *Armor*, Vol. LXXXI, No. 4, July-August 1972, pp. 9-13.

45. *TACFIRE - Operational Functional System Description*, Litton Industries, Van Nuys, California, March 1970.

46. *A Technical Description of the DMES 1000*, Litton Industries, Van Nuys, California, May 1969.

47. *The Story of VAST*, PDR Electronics Division, Harris-Intertype Corporation, Cleveland, Ohio, 1971.

48. Small, D. L., "Natural Processing in Computer Systems," *Naval Engineers Journal*, American Society of Naval Engineers, Washington, D.C., April 1973.

49. Constant, M. L., and P. L. Seely, "Computer-Mediated Human Communications in an Air Traffic Control Environment: A Preliminary Design," *Computer Communication Impacts and Implications*, Proceedings of First International Conference on Computer Communication, Washington, D.C., October 24-26, 1972.

50. *Problems Confronting the Federal Aviation Administration in the Development of Air Traffic Systems for the 1970s*, Twenty-Ninth Report by the Committee on Government Operations, U.S. House of Representatives, Washington, D.C., 1970.

51. Schwartz, E. S., *Studies in Air Traffic Control Language*, Technical Documentary Report AL-TDR-64-5, Air Force Avionics Laboratory, Wright-Patterson Air Force Base, Ohio, 1 February 1964.

52. Schwartz, E. S., "A Dictionary for Minimum Redundancy Encoding," *Journal of the ACM*, October 1963, pp. 413-439.

53. Laveson, J. I., *Simulation of Man-Machine Spoken Language Communication for Use in Air Traffic Control*, Ph.D. Thesis, Drexel University.

54. *NATOPS Crew Operators Manual (TACCO), Navy Model P-3C Aircraft*, NAVAIR 01-75PAC-1.1, Naval Air Systems Command, Washington, D.C., 15 February 1971.

55. *Naval Flight Officer Function Analysis, Volume III: P-3C Tactical Coordinator (TACCO)*, Naval Aerospace Medical Research Laboratory, Pensacola, Florida, December 1972.

56. *The Human Operator Simulator, Volume I: Introduction and Overview*, Technical Report 1046-F, Analytics, Inc., Jenkintown, Pennsylvania, 29 June 1973.

57. *The Human Operator Simulator, Volume II: Human Operator Procedures Language*, Technical Report 1046-H, Analytics, Inc., Jenkintown, Pennsylvania, 29 June 1973.

58. "Cut Coding Time 42% Try Dictating," *Computerworld*, August 15, 1973.

59. Berkeley, E. C., A. Langer, and C. Otten, "Computer Programming Using Natural Language," *Computers and Automation*, June, July, and August 1973.

60. Walker, D. E., "Automated Language Processing," in *Annual Review of Information Science and Technology*, Vol. 8, American Society for Information Sciences, Washington, D.C., 1973.

61. Woods, W. A., R. M. Kaplan, and B. Nash-Webber, *The Lunar Sciences Natural Language Information System*, BBN Report No. 2378, Bolt, Beranek and Newman, Inc., Cambridge, Massachusetts, 15 June 1972.

62. Winograd, T., *Understanding Natural Language*, Academic Press, New York, 1972.

63. Diller, T. C., *Thematic Patterning in the SDC Vocal Data Management Dialogues*, System Development Corporation, NIC #17663, SUR #94, July 1973.

64. Barnett, J., "A Voice Data Management System," *IEEE Translation on Audio and Electronics*, Vol. AU-21, No. 3, June 1973.

65. U.S. Air Force Electronic Systems Division, AFSC, *Base Communications Mission Analysis - 1985*, Vol. 1, U.S. Air Force, Hanscom Field, Massachusetts, April 1973.

66. U.S. Army Computer Systems Command, *Introduction and Staff Management Procedures for Developmental Tactical Operations Systems*, U.S. Army, Fort Hood, Texas, March 1972.

67. Pyles, R. A., *Computer Program Specification for the Management Information System for the LHA-1 Class Ship*, Litton Industries, 107954-900, Van Nuys, California.

68. Josselson, H. H., "Automatic Translation of Languages Since 1960: A Linguist's View," *Advances in Computers*, Vol. II, Academic Press, New York, 1971, pp. 2-54.

69. Sakai, T., and S. Sugita, "Mechanical Translation of English into Japanese," *Electronics and Communications in Japan*, Vol. 49, 1966.

70. Sakai, T., S. Sugita and A. Watanabe, "Mechanical Translation from Japanese into English," *J. Information Processing Society of Japan*, Vol. 10, November 1969, p. 418.

71. Olson, H. F., and H. Belar, "Phonetic Typewriter," *J. Acoust. Soc. Amer.*, Vol. 28. 1956, pp. 1072-1081.

72. Williams, C. E., and K. N. Stevens, "Emotions and Speech: Some Acoustical Correlates," *J. Acoust. Soc. Amer.*, Vol. 52(A), 1972, pp. 1239-1252.