

2/19/74

FEBRUARY 1974

MDC G5312

(NASA-CE-120163) CREW/COMPUTER
 COMMUNICATIONS STUDY. VOLUME 1: FINAL
 REPORT (McDonnell-Douglas Astronautics
 Co.) 140 p HC
 CSCL 09B
 574-188-3
 63/08
 Unclass
 16798

CREW/COMPUTER COMMUNICATIONS STUDY

VOLUME I Final Report

MCDONNELL DOUGLAS ASTRONAUTICS COMPANY

MCDONNELL DOUGLAS



Reproduced by
**NATIONAL TECHNICAL
 INFORMATION SERVICE**
 US Department of Commerce
 Springfield, VA. 22151

PHOTO COPY AVAILABLE

140



**CREW/COMPUTER
COMMUNICATIONS STUDY**

**VOLUME I
Final Report**

FEBRUARY 1974

MDC G5312

J. D. JOHANNES
PRINCIPAL INVESTIGATOR

R. R. JOSLYN
PROJECT MANAGER

MCDONNELL DOUGLAS ASTRONAUTICS COMPANY-WEST

5301 Bolsa Avenue, Huntington Beach, CA 92647

PREFACE

This report documents results of the Crew/Computer Communications Study prepared for NASA-Marshall Space Flight Center by McDonnell Douglas Astronautics Company in compliance with the requirements of Contract No. NAS8-25701. The report covers work done between 1 February 1971 and 1 February 1974.

The report is presented in two volumes: Volume I is the final report of the study; Volume II contains the hardware documentation, software documentation, and user's manual. Volume II is considered an appendix to Volume I.

If additional information is required, please contact any of the following McDonnell Douglas or NASA representatives:

- Mr. R. R. Joslyn, Project Manager
McDonnell Douglas Astronautics Company
Huntsville, Alabama
Telephone: (205) 881-8640 or 881-0611
- Mr. H. E. Pitcher, Program Development Director
McDonnell Douglas Astronautics Company
Southeastern Region
Huntsville, Alabama
Telephone: (205) 881-0611
- Mr. W. E. Schweickert, Contract Negotiator/
Administrator
McDonnell Douglas Astronautics Company
Huntington Beach, California
Telephone: (714) 896-4821 or 896-2794

PRECEDING PAGE BLANK NOT FILMED

- Mr. B. C. Hodges, Project COR, S&E-COMP-C
NASA-Marshall Space Flight Center, Alabama
Telephone: (205) 453-1385
- Mr. W. E. Parsons, Project COR, D. E.
NASA-Kennedy Space Center, Florida
Telephone: (305) 867-5632

ACKNOWLEDGMENTS

This study benefited from the excellent guidance and communications provided by Mr. B. C. Hodges of NASA-Marshall Space Flight Center and Mr. W. E. Parsons of NASA-Kennedy Space Center.

The effort described in Part I, Phase C, was performed by:

D. W. Giedt - Systems Programming and Evaluation
J. T. Hine - System Software Development
G. Montoya - Principal Investigator
R. L. Noulin - Application Software Development.

The following individuals from McDonnell Douglas Astronautics Company - Houston Operations provided the flight mechanics and procedural information needed to generate the rendezvous demonstration of the Apollo lunar module and command service module:

R. V. Glowczwski
M. J. McRae
L. M. Poole
D. C. Wammack.

The effort described in Part II, Phases A and B, was performed by:

T. J. Edwards - Software Development
C. C. Kesler - Speech Recognition Software
R. G. Runge - Principal Investigator
M. Uemura - Word Recognition Hardware

The effort described in Part III, Phase A, was performed by:

- J. T. Hine - System Software Development
- J. D. Johannes - Principal Investigator
- R. G. Koppang - PKD Support Software Development
- R. W. Parker - Experiment Scenario Development
- R. J. Perkins - Application Software Development

CONTENTS

Section 1	INTRODUCTION AND SUMMARY	1
	1.1 Background	1
	1.2 Study Objectives	2
	1.3 Study Organization	2
	1.4 Status of Study	3
	1.5 Study Results	4
Section 2	CREW/COMPUTER COMMUNICATIONS STUDY - PART I, PHASE C	7
	2.1 Objectives	7
	2.2 Task 1 - Development and Expansion of Interactive Vocabulary Structure	8
	2.3 Task 2 - Development of Typical Manned Space Mission Simulation	16
	2.4 Task 3 - Development of Crew Inter- face Routines for Crew/Computer Interaction	26
	2.5 Conclusions and Recommendations for Future C/CC Developments	32
Section 3	DESIGN AND PERFORMANCE OF A LARGE VOCABULARY DISCRETE WORD RECOGNITION SYSTEM - PART II, PHASES A AND B	35
	3.1 General Discussion of Discrete Word Recognition System	36
	3.2 Word Recognition System Configuration	43
	3.3 Results	51
	3.4 Speech Analyzer Design and Performance	61
	3.5 Word Recognition System Software	96
	3.6 Recommendations	104
Section 4	EXPERIMENT CONTROL - PART III, PHASE A	107
	4.1 Objectives	107
	4.2 Task 1 - Experiment Selection and Operations Definition	108

4.3	Task 2 - Fabrication of Portable Keyboard and Display	116
4.4	Task 3 - Interactive Software Design Requirements	117
4.5	Task 4 - Software Implementation	117
4.6	Task 5 - Experiment Demonstration Scenario	119
4.7	Recommendations and Conclusions	122
Section 5	RECOMMENDATIONS FOR FUTURE DEVELOPMENT	125
	REFERENCES	127
	BIBLIOGRAPHY	129

FIGURES

1-1	Study Organization	3
1-2	Vocabulary Structure	4
2-1	Tradeoff for Selection of Operation to be Implemented	20
2-2	Interactive Processing Sequence	21
2-3	Rendezvous Maneuvers	25
2-4	Software Organization for C/CC System	27
2-5	Operating System Flow Diagram	28
3-1	Sonogram of Digits (Male Speaker)	38
3-2	Word Recognition System – Simplified Block Diagram	40
3-3	System Performance With Modified Template Production Function	43
3-4	Word Recognition System Configuration	44
3-5	Analyzer to Minicomputer Data Format	47
3-6	Speech Analyzer Accuracy	53
3-7	Probability Distribution Versus Distance	56
3-8	Dependence of Accuracy on Vocabulary Size	58
3-9	T. J. Edwards Adaptive Testing	59
3-10	Probability Distribution	62
3-11	System Performance With Modified Template Production Function	62
3-12	Speech Analyzer – Simplified Block Diagram	64

3-13	K-158 Microphone Frequency Response and Directional Characteristics	65
3-14	Voiced Sound Spectrum	66
3-15	Equalized Frequency Response	67
3-16	Word Boundary Detector Block Diagram	70
3-17	Automatic Level Control Circuitry Block Diagram	71
3-18	ALC Gain	73
3-19	Single-Stage Filter Configuration	75
3-20	Low-Pass Filter	79
3-21	Low-Pass Filter Performance	80
3-22	Low-Pass Filter Difference	82
3-23	Adjacent Low-Pass Filter Difference	84
3-24	A_H Difference Filter	85
3-25	A_{VH} Filter	86
3-26	Analyzer Data for the Word "Four"	93
3-27	Analyzer Data for the Word "Eight"	94
3-28	Analyzer Data for the Word "Five"	95
3-29	Sonogram of Digits (Male Speaker)	98
3-30	Data Storage Format	101
4-1	S192 Surface Coverage	111
4-2	Experiment Definition	112
4-3	Experiment Support	116
4-4	Crew/Computer Communication Experiment Control Study	121
4-5	Experiment Control Scenario	122
4-6	Truth Site Photo F1234.5	123
4-7	Initialization of Discriminant Trainer, Crop Soil (S) versus Remaining Crops (R)	124

TABLES

3-1	Word Recognition System Software Modes Via Keyboard Control	50
3-2	Vocabulary List in 100-Word Speech Recognition System	54

GLOSSARY

A/D	Analog to digital
ALC	Automatic level control circuitry
ASCII	American standard codes for information interchange
A _G	Gross amplitude
A _H	High-frequency amplitude
A _L	Low-frequency amplitude
A _{VH}	Very-high-frequency amplitude
BITE	Built-in test equipment
C/CC	Crew/computer communications
CDH	Constant delta height
CPU	Central processing unit
CRT	Cathode ray tube
CSI	Concentric sequence initiation
CSM	Command service module
D/A	Digital to analog
DEC	Digital Equipment Corporation
DEC-339	CRT display
DR 11-C	Serial interface card
FET	Field effect transistor
FPE	Functional program element
HP/LP	High pass/low pass
IC	Integrated circuit

PRECEDING PAGE BLANK NOT FILMED

LED	Light-emitting diodes
LA30	Printer/keyboard
LM	Lunar module
PDP 9	DEC computer which supports the PKD
PDP 11/40	Speech processor computer
PKD	Programmable keyboard and display
REQ Lines	The interrupts
RC	Resistance capacitance
TPF	Terminal phase finalization
TPI	Terminal phase initiation
TTY	Teletypewriter
7700A	Lear Siegler Inc. display

Section 1
INTRODUCTION AND SUMMARY

This report documents the results of the following phases of the Crew/Computer Communications Study: Part I, Phase C, Part II, Phases A and B, and Part III, Phase A.

1.1 BACKGROUND

The value of man's presence in space exploration was established in the Gemini and Apollo flights during the 1960's and 1970's. The value of onboard computing capability was also established on the Apollo flights where a computer was used to supply the pilot with operational status data and the results of computational analysis needed to make critical decisions.

The Skylab, Space Shuttle, and Space Lab vehicles and those conceived for future space explorations depend on the men onboard and their access to extensive on-line computing capability to ensure the success of a mission.

Vehicles such as the earth-to-orbit Shuttle, the inter-orbital Shuttle, and the Space Lab will require command and control systems with considerable autonomy. To provide this autonomy, onboard systems must be capable of complex maneuvers, especially when the vehicle is in an attitude or position in space that precludes ground tracking and control. The need for independence from earth-based, large computational facilities for ground control functions places the crew in a position where reliance on onboard computers may be the best assurance that the mission objectives will be met.

The dependence of the crew on spacecraft computers makes it highly desirable to simplify the means of communication and the methods of interaction between the crew and the computer. The reaction and decision time on any space mission is critically short, and the computer must

therefore be an integral part of the total spacecraft system, with command authority over all critical subsystems. To make the computer responsive to the judgment and decisions of the man onboard, it must be able to accept and return information to him in a manner which conforms to his normal way of communicating. Thus, the methods used for man-machine interaction must become more anthropomorphic in character.

These considerations identified the need for a study in crew/computer communications. McDonnell Douglas Astronautics Company has been conducting this study for NASA-Marshall Space Flight Center since 1 March 1970.

1.2 STUDY OBJECTIVES

The primary objective of the Crew/Computer Communications (C/CC) Study was to develop techniques, methods, and system requirements for effective communications between man and computer. A secondary objective, established to prove and validate the results of the study, was to develop, implement, and demonstrate an operational C/CC system.

1.3 STUDY ORGANIZATION

Figure 1-1 shows the study flow organization. This flow is based on sequentially identifying requirements, defining the structure of a typical C/CC system, implementing a demonstration system, evaluating results, and formulating requirements for advanced C/CC systems.

The study was initiated by investigating the requirements for interaction between astronauts and their support computers. Hence, it was originally named the Astronaut/Computer Communications Study. However, as the work progressed, it became apparent that the results of the study and the techniques developed would be applicable not only to astronauts in spacecraft, but also to the crew of any craft which depends on the support of onboard or ground-based computers for accomplishing mission objectives. The name was then changed to reflect the broadened application, and the

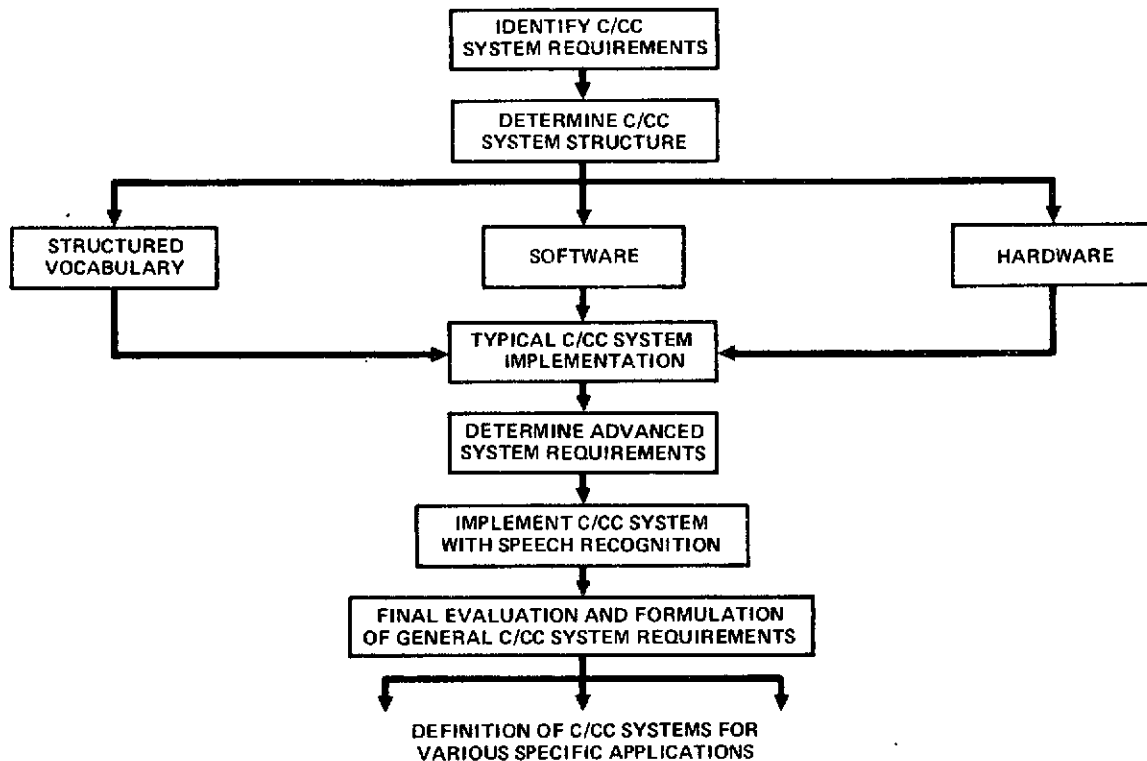


Figure 1-1. Study Organization

scope of the study was extended. The techniques developed for man-machine interaction in the course of this study are also expected to be applicable to the crew manning command and control terminals that may form part of a large computer complex used for checkout and launch of spacecraft.

1.4 STATUS OF STUDY

This report marks the completion of Part I, Phases A and B of Part II, and Phase A of Part III. Section 2 describes the activities and results of Part I, Phase C. This work was completed in October, 1972. Section 3 documents the development in Part II of the word recognition system that was subsequently delivered to NASA-KSC in August, 1973. Part III, the specific applications part of the study, was initiated during the last quarter of 1972. A technical approach for Phase A, of Part III, the C/CC system for experiment applications, was defined and its implementation is described in Section 4 of this study.

1.5 STUDY RESULTS

Building on results of Phases A and B of Part I, a representative C/CC system was implemented. In the course of developing an operational system, several new concepts were established and some of the previous concepts were altered.

Specifically, the following new concepts were developed for crew/computer communications during Phase C:

- A. Communications between man and computer take place in any one of three modes. Figure 1-2 indicates the modes, which are described as follows.

Mode 1 – The interactive and cooperative manner of performing a task by sequentially executing each discrete step of a procedure.

CR19-1

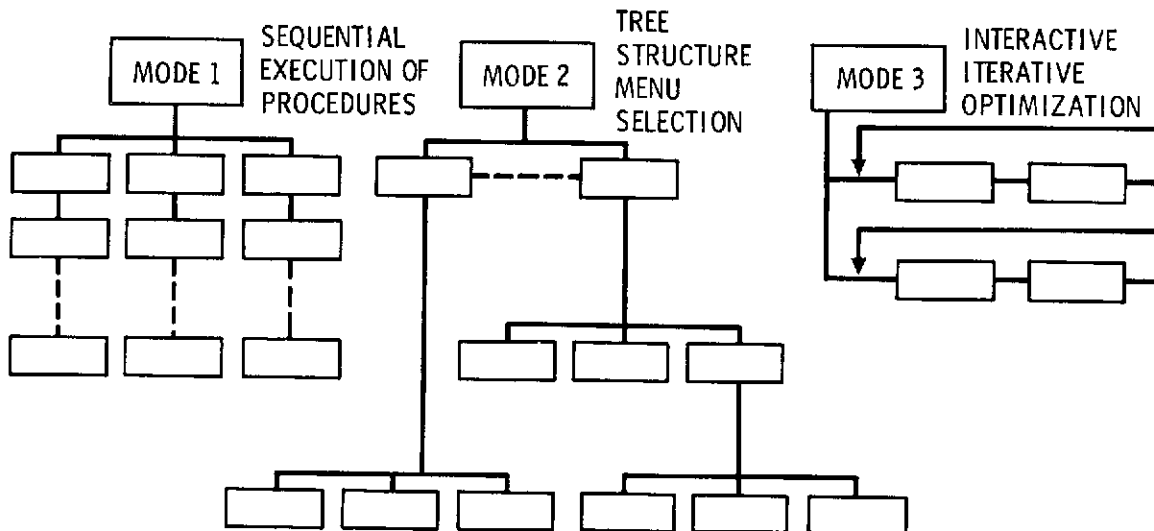


Figure 1-2. Vocabulary Structure

Mode 2 – The interactive progression through a tree-type structure to reach the point where execution of a desired, completely self-contained task may be initiated.

Mode 3 – The interactive optimization of a task or a sequence of tasks requiring the man to furnish various values for a set of parameters. These values are to be used by the computer in accomplishing the optimization. This mode is structured to be an iterative process.

- B. Most of the procedures performed onboard a spacecraft by the crew and the computer will take place in the mission phase; that is, one mission phase at a time, sequentially executed as described in Mode 1. The mission-phase type of operations is generally performed under control of the vehicle commander, and will demand most resources onboard.

Section 2

CREW/COMPUTER COMMUNICATIONS STUDY--PART I, PHASE C

This section describes the research, analysis, and development work performed to accomplish the objectives of Part I, Phase C of the C/CC Study. This phase was conducted in three major tasks. Task 1 covers the development of a vocabulary structure to be used in communications between man and computer. In Task 2, a representative spacecraft operation was selected which depends on crew/computer interaction for its completion. This task also covers the analysis and scenario definition required to develop a demonstration system to be used in evaluation of C/CC techniques derived from this study. Task 3 covers the development of the software programs required to implement the representative crew/computer communication demonstration.

2.1 OBJECTIVES

The primary objective of Phase C was to implement a representative crew/computer communications system to serve as a demonstration vehicle for evaluating the techniques and concepts developed in the course of the study.

A secondary objective, but one that yielded many valuable results, was to improve on previously established techniques and concepts of man and computer interaction. This improvement was achieved in the course of developing a working system. Also, weaknesses in basic concepts and difficulties in actual implementation of a system incorporating these concepts were identified.

Another objective related to Task 3 of this phase was to define and develop a software operating system capable of supporting efficient crew/computer communications.

PRECEDING PAGE BLANK NOT FILMED

2.2 TASK 1 – DEVELOPMENT AND EXPANSION OF INTERACTIVE VOCABULARY STRUCTURE

Task 1 covers the research and analysis required to develop a vocabulary structure for communications between man and computer. The structure was configured in accordance with the following criteria established in Phase B:

- A. Permit communications between man and machine using conversational language instead of numerically coded commands and requests.
- B. Permit on-line interactive cooperation between the crew and computer in accomplishing mission tasks.
- C. Provide an efficient method of implementing all the onboard operations identified by the study.
- D. Allow manned access to any one procedure without subjecting the man to an undesirable long series of selection decisions.
- E. Permit the man to be the ultimate controller of the system by giving him the software controls to exercise procedural options at each realizable decision point. Options given should be such that a man might impose his judgment on the system wherever feasible. This feature will allow him to apply his training, background, and ability in evaluating simultaneously occurring conditions in the system.

The vocabulary structure established during this study meets these criteria.

2.2.1 Vocabulary Structure

Three modes of structuring interactive communications were employed in the demonstrations developed for this phase. The basic modes are as defined in Section 1.5. In implementing the scenario, the modes were used in two different ways. First, a mode was selected as most representative of the overall flow of operator-computer interaction. Next, modes were selected for compatibility with the requirement of subtasks. This technique of combining mode structures was employed throughout the demonstration scenario.

2.2.2 Programmable Command Keyboard Concept

To permit conversational interchange of information between man and computer, a variable nomenclature display and operator input device was needed. This would be a device whose nomenclature and function could be altered depending on the task being performed. If such versatility were to be obtained from programmable displays and switches, the amount of hardware to implement all control functions could be reduced. Thus, McDonnell Douglas developed the concept for a programmable command keyboard.

Several ways of implementing programmable control functions were evaluated. First, individually lighted switches were analyzed; however, they were discarded because the number of words that can be displayed is usually limited by the physical size of the display surface or of the electrical parts required to store the various words. An electromechanically advanced tape, presenting one command or function at a time on the face of the switch, was also evaluated. The drawback in this system was delay in accessing the desired function. Although this system had more capacity than the lighted switches, the capacity was still limited and the function nomenclature difficult to alter.

An overlay technique on a programmable display seemed to offer the best promise of fulfilling the requirements. A photoelectric matrix overlaid on a cathode ray tube (CRT) display was developed at Draper Laboratories of The Massachusetts Institute of Technology. This device was evaluated for integration into the C/CC system. However, the matrix system was also discarded because of the photoelectric, mechanical, and electronic complexity involved. The cost of development of such a system exceeded the funds allocated to complete the study, while the bulkiness and frailty of the CRT and ancillary support hardware would preclude its use in a spaceborne system.

McDonnell Douglas, realizing the need for a compact interactive terminal to support the demonstration of C/CC techniques developed in this study,

established the requirements and initiated design, development, and assembly of an engineering model of the programmable keyboard and display (PKD). This device is further described in Appendix B.

The PKD provides the means for the operator to select any one of 16 programmed options at any juncture in his procedure. The options are displayed, under computer control, in a conversational language. The man selects the desired option by directly touching the face of the display containing the words describing this option. A unique code is sent by the PKD to the computer indicating which area of the display was selected. The computer responds by carrying out the action indicated by the selected option.

Additionally, the PKD has 12 mode-control switches to select program controlling options which are independent of the particular task at hand. These switches allow the man to direct the execution control or executive program. The mode control or special function switches and programmable switches provide for complex, wide-ranging interchanges of information between computer and operator.

2.2.3 Types of Operations

Three types of operations, representing the primary categories of onboard activity, were developed in this study. They are: (1) mission phase, (2) function category, and (3) interactive timeline. These proved adequate for introducing the subsets of crew activity which formed the background for the simulated mission. The goal was to define top-level entry to all primary onboard activities.

Mission Phase

The mission phase is the type of operation executed by the spacecraft commander in accomplishing primary mission tasks. The tasks presented to the crewmen are selected from the mission timeline by the operating system.

For our Apollo lunar module and command service module (LM/CSM) rendezvous demonstration, selections from lunar orbit operations through reentry are presented.

Initially, a multi-path structure is employed to allow a choice between the mission phases presented. When a specific task is reached, the structure becomes sequential. At all decision points in the structure, the crewman has the option of backing up, continuing, or returning to the top level of the selection tree through PKD switch selections.

Selection of mission phase alerts the onboard system to provide all system resources required in performing a mission-oriented task. All priorities on resources are adjusted to meet these requirements.

Function Category

While mission phase operations are most likely to be used by the commander of the spacecraft, function category operations provide for tasks to be performed by the specialist onboard. When executing mission phase operations, the tasks performed may affect every system in the spacecraft. In function category operations, only the equipment, systems, and resources associated with a specific function category are affected.

The nine function categories, as identified in Phase A and described in Appendix A for Phases A and B, are: mission control, data management, communications, flight control, guidance and navigation, experiments, maneuver management, operational status, and mission-independent crew function. Two function category definitions were revised as a result of work performed in Phase C. The timeline analysis, generation, and modification were removed from mission control and made a unique type of operation called "interactive timeline." Additionally, to be consistent with other industry documentation, the name of operational status was changed to checkout.

Interactive Timeline

The third type of operation performed onboard is the interactive timeline mode, as indicated above. This mode permits the crewman to review the mission profile and to vary it if necessary.

The interactive timeline mode presupposes an onboard timeline projection and analysis program for long-duration spaceflights that need some type of timeline analysis to provide automatic scheduling and system resource allocation. It is anticipated that changes will be made in the mission timeline. To change the timeline without interfering with scheduled tasks, the computer will perform conflict analysis on resource demands. If a conflict does exist, the crewman will be notified so that he may alter priorities or introduce other timeline variations. This type of operation will generally be performed in the interactive, iterative optimization mode of communication.

2.2.4 Modes of Performance

To increase the scope of crew/computer communications, two modes of performance were defined: the execution and simulation modes. These added performance modes expand the crewman's overall system control.

In the execution mode, the vehicle system is exercised by the procedure using all necessary system resources. The simulation mode will appear to the operator to be exactly the same as the execution mode; however, while procedures are being performed in the simulation mode, no hardware is affected. Only sufficient resources to effect the simulation are required. This mode provides the training and skill maintenance capability desirable for long-duration orbital or interplanetary mission vehicles.

2.2.5 Analysis Capability

Prior to execution or simulation of a selected task, the task can be analyzed by the pilot or commander of a spacecraft to evaluate the effect that various parameters have on the task. Once a set of optional parameter values is selected, the man can store these values in the data base as the nominal values to be used by the computer in actual task execution.

For example, prior to initiating rendezvous, the crewman may enter various times of ignition for terminal phase initiation and analyze the effect that these various times will have on the actual burn. The criteria for selecting optimal parameters will depend on the status and configuration of the vehicle at different points in the mission.

2.2.6 Special Functions

The special function switches (mode control switches) provide the operator with control of the software supporting an interactive system. Signals from these switches cause the execution control software to respond in a pre-defined manner with the initiated action being common to all application programs, although the actual steps or programs executed are peculiar to the task being performed. This capability covers functions common to interactive crew/computer communications. The special function switches are dedicated controls to allow the crewman to alter the system operating mode at any point in time.

The functions performed by the switches were derived in two ways. Initially, the functions were taken from existing systems such as the McDonnell Douglas S-IVB automatic checkout equipment. Later, during development of the rendezvous scenario and coupled with an examination of proposed systems, the functions were redefined or amended. The current actions defined for the special function switches are as follows:

- A. Monitor – Return to the top level or monitor state of the vocabulary structure and select one of the three types of onboard operations (mission phase, function category, or interactive timeline), as already described.
- B. Backup – Retreat to the previous step, stop point, or decision level in this procedure. In performing analysis or checkout tasks, this special function key permits the man to repeat a particular calculation with different values assigned to the parameters involved, or to repeat the execution of a checkout procedure.
- C. Manual – Execute the STOP procedure if required, then go to the operator's selected task. The manual key has been included to permit the knowledgeable operator to jump from any branch in the structure of the vocabulary to any other branch, whether or not

the other branch is in the same operating mode. Use of this key would generally require possession of a "road map" indicating the points in the structure where a desired task may be initiated.

Each of these points will be given a numerical label on the map and some indication as to whether it is an allowable entry point in the operational procedure. Activation of MANUAL is followed by keying in the numerical code associated with the desired point. Upon activation of ENTER following the numerical entry, the display associated with the desired task appears on the crew terminal display (the PKD).

The manned function switch can be a powerful tool in the performance of operations from a crew/computer communications terminal. Using this tool, the operator may perform many varied and even unrelated tasks in a random fashion without following a preprogrammed sequence of decisions, much as he would if he had a complete set of control panels filled with manually operated switches to control every component in the system. However, this mode of operation does require thorough knowledge of the system to ensure achieving the desired results and prevent development of dangerous conditions. For example, there are points in some procedures that may not be entered without previously setting some required initial condition. By the same token, there are conditions in certain procedures that would not allow leaving that procedure until other safing steps are taken.

- D. Emergency Stop – Execute the current EMERGENCY STOP routine(s) posted for the task being accomplished. The posted routine should place the equipment or resources used in that task in a safe mode or condition. Since the task may exercise or control different parts of the flight system, the routine must ensure that other tasks are not affected.
- E. Stop – Orderly cessation of current task, after which exit can be made to other activities; i. e., stop orbit calculations in the mission

phase branch and go to checkout in the function category branch. The STOP function provides the crewman with a fast, orderly method of shifting operational pursuits. The function is particularly useful in halting automatic sequences at a safe point when it is no longer desirable to complete the sequence.

- F. Proceed – Execute the next step in the current procedure whenever execution has been halted. This key will cause the computer to move forward in the performance of a procedure, especially in the mission phase mode of operation. This key may also be thought of as an execute key in that once initial conditions for a procedure are met or initial values are entered and the procedure is ready to be executed, the PROCEED key will initiate the execution of that procedure.
- G. Return – The nominal flow of a procedure may be interrupted by requesting performance of other tasks such as display of a set of data or execution of a checkout function. The point at which this flow is interrupted is stored along with parameter values pertinent to the system status. This interruption may be initiated by occurrence of contingency conditions or activation of special function keys such as STOP, MANUAL, or DATA. After the desired task has been completed, RETURN must be activated to restore the system to the conditions existing when the interruption occurred.
- H. Data – Display dynamic data relevant to the current task. Data being displayed have been formatted to give the crewman dynamic information he requires which is peculiar to the task being performed. Once the crewman assimilates this information, he may continue execution of the procedure by pressing PROCEED or RETURN. PROCEED causes the active procedure to move forward in execution. RETURN brings the system back to the point of exit, with the same information displayed as before the special DATA request was made.
- I. Clear – Clears the field currently activated for numeric entry from the numeric keyboard.

- J. Enter – Indicates operator acceptance of the content of the numeric field. The active program stores the value for future use and proceeds to the next step.

2.2.7 Numeric Inputs

The Phase B study pinpointed the need for inputting numerics to accomplish tasks such as identifying equipment and assigning parameter values. Phase C validated those findings and further justified the need for numerical inputs for use in the analysis mode.

Numerical inputs are postulated for all three types of onboard operations. Generally, the inputs occur in the lower levels, at the point where operations are being performed on a select set of parameters or complement of equipment.

Two of the special function switches, ENTER and CLEAR, have an important role in numeric entry. These switches are employed to give the crewman or operator additional control at entry time. The CLEAR switch allows the operator to restart an entry which is not satisfactory. The ENTER switch provides the operator with a means of final approval after the entry is completed, giving the operator an opportunity to reverify his input before proceeding.

2.3 TASK 2 – DEVELOPMENT OF TYPICAL MANNED SPACE MISSION SIMULATION

After the architecture of a C/CC system was defined, the system components were brought together in a selected representative spacecraft operation. The resulting system was used as a demonstration vehicle to validate the concepts and techniques for man-computer interaction developed in the course of the study.

2.3.1 Candidate Tasks for Simulation

A variety of simulation tasks was initially considered to demonstrate the operational value of the C/CC system developed by this study, including the unique PKD hardware characteristics and the software required to

support the vocabulary structure developed. Reentry of a space shuttle was considered because of the high workload for both the pilots and guidance computer and the need for the two to communicate clearly and concisely. Rendezvous between a shuttle and space station was also considered as a good candidate for the same reasons, but with the main workload concentrated at each of the thrusting sequences. The rendezvous of the Apollo lunar module and command service module had the additional advantage of being well defined and completely documented. Control and analysis of onboard experiments were under study by the Computation Laboratory utilizing an IBM 7094 that was connected to the simulation computer (PDP-9). The heavy interactive requirements of this mission would thoroughly exercise all the capability of the PKD and structured vocabulary. Checkout of a Delta tug was also considered to be a likely candidate for an interactive system of this type.

2.3.2 Criteria for Selection

With five strong candidates for the demonstration task, a tradeoff study was conducted to make the optimum choice. The following five key elements were determined to be the tradeoff criteria, with each element being weighted according to its importance.

- A. Demonstrated Capability in Field – A straight forward means of implementing the simulation task is vital to avoid distractions such as questioning whether data being displayed is invalid or ambiguous. New computer interactive techniques from an authoritative source are likely to be seriously considered for future manned spacecraft.
- B. Procedures Defined and Documented – A reference procedure is required to establish the merits of a new and innovative technique for solving complex interactive problems. If the reference is well defined and documented, a much larger audience will be receptive to the special attributes of the new approach.
- C. Hardware Requirements – Since a Digital Equipment Corporation PDP-9 would be used as the simulation computer, the development of the total system would be impeded if there were a need for any device not already connected to this computer but required for the simulation.

- D. Software Requirements – It is detrimental to select a simulation task that is too complex to model accurately on a small computer or to oversimplify a complex task and make the results too general.
- E. NASA Interest – Opinions were solicited from various NASA personnel as to which one of the candidate simulations was most relevant to their work. New techniques for performing some of the tasks were also received from some NASA personnel.

2.3.3 Tradeoff and Selection

- A. Grading – Each of the simulation tasks was graded on how it satisfied the five criteria. The criteria grades are defined as follows:
 - 4 = Excellent – Selected simulation task completely satisfies the criteria (i. e. , procedures that are documented in a manner so that they can be used directly).
 - 3 = Very Good – Simulation task satisfies all major elements of the criteria with only slight discrepancies (i. e. , procedures require slight rewriting).
 - 2 = Good – Simulation task satisfies most major elements of the criteria (i. e. , procedure requires rewriting for application, but is nevertheless a good outline).
 - 1 = Fair – Simulation task satisfies only some of the criteria (i. e. , procedure is only vaguely related to the simulation task).
 - 0 = Unacceptable – Simulation task satisfies none of the criteria (i. e. , no procedure exists).
- B. Weighting Factor – Since all the criteria for selection are not equally important, a weighting factor was introduced. The product of the grade and weighting factor yield the score for that task. The range of values for this factor is 1.0 down to 0.0. A score of 1.0 would be given to the criterion that is absolutely essential as defined in performing a task, while the illogical value of 0.0

would be given to the criterion that is irrelevant to solving the problem. It can easily be seen that all rational values will lie between 0.0 and 1.0, while most values will be greater than 0.5 (otherwise, they are probably poor criteria for selection).

- C. Tradeoff – Figure 2-1 shows the results of the evaluation of the proposed candidates. The Apollo LM/CSM rendezvous scored highest and was therefore selected to serve as the demonstration vehicle for evaluation of study results.

2.3.4 Implementation Approach

The initial hardware interface for the study was the Shuttle crew member operating a CRT display system (the DEC-339), and a keyboard. After implementing a preliminary test scenario on the DEC-339 display and associated special function keyboard, it became apparent that the major part of the time the crew member would be either making menu selections for the software modules needed or utilizing only the alphanumeric display capability of the CRT. It was also noted that to interface with the previously defined functional category of experiments, a dedicated control panel would be required for each experiment or a general-purpose multifunctional interface terminal. Having identified the required attributes of the crew/computer communications terminal, and finding no such device in existence, it was decided that one would be fabricated.

After integration of the programmable keyboard and display with the C/CC system, the capability existed for graphical output on the CRT, numeric input, special function key operations, and most importantly, a computer-controlled display capability in the PKD. The system in this configuration possessed all the attributes of the complete crew/computer communications system identified in the Phase A Final Report. The interactive processing sequence is shown in Figure 2-2.

On completion of the rendezvous scenario, it was noted that there was no need for a graphic capability for the portion of the scenario implemented in this demonstration. The CRT was at this point eliminated from the system

	DEMONSTRATED CAPABILITY IN FIELD			PROCEDURES DEFINED AND DOCUMENTED			HARDWARE REQUIREMENTS			SOFTWARE REQUIREMENTS			NASA INTEREST			TOTAL
	G	W	S	G	W	S	G	W	S	G	W	S	G	W	S	
SHUTTLE REENTRY	4	0.8	3.2	3	0.8	2.4	1	0.7	0.7	3	0.8	2.4	2	0.9	1.5	10.5
RENDEZVOUS SHUTTLE AND SPACE STATION	4	0.8	3.2	3	0.8	2.4	3	0.7	2.1	3	0.8	2.4	3	0.9	2.7	12.8
RENDEZVOUS LM/CSM (APOLLO)	4	0.8	3.2	4	0.8	3.2	3	0.7	2.1	4	0.8	3.2	4	0.9	3.6	15.3
EXPERIMENT CONTROL AND ANALYSIS	2	0.8	1.6	2	0.8	1.6	2	0.7	1.4	2	0.8	1.6	4	0.9	3.6	9.8
DELTA TUG CHECKOUT	3	0.8	2.4	3	0.8	2.4	4	0.7	2.8	4	0.8	3.2	2	0.9	1.8	12.6

G - GRADE
W - WEIGHT
S - SCORE

Figure 2-1. Tradeoff for Selection of Operation to be Implemented

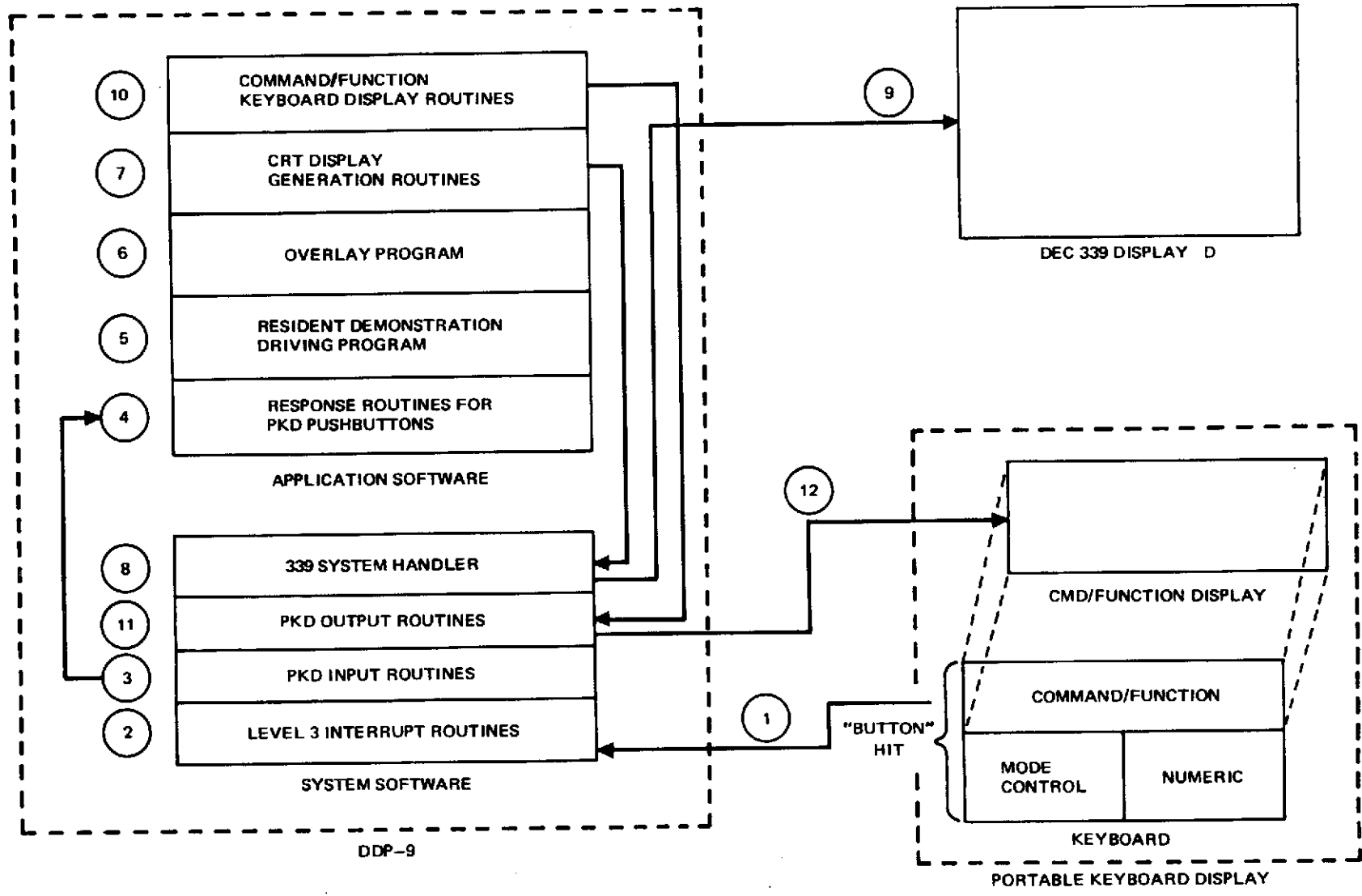


Figure 2-2. Interactive Processing Sequence

and all the menu selections were converted to displayed key selections on the programmable command keyboard. All outputs from software modules were displayed on the programmable command keyboard display in much the same manner as if it were a CRT.

Future applications that require a graphics capability will probably incorporate a CRT in the system, or if resolution requirements allow it, a plasma display. Data analysis for image processing of experiments and trend control displays would require such graphic display devices.

2.3.5 LM/CSM Rendezvous - Scenario

The selected onboard task covers the period from the end of lunar landing operations through docking of the lunar module ascent stage with the command service module (CSM).

It is assumed that the command module pilot of the CSM is the crewman involved in communications with the onboard computer. The first display following selection of the mission phase mode is therefore the set of mission phases on which the pilot may want to work. This set starts with his most recently active phase, lunar orbital operations, and covers activities up to reentry.

The rendezvous technique used requires that the lunar module be the active vehicle and the CSM the passive target. Only as a contingency can the CSM become active, as in the case where failures in the lunar module subsystems were to prevent it from completing the rendezvous. The displays used in the demonstration show position and velocity vector parameters of the lunar module relative to the CSM in a set of curvilinear coordinates centered on the lunar module. This condition may be somewhat misleading because the displays shown are those presented to the pilot; however, it must be remembered that he is acting as a passive monitor of the progress of the maneuvers. In this position, the CSM computer is used only to provide backup calculations and unless a problem arises, will not be used to provide actual guidance and control of the vehicle.

Only the initialization and analysis portion of the LM/CSM rendezvous procedure was implemented. This portion covers the procedure up to completion of setup for initial conditions. No other steps were implemented because the steps covered adequately represent the techniques of communication developed during the study.

To exercise the system's capability in a function category type of operation, a checkout procedure for a Space Shuttle payload was implemented. This particular sequence, operated in Mode 2 as described in Section 1.5, covers checkout operations performed aboard a Space Shuttle which is already in orbit and ready to deploy its payload or on the ground prior to Earth launch. For this portion of the scenario, it is assumed that the action takes place at the payload control station of the Shuttle. Extensive built-in test equipment (BITE) is expected to be used in such complex payloads as an orbit-to-orbit tug. The capability achieved through BITE is utilized to complete the checkout task.

2.3.6 Analysis of Rendezvous Maneuvers

To minimize the fuel consumed during the LM/CSM rendezvous, each of the five thrusting maneuvers must be correlated. Because of the complexity of the orbital mechanics involved, a computer is used to calculate what the burn requirements are for each of the thrusting maneuvers. Although the Apollo can calculate the burn requirements for rendezvous, it is not possible to look at the effects of changing selected parameters. This capability would be a desirable feature on a vehicle such as the Space Shuttle which performs a large variety of orbital operations. Due to the interest expressed by several NASA personnel, it was decided that the orbital analysis capability would be implemented during this study.

The Houston Operations personnel of McDonnell Douglas were contacted to define the necessary equations to implement the above capability. Because the coelliptic sequencing technique is a more complex task to interface with than a direct ascent rendezvous, the coelliptic technique was used in the demonstration. After carefully studying the Apollo and Shuttle orbital

maneuvering requirements, the following assumptions were made for the mathematical model:

- A. Both the active and passive vehicles are in circular orbits.
- B. Concentric sequence initiation (CSI) and constant delta height (CDH) maneuvers are 180 degrees apart.
- C. Out-of-plane distance and velocity are small.
- D. Impulse burns.
- E. Spherical earth and moon.
- F. No burn or navigation errors.

By making these assumptions, it was possible to utilize a modified version of the Clohessy-Wiltshire equations to solve the change in velocity required for each burn. These equations were combined with a group of routines used to calculate the time of next burn, time between burns, and out-of-plane data, thereby yielding a program to calculate all of the necessary maneuvering data.

All five thrusting maneuvers are identified in Figure 2-3. The first of these maneuvers is CSI. It raises the lunar module orbit's apogee to pass through the CDH altitude and eliminates the out-of-plane velocity components. Ninety degrees of orbit travel later, a plane-change maneuver is executed to eliminate the remaining out-of-plane velocity. The third maneuver, CDH, is accomplished by thrusting in the direction of the velocity vector and is used to circularize the active vehicle's orbit at the passive vehicle's orbital radius minus some change in height. When the orbital phasing angle between the active and passive vehicles is the desired value, the terminal phase initiation (TPI) maneuver is executed, causing the active vehicle's orbit to cross the passive vehicle's orbit after 130 degrees of orbital travel. Thrusting is again along the velocity vector. If the maneuvers are performed correctly, the active and passive vehicles will arrive at the orbit crossing point together. At this point, it is necessary to execute the last thrusting maneuver, terminal phase finalization (TPF), which is used to brake the active vehicle and make both orbits coincide.

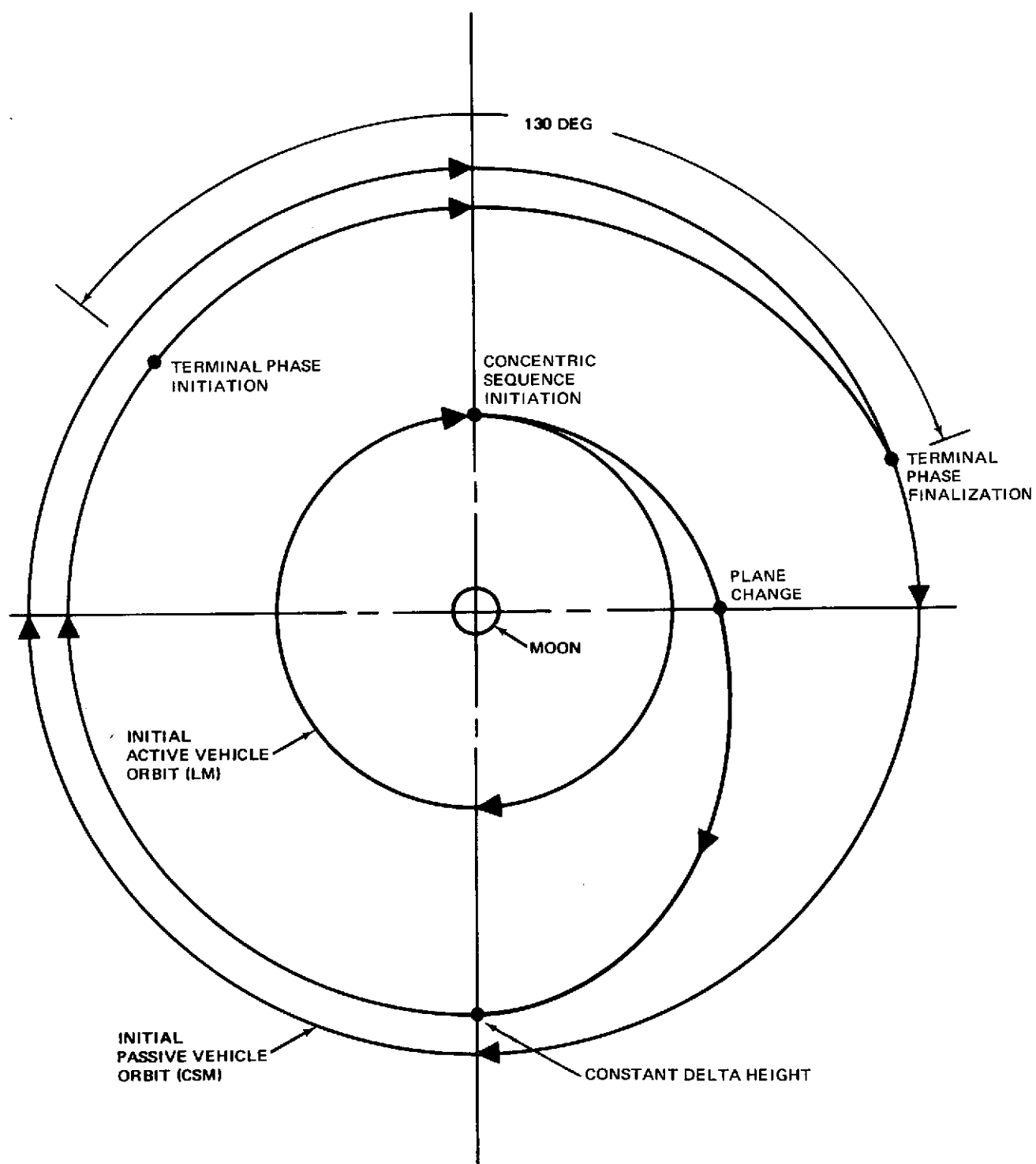


Figure 2-3. Rendezvous Maneuvers

2.4 TASK 3 – DEVELOPMENT OF CREW INTERFACE ROUTINES FOR CREW/COMPUTER INTERACTION

This section contains descriptions of the computer programs developed for interactive operation of the communications methods established by the study. The software is functionally organized as described in Figure 2-4.

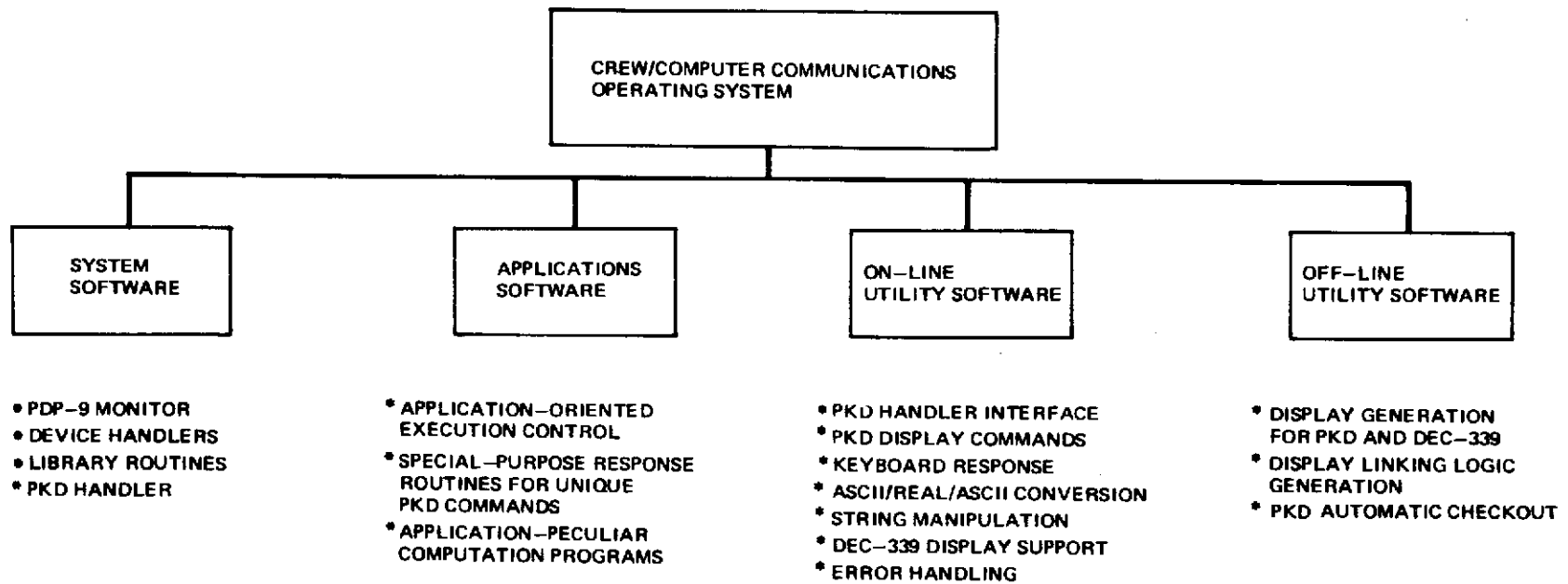
The software was developed on the PDP-9 computer provided by NASA at the Marshall Space Flight Center. The programs are in both FORTRAN IV and assembly language. Detailed program descriptions and a user's guide appear in Appendix A.

2.4.1 Operating System

From the results of this study, especially from the experiences gained in preparing a scenario, it is possible to select an operating system which would be effective for implementing the crew/computer communications methods described in this study. The operating system chosen is a table-driven interactive display executive. This executive is directed by the results of interrogating certain tables describing symbolically what actions are to be taken for a given operator input.

The basic sequence of a table-driven executive, from receipt of operator input, is as follows: (See Figure 2-5)

- A. Branch on type of input.
- B. Test value of input, screening out erroneous or meaningless values.
- C. Perform table lookup using input value.
- D. Reply by performing operations indicated by table, including one or all of the following:
 1. Change modes.
 2. Process parameter input.
 3. Establish display/programmable selections to be presented next.
- E. Present current programmable selections, then go to A.



27

- EXISTING SOFTWARE
- * SOFTWARE GENERATED FOR C/CC SYSTEM

Figure 2-4. Software Organization for C/CC System

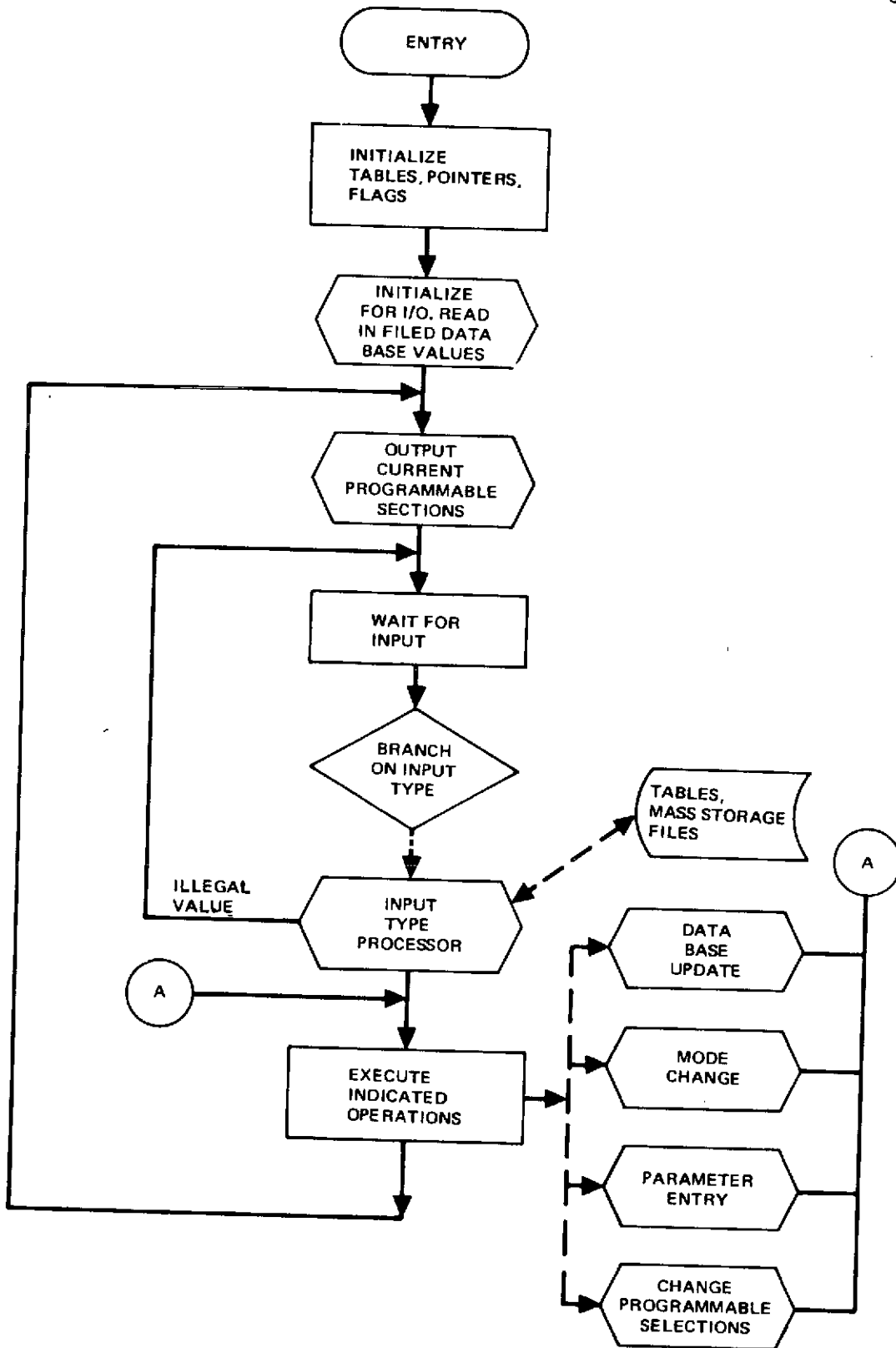


Figure 2-5. Operating System Flow Diagram

The tables required are of three types, one for each distinct input class – programmable command, special function, and parameter entry. The programmable command table permits progression to the next level or branch in the selection tree, which, in turn, may require execution of a subprogram relevant to the last input. The special function table contains pointers to subroutines or in-line code which accomplish the requested function. The parameter entry table points to or contains the parameter description; i. e., range, format, display location, and its location in the data base when applicable.

At the point in the selection tree where a specific application is to be executed, the operating system will pass control to the application control program. Programs which perform tasks unique to the application are thereby engaged within the operating system but do not burden the system.

The technique of table control can also be employed within the application software. In fact, the rendezvous application did employ this technique of control (see Section 2.4.3). This commonality of technique can substantially reduce the cost of developing applications by sharing the off-line table and display generation software used by the operating system.

Two additional advantages of a table-driven operating system are a shortened time for application programming and ease of checkout. Both facilitated the development of the crew/computer communication methods.

2.4.2 System Software

The system software for this study was provided with the Digital Equipment Corporation's PDP-9 computer. The software supplied for developing crew/computer communications consists of an advanced monitor or operating system, input/output processing system overlay generator, and a subroutine library containing common mathematical functions.

The advanced monitor system and the library routines were adequate for the communications application. However, the input/output processing

routines, commonly called handlers, did not cover the PKD. A PKD handler was designed in accordance with ground rules for handler design provided by Digital Equipment Corporation. After development and checkout, the PKD handler was appended to the existing input/output processing package and became a permanent part of the system.

The advanced monitor system is used as is for start-up, off-line program development, and loading of the demonstration programs prior to execution. Once the demonstration is loaded and executing, it retains control until terminated by the computer operator.

Near the completion of the rendezvous, the available core was exceeded. To solve this problem, the PDP-9 overlay capability was employed. This system software allows lengthy programs to be divided into self-contained overlay modules stored in disks. Each overlay contains all routines used in that module except for the I/O handlers, which are always stored.

Each overlay is labeled with a file name and called into core by that name, providing complete freedom of transfer between overlays. The system response time is very small, so small that only a knowledgeable observer can detect the almost imperceptible delay due to loading a new overlay.

2.4.3 Application Software

A table-driven application executive was developed to implement the structured vocabulary. The tables for this executive are generated with the off-line utility routine GENFIL. They contain display file names, keyboard response pointers, and keyboard status. With a table structure, execution control can be readily modified simply by changing the table data. In addition, extensive new capability can be added to the executive by simply adding new elements to the table and including the decision and execution code in the executive. These code additions have proven to be rather small (five or six instructions per new function), yet they yield very impressive improvements in system performance. See Appendix A for a detailed description of the GENFIL routine.

The executive monitors the PKD switches so that when any switch is pressed, the command sent to the computer is checked for errors and, if valid, a branch is executed to the proper support code. If an overlay of support code is required, it would be loaded and executed. The executive would then return to monitoring the PKD switches.

The selected application necessitated the development of several special-purpose software modules. The response routine for the EMERGENCY STOP special function key is an example of such a module. On a spacecraft, this key would be used to initiate execution of a previously named automatic safing routine. These routines are unique for an application and in general will be of little value in specific applications.

All in-line codes required to support specific displays (e. g., rendezvous calculations) are partitioned into core overlays. By placing the general display linking the code on the disk, and locating all the special display codes in overlays, it has proven possible to have more core available for overlays, thereby increasing the likelihood of the next required block of in-line code fitting into the present overlay. Both open and closed routines are located in these overlays.

2.4.4. On-Line Utility Software

A group of routines was developed to perform specific tasks during the execution of the demonstration program. Since these closed routines can be called at any time, they must be stored in the core to achieve an acceptable response time. These general-purpose routines were indispensable in rapidly implementing a new interactive application. A detailed description of these routines is included in Appendix A.

The following functions are performed by this group of routines:

- A. PKD Handler Interface
- B. PKD Display Commands
- C. Keyboard Response
- D. ASCII/REAL/ASCII Conversion

- E. Character String Manipulation
- F. DEC-339 Support
- G. Error Handling.

2.4.5 Off-Line Utility Software

At present, two principal functions are satisfied by off-line routines. The first is generating the display data and its associated linking logic. Since both the menus and label portion of the data displays are predetermined, this information must be generated and stored on a disk before the application program is executed. The second function performed by this group is the automatic checkout of the PKD. Software was developed to troubleshoot problems in the PKD and check out new or modified PKD's.

It should be noted that none of the routines in this group is used during the execution of the application program. A detailed description of these routines is also included in Appendix A.

2.5 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE C/CC DEVELOPMENTS

This system was demonstrated to other NASA centers and to various research and applications departments within McDonnell Douglas. Persons exposed to the demonstration found the concepts developed by this study valid in contributing to an efficient system of information exchange between a computer and a human operator.

Although the sequence of decisions required to reach a particular task may seem at times long, tedious, and unproductive, an operational system does not have to be so. By its very nature, the demonstration was filled with instructions for an uninitiated audience. These didactics are distracting and unnecessary to the well-trained operator. It is expected that an operational system would not require as many decisions to be made before reaching the active portion of the procedure as are currently used in our demonstration system.

Additionally, the use of the MANUAL special function capability allows the trained operator to enter an operational procedure at any one of the allowable entry points without making extensive preparations. As stated before, use of this capability should be left only to the operator thoroughly familiar with the systems and the structure of the procedure he wishes to perform.

More effective communications can be obtained when the medium used is made more anthropomorphic. Man's most natural mode of communication is speech. Any system that is to claim efficiency in communication with humans must be designed around the use of speech as a medium. Advanced systems of crew/computer communications must therefore include speech recognition capability to permit the computer to recognize and respond to "its master's voice." Part II of this study was concerned with this facet of man-machine communications.

Part III of this study was directed toward development of C/CC systems for experiment control applications. Just as various methods of communication were implemented in the representative system of Part I, Phase C, depending on the task to be carried out, so it is expected that when other specific applications are selected, the C/CC system used will have particular attributes for each application. Yet with the general-purpose software and hardware now in hand, additional applications can be developed without concern for background or supporting activities.

Section 3

DESIGN AND PERFORMANCE OF A LARGE VOCABULARY DISCRETE WORD RECOGNITION SYSTEM – PART II, PHASES A AND B

This report covers the word recognition work accomplished on the Crew/Computer Communications study. The objective of this work was the development, construction, and test of a 100-word vocabulary, near-real-time word recognition system. Additional goals included reasonable replacement of any or all of the 100 words in the vocabulary, rapid learning of a new speaker, storage and retrieval of training sets, verbal or manual single-word deletion, continuous adaptation with verbal or manual error correction, on-line verification of vocabulary as spoken, system modes selectable via verification display keyboard, relationship of classified word to neighboring word, and a versatile input/output interface to accommodate a variety of applications.

A further goal of this work was to identify methods of improving cost and performance of this system and to delineate promising avenues of technology related to speech recognition systems.

All objectives of this work have been successfully completed. Typically, the word recognition system is capable of classifying 100 words with an accuracy of 97.7 percent and a classification time of less than 0.9 second per word.

Although the ideal continuous speech recognition system has not yet been designed, a useful and expandable large-vocabulary word recognition system has been successfully developed and is now under evaluation at Kennedy Space Center.

This portion of the Crew/Computer Communications study was to evaluate the feasibility of using the word recognition system as a medium of communications. This effort, under the direction of B. C. Hodges of the computation

Laboratory at NASA Marshall Space Flight Center, was an outgrowth of the studies which were directed by Mr. Hodges over the past three years.

This work was performed with contractual support from two centers of NASA. Initial research and feasibility culminating in design, development, and testing of a 27-word recognition system was supported by Marshall Space Flight Center. Extension of the technology to a 100-word recognition system was sponsored by Kennedy Space Center under the direction of W. E. Parsons, Chief of Systems Engineering, and with active participation of G. Wood, also of Systems Engineering. Initial development for the 27-word recognition system began in June, 1972. The 100-word recognition system was delivered to Kennedy Space Center in September, 1973.

3.1 GENERAL DISCUSSION OF DISCRETE WORD RECOGNITION SYSTEM

Acoustic speech signal analysis is a building block for any speech recognition system. Some familiarity with the production and spectral analysis of speech may clarify the approach selected for the word recognition system.

3.1.1 Speech Generation and Spectral Analysis

Fundamentally, one speaks by modulating the flow of air which passes from the lungs through the windpipe, vocal chords, throat, nose, and mouth cavities (Reference 1). There are four ways to modulate this air flow—vocal chord, cavity, frictional, and stop modulation. The smallest unit of speech is called a phoneme. Potter and Kopp (Reference 1) list 39 English phonemes which are noncombinatorial. Of these, 31 have vocal chord modulation and are called voiced sounds. The vocal chord pitch of a male speaker generally is between 80 and 150 Hz and rarely exceeds 180 Hz (Reference 2). The waveform generated by the vocal chord vibration is roughly triangular in shape, having a duty cycle of one-third to one half the pulse period. Vocal chord vibration generates a line spectrum with harmonics at the pitch rate decaying at a rate of 12 dB per octave. Of the 31 voiced sounds, 28 are cavity or cavity plus frictionally modulated and do not involve stop modulation. Exclusive cavity modulated sounds are commonly known as vowel and

vowel-like sounds. Sonograms, as typified by Figure 3-1, show that all cavity-modulated speech sounds have spectral envelopes characterized by three or four spectral peaks in the region below 3,600 Hz for male speakers (Reference 3). These peaks represent the resonances of the vocal cavity and are commonly called formants. The vertical striations in Figure 3-1 illustrate the vocal chord pitch variations.* Notice that the cavity resonances or formants vary slowly in relation to the pitch rate. Nominally, male pitch frequency is about 132 Hz while formants typically vary at a rate less than 8 Hz. It is apparent that real-time formant extraction provides a significant portion of the information content in speech at a reduced data rate from the original acoustic signal.

Fricative modulation is characteristic of the group of sounds such as /s/, /f/, /ʃh/, /θ/, and /h/.** These unvoiced fricatives are produced by placing the articulators to form a small opening or constriction through which air must pass. The turbulent air yields a continuous rather than line spectrum and broad resonance bandwidths relative to formants. Figure 3-1 shows that /f/ as in "four" has a heavy concentration of energy around 7 kHz while /s/ as in "six" and /θ/ as in "three" have two major areas of energy concentration at about 4 kHz and 7 kHz. These sounds are most easily identified by measurement of broad-band energy and detection of the absence of voicing.*** It appears then that broad-band spectral analysis will accommodate detection of the five unvoiced fricatives, while formant extraction will describe reasonably well 28 cavity or cavity plus frictionally modulated speech sounds.

*Sonogram from a paper by G. L. Clapper, "Automatic Word Recognition," IEEE Spectrum, August 1971.

**Examples of unvoiced fricatives are /s/ as in six, /f/ as in four, /ʃ/ as in she, /θ/ as in three, and /h/ as in he. All symbols in this report surrounded by virgules will indicate phonemes.

***Five additional voiced fricatives are /ʒ/, /v/, /ð/, /z/, and /ʒ/. These sounds are reasonably well defined by the formants since they are produced by a combination of vocal cord, cavity and frictional modulation. When vocal modulation is added /f/ becomes /v/, /s/ becomes /z/, /ʃ/ becomes /ʒ/, /h/ becomes /ʒ/, and /θ/ becomes /ð/.

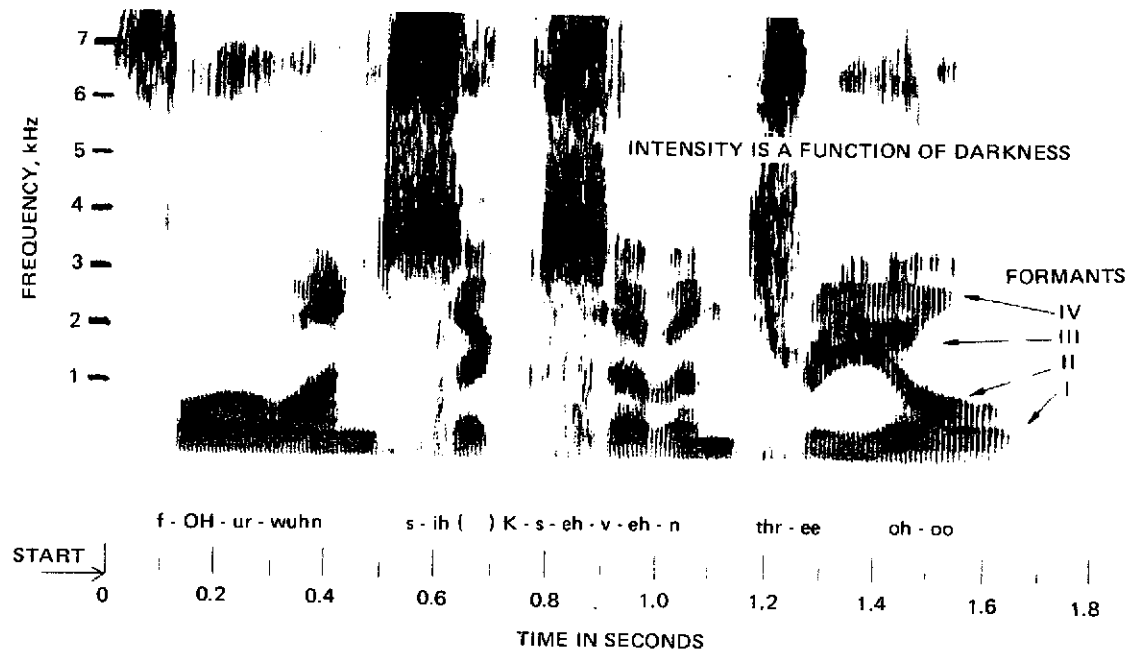


Figure 3-1. Sonogram of Digits (Male Speaker)

Six sounds, called plosives or stops, remain to be considered. The plosives are /p/, /t/, and /k/ which are unvoiced and /b/, /d/, and /g/ which are generally preceded by voicing. These sounds are made by stopping the breath flow at some point in the articulatory tract, building up breath pressure, and then rapidly releasing the breath. Once breath is released, these sounds are extremely short compared with all other phonetic sounds and yield a broad-band flat spectrum. Examples of these plosives are /t/ as in eight and /d/ as in send. The class of plosives in isolated words is usually identified by the absence of energy above the fundamental pitch for words ending in plosives. Frequently, ending plosives are not exploded. Initial plosives may be identified by the manner in which they influence the next phonetic sound.

3.1.2 Word Recognition System Description

An analyzer capable of extracting most of these significant measures for speech has been developed by McDonnell Douglas. Machine recognition of speech requires a higher level of processing than can be attained by speech analysis. A realistic approach to this higher-level processing is provided by a speech processor, which includes a verification display with keyboard to detect any machine errors arising from noise or human factors, and it allows various computer software options via the keyboard. A simplified block diagram of the system is shown in Figure 3-2.

The word recognition system was designed to operate on isolated utterances of from 0.2 to 1.3 seconds with no pause between phonemes or adjoining words exceeding 0.28 second. When a word is spoken into the microphone, the speech analyzer makes eight measurements of the acoustic signal every 8 milliseconds and transmits these data to the computer. To obtain high-accuracy word recognition, the eight real-time measurements provided by the analyzer must be relatively similar for the same word when repeated by the same speaker and yet provide a unique pattern in multi-dimensional measurement space for words differing by only a single phonetic sound. The latter requirement assumes that the vocabulary may be any unique 100 words. The speech analyzer measurement set selected includes the first three formants, provision for measuring amplitude of the speech signal below 600 Hz, amplitude of the signal between 3,700 and 5,000 Hz,

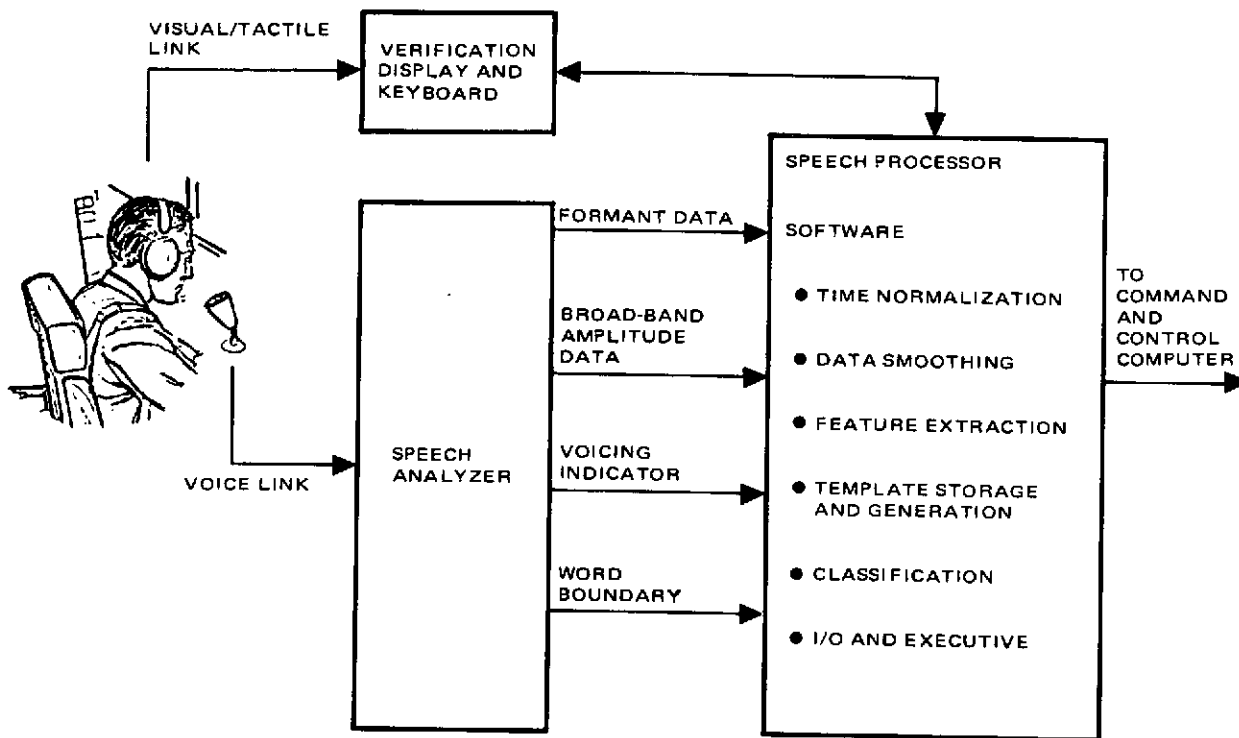


Figure 3-2. Word Recognition System – Simplified Block Diagram

amplitude of the signal between 6,000 and 7,600 Hz, and gross amplitude of the speech essentially from dc to 7,600 Hz, a voiced/unvoiced binary indicator, and an indication of the beginning and approximate duration of the word. In addition, the speech analyzer provides spectrum equalization to enhance detection of the formants. The method used for formant detection is speech spectrum segmentation. Segmentation boundaries are sequentially varied as required, and merging formant frequencies are accommodated. The analyzer also provides automatic level control circuitry to accommodate the dynamic range of speech and the variation in speaking level by various talkers, and to accentuate consonants which are frequently much smaller in amplitude than vowels.

The main purpose of the speech processor is to store a representative pattern for each word in the vocabulary, compare these stored patterns or templates with each spoken word, and classify and display a written verification of the spoken word. The classification software first determines the end of the word and total word duration. Next, the raw data are

time-normalized with a linear time base into a fixed matrix of data points, with a data compression of the raw data, typically by a factor of 4. All raw data within a normalized sample window are then averaged together to provide data smoothing. For each sample window, binary amplitude features and voicing are extracted. The resulting data per window contains 31 bits—16 for the three formants, 12 for the amplitude features, and 3 for voicing. Two bits to the right of each integer measurement are used for averaging and adaptive template generation. An additional 48 bits for each spoken word are used for storing the word duration, average template weight, class identification number, and template index number. The total storage per time-normalized word depends upon the selected samples per word. The software provides 16, 24, or 32 normalized sample windows per word with corresponding storage per word of 544, 792, and 1,040 bits.

The absolute difference between corresponding elements of the stored templates for each word and the incoming word template is then computed. The sums of the differences of the formant, voicing, amplitude features, and timing are independently weighted. These four sums are then combined linearly. The stored template having the smallest sum is then selected as the incoming word.

The speech processor software provides a variety of modes to be selected via the display keyboard. The modes included are (1) enter new vocabulary, (2) display current vocabulary, (3) replace spelling of a word, (4) replace spelling and training set of one word and retrain "n" iterations, (5) training mode which displays the word to be spoken, (6) adaptive training while classifying, (7) operation mode which displays last word spoken, (8) new speaker which allows storage or retrieval of a training set via paper tape, (9) sentence mode which displays words in the sequence spoken separated by a space, (10) distance mode which displays computed distance of n templates closest to the spoken word, number of templates per word and average weighting associated with each template, and (11) numeric mode which recognizes only digits or algebraic characters.

In the training mode, a manual delete is provided to erase the last word spoken in the event of coughing, noise, or other unintended sound. In the sentence mode, all errors may be deleted verbally prior to final page composition. In the adaptive training mode, training templates may be modified under verbal control in the event of an error. In adaptive training, all correct classifications automatically modify the corresponding correct training template to accommodate long-term speaker variability due to fatigue, background variation, colds, and the like.

3.1.3 Performance

The accuracy, speed, and computer storage requirements depend on conditions imposed on the system by parameter options, modifications in the vocabulary, speech pronunciation consistency, and cooperation of the user. Parameter changes include the time-normalized samples per word and the stabilizing factor in template production. Modifications of the vocabulary include vocabulary size and structure.

The system has been tested on a small population of speakers. The 100-word NASA vocabulary consisting of the 10-digit, 26 words of the ICAO alphabet (alfa through zulu) and 64 control verbs (enter, stop, turn-on, etc.), was commonly employed. A sample of the accuracy of performance as a function of 100-word vocabulary iterations is shown in Figure 3-3. Notice particularly the rapid learning of the system, attaining 93 percent on the second iteration. Note also the sharp decline in performance between successive days and rapid recovery. The long-term average of iterations 2 through 36 is 97.7 percent. The average of the last 10 passes is 98.7 percent; that is, 13 errors in 1,000 words. The amount of time required to classify each word varies with the number of normalized samples per word. For 16, 24, and 32 samples per word, the times were 0.6, 0.87, and 1.54 seconds, respectively. Structured vocabularies may be used to further improve recognition accuracy. Since the system provides for multiple templates per word, more than 100 templates are required for the 100-word vocabulary. For the results shown in Figure 3-3, 133 templates were required. Hence, template storage required was (133 templates) x (792 bits) = 105,336 bits.

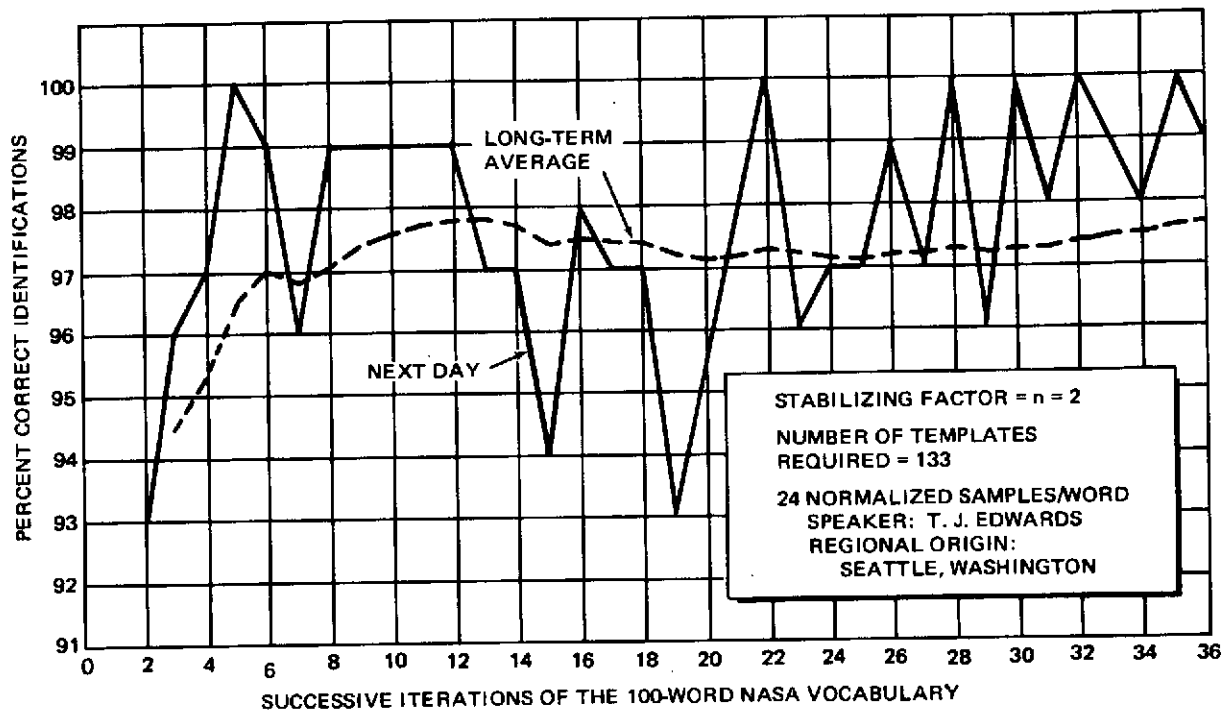


Figure 3-3. System Performance with Modified Template Production Function

3.2 WORD RECOGNITION SYSTEM CONFIGURATION

The word recognition system is constructed with the subunits illustrated in Figure 3-4.

3.2.1 Speech Analyzer

For each acoustic signal representing the input word, the speech analyzer makes measurements on the signal at uniform time intervals, the result being a set of digital numbers. The average duration of a word is 680 milliseconds. A short word such as "top" has an average duration of 400 milliseconds, while a long word such as "originate" has an average duration of 1 second. When a word is spoken into the microphone, the speech analyzer makes eight measurements of the acoustic signal every 8 milliseconds and transmits these data to the computer. Hence, on the average 680 divided by 8 or 85 samples per word are sent to the computer.

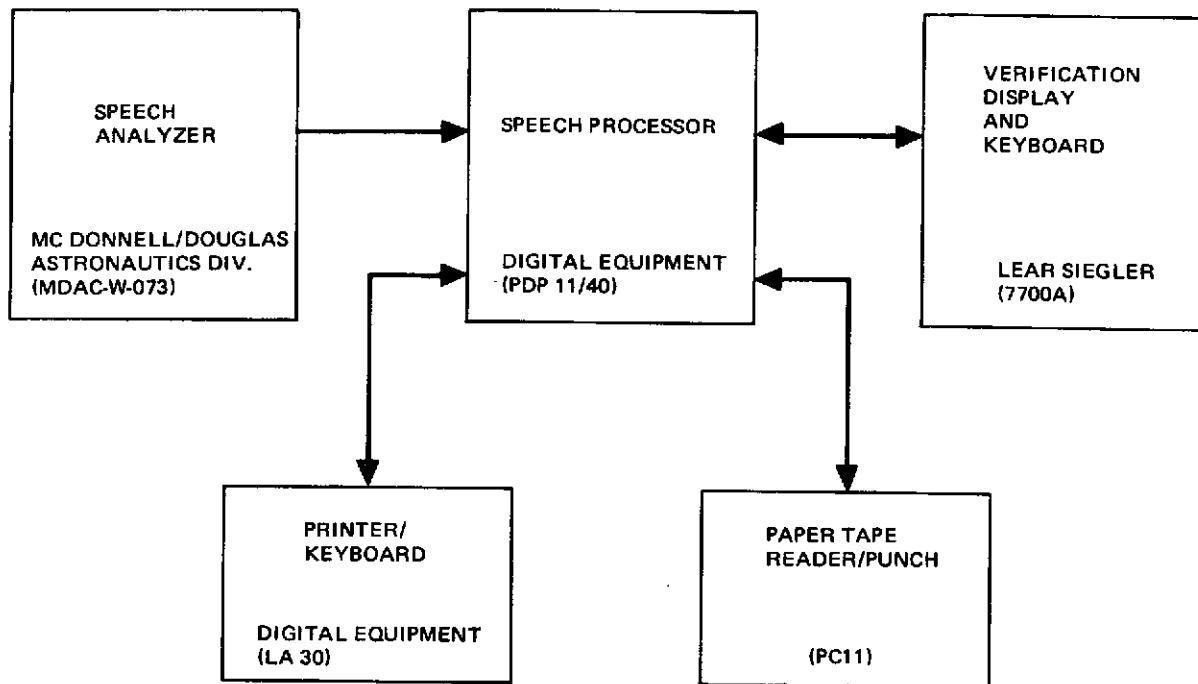


Figure 3-4. Word Recognition System Configuration

To obtain high-accuracy word recognition, it is essential that the eight real-time measurements provided by the speech analyzer be relatively similar for the same word when repeated by the same speaker and yet provide a unique pattern in multidimensional measurement space for words differing only by a single phonetic sound. The latter requirement assumes that the 100-word vocabulary may be any unique 100 words. In addition, the measurement set used by the speech analyzer should provide a reduction in bit rate over that obtained by direct digitization of speech data. The analyzer provides 4,375 bits per second while direct speech digitization requires approximately 50,000 bits per second (7,000 samples per second times 7 bits per sample).

Additional data reduction could be obtained by doubling the sampling time (8 to 16 milliseconds) and omitting redundant speech samples by use of nonuniform sampling intervals for bandwidth compression applications of the speech analyzer.

Of the 39 English speech sounds which are not combinatorial, 28 sounds have spectral envelopes characterized by three spectral peaks in the region below 3,600 Hz (Reference 2). These peaks represent the resonances of the vocal cavity and are commonly called formants. The speech analyzer determines in real-time the frequency location of the three major formants and generates a digital code indicative of each frequency band. In locating these formant frequencies, the analyzer segments the spectrum between 0 and 3,600 Hz into three overlapping bands. The range of each band and hence each segment may be translated. The selected band range for the majority of male speakers is as tabulated below.

Formant	Frequency Range (Hz)	Digital Code Range (Bands)	Bits/Sample
F1	180 to 890	0 to 9	4
F2	750 to 2,520	0 to 7	3
F3	1,500 to 3,580	0 to 5	3

Of the 31 voiced sounds, — that is, when the vocal cords vibrate — 28 are particularly well described by the measurement of the formants (exceptions are b, d, g). The analyzer also provides an additional bit for each sample to indicate the presence or absence of voicing and may be used to weight the relative value of the formant data (Reference 4).

In addition, the analyzer measures the gross amplitude, A_G , of the speech signal (6 bits), the low-frequency amplitude, A_L , of the signal below 600 Hz (6 bits), the high-frequency amplitude, A_H , of the signal between 3,700 and 5,000 Hz (6 bits), and the very-high-frequency amplitude, A_{VH} , of the signal between 6,000 and 7,600 Hz (6 bits).

A_G is useful in defining word boundaries and normalizing other amplitude measures, and can also be used for detecting a pause at the end of a word followed by a short burst of energy, thereby identifying a stop consonant. A_L , A_H , and A_{VH} are used by the existing computer software to form ratios

with A_G. The purpose of these measures is to aid in describing the 11 phonetic sounds* not clearly defined by the formant data and to assist in identifying the nasal consonants (/m/ as in measure, /n/ as in won, and /ng/ as in displaying). (See Reference 4.)

The aforementioned amplitude, formant, and voicing pattern over the entire duration of the word is input to the computer via the general-purpose interface designated by Digital Equipment Corporation as the DR11-C. Three of these interface channels are provided with the system and are housed in the PDP11/40, the speech processor for the word recognition system.** Each channel has the capacity for parallel input or output of 16 bits and uses standard transistor-to-transistor logic levels. Each output is provided with a holding register. Each input has two interrupts termed REQ-A and REQ-B, with the latter having the lowest priority. The entire data set is sent to the mini-computer from the speech analyzer; each sample interval is in the format illustrated in Figure 3-5.

Of the three available DR11-C's, the inputs of only two of the interfaces are employed. This leaves the third DR11-C available for monitoring and controlling peripheral equipment via the word recognition system.

3.2.2 Speech Processor and Peripherals

The PDP11/40 speech processor selected for the word recognition system is made by the Digital Equipment Corporation. The processor is used to:

- A. Store and generate prototypes of each word in the vocabulary.
- B. Store the alphanumeric representation of each word.
- C. Perform time-normalization and amplitude feature extraction of each incoming word.

*The 11 sounds not accurately described by the format data are: the unvoiced fricatives /f/ as in four, /h/ as in hotel, /s/ as in six, /θ/ as in three, /ʃ/ as in dimension, function, and option, and the stop consonants /p/ as in papa and stop, /t/ as in sight, /k/ as in kilo or yankee, /b/ as in debug or bravo, /d/ as in delta or send, and /g/ as in golf or begin. The examples are from the selected 100-word vocabulary.

**The PDP11/40 was specified by the contract. The speech analyzer can be configured to operate with a variety of computers. In the MDAC laboratory, for example, the analyzer is mated to an in-house developed minicomputer as well as an XDS-930 computer.

The PDP11/40 options selected include extended memory, 16,384 16-bit words, with 0.9-microsecond cycle time, signed integer multiply and divide with arithmetic shifts (KE11-E), a hardware bootstrap loader (BM792-YA), three general-purpose serial interfaces (DR11-C), and an asynchronous serial interface, the DL11-C. The DL11-C provides the interface between the 7700A data terminal and the PDP11/40. This interface and the 7700A are set for 9600 baud.

The PDP11/40 has proven to be an excellent machine as a building block of the word recognition system. Some of the features of the PDP11/40 are:

- A. Byte processing. This is particularly important to the word recognition system since any of the eight measurements made by the speech analyzer consist of less than eight bits. Hence, efficient storage and processing of these data is possible.
- B. Six general-purpose registers (excluding registers 6 and 7). These registers can be used for accumulators or addressing.
- C. Eight addressing modes. The modes include register, auto-increment, autodecrement, or indexing with both direct or indirect addressing. These modes may be used with any of the six general-purpose registers, R0 through R5.
- D. Double operand instructions. This feature facilitates programming, since any two consecutive locations in the core may be addressed by a single instruction (for example, ADD A, B).
- E. Modular chassis design allows for ease of adding peripheral equipment and the compatible interface. When the general-purpose parallel interface is used, both the input data and output are given an address. In addition, the status register associated with these interface units is addressable. The status register controls whether interrupts from external equipment will be accepted or not. Also, signals are provided to initialize external equipment under program control.

F. High speed facilitates real-time word recognition. The time in microseconds required by some of the typical instructions is tabulated below.

Instruction	Execute/Fetch Time	Source Time	Destination Time	Total
TST	0.99	0.00	0.00	0.99
BNE	1.76	0.00	0.00	1.76
ADD	1.60	0.84	0.00	2.44
SUB	1.60	0.84	0.00	2.44
MUL	8.88	0.84	0.00	9.72
DIV	11.30	0.84	0.00	12.14

G. Hardware traps are provided which detect software errors.

The capacity of the verification display and keyboard used by the system (the Lear Siegler 7700A) is 25 lines by 80 characters. At the 9600 baud transfer rate, the entire 2,000 characters may be presented in less than 2 seconds. The terminal transmits data to the computer via the DL11-C in serial form containing one start bit, seven data bits, even parity, and one stop bit. Optical coupling for both transmit and receive is provided between the DL11-C and the 7700A which alleviates grounding and common mode problems. The terminal can transmit blocks of data (edit mode) or single characters (conversation mode). The edit mode is particularly useful when a large vocabulary list is initially being typed since it allows on-line correction prior to transmission to the computer. In addition, the terminal provides direct program control over the cursor position, allowing random access to any character. Under program control, the keyboard can be activated or deactivated. This display has also been found useful in graphically displaying on-line, formant versus time plots.

The function of the keyboard is to allow the operator to select the various modes of operation provided by the word recognition system software, input the spelling of any one or all words in the vocabulary, and assist in system training. The various modes selectable via the associated keyboard entry are tabulated in Table 3-1.

Table 3-1
WORD RECOGNITION SYSTEM SOFTWARE MODES
VIA KEYBOARD CONTROL

VC:	Enter new vocabulary
DV:	Display current vocabulary
RP:	Replace spelling of one word in vocabulary
RT:	Replace spelling and training template of one word in vocabulary with a new word and train (enter number of training passes)
NS:	New speaker - used for reading new speaker from tape with reader on, or punching tape of last speaker with reader off
TM:	Training - displays word to be spoken during training
OM:	Operational mode - displays last word spoken
AT:	Adaptive test and train-allows on-line verbal correction of misclassified word*
DM:	Distance mode - presents spelling and computed distance of "n" prototypes closest to the word input, displays number of prototypes per word, and the average weight associated with each prototype
SM:	Sentence mode - displays spoken words in sequence, each word separated by a space [Begin = (, Delete = Erase last word, End =)]
NM:	Numerical mode - allows for recognition of digits and algebraic characters only, sets up calculation of numerical data based on algebraic operators (+, -, ÷, *)

* In the AT mode, two words are added to the 100-word vocabulary. They are "ssssss" and "ERASE." In the event of a word error -A- the sound "ssssss" followed by the correct classification -B-, followed by "ssssss," causes the software to modify the prototype storage of the correct classification -B- by the original time-normalized error word "A." All correct classifications of a word automatically augment the prototype associated with that correct classification.

The exact modes of operation are displayed at all times in the lower left-hand corner of the display.

The PC11 high-speed paper tape reader uses fanfold paper and operates at 300 characters per second, while the punch operates at 50 characters per second. When a training tape exists, a new speaker may be entered via the reader in less than three minutes or a training tape punched in less than eight minutes.

The LA30 writer is used primarily to generate new software, but it is also useful for permanently recording the data from the speech analyzer over the interval of a word. The LA30 is especially valuable when used in conjunction with the paper tape edit software (ED11). The LA30 writes at a rate of 30 characters per second.

3.3 RESULTS

This section describes the results obtained with the existing speech analyzer and PDP11/40 software. Also included are more recent results using the McDonnell Douglas XDS-930 computer. The accuracy, speed, and storage of the word recognition system depend on conditions imposed on the system by parameter options, modifications in the vocabulary, and the speech pronunciation consistency of the user. Parameter changes include samples per word and the stabilizing factor in template production. Modifications in the vocabulary include vocabulary size and structure. A small population of speakers was investigated.

The classification accuracy and speed may be altered by changing the number of time-normalized samples per word. The current options available are 16, 24, and 32 samples per word. Lowering the samples per word tends to reduce accuracy slightly but offers advantages for storage. This variable may be traded as required by the application.

A plot of accuracy versus samples per word using the 100-word data from the speech analyzer but a slightly different classification technique

simulated on the McDonnell Douglas XDS-930 computer is shown in Figure 3-6. The speaker was Carl Kesler and results are averaged over repetitions 26 through 35 of the 100-word vocabulary with the adaptive training mode active. These results illustrate that as the samples per word increase, the accuracy increases in a fashion similar in appearance to the function $(1 - e^{-x})$, while speed and storage requirements increase in essentially direct proportion with the normalized samples per template.

It should be noted that a considerable reduction in decision time was achieved when the PDP11/40 was used, since the cycle time of the XDS-930 is 1.75 microseconds as opposed to 0.9 microsecond on the PDP 11/40 and the XDS-930 has only two accumulators, as opposed to the PDP 11/40's six.

The remainder of the results were obtained using the PDP 11/40 software delivered to NASA.

In classification of each incoming word, the distance between the incoming word and all templates or prototypes is computed. Ideally, the distance between repetitions of the same word would be zero. However, variation in pronunciation of the same word causes the distance between repetitions of the same word to be greater than zero. For example, using 32 samples per word and if the word spoken by a particular speaker is "echo," it will have an average classification distance of 347 — as low on a single repetition as 338 or as high as 355. Hence, the distance between the average "echo" (template) and the incoming word is a maximum of 8. Since the distances between all templates and the incoming word are computed, the second-smallest-distance member of this ordered set can be determined and is called the nearest neighbor. When the word "echo" is spoken (typically, 347) by a particular speaker, the nearest neighbor is "manual," which has an average distance of 742. Hence, the distance between "echo" and its nearest neighbor "manual" is 395, while the distance between repetitions of "echo" is 8. Such a wide separation between the same word and its nearest neighbor is not always possible. This separation is for the largest part dependent on the vocabulary. The vocabulary selected by NASA for testing the 100-word recognition is listed in Table 3-2.

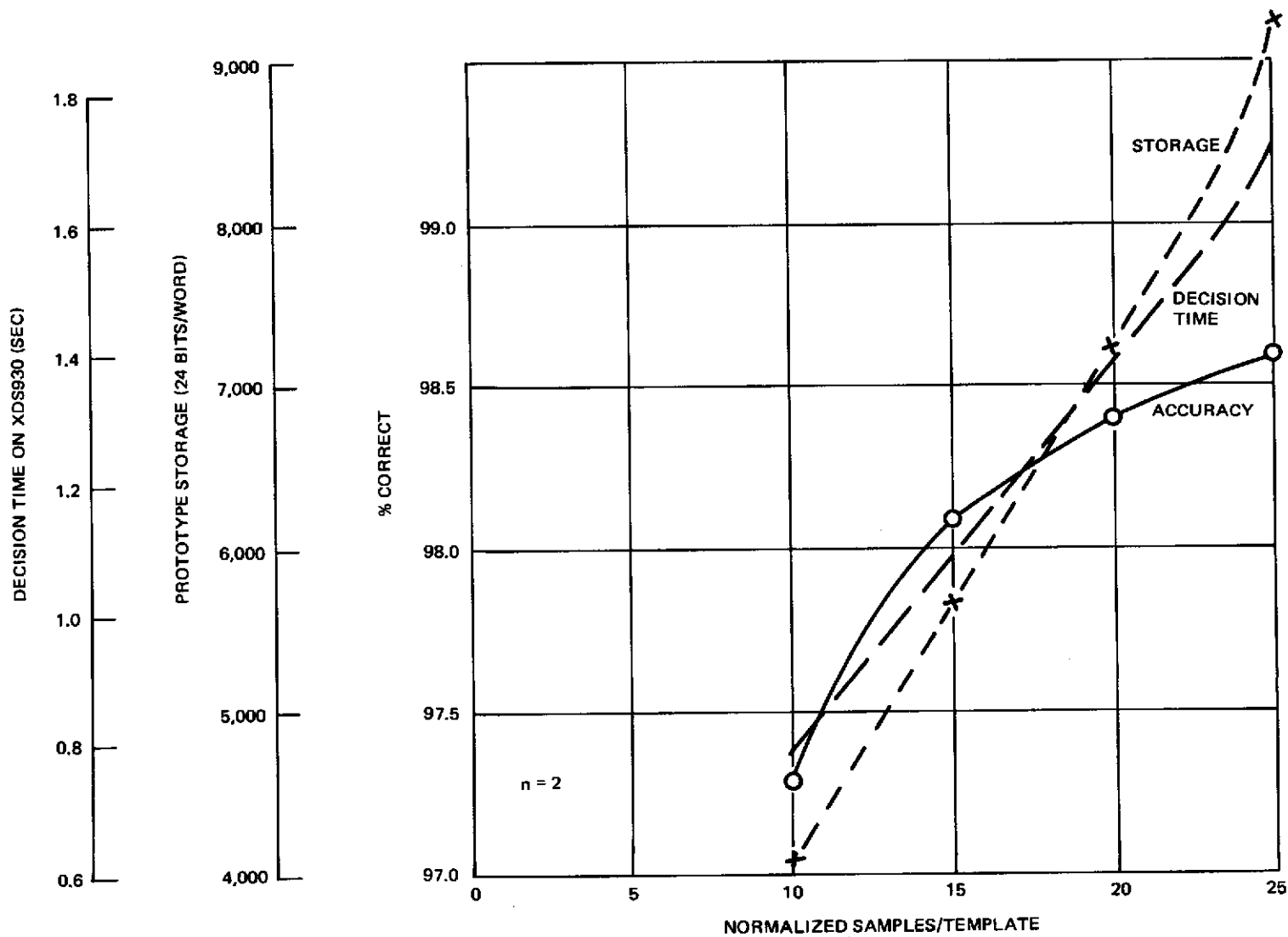


Figure 3-6. Speech Analyzer Accuracy

Table 3-2
 VOCABULARY LIST IN 100-WORD SPEECH
 RECOGNITION SYSTEM

Zero	Papa	Delete	Originate
One	Quebec	Develop	Parameter
Two	Romeo	Dimension	Perform
Three	Sierra	Display	Print
Four	Tango	Dump	Program
Five	Uniform	Emergency	Point
Six	Victor	End	Pulse
Seven	Whiskey	Enter	Read
Eight	X-Ray	Execute	Record
Niner	Yankee	Function	Reset
Alfa	Zulu	Go-To	Resume
Bravo	After	Halt	Rubout
Charlie	Alter	Hold	Sample
Delta	Apply	Immediate	Save
Echo	Assign	Insert	Send
Foxtrot	Average	Issue	Set
Golf	Begin	Manual	Start
Hotel	Calibrate	Measure	Stop
India	Call	Minus	Terminate
Juliett	Cancel	Modify	Test
Kilo	Change	Monitor	Trace
Lima	Clear	Off	Top
Mike	Close	On-Line	Turn-On
November	Compile	Open	Up
Oscar	Debug	Option	When

Adaptive Test/Train Control Words: ssssss
 Erase

In experimenting with the word recognition system, an attempt was made to improve the recognition accuracy obtained by speaker R. G. Runge by changing selected words in the vocabulary; that is, ZERO to OH, BRAVO to BAKERY, GOLF to GOLDEN, KILO to KILOGRAM, NOVEMBER

to NELLY, TANGO to TANGERINE, UNIFORM to UMBRELLA, VICTOR to VICTORY, WHISKEY to WHALE, YANKEE to YANK, ZULU to ZEBRA, AFTER to AFTERWARD, ALTER to ALTERING, DISPLAY to SHOW, IMMEDIATE to IMMEDIACY, OFF to TURN-OFF, OPTION to OPTIONAL, PRINT to PRINTS, SEND to SENDING, UP to UPWARD, and WHEN to WHENEVER. The system was first tested on the original NASA vocabulary list and then on the modified vocabulary with the 32-sample-per-word option used for both tests. For both experiments, seven training passes (100 word list repeated seven times) were input to the system, then 10 test passes were input without adaptive training while the tests were made, then seven additional training passes were made, followed again by 10 test passes. Hence, each test was independent and consisted of 20 repetitions of each of the 100 words plus the two key control words. For the unmodified vocabulary, 223 errors were made in classifying 2,040 words, or 89.07 percent were correct. For the modified vocabulary, 89 errors were made for the 2,040 words, yielding an accuracy of 95.72 percent. It appears that selective vocabulary replacement, empirically based on the speaker errors, can significantly improve system performance.

With the modified vocabulary, an investigation was conducted using the distance mode to determine how well the same word compares to itself and to its nearest neighbor when it is repeated. The average distance of each word and its nearest neighbor was approximated by speaking each word twice and recording the distance of the spoken word and its nearest neighbor. Next, the distance difference between the same word and its nearest neighbor was recorded for all members of the vocabulary list. In this manner, it is possible to estimate the probability distribution for the difference between repetition of the same word and the average difference between the incoming word and its nearest neighbor on the common difference distance scale. The results are shown in Figure 3-7. The cross-hatched area is where a word and its nearest neighbor intersect, the region that causes errors to occur. On the last 10 repetitions of the vocabulary by R.G. Runge, an experimental accuracy of 95.48 percent was obtained, 47 errors in 1,040 words. From the joint probability distribution of the same word and

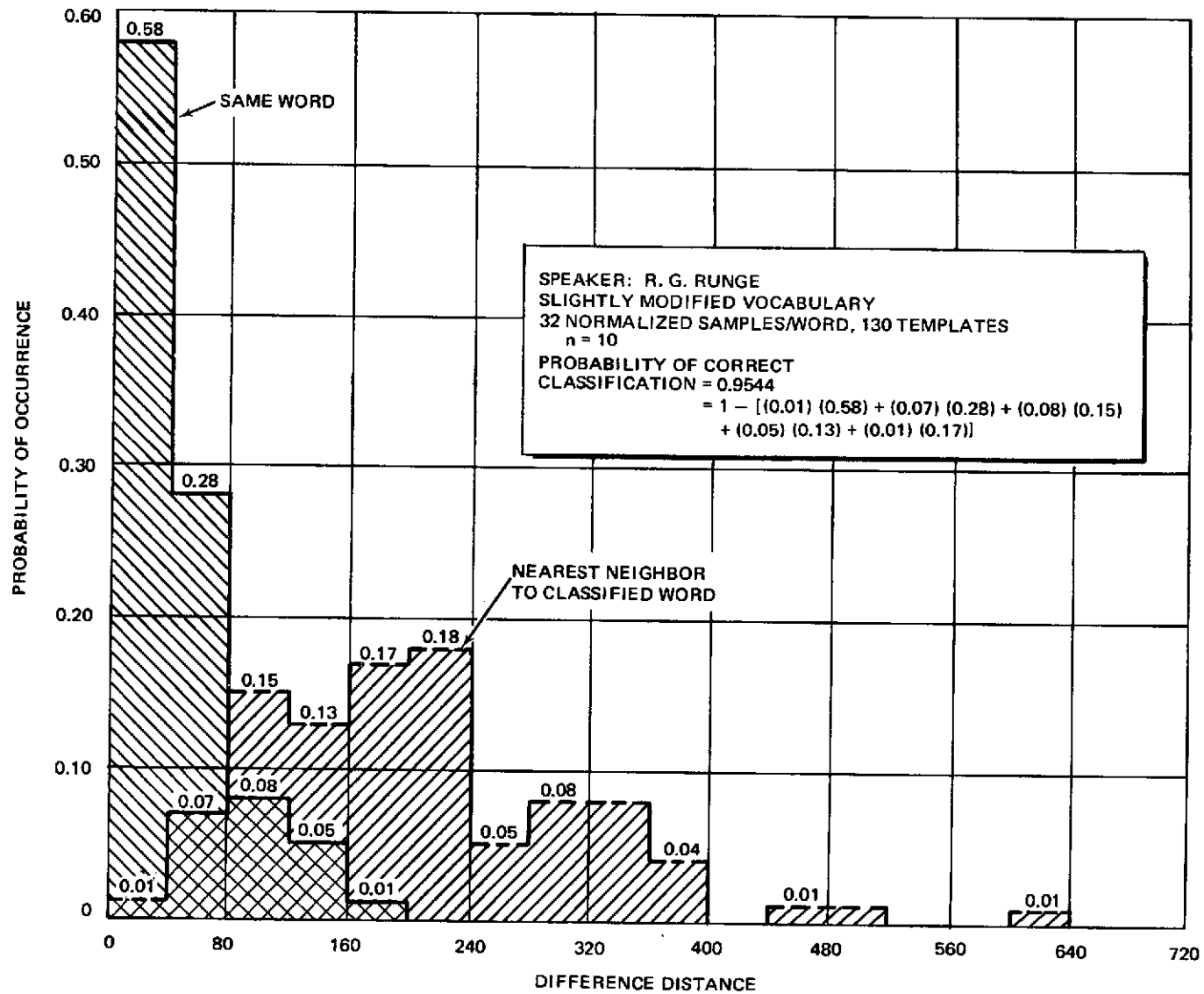


Figure 3-7. Probability Distribution Versus Distance

its nearest neighbor. an accuracy of 95.44 percent would be anticipated. It appears that the same word when repeated yields a rather compact distribution while the nearest neighbor distribution is somewhat broad and dispersed. This suggests that careful attention should be given to vocabulary selection if optimum accuracy is to be obtained. For example, for R.G. Runge, "three" is often confused with "charlie," "parameter" with "monitor," "delta" with "alfa" or "papa," and "assign" with "seven". Carefully selected replacements for these words would improve recognition performance, as previously suggested.

Recognition accuracy is also a function of vocabulary size. A test was made using the first 50 words of the modified vocabulary by speaker R.G. Runge. The first 50 words were repeated twice in each 100-word training pass. Next, a 25-word vocabulary consisting of the months of the year, the digits zero through nine and the words "begin," "end," and "delete" were used to train the system. The 25-word system was trained as a 100-word system with each training pass consisting of four repetitions of the 25-word vocabulary. Testing was performed in the nonadaptive mode. Seven training passes were used for the 100, 50, and 25 word vocabulary. Interpolating the results, Figure 3-8 illustrates the dependence of vocabulary size on system accuracy.

The speed of response to the word recognition system is a function of the time-normalized samples per word. Measurements of response times versus the samples per word option are approximately:

Samples/ Per Word	Response Time (sec)	Number of Templates During Timing Test
16	0.60	110
24	0.87	106
32	1.54	130

This again portrays the linear relation between sample per word and response time.

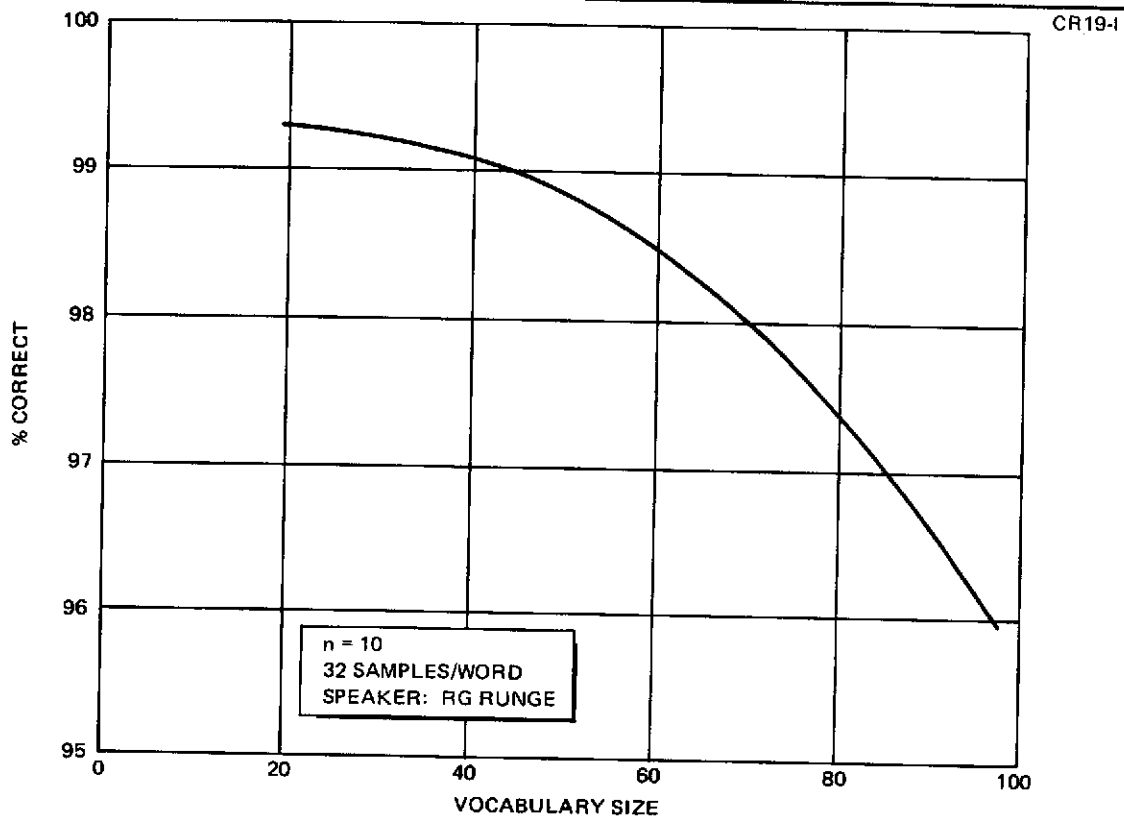


Figure 3-8. Dependence of Accuracy on Vocabulary Size

A comparison of accuracy and samples per word was made by R. G. Runge using 16 and 32 samples per word on the 25-word vocabulary, each using seven training passes and nonadaptive test passes. The 16 samples per word, 25-word vocabulary yielded 97.78 percent accuracy with six errors in 270 words, while the 32 sample per word system yielded an accuracy of 99.26 percent, as previously described, with only two errors. The only confusion for the 32 sample per word system was between "delete" and "eight" and "vice versa." Experiments by R. G. Runge with syllable stress indicate the system performance is relatively independent of this variation.

The accuracy of the system is somewhat variable, depending on the speaker. The system appears to yield its best performance when used continuously in the adaptive training or AT mode. Figure 3-9 shows the results obtained by T. J. Edwards using 16 and 24 samples per word, continuously adaptively testing, and spread over two different days.

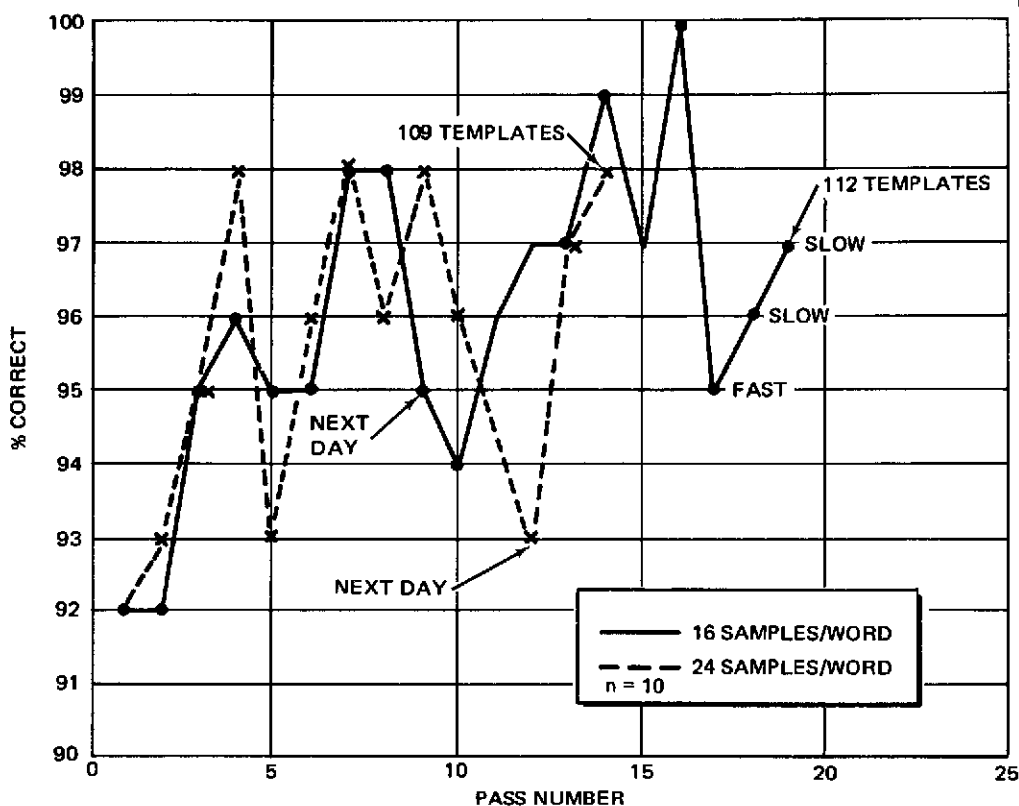


Figure 3-9. T. J. Edwards Adaptive Testing

The results are presented graphically in Figure 3-9 as a function of iteration or pass number and were obtained using the 100-word NASA vocabulary. Initially, two training passes were used before entering the operational mode.

The third from the last pass of the 16 sample per word graph shows the result obtained when the words are intentionally spoken quickly. The last two passes are the results when the words are intentionally spoken slowly.

The average accuracy for all 19 passes using 16 samples per word is 96 percent; however, the average accuracy from pass 7 through 16 is 97.1 percent, — perhaps a more realistic figure, since it allows the system a more reasonable number of training passes and the intentional fast and slow passes are not included.

The words causing the 76 errors out of the 1,900 words spoken by T.J. Edwards are listed below, followed by the number of times the given word was in error.

Eight	5	Zulu	2	Apply	1
Delta	5	Call	2	Cancel	1
Golf	5	Debug	2	Dimension	1
Print	5	End	2	Enter	1
		Set	2	Function	1
Dump	4	Stop	2	Minus	1
Halt	4			Monitor	1
Delete	4	One	1	Point	1
		Two	1	Reset	1
Manual	3	Lima	1	Sample	1
Save	3	Papa	1	Start	1
Send	3	Quebec	1	Top	1
		Victor	1	When	1

The results of various other speakers, using the 16 sample per word configuration, are listed below:

	Training Passes	Test Passes	Accuracy (%)
Mas Uemura (MU)	7	13	95.15
Dan Nikodymn (DN)	5	4	94.00
Marion Olsen (MO)	5	1	94.00
Walt Parsons (WP)	5	3	95.33

An important feature of the speech analyzer design is the ability to accommodate men, women, and children. (This is accomplished in the hardware with a frequency range selector switch.) Marion Olsen was our only female speaker.

As described previously, a new template is created for a word only if that word is incorrectly classified and it is not one of "n" closest templates to the word input. This value of "n" can realistically range from "1" to "10." All previous results up to this discussion have used an "n" value of "10," which quickly stabilizes the number of templates required to represent

each word; that is, fewer templates are required to represent the 100-word vocabulary. However, it has been demonstrated in simulation tests that the smaller the value of "n," the higher the accuracy of the system; therefore, an "n" value of "2" was utilized in making the test depicted in Figures 3-10 and 3-11. Both theoretically and empirically, these graphs show that the expected long-term classification response of the system is approximately 97.7 percent. Using the average of the last 10 passes, 27 through 36, an accuracy of 98.7 percent is obtained. This is indeed better than has previously been obtained with an "n" value of "10," and the increase in the number of templates is not at all prohibitive. It should be acknowledged that an "n" value of "1" was also tested, but the number of templates required was well over 215 and the system accuracy was very slow in converging. Therefore, an "n" value of "2" seems to be the optimum system configuration for template production.

An interesting experiment which has not been performed would be to train the system on more than one speaker. Currently, the system has never exceeded 130 templates for a single speaker, using a stabilization factor of $n = 10$. The software system design allows for 224 templates with 24 samples per word. Since at least 94 prototypes or templates are available, it is highly probable that the system will accommodate at least two speakers simultaneously, provided they alternate while training the system. This is apparent from the fact that over half of the words are generally correctly classified for a second speaker, which should require less than 65 templates for the remaining half of the vocabulary.

3.4 SPEECH ANALYZER DESIGN AND PERFORMANCE

This section discusses the design and performance of the speech analyzer and shows examples of the output data.

The speech analyzer consists of 12 functional blocks: (1) microphone and pre-amplifier, (2) speech spectrum equalizer, (3) automatic level control, (4) filter network with rectification and low-pass filtering, (5) difference amplifier and low-pass filters, (6) frequency range multiplexer, (7) formant extraction network, (8) amplitude digitizer and interface logic, (9) word boundary detector, (10) voicing detector, (11) timing and control network,

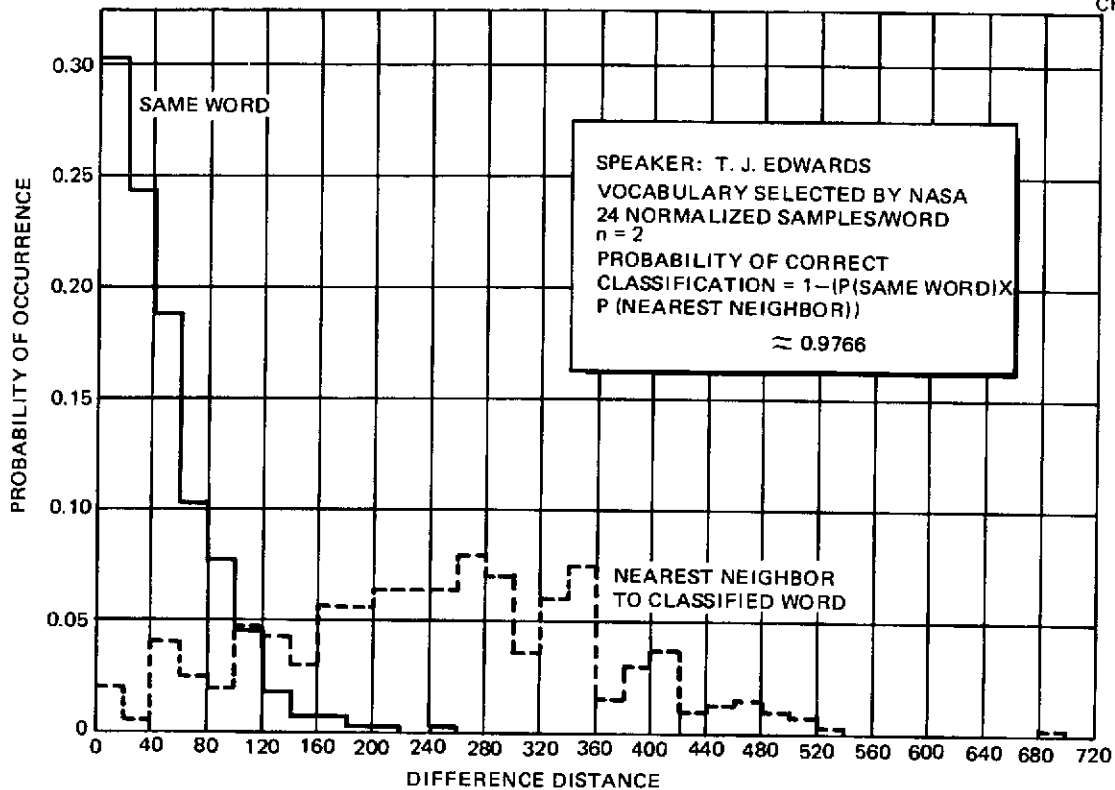


Figure 3-10. Probability Distribution

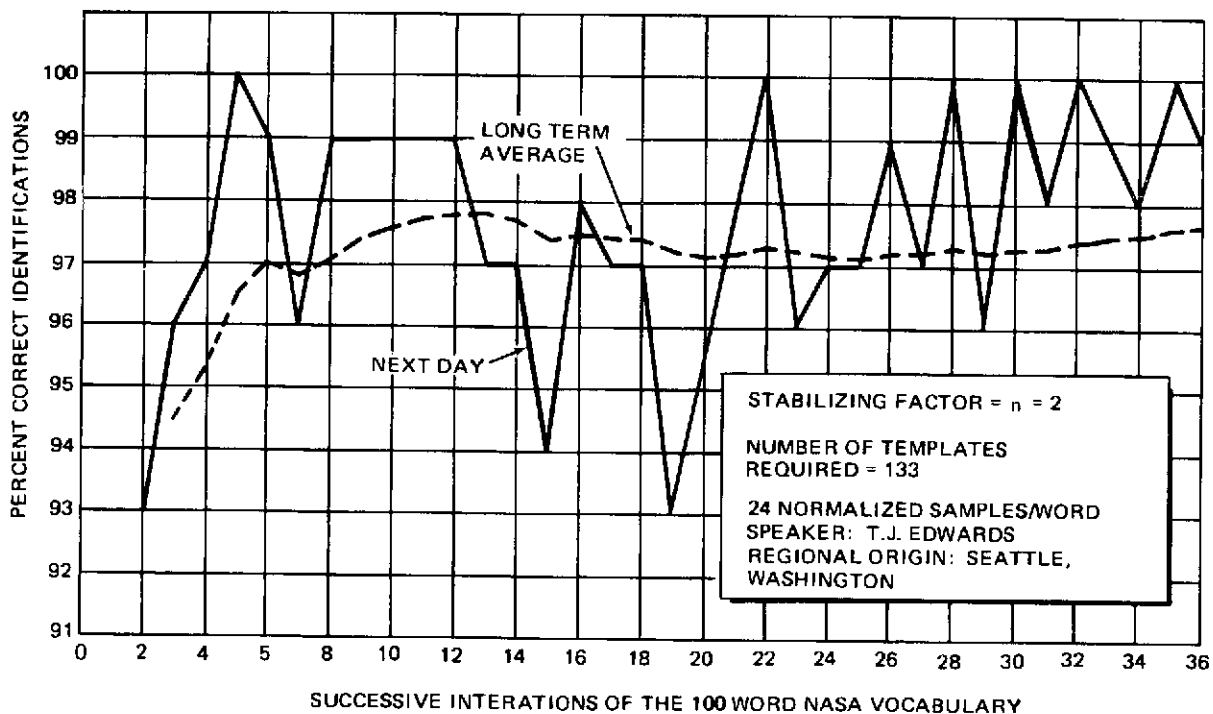


Figure 3-11. System Performance with Modified Template Production Function

and (12) data display. The connection of these functional blocks is illustrated by the simplified block diagram in Figure 3-12.

The system is constructed on 25 circuit boards. Fifteen of the boards, containing mostly redundant circuitry, are printed circuit boards with the remaining circuitry hand-wired. The first board in the system contains the preamplifier spectrum equalizer, band-limiting, low-pass filter, and the word boundary detection circuitry.

3.4.1 Microphone and Preamplifier

The microphone selected for the system is manufactured by AKG Microphones (Model No. K-158). The microphone impedance is 200 ohms and sensitivity is -54 dB. This unit operates on the differential in acoustic pressure and hence rejects ambient noise, particularly in the low-frequency region. The frequency response of the microphone to a source 2 inches distant and to plane waves from a distance of 3 feet as well as the directional selectivity for various frequencies is shown in Figure 3-13.

This microphone supplies a two-stage, low-noise-differential input preamplifier which has an input resistance of 10 kilohms. The first stage of the preamplifier has a gain of 100 and is constructed using a Fairchild UA725 integrated circuit, which has exceptionally low input noise current. The positive and negative supply of this amplifier is provided by two 12-volt zener diodes to further enhance the superior power supply rejection of the UA725.

The second-stage amplifier has a closed-loop gain of approximately 2.5 and is ac-coupled to the first stage. The overall undistorted bandwidth is 1 Hz to 35 kHz.

The microphone-to preamplifier connection may be momentarily or continuously disconnected from the preamplifier input terminals by manual controls incorporated in a hand-held microphone switch.

3.4.2 Speech Spectrum Equalizer

When the vocal cords vibrate, the glottal waveform produced is roughly triangular in shape with a pulse width, t_1 , which varies from one-third to

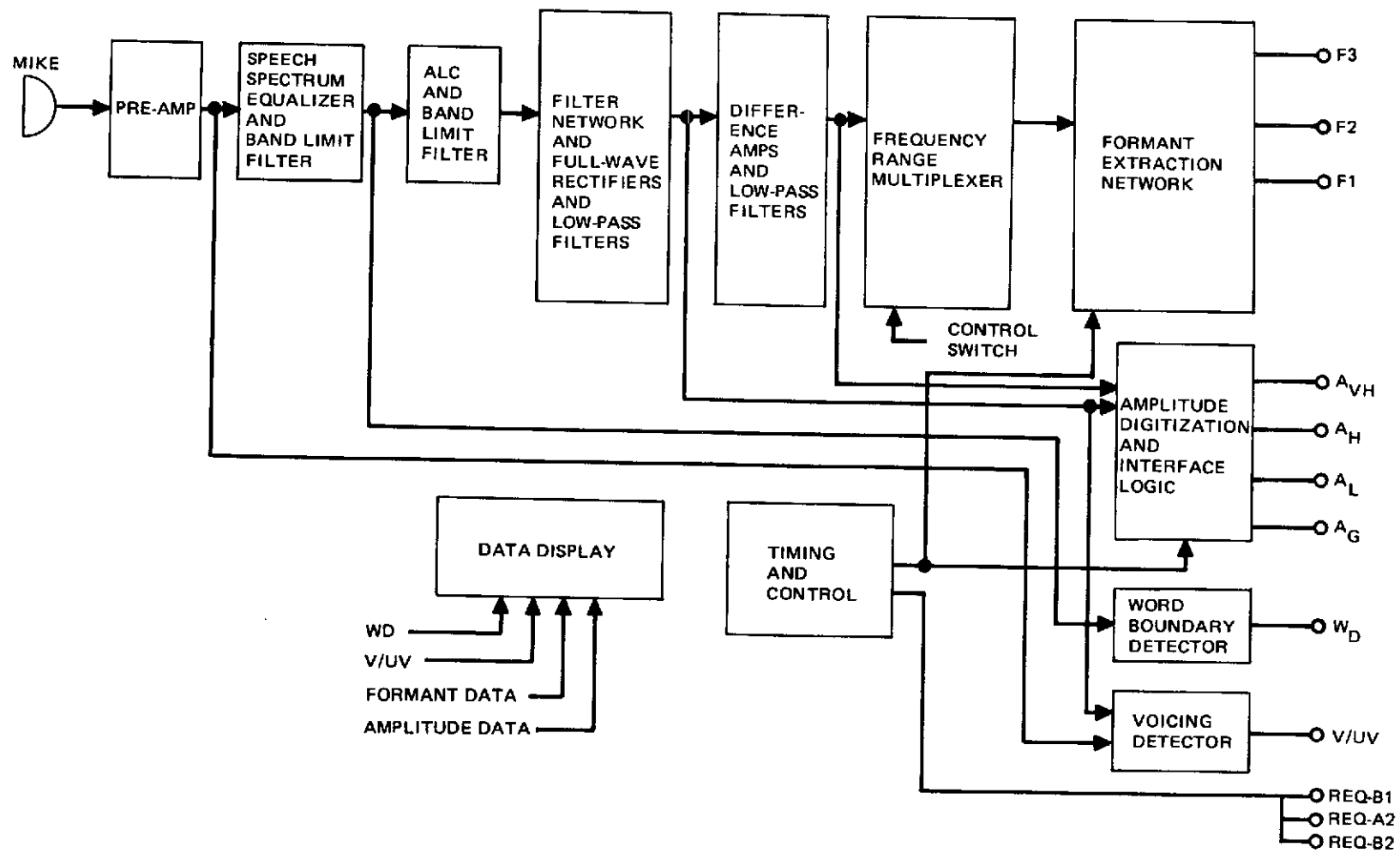


Figure 3-12. Speech Analyzer—Simplified Block Diagram

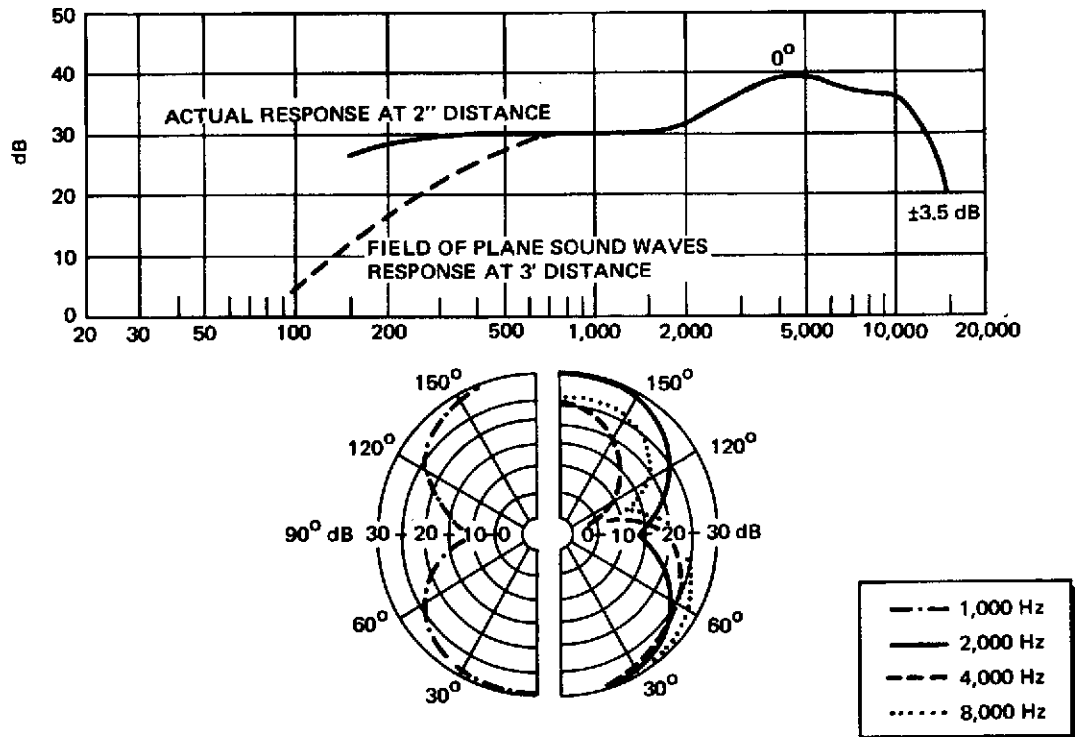


Figure 3-13. K-158 Microphone Frequency Response and Directional Characteristics

one half the pitch period, T . This yields a spectrum whose harmonic coefficients may be approximated by (Reference 5)

$$F\left(\frac{2n\pi}{T}\right) = \left[\frac{\text{Sin}\left(\frac{\pi n t_1}{T}\right)}{\left(\frac{\pi n t_1}{T}\right)} \right]^{-2}$$

This function decreases by a value of $(1/n)^2$ or at a rate of -12 dB per octave. If $t_1/T = 1/2$, then it has zeros for all even values of n . If $t_1/T = 1.3$, zeros occur when n is a multiple of 3. In either case, the rate of roll-off remains constant. A male speaker typically has a pitch of around 125 Hz with the third formant reaching as high as 3,650 Hz. Our interest is in detecting formants up to the 29th harmonic of the fundamental male pitch, where the 29th harmonic is down relative to the second harmonic (first possible formant peak in the F1 band) by a factor of approximately 210.

The equalizer generates the inverse of this roll-off and hence flattens the spectrum for voiced sounds. The importance of the equalizer in assisting in the detection of the spectral peaks associated with each formant can be easily understood if the system is considered with and without the equalizer (Figure 3-14). Without the equalizer, the average amplitude in F2 band 1 would certainly exceed the amplitude in F2 band 5. With the equalizer, the correct peak, F2 band 5, may be detected.

The equalizer is placed prior to installing the automatic level control circuitry (ALC). This serves two purposes: (1) it prevents the input signal to the ALC from dropping below the quantization level of the A/D converter, and (2) it tends to provide additional accentuation of higher harmonics.

The frequency response of the equalizer is shown in Figure 3-15. The rate of rise can be controlled by varying R_2 . Reducing R_2 to zero increases the rate of

CR19-1

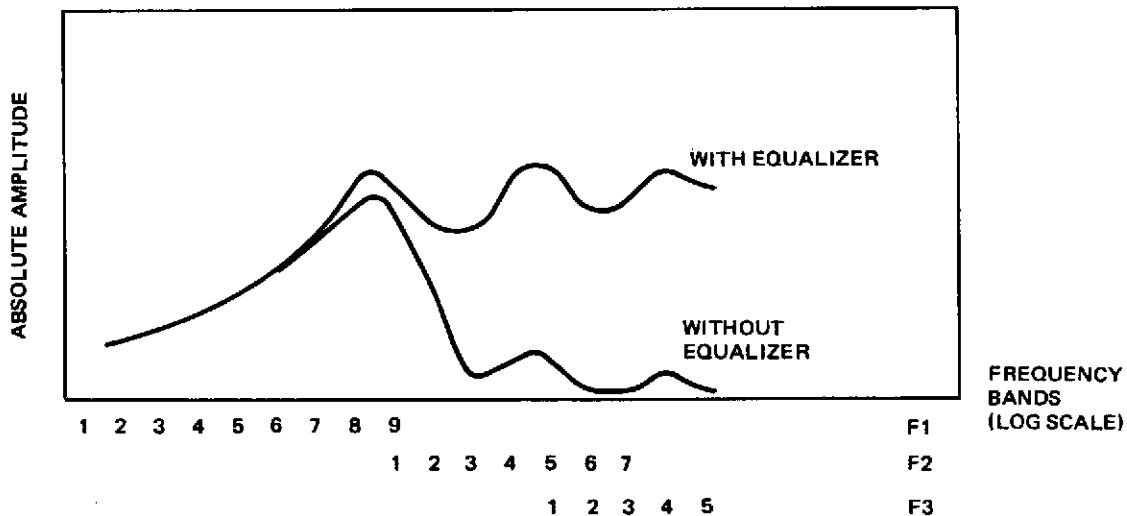


Figure 3-14. Voiced Sound Spectrum

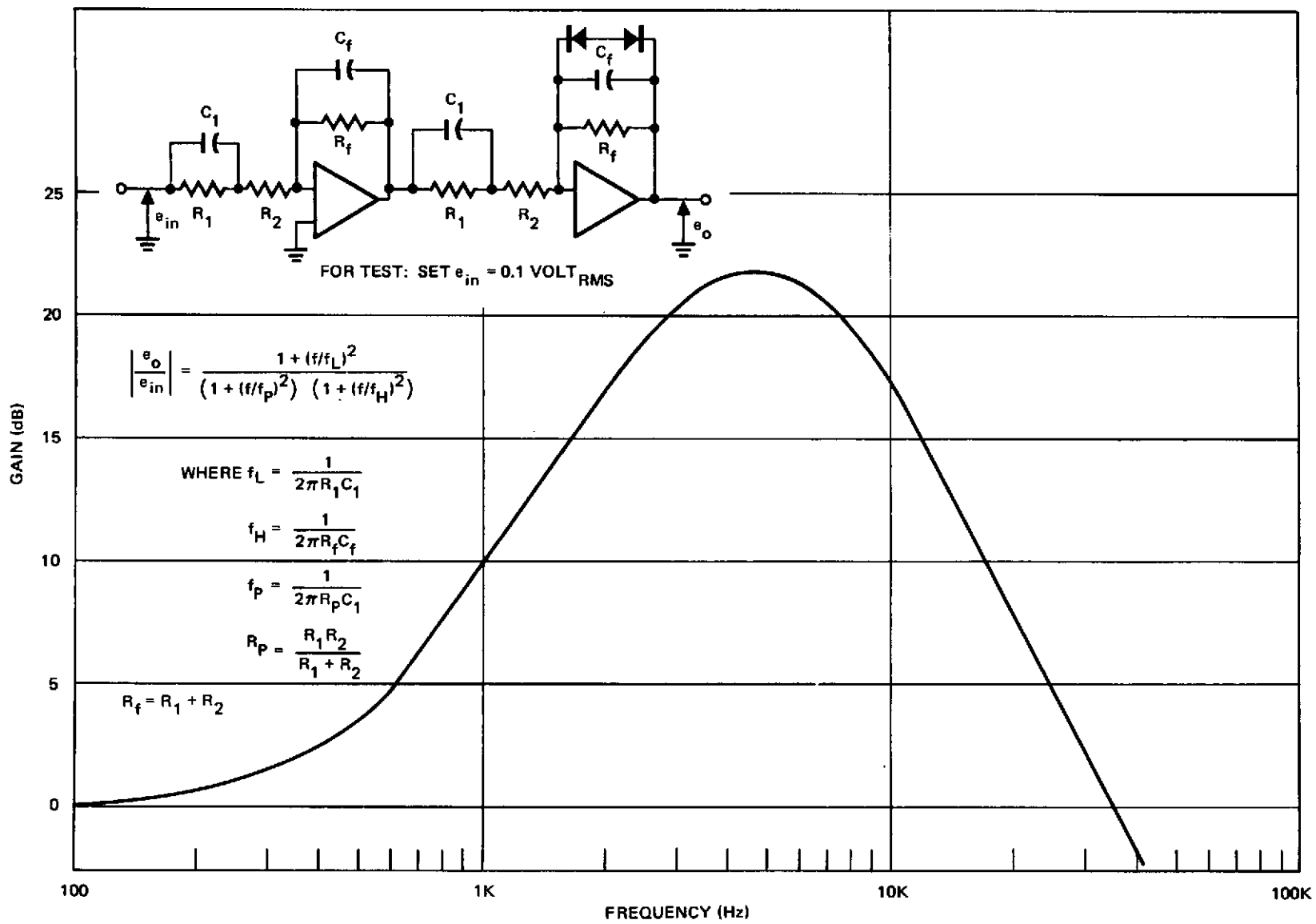


Figure 3-15. Equalized Frequency Response

rise of the equalizer to +12 dB per octave. Since this circuitry is not intended to emphasize components beyond the third formant (approximately 3,600 Hz), it is designed to flatten out at 4.5 kHz and is down 3 dB at 8.5 kHz. This ensures that transients associated with stop consonants do not produce sufficient overshoot to cause the amplifiers to saturate. In addition, zener clamping is provided in the last stage to ensure that the full-scale voltage of the A/D converter associated with the ALC is not exceeded (± 5 volts).

3.4.3 Band Limiting Filter

This filter is a conventional eight-pole Butterworth with a cutoff frequency of 6,727 Hz. This circuitry is contained on board No. 1. The Butterworth was chosen here despite the superior roll-off of the Chebyshev, since the Butterworth has less overshoot and is maximally flat. This circuitry is constructed using a cascade of four universal active filters. (A detailed discussion of these universal active filters is presented later.) The same universal active filter is used for the Butterworth and Chebyshev filters. The four stages are arranged in order of increasing Q to allow maximum input signal without saturation. The Q of each stage, in order, is 0.5098, 0.6013, 0.9000, and 2.5629. The magnitude transfer relation for each stage at the low-pass output is:

$$M = \left| \frac{e_o}{e_{in}} \right| = \frac{1}{\left[(1-u^2)^2 + \left(\frac{u}{Q} \right)^2 \right]^{1/2}}$$

where

$$u = f/f_o$$

Note that when $u = 1$, then $M = Q$. The magnitude response is almost identical to that computed when adjustments are made for Q from the first to the fourth stage in a cascaded manner. The measured frequency response of this filter is as follows:

Frequency (Hz)	M
25	0.996
2,000	0.990
3,000	0.986
4,000	0.982
5,000	0.978
6,000	0.920
6,300	0.860
6,500	0.802
6,727	0.707
7,242	0.500
8,000	0.250
9,000	0.100
11,669	0.010

This filter ensures that the A/D converter in the ALC will not yield aliasing problems for the 25-kHz sampling rate.

Two filters of this type exist. One supplies the input to the ALC; the other, which performs identically, provides smoothing of the ALC D/A output prior to input to the filter network.

3.4.4 Word Boundary Detector

The word boundary circuitry consists of a gain scaling amplifier, full-wave rectifier, resettable integrator, comparator, and three one-shot multivibrators, one of which is retriggerable. A block diagram of this circuitry is given in Figure 3-16. When the retriggerable one shot, termed W_D , fires, the leading-edge signals the beginning of a word. The detector remains active provided no pauses exist in excess of 280 milliseconds during a word. Threshold, θ_{W_D} , may be adjusted as required to satisfy the ambient background; however, the trailing edge of W_D goes beyond, in time, the actual end of a word. The software looks backward from the fall of W_D and considers the end of the word to be where A_g is less than six counts. The word

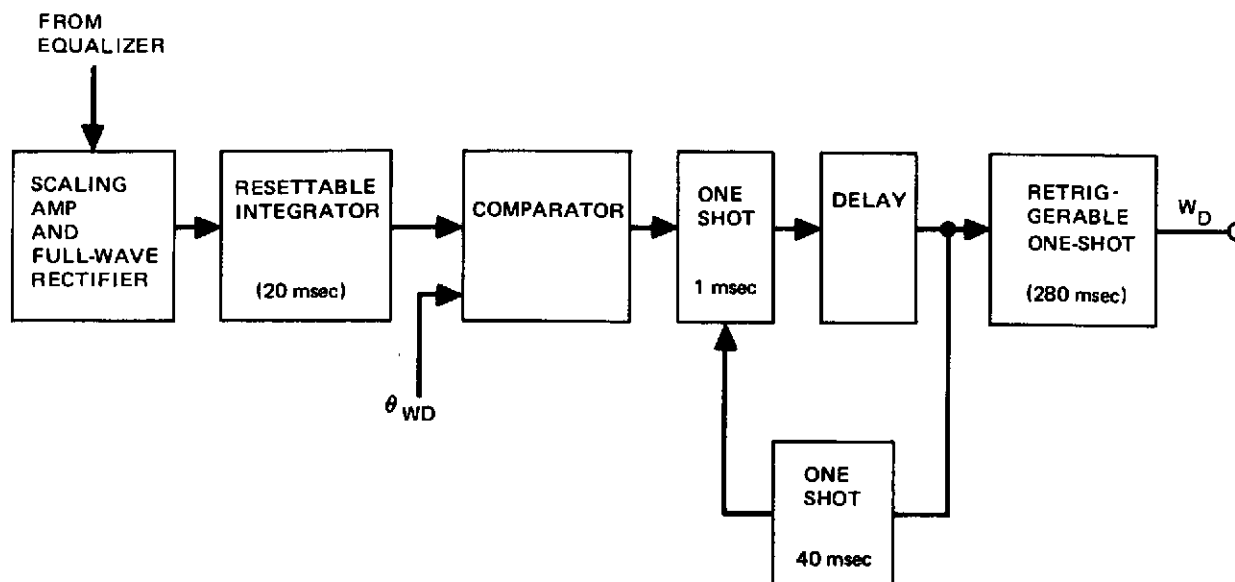


Figure 3-16. Word Boundary Detector Block Diagram

boundary circuitry precedes the ALC, which delays the speech signal by 20.48 milliseconds. Due to the ALC delay, the beginning of speech data, F_1 , F_2 , F_3 , etc., always lags behind the detection of the beginning of a word.

3.4.5 Automatic Level Control Circuitry

A block diagram of the ALC network is shown in Figure 3-17. This circuitry is hand-wired and located on two circuit boards, 2 and 3.

The purpose of this circuitry is to accommodate the dynamic range of speech, normally considered to be 30 dB, to accommodate the variations in speaking level by various talkers (± 9 dB relative to the average), and to accentuate consonants which are from 12 to 20 dB smaller than vowels (References 1, 6, and 7). An increase in consonant amplitude increases the consistency with which the formants of voiced consonants may be detected.

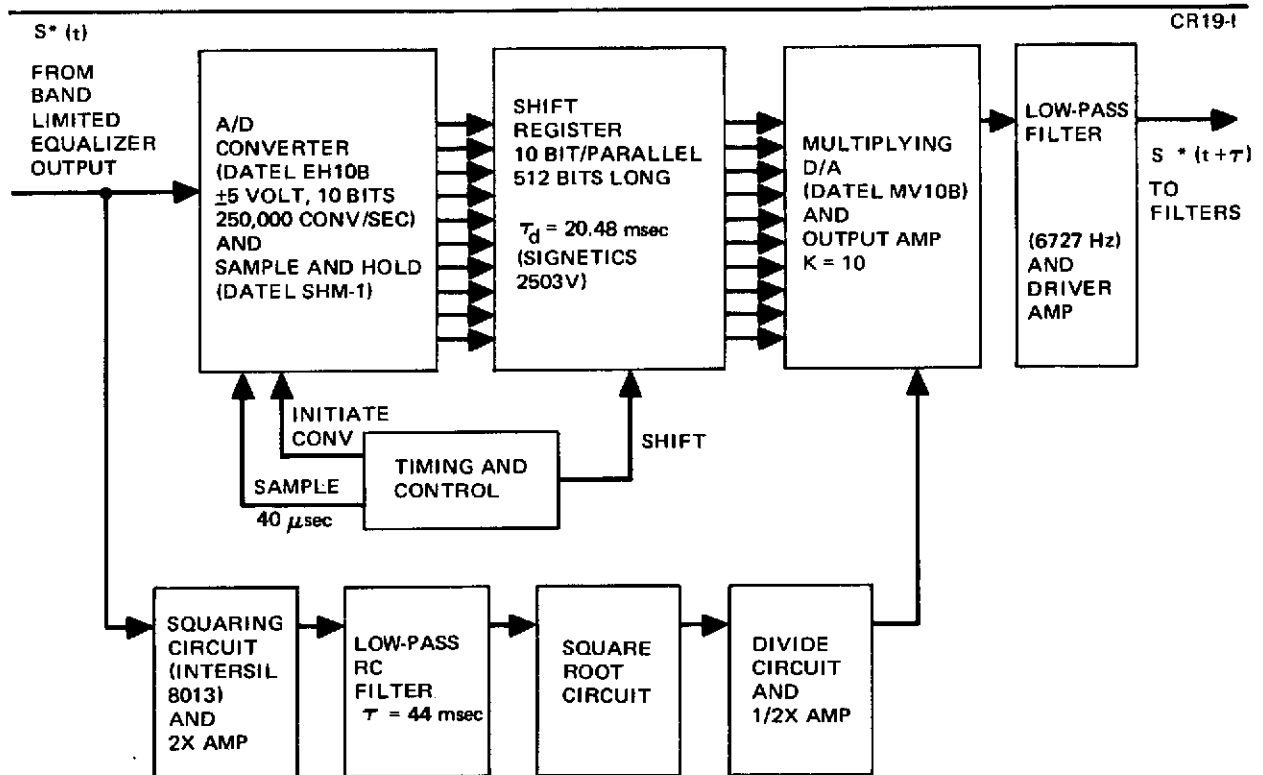


Figure 3-17. Automatic Level Control Circuitry Block Diagram

The ALC circuitry must be fast-acting to accommodate a vowel followed by a consonant. The method selected for automatic level control is to operate on the incoming equalized speech signal according to the relation

$$S_0^*(t + \tau) = \frac{K S^*(t + \tau_d)}{S^*(t)_{\text{RMS}}}$$

Implementation is accomplished using an A/D converter, shift register, and multiplying D/A converter, which delays the incoming speech signal by 20.48 milliseconds. In parallel with this circuitry, the function $1/S^*(t)_{\text{RMS}}$ is generated and input to the D/A. Multiplication is accomplished since the D/A reference voltage is a variable. Note that $S^*(t)_{\text{RMS}}$ has time to be formed prior to multiplication by $S^*(t + \tau_d)$. The A/D sampling rate of 40 microseconds corresponds to a channel bandwidth of 12.5 kHz, well beyond that required for speech. A small constant voltage is supplied to the divider to ensure that division by zero does not occur; hence, the allowable

reference input to the D/A will not be exceeded and the divider circuit saturated. The performance of this circuitry with a sinusoidal input over the circuit bandwidth is given by

$$e_o(t + \tau_d) = \frac{(\sqrt{2}) [2 e_{\max} (t + \tau_d)]}{\sqrt{e_{\text{inRMS}}^2 + K}}$$

where

$$K = 0.0122$$

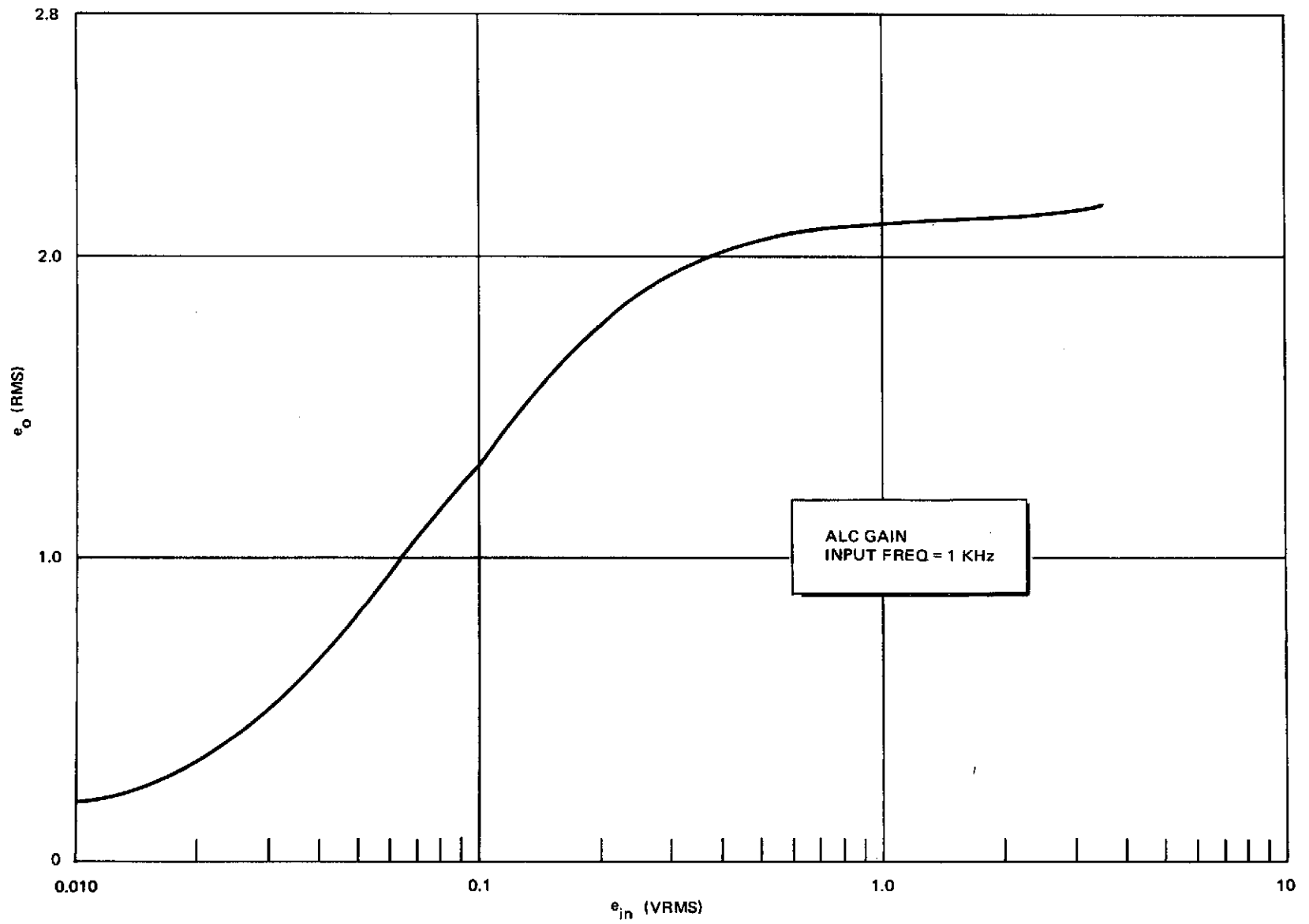
Note that when $e_{\text{inRMS}} = 0.35$ volt, then $e_{\text{inRMS}}^2 > (10 \cdot K)$; hence, the denominator is essentially equal to e_{inRMS} beyond this input voltage.

A graph of the measured performance of this circuitry is shown in Figure 3-18. A 30-dB change in the input from 0.1 to 3.16 volts is accompanied by an output voltage change of 4.6 dB.*

The output of the D/A is followed by a low-pass Butterworth filter which smooths the discrete steps of the D/A output. This filter is identical in design and performance to the 6,727-Hz Butterworth filter which follows the equalizer and has been previously described. The 3-dB bandwidth of the ALC is 6,465 Hz, slightly below the 6,727-Hz cutoff of the filter due to the X10 scaling amplifier preceding the filter and driver amplifier following it.

Proper selection of the averaging time, τ , in generating $S^*(t)_{\text{RMS}}$ minimizes any signal distortion occurring at the beginning or end of the word. A τ too small causes amplitude distortion only at the end of a word, whereas with a τ too large, amplitude distortion occurs only at the beginning. The selected τ of 44 milliseconds minimizes the magnitude of any amplitude distortion and restricts its duration to less than 20 milliseconds.

*At low frequencies, where the equalizer gain is essentially unity (100 Hz), 0.100-volt rms input to the ALC corresponds to a microphone signal of 0.4 mVRMS.



73

Figure 3-18. ALC Gain

3.4.6 Filter Network, Full-Wave Rectifiers, and Difference Amplifiers

The two functional blocks following the ALC are the filter network with associated full-wave rectifiers and the difference amplifiers. The filter network contains 21 low-pass filters having 0.18-dB ripple and using an 8-pole Chebyshev design. The low-pass filters are spaced at 1/4 octave** logarithmic increments covering the range from 176.8 to 5,657 Hz. In addition, a 6,727-Hz, high-pass Butterworth filter is provided. The HP filter is cascaded with ALC 6727 LP. This cascade is rectified to provide A_{VH} . The ALC filter output is rectified to provide A_G . The rectified 596-Hz filter is used for A_L . A total of 23 rectifiers is used. The output from 21 adjacent rectified filters is differenced to generate 20 logarithmically spaced bandpass filters. An additional difference of the rectified 4,757-Hz filter from the rectified 3,364-Hz filter is taken for measurement of A_H . Four additional second differences are formed. The second differences are used in controlling the bandwidth of the formant extraction network at crossover boundaries. A total of 25 differences is formed. Boards 5 through 11 contain the filters, boards 12 and 13 the rectifiers; and boards 15 through 17 the difference amplifiers.

The logarithmic filter spacing provides considerably more detail in the first formant band than either the conventional 300-Hz Sonogram or the Koenig frequency scale. This spacing, due to increasing bandwidth with increasing frequency, allows a relatively fixed formant coding independent of minor pitch variations for the same sound for a single speaker.

Low-pass filter differencing has four major advantages:

- A. Any two members of the filter set may be differenced to yield 210 unique bandpass filters.*
- B. A significant cost-saving is obtained using low-pass filters and difference amplifiers as opposed to the equivalent high pass/low pass combination or bandpass, since additional filters are more expensive than low-cost integrated circuit amplifiers and do not require tuning or alignment.

* Number of combinations of 21 filters taken two at a time
= $21! / (2!)(19!) = 210$

**1/4 octave spacing yields $f_{n+1} = 2^{1/4} f_n = 1.189207 f_n$

- C. Weighting differences provides a method for rejection of any spectrum which is continuous over the spectrum of speech.
- D. Both a bandpass and low-pass filter are available simultaneously.

Low-Pass Filter Discussion

The Chebyshev filter was selected for two reasons: (1) maximum roll-off rate without ripple in the stop band, and (2) unity gain in the passband up to and including the design cutoff within a selectable ripple tolerance. These characteristics yield the closest approximation to an ideal bandpass filter when differenced.

Each filter is composed of four stages. Each stage uses the Kinetic Technology Inc. FS-51 universal active filter to implement each complex pole pair. * A schematic of the circuitry contained by the FS-51 and external components is shown in Figure 3-19. R_1 , R_2 , R_3 , R_Q , and R_B are external

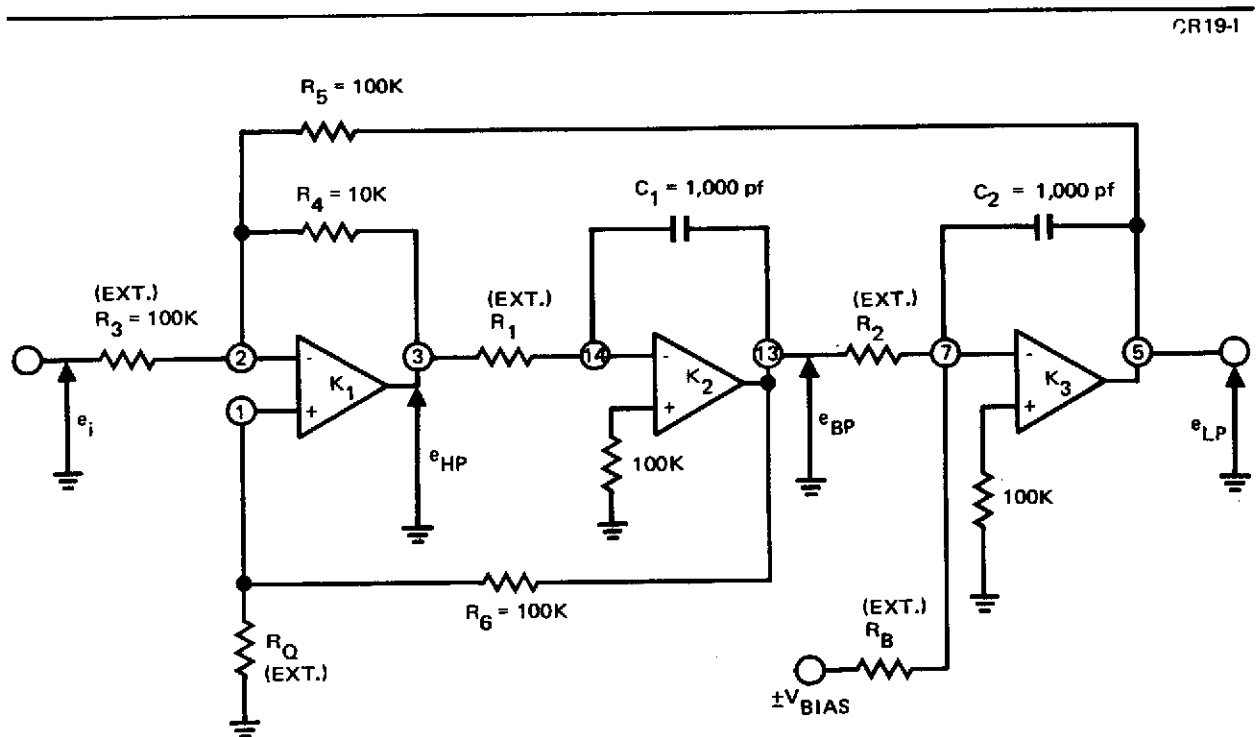


Figure 3-19. Single-Stage Filter Configuration

*These hybrid active filters are also manufactured by Beckman Instruments under the name 821 Universal Active Filter.

components; all other components are contained by the FS-51 hybrid integrated circuit. R_B is used to cancel any offset which may occur, particularly to low-frequency filters where large external resistors are used.

The general transfer function of this circuitry may readily be calculated in the following manner.

Let

$$w_1 = 1/R_1 C_1, w_2 = 1/R_2 C_2, A_1 = R_4/R_5, A_2 = R_4/R_3,$$

and

$$A_3 = R_Q/(R_6 + R_Q)$$

Inspection of the circuit shows that

$$\begin{aligned} e_{BP} &= \frac{-e_{HP}}{SR_1 C_1} = -\frac{w_1 e_{HP}}{S} \text{ and } e_{LP} = -\frac{e_{BP}}{SR_2 C_2} = -\frac{w_2 e_{BP}}{S} \\ &= \frac{w_1 w_2 e_{HP}}{S^2} \end{aligned}$$

The high-pass amplifier output may be expressed as

$$e_{HP} = A_3 e_{BP} + A_2 (A_3 e_{BP} - e_i) + A_1 (A_3 e_{BP} - e_{LP})$$

Expressing e_{HP} and e_{BP} in terms of e_{LP} yields

$$\begin{aligned} \frac{S^2 e_{LP}}{w_1 w_2} &= -\frac{A_3 S e_{LP}}{w_2} - \frac{A_2 A_3 S e_{LP}}{w_2} - A_2 e_i - \frac{A_1 A_3 S}{w_2} e_{LP} \\ &\quad - A_1 e_{LP} \end{aligned}$$

Collecting like terms and rearranging yields

$$\frac{e_{LP}}{e_i} = \frac{-A_2 w_1 w_2}{S^2 + w_1 A_3 (1 + A_1 + A_2) S + A_1 w_1 w_2}$$

Substituting e_{LP} in terms of e_{BP} gives

$$\frac{e_{BP}}{e_i} = \frac{A_2 w_1 S}{S^2 + w_1 A_3 (1 + A_1 + A_2) S + A_1 w_1 w_2}$$

Substituting for e_{BP} in terms of e_{HP} gives

$$\frac{e_{HP}}{e_i} = \frac{-A_2 S^2}{S^2 + w_1 A_3 (1 + A_1 + A_2) S + A_1 w_1 w_2}$$

These general equations take on a simple standard form when it is noted that for the particular application we let

$$w_1 = w_2; w_0 = \sqrt{A_1 w_1 w_2} = \sqrt{A_1} w_1 = \frac{w_1}{\sqrt{10}} = \frac{1}{\sqrt{10} \cdot R_1 C_1}$$

and

$$A_1 = A_2 = 1/10$$

Hence, the low-pass filter transfer function may be expressed as

$$\frac{e_{LP}}{e_i} = \frac{-w_0^2}{S^2 + (1.2)(\sqrt{10}) \cdot A_3 w_0 S + w_0^2} = \frac{-w_0^2}{S^2 + \frac{w_0 S}{Q} + w_0^2}$$

Therefore

$$Q = \frac{1}{(1.2) \sqrt{10} A_3} = \frac{R_6 + R_Q}{(1.2) (\sqrt{10}) (R_Q)} = \frac{R_6 + R_Q}{3.7947 R_Q}$$

In design of any filter, two resistors must be determined, $R_1 = R_2$ and R_Q . These resistors may be selected given f_0 and Q , which are determined by the Chebyshev design, as shown below. Note from the circuitry that $C_1 = C_2 = 1,000$ pf and $R_6 = 100$ K. Hence

$$R_1 = \frac{1}{\sqrt{10}} \cdot \frac{1}{(2 \pi f_0)} \cdot \frac{1}{C_1} = \frac{50329.2}{f_0} \times 10^3$$

$$R_Q = \frac{100 \times 10^3}{(3.7947Q - 1)}$$

Note that at $\omega = \omega_0$, $e_{LP}/e_i = j Q$, $e_{BP}/e_i = Q/\sqrt{10}$, and $e_{HP} = -jQ/10$.

The performance of the individual Chebyshev stages, required cutoff frequency, Q of each stage, and the resulting filter performance are shown in Figure 3-20.

Figure 3-21 shows the measured frequency response of four of the eight-pole low-pass filters plotted in dB. From the graph, it is difficult to determine the frequency locations and magnitude of the ripple in the passband. Measurements of these parameters indicate that the maximum peak of the filter occurs at $(0.566) (f_0)$ and has a magnitude on the average of $+0.34$ dB. The maximum negative valley occurs at $(0.935) (f_0)$ and has an average magnitude of -0.10 dB.

The maximum obtainable roll-off rate of an eight-pole Chebyshev filter is 90.31 dB per octave. To obtain this roll-off requires 3 dB of ripple in the passband. This is an intolerable amount of ripple for the desired application. The compromise ripple selected, 0.18 dB, yields a roll-off rate of 76.3 dB per octave. This is 28 dB per octave better than can be achieved with the Butterworth design.

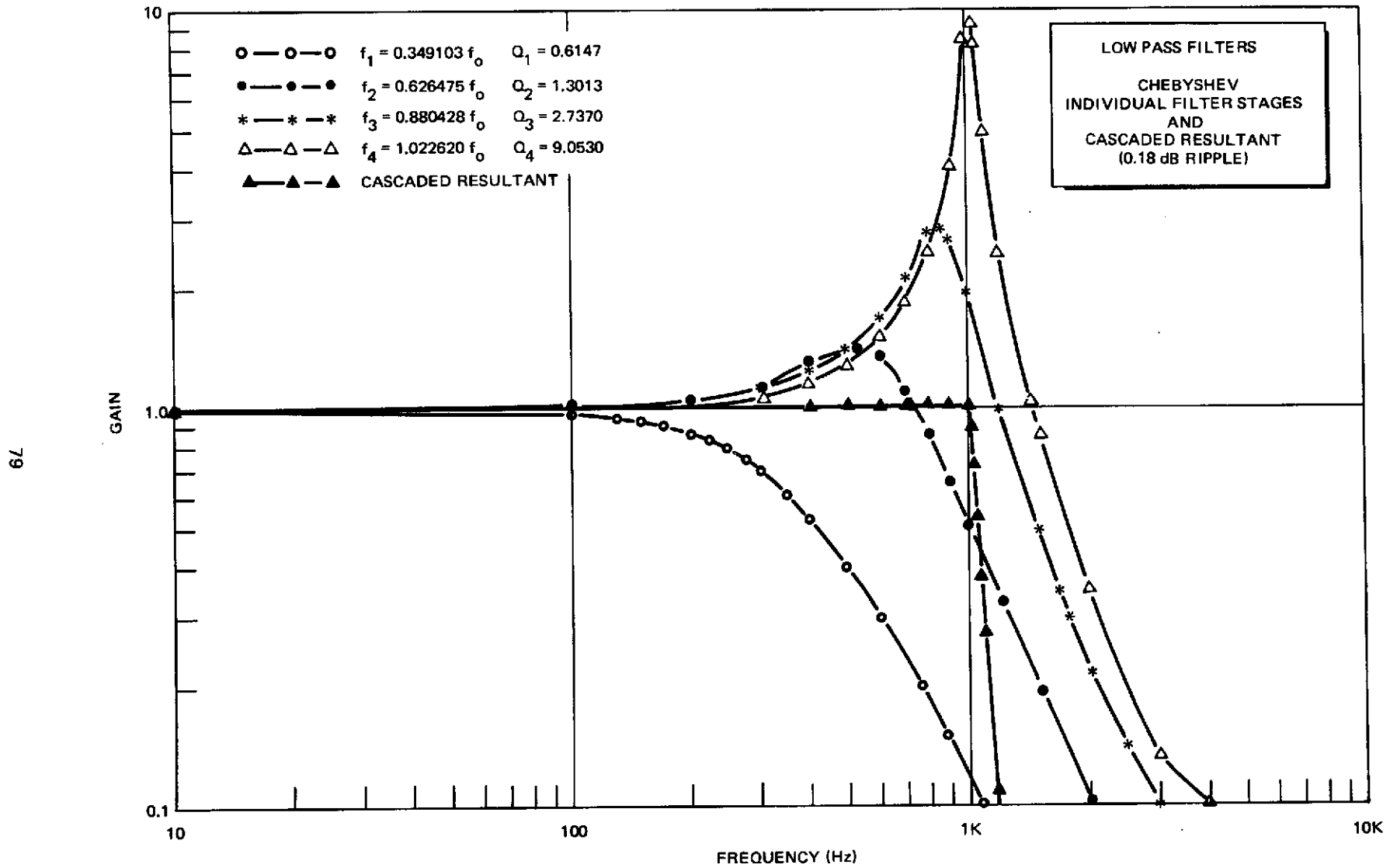


Figure 3-20. Low-Pass Filter

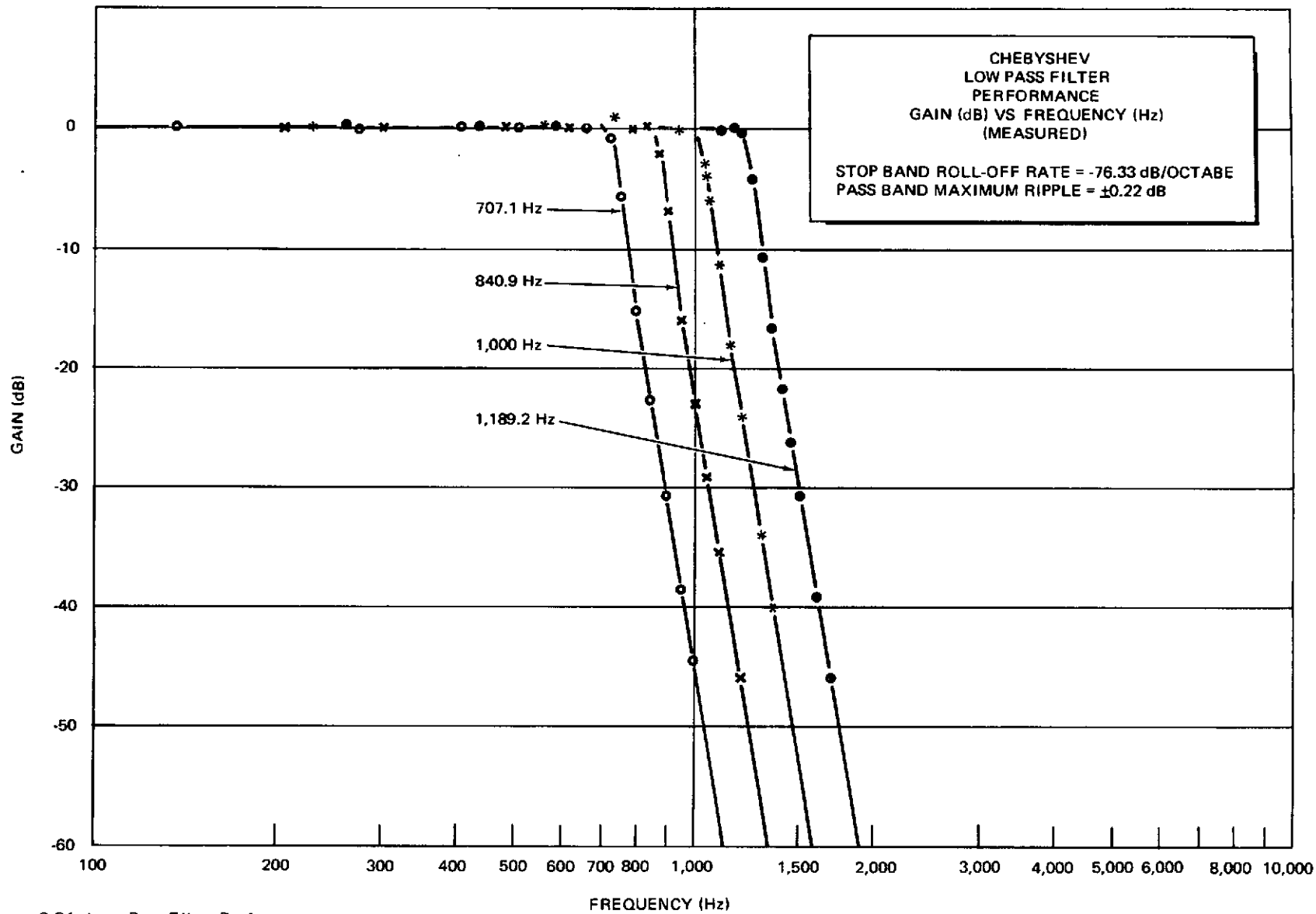


Figure 3-21. Low-Pass Filter Performance

Full-Wave Rectifiers

Each filter is followed by a full-wave rectifier. Full-wave rectification was selected to obtain the maximum average signal for each filter. Each rectifier circuit uses a dual IC operational amplifier. The configuration allows very small signals, as well as large, to receive undistorted rectification over the speech signal spectrum.

Each rectifier averages the rectified filter output by means of a resistance capacitance (RC) network. The RC time constant of the rectifier is 11.2 milliseconds. The time constant of the rectifier associated with A_{VH} is shorter than the standard time in order to accommodate the short-duration fricatives it senses.

Difference Amplifiers

The difference amplifiers yield bandpass filters. The advantages of this approach have been previously described. A disadvantage is that unwanted ripple occurs outside the bandpass in the low-frequency region. A graph is presented in Figure 3-22 of the gain versus frequency of the difference amplifier where

$$D_{n+1} = A_{n+1}(f) - A_n(f)$$

where

A_n and A_{n+1} are the rectifier output signal

and

D is used to denote difference

Notice that positive peak ripple of 2 percent can occur. Positive ripple can hamper detection of the formants. For this reason, the system weighs $A_n(f)$ by an additional 4 percent; that is,

$$D_{n+1} = A_{n+1}(f) - 1.04 A_n(f)$$

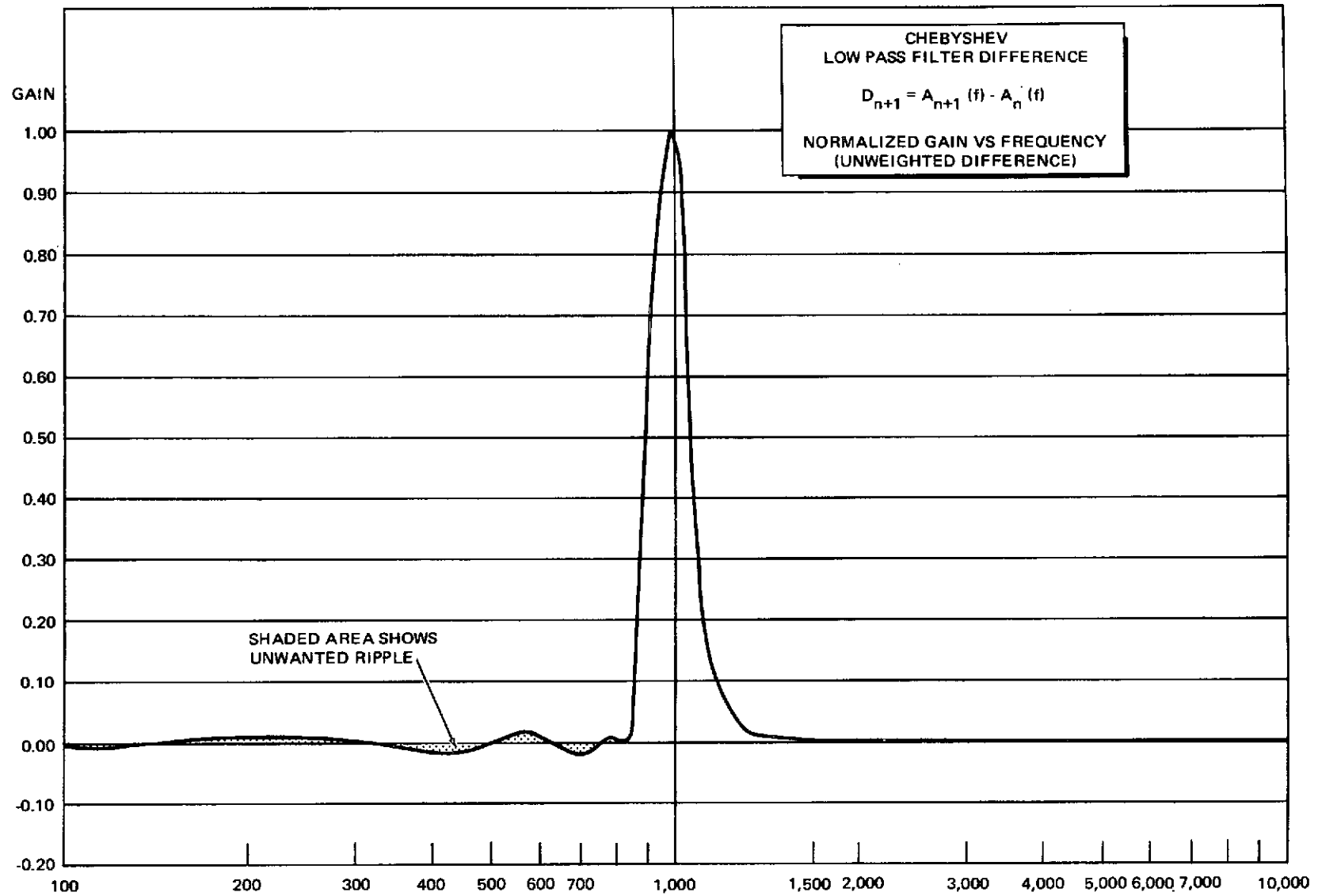


Figure 3-22. Low-Pass Filter Difference

Adjacent differences cross over at approximately 6 dB as shown in the next graph (Figure 3-23). This crossover may be calculated from the fact that at the crossover

$$D_{n+1} = D_n$$

The rising portion of $D_{n+1} \cong 1 - A_n(f)$, while the falling portion of D_n in this frequency region is approximately $A_n(f)$. Hence

$$1 - A_n(f) = A_n(f)$$

$$A_n(f) = 1/2$$

or

$$20 \log_{10} A_n(f) = 20 \log_{10} 1/2 = -6 \text{ dB}$$

The second graph of the difference (Figure 3-23) is plotted with dB vs frequency. This graph shows roll-off rates in dB as well as the overlap and crossover of adjacent difference amplifiers.

Each difference amplifier provides additional smoothing of the filter data by means of an RC network. The RC time constant of the difference amplifiers is 24 milliseconds.

Each difference amplifier has a gain of 5.1. This enhances detection of weak formants, but is chosen so that even if the ALC output is a single frequency with maximum output ($2 v_{RMS}$), the difference amplifiers will not be saturated.

Additional Filters

The only filters not discussed in detail are those associated with A_V and A_{VH} . The frequency response of these two filters is shown in the following two graphs: A_H is shown in Figure 3-24 and A_{VH} in Figure 3-25.

84

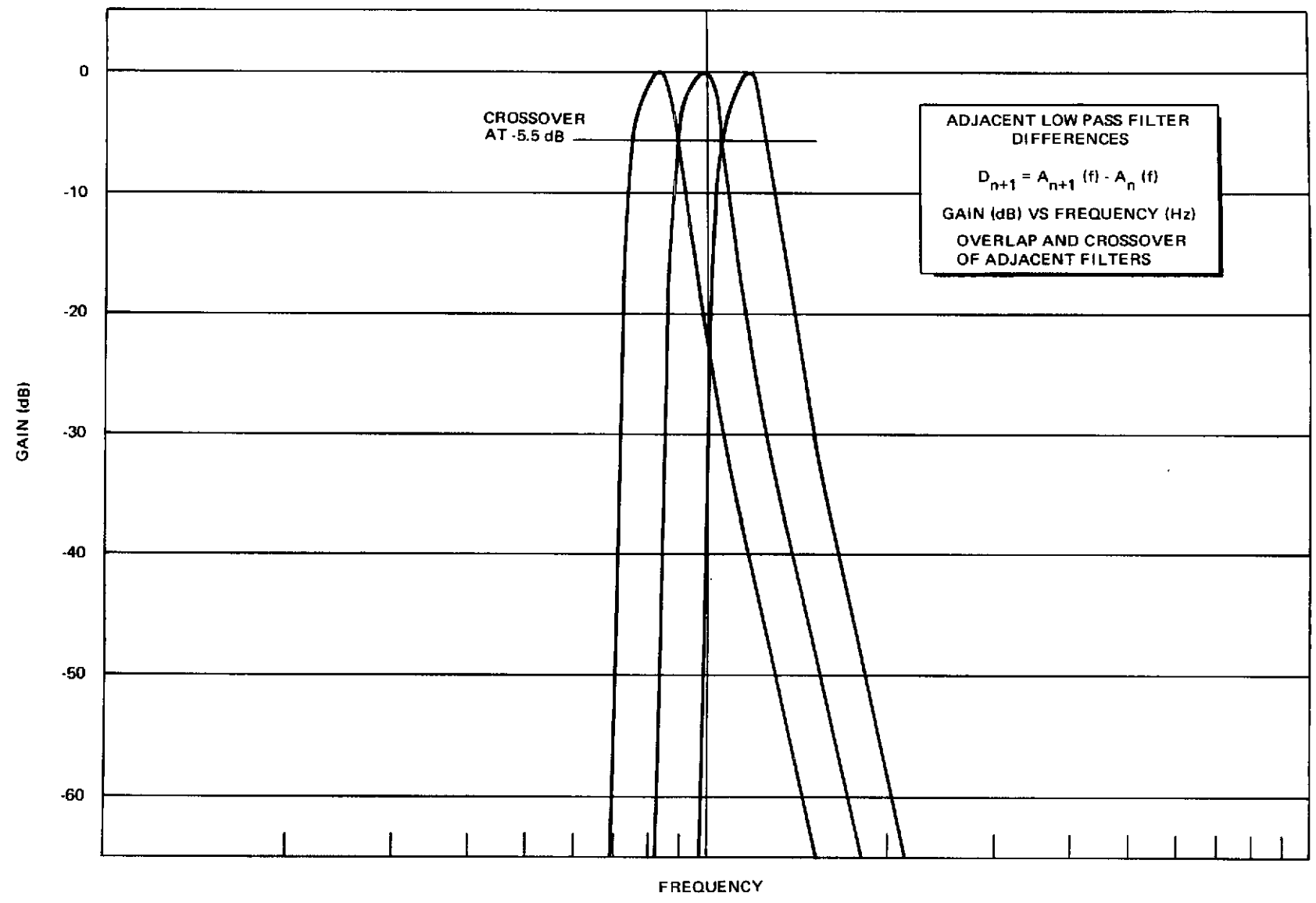
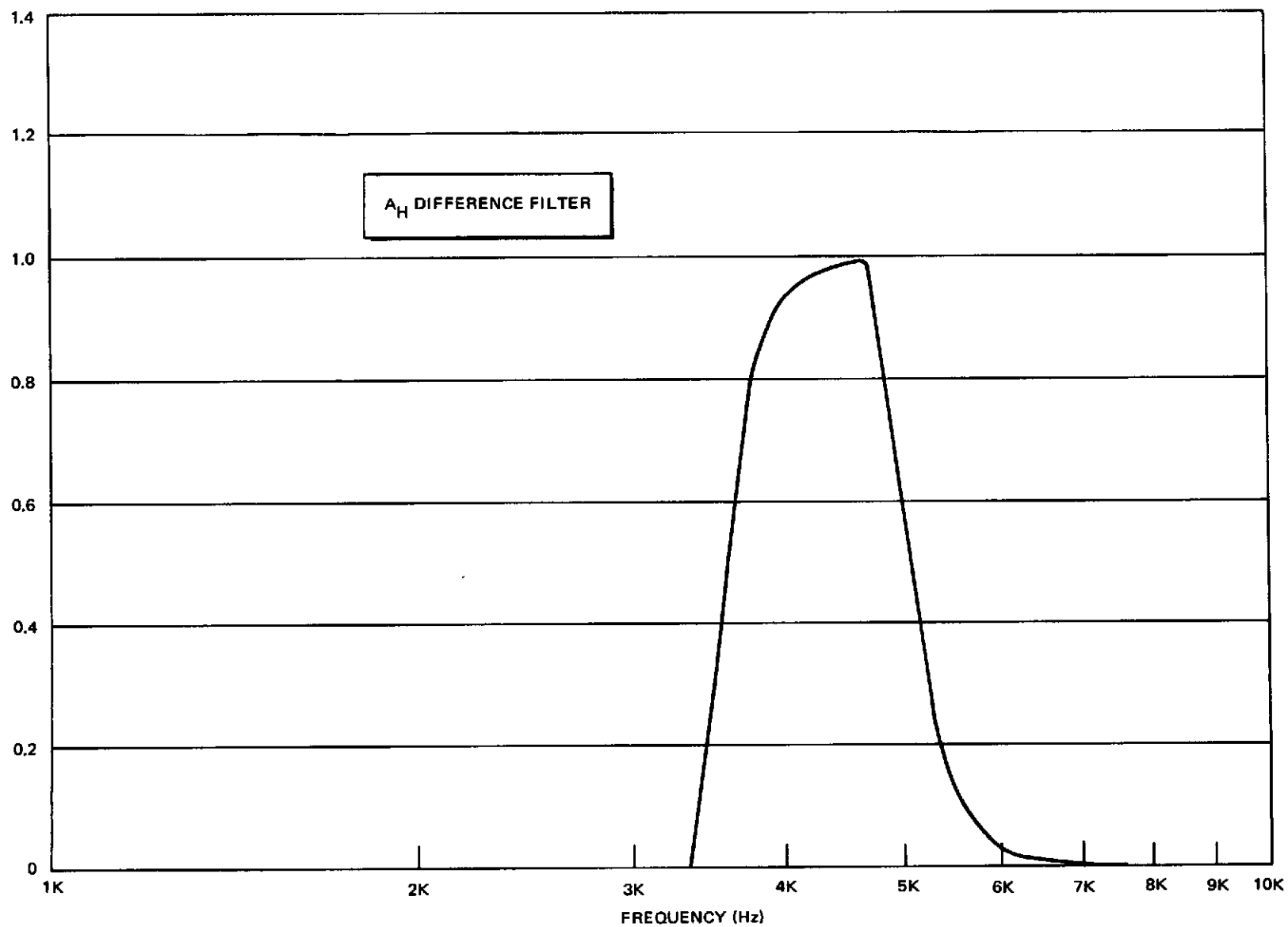
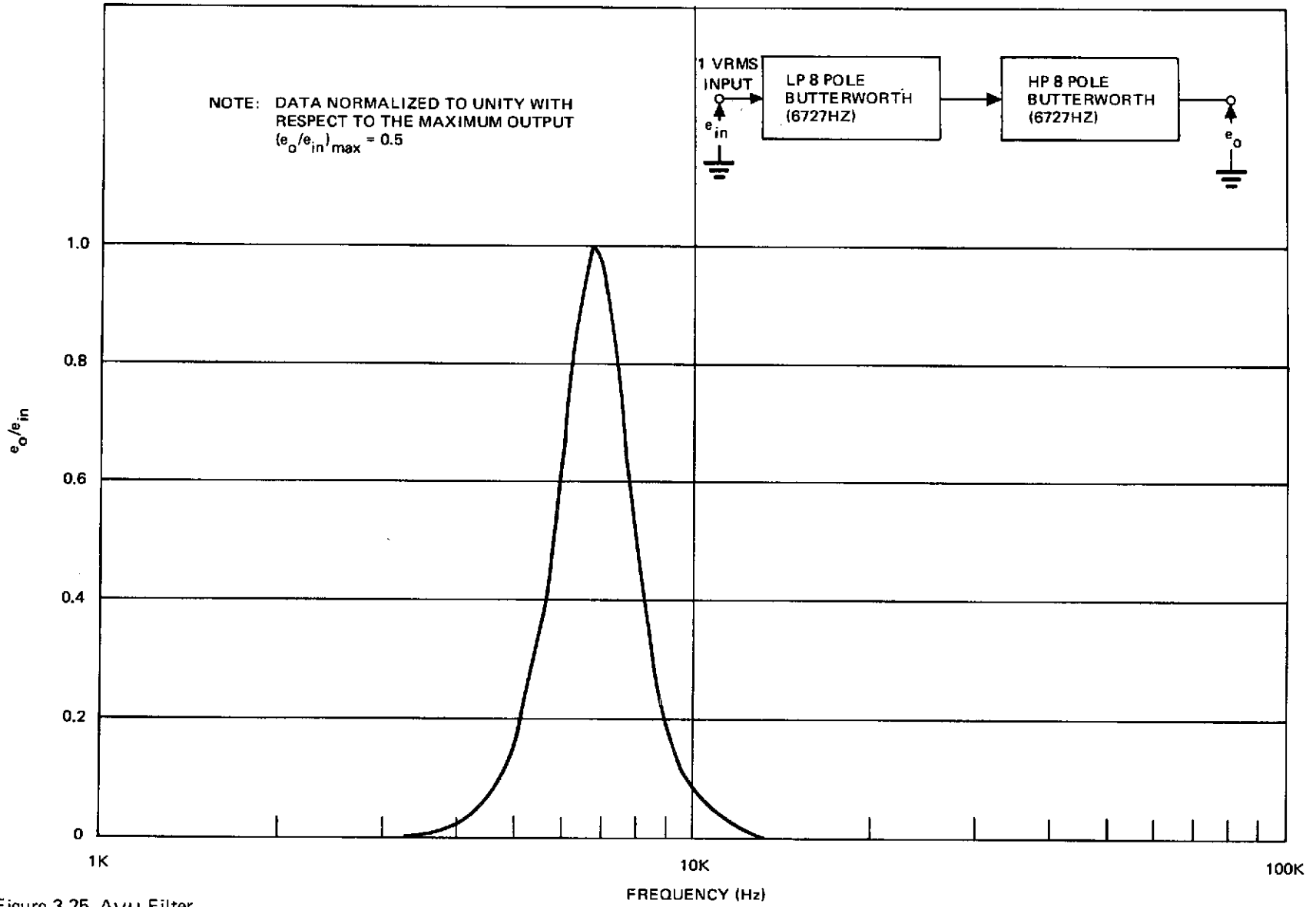


Figure 3-23. Adjacent Low-Pass Filter Differences

Figure 3-24. A_H Difference Filter



98

Figure 3-25. AVH Filter

3.4.7 Frequency Range Multiplexer

This circuitry contained by board 21 receives its input from 20 difference amplifiers. The output of this network is 19 analog signal voltages which are operated on by the formant extraction circuitry. The purpose of these multiplexers is to allow 1/4-octave translations of the difference amplifier filter bands which input to the formant circuitry. Analog FET switches are used to accomplish the multiplexing. Currently, control of the range switching is accomplished manually via a three-position rotary switch on the front panel. The table below shows how the switch positions may be used to accommodate a wide variety of speakers.

Switch Position	Effective Filters to Formant Network	Formant Range		
		F ₁ (Hz)	F ₂ (Hz)	F ₃ (Hz)
Low	1 to 18	185 to 890	750 to 2,525	1,505 to 3,580
Medium	2 to 19	225 to 1,060	890 to 3,010	1,790 to 4,260
High	3 to 20	270 to 1,265	1,060 to 3,580	2,120 to 5,110

The following table shows the mean fundamental and formant frequencies for 33 men, 28 women, and 15 children obtained for 10 vowel sounds as measured by Peterson and Barney in Reference 8.

	Average	Men (Hz)	Women (Hz)	Child (Hz)	Ratio	
					Women/Men	Children/Women
Region of Interest	F ₀	132.2	223.0	264.3	1.686	1.185
	F ₁	502.0	575.0	671.0	1.145	1.167
	F ₂	1,420.0	1,694.0	1,928.0	1.192	1.138
	F ₃	2,386.0	2,783.0	3,266.0	1.166	1.174

When the ratio between men and women and the ratio between children and women is compared to $1/4$ octave = 1.1892 translations, it is apparent that the system is nearly matched for accommodations of the total set by use of the range multiplexer.

3.4.8 Formant Extraction Network

This network segments the speech spectrum (Reference 3) into three bands, detects the maximum filter amplitude in each segment, and produces a digital code indicating the largest filter amplitude in each segment. The segment boundaries are allowed to overlap; that is, the end of F_1 is also contained by the beginning of F_2 , while the end of F_2 is contained by the beginning of F_3 . The formants are found in sequence; F_1 is followed by F_2 , which is followed by F_3 . If the end of the F_1 segment is selected, the beginning of F_2 is raised; similarly, the beginning of F_3 depends upon finding F_2 first. The maximum counts for each segment are 9, 7, and 5, corresponding to F_1 , F_2 , and F_3 .

A maximum in any segment may be detected provided any input to the segment is greater than any other member of the segment by any voltage from 2 millivolts to 12 volts. This is a dynamic range of approximately 56 dB.

Each network has a threshold. No formant is detected unless at least one member of the segment exceeds this threshold. If no member exceeds the threshold, a zero code occurs at the output.

To ensure that consistent bandwidths are obtained for each segment, second-difference filters are employed at the end of each segment and the beginning of the next. It is not necessary to provide this circuitry at the beginning of F_1 .

Each network has a storage register to retain the last formant count while a new count is being determined. Each formant network compares the newly generated count against the stored count to determine if a change has occurred. Data may be transmitted only on changes for compression purposes or may be sampled at the end of each sequence. The latter mode is currently used by the word recognition system software.

When voicing occurs and F_2 does not exceed the threshold, $F_1 = 9$ fill logic causes $F_2 = 1$. Also, when voicing occurs and $F_2 = 5$, this causes $F_3 = 1$ and $F_2 = 6$ causes $F_3 = 2$, while $F_2 = 7$ causes $F_3 = 3$. This fill logic is used as voiced sounds usually contain at least three formants; however, when any two formants occupy a boundary region and are closely spaced, this information would be lost due to the boundary switching network.

The formant circuitry is contained on boards 21, 22, and 23, respectively, for F_1 , F_2 , and F_3 . The fill logic is contained on board 25. Each formant network is capable of responding accurately in less than 200 microseconds, although with the current 8-millisecond sampling rate, 1.6 milliseconds are allocated for detection of the maximum in each segment.

The following table shows the allowed count range of F_2 as a function of F_1 and the allowed count range of F_3 as a function of F_2 . These ranges are dependent on the boundary switches. In addition, this table shows the operation of the fill logic for F_2 as a function of F_1 and F_3 as a function of F_2 . The ranges shown do not include the zero count.

F_1 Count	Allowable F_2 Range	F_2 Fill if $F_2=0$	F_2 Count	Allowable F_3 Range	F_3 Fill if $F_3 = 0$
1	1 to 7	None	1	1 to 5	None
2	1 to 7	None	2	1 to 5	None
3	1 to 7	None	3	1 to 5	None
4	1 to 7	None	4	2 to 5	None
5	1 to 7	None	5	3 to 5	1
6	1 to 7	None	6	4, 5	2
7	1 to 7	None	7	5	3
8	2 to 7	None			
9	3 to 7	1			

The next table lists the frequency range, corresponding formant code, and the bandwidth associated with each code. This table is a function of the selected frequency range, LOW, MED, or HIGH. These measured data are obtained using two sinusoidal generators of equal amplitude summed together as the input to the system.

Frequency Range (Hz)			F ₁ Count	F ₂ Count	F ₃ Count	Bandwidth (Hz)		
Low	Medium	High				Low	Medium	High
185-225	225-267	267-318	1			40	42	51
225-267	267-318	318-376	2			42	51	58
267-318	318-376	376-445	3			51	58	69
318-376	376-445	445-532	4			58	69	87
376-445	445-532	532-633	5			69	87	101
445-532	532-633	633-750	6			87	101	117
532-633	633-750	750-890	7			101	117	140
633-750	750-890	890-1060	8			117	140	170
750-890	890-1060	1060-1265	9	1		140	170	205
890-1060	1060-1265	1265-1505		2		170	205	240
1060-1265	1265-1505	1505-1790		3		205	240	285
1265-1505	1505-1790	1790-2120		4		240	285	330
1505-1790	1790-2120	2120-2525		5	1	285	330	405
1790-2120	2120-2525	2525-3010		6	2	330	405	485
2120-2525	2525-3010	3010-3580		7	3	405	485	570
2525-3010	3010-3580	3580-4260			4	485	570	680
3010-3580	3580-4260	4260-5110			5	570	680	850

3.4.9 Amplitude Digitization and Interface Logic

Four amplitudes, A_G, A_L, A_H, and A_{VH}, are all digitized, each to six bits. Each digitizer is preceded by a scaling amplifier to ensure that all signals are in the range from 0 to 10 volts and to provide offset biasing to reject background noise. The scaling amplifiers and digitizers are located on board 28. All converters are simultaneously requested to start conversion. The conversion time is 50 microseconds maximum following the command to start. This relatively fast conversion rate with data changing at a rate less than 5 Hz eliminates the need for a sample and hold preceding the A/D converters.

Each formant circuit board also has provisions for digitally encoding formant amplitude within 50 microseconds following registration of the formant frequency counts.

Interface logic is provided on board 30 to allow digital data multiplexing. The digitized amplitude data are input to the PDP 11/40 via DR11C-2 in two 12-bit blocks using the two-interrupt lines REQA-2 and REQB-2. The sequence of this data entry is A_G and A_L (12 bits) followed by A_H and A_{VH} (12 bits).

3.4.10 Voiced/Unvoiced Detection

The voiced/unvoiced detection circuitry is designed using three criteria. Voicing is indicated if (1) the speech signal contains energy in the region from 60 to 290 Hz; (2) the signal in the 60 to 290 Hz region is periodic — that is, in a time interval of 60 milliseconds, four positive zero crossings must occur; and (3) the energy in the A_L band (approximate 0 to 600 Hz) must be greater than a threshold, θ_L . The value of θ_L chosen is approximately equal to one-seventh the maximum value of A_L during each word. Alternately, this comparison could be based on A_L greater than a fraction of A_G , and hence this decision becomes relative rather than absolute.

This detector is well synchronized with any formant activity in F_1 or F_2 , since the energy in A_L comparison follows the ALC while frequency and periodicity are sensed prior to the ALC.

3.4.11 Timing and Control

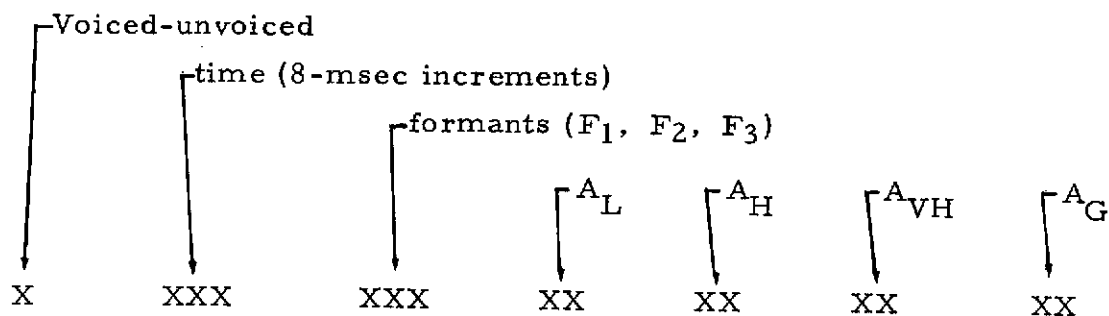
Timing and control are provided to allow the sequential boundary selection for F_2 and F_3 , strobe the final formant values into a holding register, start conversion of the amplitude digitizers, and send interrupts to the minicomputer. Interrupts are sent to the computer only if the word duration, W_D , has been detected. One additional interrupt is given when the word duration detector fails. Control is also provided to ensure that the three interrupt lines, REQB-1, REQA-2, and REQB-2, occur in that order. Timing and control is contained primarily on board 27; however, REQ lines are generated on boards 25 and 30. The 8-millisecond sample interval may be adjusted by modifying the clock circuitry.

3.4.12 Display

The display used by the system is simply an array of 40 light-emitting diodes (LED's). These diodes are driven by inverter/drivers contained on boards 26 and 29. The display allows monitoring of all of the system output data. The LED's associated with F₂ count 8 and F₃ count 8 should never be active unless a hardware failure occurs; hence, only 38 LED's are needed to display the output data. This display has been found to be a valuable aid in adjusting analyzer thresholds to meet a given environment and for detecting hardware malfunctions.

3.4.13 Analyzer Output Data

Figures 3-26 through 3-28 represent the actual data as received by the word recognition system computer from the speech analyzer. The graphs show a time versus frequency plot of the formants where frequency is in terms of formant bands and time (the last three lines) is in 8-millisecond increments. The numbers which depict the formants are the amplitude measures A_L for F₁, A_H for F₂, and A_{VH} for F₃, normalized from 0 to 9 by the largest occurrence of each amplitude measure. The actual sampled amplitude, A_G, is represented in the first two horizontal lines under the graphs. Below these two lines is the voicing indication represented by a "V" for each time sample when voicing was present. The three two-digit numbers at the lower right of the graphs indicate, from top to bottom, the maximum amplitudes over the entire word of A_{VH}, A_H, and A_L which were utilized in normalizing the amplitudes in the graphs. The raw data below the graphs were utilized to produce the graphs and are recorded as follows



3.5 WORD RECOGNITION SYSTEM SOFTWARE

This section covers the software recognition approach utilized by this discrete word recognition system. The main subjects to be covered are prototype or template generation, word classification, and adaptive training.*

3.5.1 Data Compression and Feature Extraction For Template Generation

The speech analyzer measurement set for each word in the vocabulary must reside in a digital computer memory for comparison with all incoming words. To facilitate this task and to reduce data storage, the measurement set is reduced to a matrix, or template, of elements which correspond to a given word. This template is the only representation of a word and is an important part of the recognition system.

3.5.1.1 Time Normalization and Data Smoothing

To reduce every word to a fixed matrix of data points, a time normalization scheme is utilized which tries to overcome the inherent variability in the time length of a word as well as to reduce the word to a standard length for computational purposes. This was accomplished by using a linear time-base.

This linear time-base approach consists of dividing the sampled raw data over the duration of a word into a fixed number of equal time intervals. The number of these time windows per word may be 16, 24, or 32, as selected by the user. All the raw data within each normalizing sample window (including the boundaries) are then averaged together to produce a mean measurement value for each window, reducing the effect of any particularly noisy sample point and smoothing the data, which provides a degree of time alignment. The tabular data presented below indicate the number of points contained by a window for various minimum, typical, and maximum word lengths and the dependence of the number of raw data points per window on the selected normalized sample per word.

*A significant portion of this section is also described by Reference 4.

Word Length (msec)	Raw Speech Samples	Normalized Samples/Word	Average Sample Points/Window
Minimum 400	50	16	3.1250
Average 680	85	16	5.3125
Maximum 1,000	125	16	7.8125
Minimum 400	50	24	2.0833
Average 680	85	24	3.5417
Maximum 1,000	125	24	5.2083
Minimum 400	50	32	1.5625
Average 680	85	32	2.6562
Maximum 1,000	125	32	3.9062

3.5.1.2 Formant Data

The formant information is stored directly following the time-normalization data. Due to the number of filter outputs for each formant band, six bits are allotted for each time-normalized data point in formant 1, and five bits for each time-normalized data point in formants 2 and 3. This yields exactly one PDP 11/40 16-bit word. All formant information is stored with a B value of 2: that is, two binary bits to the right of the integer portion of a formant value.

3.5.1.3 Amplitude Feature Extraction.

These features depend on the ratio of A_L/A_G , A_H/A_G , and A_{VH}/A_G . The utility of these features will be apparent when considered with respect to the Sonogram in Figure 3-29.*

The first feature is called AMI, and attempts to separate certain unvoiced fricatives such as /f/ as in "four" from the sibilants** and voiced sounds using the following binary decision. [In the following, K = time-normalized samples per word (16, 24, 32).]

*This Sonogram is taken from a paper published by G. L. Clapper, Automatic Word Recognition, IEEE Spectrum, August 1971.

**There are six sibilants. Three are unvoiced: /s/ as in "six," /ʃ/ as in "she," and /tʃ/ as in "church," and three voiced: /z/ as in "zoo," /dʒ/ as in "judge," and /ʒ/ as in "azure."

$$AM1(i) = \begin{cases} 1 & \text{if } \frac{A_{VH}}{A_G} \gg \frac{A_H}{A_G} \text{ and } \frac{A_{VH}}{A_G} > \frac{A_L}{A_G} \\ 0 & \text{otherwise} \end{cases}$$

$i = 1, \dots, K$

From the Sonogram, this is most certainly the case for /f/ where AM1 would equal "1."

The second amplitude feature, AM2, determines when voicing is probably present for the nasals as well as for the vowels and other voiced sounds.

$$AM2(i) = \begin{cases} 1 & \text{if } \frac{A_L}{A_G} > \frac{A_H}{A_G} \text{ and } \frac{A_L}{A_G} > \frac{A_{VH}}{A_G} \\ 0 & \text{otherwise} \end{cases}$$

$i = 1, \dots, K$

Note from the Sonogram for the word /wuhn/, that is "1," that the /n/ sound apparently has two formants in the first formant band. This is likely to cause noisy formant data from the speech analyzer. In addition, formant band 2 is apparently missing. The presence of AM2 and F2 and possibly the absence of F1 may indicate a nasal sound as in the /n/ in seven.

CR19-1

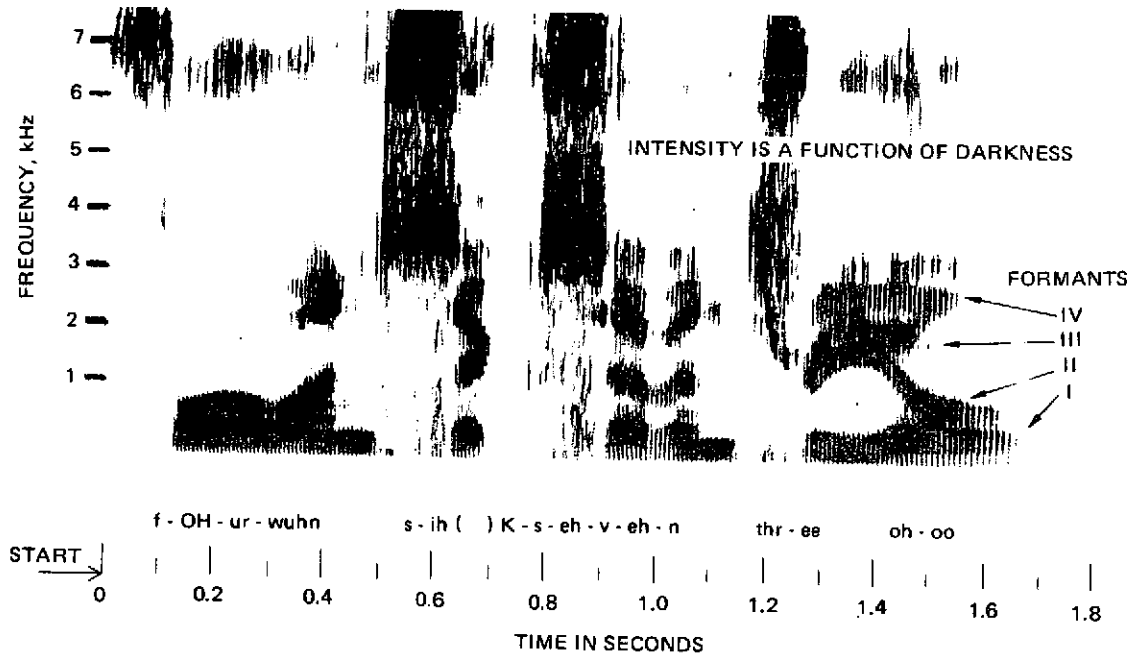


Figure 3-29. Sonogram of Digits (Male Speaker)

A third feature, AM3, attempts to separate some of the sibilants as the sound /ʃ/ from /s/.

$$AM3(i) = \begin{cases} 1 & \text{if } \frac{A_H}{A_G} \gg \frac{A_{VH}}{A_G} \text{ and } \frac{A_H}{A_G} > \frac{A_L}{A_G} \\ 0 & \text{otherwise} \end{cases}$$

$i = 1, \dots, K$

In general, /s/ has a very strong energy content around 7.6 kHz while /ʃ/ has peak energy concentration around 4.0 to 5.0 kHz (Reference 9). AM3 is likely to be "1" for the sound /ʃ/ as in shack where AM3 would be "0" for the /s/ in sack.

The fourth amplitude feature, AM4, determines whether a sibilant is most likely present.

$$AM4(i) = \begin{cases} 1 & \text{if } \frac{A_{VH}}{A_G} \gg \frac{A_L}{A_G} \text{ and } \frac{A_H}{A_G} \gg \frac{A_L}{A_G} \\ 0 & \text{otherwise} \end{cases}$$

$i = 1, \dots, K$

Features AM1 through AM4 are stored for each sample window as a one or zero and carry an additional two bits of precision for purposes of averaging. Hence, these four binary features require 12 bits of storage for each sample window.

3.5.1.4 Voicing

The next block of data extracted from each sample window is the voiced-unvoiced structure. As discussed previously, the speech analyzer determines if the sound is voiced at each sample interval. In addition, as the sampled data is brought into the computer, and before time-normalization, the ratio of A_L to A_{VH} and A_H is calculated for each sample and if A_{VH} or A_H is greater than A_L , then a voicing indicator - if present - is removed if it is not preceded by another voiced sample. And, at the end of a string of voiced samples, if A_L is greater than A_H or A_{VH} , then a voicing indicator is added to the following data samples until finally A_L is less than A_H or A_{VH} . This software editing seems to make the voicing indication more reliable. The voicing indicator is also stored as a binary bit with two extra bits for precision (three bits total per window).

3.5.1.5 Word Length

The final measurement in the creation of the word template is the time length of the discrete word. This time length represents the number of 8-millisecond intervals — the rate at which the speech signal is sampled by the speech analyzer — which occur from the instant when A_G first rises above threshold until it finally falls below the same quiescent threshold. Information thus obtained is stored as an 11-bit value. The least significant two bits are considered to be the fractional portion of this number.

3.5.1.6 Data Storage Format

The data generated from each time-window during time normalization and its format are shown in Figure 3-30.

3.5.2 Classification

Due to the necessity of a near-real-time algorithm for the recognition of discrete words, the classification software is relatively simple but effective for the recognition of 100 different word classes. The classification scheme basically consists of the recognition process of template matching.

Template matching is performed only on four measurements of the word: (1) formants; (2) the feature amplitudes AM1, AM2, AM3, and AM4; (3) the voiced-unvoiced structure; and (4) the time-length of the word. These measures have proved adequate to achieve a reasonable level of accuracy.

The formant comparison between the input word's template and a stored template is a simple first-order distance measure calculated as follows:

$$C_1(K) = \sum_{j=1}^k \sum_{i=1}^3 |F_{i,j} - F_{i,j}^{\wedge}|^*$$

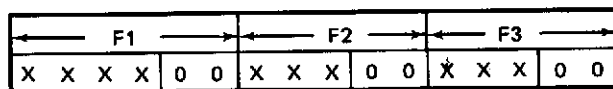
The voiced-unvoiced structure comparison is merely a weighted binary decision as follows.

* In the equations, "i" corresponds to the three formants, "j" corresponds to the sample window number, " \wedge " refers to the sorted template data, "V" refers to the voiced-unvoiced structure data, "K" corresponds to the template index, "W" corresponds to a static weight assignment, and "k" refers to the sample windows per word.

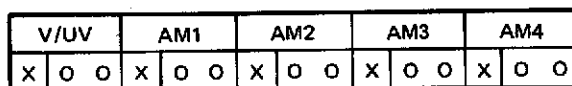
DATA STORAGE FORMAT

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0

FORMANT DATA

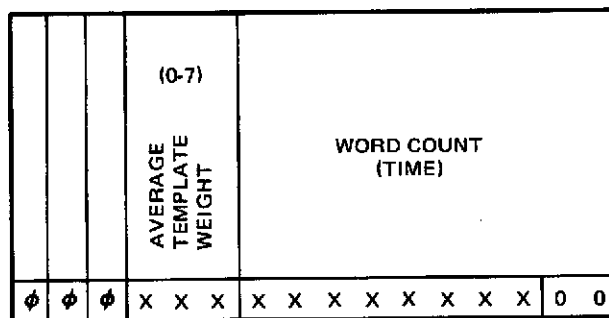


AMPLITUDE AND VOICE DATA



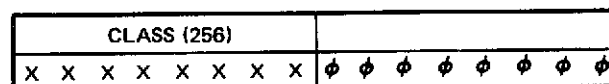
DATA STORAGE/NORMALIZE TIME SAMPLE

TIMING AND WEIGHTING DATA

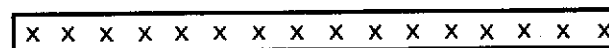


TIMES SAMPLES/WORD	TEMPLATE STORAGE/WORD
16	34
24	50
32	65

IDENTIFICATION



TEMPLATE INDEX NO.



DATA STORAGE/WORD

TEMPLATE STORAGE/WORD = 3 + 2 * SAMPLES/WORD

- X → INTEGER DATA
- 0 → FRACTIONAL DATA
- φ → UNUSED

101

Figure 3-30. Data Storage Format

$$C_2(K) = \sum_{j=1}^k \left| V_j - \hat{V}_j \right| W_5$$

The third comparison to be made concerns the four amplitude measures AM1 through AM4. It also consists of a weighted binary decision function.

$$C_3(K) = \sum_{L=1}^4 \sum_{j=1}^k \left| AM_L(j) - \hat{AM}_L(j) \right| W_L$$

The final comparison performed concerns the length of the word. For all its simplicity, it is an important measure.

$$C_4(k) = \frac{1}{\hat{T}} \left| T - \hat{T} \right| W_6$$

Having performed all of these measures in sequence, the next operation is their summation, providing a simple linear distance measure between the input word and each sorted template.

$$C_T(K) = C_1(K) + C_2(K) + C_3(K) + C_4(K)$$

Completing the total distance measure for all templates, the final procedure in the classification process is the scanning of the C_T matrix for the smallest value which will point to the word template that best matches the input word's template.

Having selected the best response to a discrete word input, we have nearly completed the recognition system from word input to decision output. The only structure that remains for discussion is the adaptive training scheme.

3.5.3 Adaptive Training Software

Each template, once having been created, is not static. As a speaker uses the word recognition system, the templates slowly change to better conform to the average structure of each word in the vocabulary. This convergence has been accomplished by a weighted averaging, element for element, of the input template with the correctly chosen stored template response. However, if an error was made by the recognition software, some basic decisions must be made by the adaptive training network.

The first decision made by the adaptive training network is whether or not a template from the same class as the input word is within the set of "n" templates with the smallest values in the C_T matrix, a stabilization factor reducing the number of templates required to represent any given word.* If this is the case, the input word template is averaged into that template with the smallest value within the above C_T subset corresponding to the input word class. If this is not the case, another decision must be made — to automatically create a new template to better represent that input class or to eliminate some other template first.

If the computer memory will not allow an additional stored template, then the adaptive training software searches all of the word classes for the additional template in least use in the word class containing the most templates; that is, the software searches out the one variation in all of the words in the vocabulary which is least likely to occur. This template is then removed and the input word template stored in its place.

If, however, the computer memory will allow the input of another template, then another stored template will be created from the input word template to better represent that word's class. This occurs most often for those words ending in the stop-consonants /p/, /t/, and /k/, which sometimes are imploded by the speaker. Usually two templates will represent these words; one containing the word with the ending stop-consonant pronounced, the other

*As the stabilization factor "n" is reduced, it was found by C. C. Kesler that the word recognition accuracy increases. However, a reduction of "n" also requires additional training to yield stable classification accuracy and increase template storage. Experiments to date indicate that $n = 2$ is nearly optimum. With $n = 1$ template storage, decision time and rate of convergence are excessive. With $n > 10$, accuracy is significantly degraded. Software could be developed which reduces "n" as a function of the vocabulary iterations if more rapid convergence is required.

without. A word class may contain as many variations of a word as necessary to correctly identify the word, but our results show that the creation of additional templates beyond the required 100 templates is not excessive.

Another factor which leads to the stabilization of the templates is the penalizing of templates which make errors. Each time a classification response is incorrect, the template creating the erroneous response is penalized by reducing the averaging weight associated with that template by ".1". Eventually, if a template should create enough errors so that its associated averaging weight is zero, and at least one other template covers the corresponding word class, then the template will be eliminated. As the weights associated with a template range from 1 to 7, reducing a weight by ".1" also has quite an influence on the weight of the next word averaged into that template.

The previous text has covered the entire speech recognition system software from a recognition standpoint. The remaining software for user input/output is covered in the reference in Appendix E, both in flow charts and in the system operations instructional manual.

3.6 RECOMMENDATIONS

Further research speech recognition systems should include decreasing system response time, storage requirements, and cost; addition of a reject class; higher-speed automatic level control, improvement in extraction and identification of plosives, and improved performance in a noise environment. Longer-range recommendations for additional research are evaluation of speech recognition system performance over standard telephone channels, generalization across a large speaker population, utilization of a common analyzer by a group of speakers, continuous speech, and very large vocabulary word recognition systems.

Speech has much to offer as a media for control, data entry, and inquiry. It leaves the hand and eyes free for other activities. It can be used in the dark and does not require a writing implement. It provides a common natural language base. For example, an operator unfamiliar with a given spacecraft

or aircraft could readily utilize a system without knowledge of exact position of various switches, dials, knobs, keys, and other items. In addition, speech provides an additional data link between man and computer. This additional channel may enhance the reliability of the system by redundancy. Speech also allows mobility while maintaining control. It is believed that this particular system will find use by NASA engineers in component and subsystem checkout and in space-flight monitor and control. Spaceborne applications for astronauts should include direct verbal subroutine selection for control of spacecraft experiments, request of selected computer computations, selection of various forms of image data for comparison, and numerous others.

Section 4
EXPERIMENT CONTROL, PART III, PHASE A

This section describes the research, analysis, and development work performed in Part III, Phase A, of the Crew/Computer Communications Study. This phase was accomplished in five major tasks. In Task 1, the candidate experiment was selected and experiment operations defined. In Task 2, the second prototype PKD was designed and fabricated. Interactive software design requirements were completed in Task 3. In Task 4, implementation of interactive software was completed. In Task 5, the experiment operations demonstration was performed and evaluation and documentation completed.

4.1 OBJECTIVES

The application of advanced crew/computer communications techniques on a Space Shuttle is an expansion of the capability developed with the structured vocabulary, software routines, and hardware as defined in Part I, Phases A and B, of this contract.

Utilizing improved experiment control techniques, it was possible to display tutorial information, computer-aided checkoff lists, and data in an easily discernible form — plain English. In addition, techniques were developed to show how onboard data compression and analysis techniques can be used. Overall, a significant improvement in capability and performance was realized for future manned spacecraft by utilizing the concepts developed during this phase of the contract.

The general vocabulary, software, and hardware requirements to implement the communication technique developed on the spacecraft were established during this study phase.

PRECEDING PAGE BLANK NOT FILMED

4.2 TASK 1 -- EXPERIMENT SELECTION AND OPERATIONS DEFINITION

4.2.1 Criteria for Selection

In order to simulate the onboard operation of a Space Shuttle experiment, it was necessary to define both the configuration of the experiment and its operating procedure. If the simulation candidate has or will be flown on Apollo or Skylab, detailed operating manuals of the experiment are available for reference. Additional data analysis routines will contribute substantially to the fidelity of the simulation and demonstrate the interaction requirements for onboard data reduction.

The following nine attributes have been identified as being either required or desirable for the selected experiment candidates.

Criteria	Required	Desirable
Operating Procedures Available	X	
Configuration Defined (Preliminary)	X	
Space Shuttle Experiment Candidate	X	
Simulation Fidelity	X	
Operational Experience (Apollo, Skylab)		X
Sample Data Available		X
Data Reduction Techniques Defined		X
Data Reduction Techniques Operational		X
Data Reduction Routines Available		X

The more attributes the simulation candidate possesses, the greater will be the content and accuracy of the simulation. It will then be possible to concentrate effort on optimizing information transfer between the crew and the onboard computer. The time frame for this study necessitated selecting a demonstration experiment that possessed most or all of the attributes listed.

4.2.2 Experiment Selection

The 25 functional program elements (FPE's) identified in the NASA Blue Book were broken down into 56 sub-FPE's for the Shuttle Orbital Applications and Requirements (SOAR) Study conducted for MSFC by McDonnell Douglas under Contract No. NAS8-26790.

In reviewing the FPE's and comparing them with similar lists for Apollo and Skylab, it became apparent that only Skylab would provide a useful comparison base for crew/computer interface studies. In the Skylab, there are many experiments that are similar or identical to those planned for the Space Shuttle.

A major objective of this study was to demonstrate the greatly expanded capability in experiment checkout, control, and data analysis that will be realized by using a structured vocabulary as defined in the previous phase of this contract. Because of the limitations of such parameters as computer core and CPU size, I/O capability, and storage capacity, as well as mission requirements, onboard computers have not been used in the past for experiment data analysis. The astronomy, physics, and earth survey experiments will benefit substantially from an extensive onboard data analysis capability during extended missions.

Computer Science Corporation has been under contract to NASA-MSFC to develop image processing and data analysis techniques for multispectral scanners. After several meetings with Computer Science Corporation and NASA, it was established that NASA's newly developed techniques would be useful in the data reduction task of this experiment. The application they have concentrated on is the Earth survey, although many of the same techniques could be applied to the spectrometer experiments in astronomy.

Further investigation of the multispectral scanner, referred to as Experiment S192, revealed that it satisfied all of the selection criteria. Its operating procedures for Skylab were available, and its configuration was defined. Both Skylab and the Space Shuttle have this experiment onboard. The fidelity of the simulation should be excellent due to the degree

of interaction, the availability of operating data reduction routines, and the availability of an existing data base for multispectral sensors. In addition, it was found that the Earth survey category would be the most interesting demonstration for an audience ranging from principal investigators to program managers because the end result of running the experiment — to identify crop resources — could be readily seen. The multispectral scanner experiment was therefore selected as the demonstration experiment for this phase of the contract.

Specific tasks performed by this experiment include land-use mapping, resource recognition, natural disaster evaluation, and ocean resource evaluation. The scenario developed features resource recognition based on techniques developed by Computer Science Corporation.

The S192 instrument has 13 spectral bands from 0.4 to 12.5 micrometers. The system gathers quantitative high-spatial-resolution, line-scan imagery data on radiation reflected and emitted by selected ground sites in the US and other parts of the world.

The system is a conical line-scan system with spectral separation accomplished in a dispersive manner. Each channel of the system is radiometrically calibrated approximately 100 times per second. The high data rate of this equipment is not compatible with the standard recording speed of an Earth Resources Experiment Package tape recorder, so the recorder speed is increased to 60 inches per second during S192 operation. Spatial characteristics are 5.5-degree half-angle scan, 40-nautical-mile swath width, 0.182-milliradian-square instantaneous field of view, yielding a 260-foot resolution at an orbital attitude of 235 nautical mile. Figure 4-1 illustrates the S192 surface coverage.

4.2.3 Experiment Operations Definition

Although the multispectral scanner was the experiment selected, the tasks performed in executing this experiment must be representative of the experiment category in general. The following eight functions were

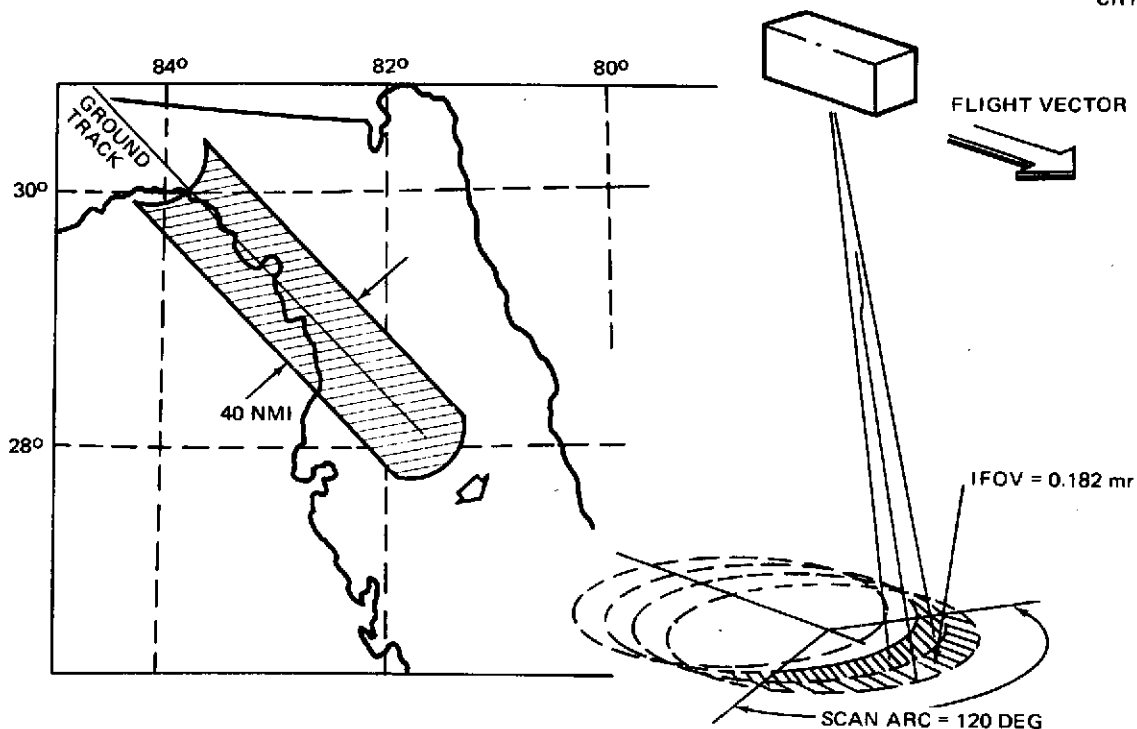


Figure 4-1. S192 Surface Coverage

identified as being typical of most onboard experiments: checkout, experiment startup (initialization), experiment operation (run experiment), power up, data analysis, data transmission, experiment termination, and experiment planning. The flow of the functions in normal operation is illustrated in Figure 4-2.

4.2.3.1 Checkout

Onboard checkout includes verifying the operational status of both the spacecraft's experiment support subsystems as previously defined and the experiment package itself. An automatic checkout capability was assumed for the demonstration of the Shuttle subsystems that support experiment operation, while a computer-aided manual checkout technique was demonstrated to illustrate this means of establishing the operational readiness of the experiment package.

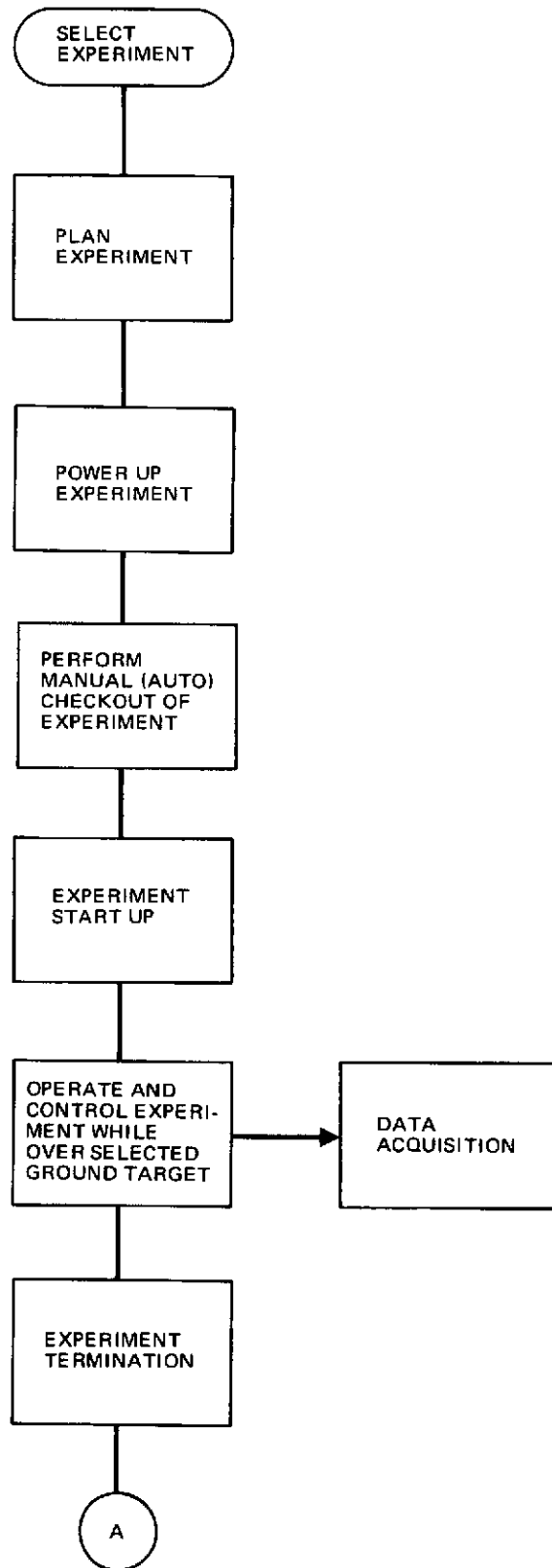


Figure 4-2 (Page 1 of 2). Experiment Definition

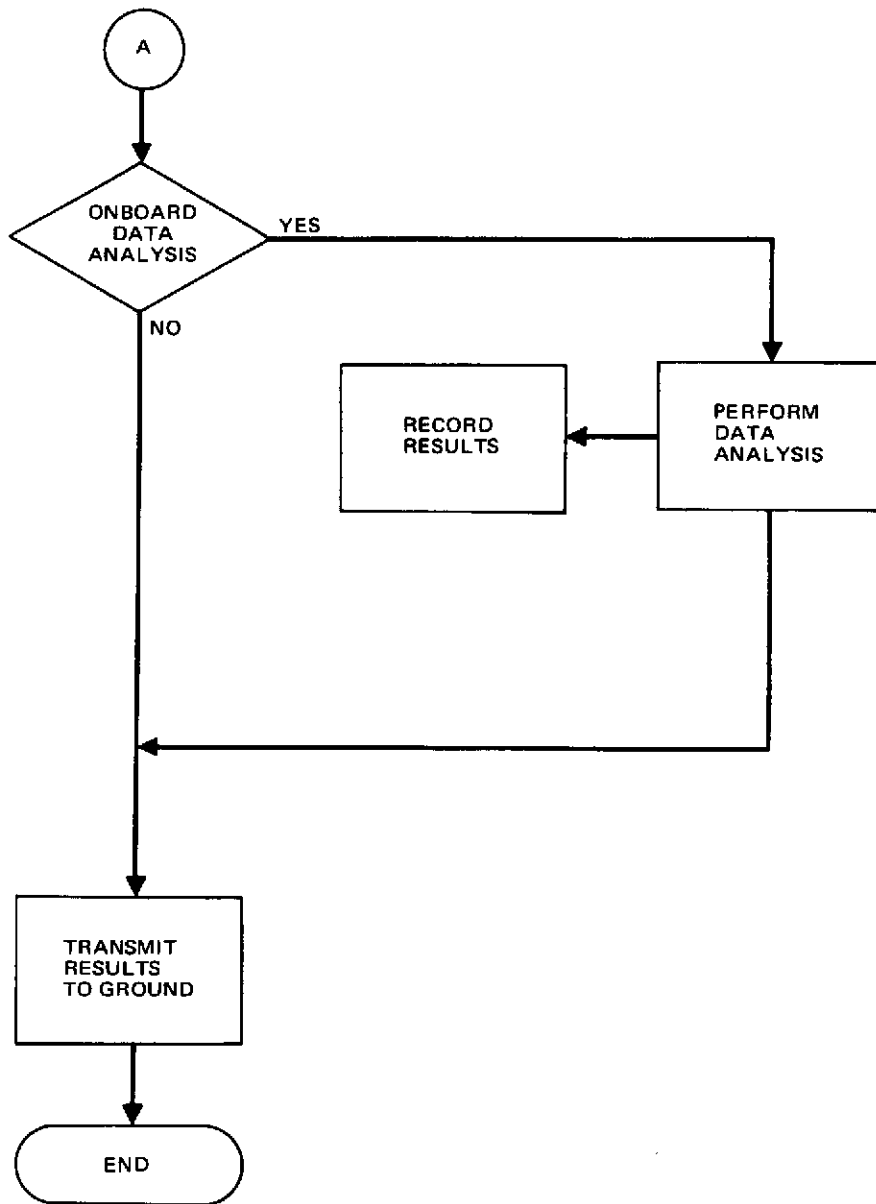


Figure 4-2 (Page 2 of 2). Experiment Definition

4.2.3.2 Experiment Startup

Experiment startup will include all activities required to configure and prepare a payload for operation (e. g., opening viewing doors, loading film, and performing initial calibration).

4.2.3.3 Experiment Operation

During this operational task, the experiment will be controlled by an onboard scientist who will have access to all controls for that experiment. The controls would include the selection of sensor frequency bands, pointing operations, monitoring output signals from the sensors, and the selection of the modes of operation.

The programmable keyboard display (PKD) controls both the experiment and the operating environment of the onboard computer. All experiment operation selections are done on the programmable command keyboard. Data entries can be made at the PKD by keying in numbers on the numeric keyboard. Special function keys are available for controlling the software, changing modes of operation, and performing safing operations on a malfunctioning system.

4.2.3.4 Power Up

The power-up task is performed before the experiment is started. This task includes powering-up equipment which must be ready before starting a time-phased, time-dependent run such as a fly-over of a ground truth site.

4.2.3.5 Data Analysis

After the desired data are gathered by making an orbital pass over the selected ground area, the data analysis phase is initiated. During this phase, a sophisticated onboard interactive computer terminal is required. Extensive crew interaction is needed to classify and analyze the raw sensor data. This function is supported by a graphic device (CRT) and the programmable keyboard display.

The CRT is used for analog signature analysis of the sensor output and display of the data analysis results. These results include both histograms

of classified data and data correlation plots used to select optimum frequencies for discriminating ground data. On completion of this activity, the discrimination process is optimally trained for crop surveys.

4.2.3.6 Data Transmission

The data transmission task includes reconfiguring the communications subsystem as required and transmitting either raw or reduced data to the ground. As such, it is considered to be an activity of the spacecraft operating system rather than the experiment itself.

4.2.3.7 Experiment Termination

The last operational task involves safing the experiment, closing viewing ports, powering down the experiment, and performing all other functions to put the experiment in a dormant state.

4.2.3.8 Experiment Planning

The successful operation of an in-flight experiment requires careful planning. The experiment to be run is part of the overall mission timeline. Before the experiment can be conducted, the spacecraft must be in the proper orbit and the vehicle attitude must provide the desired field of view over designated ground areas. The experiment operator requires information on the flight schedule and status of the experiment to be performed to ensure proper use of spacecraft resources. This planning function fulfills that requirement and additionally gives the operator the means to update plans. When updating, the operator can use any contingency plans such as deploying backup instruments or operating in alternate modes and then verifying the operational readiness of the configuration.

In addition to supporting the previously mentioned activities, the onboard computer will support payload operations by performing checkout of all onboard subsystems, automatic safing of failed systems, supporting orbital analysis calculation, and by providing a data processing capability for all onboard data compression and analysis. Experiment payloads may also require the support of subsystems that are unique to the payload bay, such

as the manipulator arms. Figure 4-3 identifies the spacecraft subsystems that will be used to support experiment payloads similar to S192.

4.3 TASK 2 – FABRICATION OF PORTABLE KEYBOARD AND DISPLAY

The Crew/Computer Communications Study included the fabrication and delivery to NASA/MSFC of a portable keyboard and display. Delivery to NASA is scheduled for the conclusion of the study. This subsection describes the fabrication of this unit.

4.3.1 Design

The PKD model built during this study was the second prototype PKD to be constructed. It was of the same design as the existing unit except for some improvements to eliminate minor mechanical interference and wiring problems. All functional requirements were identical to those of the first unit. The unit was built according to existing documentation and by verbal instruction from the design engineers. No formal drawings were released by McDonnell Douglas; however, a chassis assembly sketch and overall

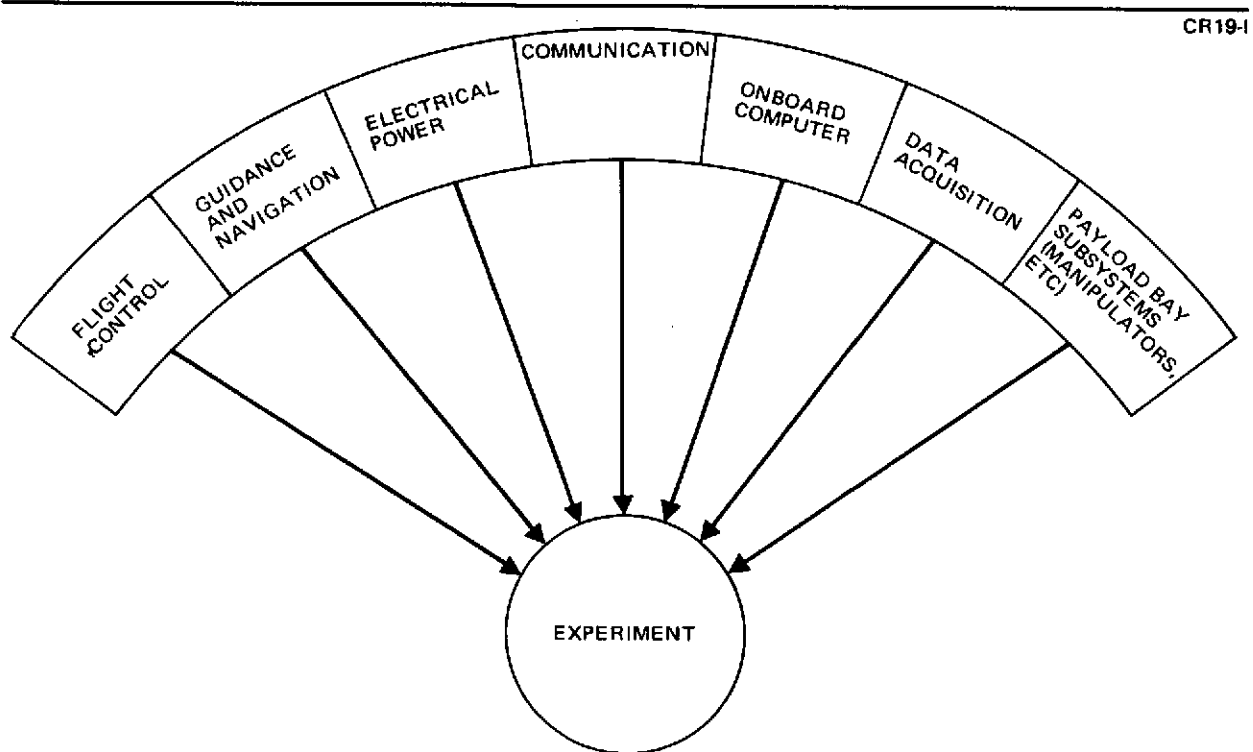


Figure 4-3. Experiment Support

wiring diagram was delivered with the unit. The unit was designed to operate in a normal air-conditioned laboratory environment, consistent with the prototype nature of the unit.

4.3.2 Assembly

The assembly of the PKD was on a minimum-cost basis consistent with good engineering design practices and unit appearance. No requirements were specified for parts selection or assembly.

4.3.3 Test

Proper operation of the PKD was verified after assembly and during software development and demonstration. No formal quality assurance coverage was applied to the unit.

4.4 TASK 3 – INTERACTIVE SOFTWARE DESIGN REQUIREMENTS

4.4.1 PKD Display Definition

During Task 1, seven experiment operational phases were identified. These included checkout, initialization, operation, data acquisition, data analysis, data transmission, and termination. Approximately 50 displays were defined for the checkout, initialization, operation, data acquisition, and termination phases. These displays are used to simulate in detail the operation of an experiment (e. g., setting switches or verifying gain controls are within allowable limits).

4.4.2 Backspace Key Requirement

After carefully studying the results of Part I, Phase C, of this contract, a number of items were identified that would improve the crew/computer communications techniques. A new special function key, BACKSPACE, was added to the repertoire. This enables a correction to be made to input data without clearing the whole data field.

4.5 TASK 4 – SOFTWARE IMPLEMENTATION

As a result of the experience gained in implementing and operating the LM/CSM rendezvous, revisions were made in the previously designed software to provide a more general operating system and one that was

inherently more flexible in supporting applications as programmed on a PDP-9 computer. The following sections discuss the changes made in the existing software. Details on the design and use of the new or revised programs for this task are available in Appendix C.

4.5.1 Revision of BACKUP Function Key

Improvements were made in use of the special function key BACKUP. In the previous version, a push-down stack was kept of the display file names as they were selected on the PKD, and when the BACKUP key was pressed, the last entry in the stack was read out, and that file was redisplayed. A problem arose when at a lower level, several paths were taken in succession from some nodal point. It was necessary with this technique to back up through each of the branches of the tree. As a result, the user saw successive displays appearing that worked back up toward the top of the menu tree with a sudden change in direction back down the tree. It was necessary to back up through all of the branches that were executed at each nodal point. A more logical way to perform this task would be to have the BACKUP key return the user to a predefined point in the menu tree. This new technique was implemented by defining the BACKUP pointer for each display in the display generation routine, FGEN.

4.5.2 File Generation

As the experiment reached the point of requiring 40 to 50 basic PKD displays, the constrictions of the PDP-9 file storage became evident. In order to make better use of the disk files with a maximum number of 48 file names imposed by the system software, groups of 10 selected displays were merged under one file name. The file generation program has also been redesigned to include a comprehensive set of tables which interface with the PKD executive. The tables contain a complete guide to actions to be taken in response to all operator inputs. A scratch data table was added to provide dynamic manipulation of an existing PKD display. A program table was also added to allow frame-related tasks to be performed independent of an application executive.

4.5.3 Executive Program

After analyzing the software requirements defined in Task 3, it became evident that the current executive program, designed for the rendezvous activity, was not versatile enough to support the experiment control package. The new executive would necessarily have to be more general-purpose and would eliminate in-line codes specific to the rendezvous scenario.

A new table-driven executive was developed using structured program techniques. The new program is in conjunction with the expanded table capability of the new file generation program, FGEN. The new executive allows display of a related code to be imbedded in the PKD frame program table and execution to be controlled by the executive. When a specific application is to be executed, the executive passes control to the application program, LENK. The advantages of the new executive are a shorter application programming time, shorter checkout time, and easier editing.

4.6. EXPERIMENT DEMONSTRATION SCENARIO

4.6.1 Background

The S192 experiment scenario, described in Appendix D, covers the execution of a spacecraft experiment from checkout through completion of data analysis. The experiment scenario (Figure 4-4) contains 58 displays. A meaningful subset of these was coded for the data analysis task. The criterion used in selection was one of demonstration time, which was not to exceed one hour and would preferably not exceed one half hour. Use of the CRT display and the CRT light pen is simulated in the demonstration.

The experiment functions selected for demonstration were power up and data analysis. Power up was selected to illustrate implementation of check-listing procedures using the computer as a prompter. The data analysis task with its complex interfacing of the experiment, the spacecraft, and the experimenter were chosen to show the capability of the PKD and the operating system software.

4.6.2 Demonstration Scenario

The multispectral scanner experiment was integrated into the existing operational categories defined in Task 1 under the function category experiments, as shown in Figure 4-4. The experiment log was expanded and defined in detail to fill the gap between categories of experiments and the selection menu for a specific experiment such as S192.

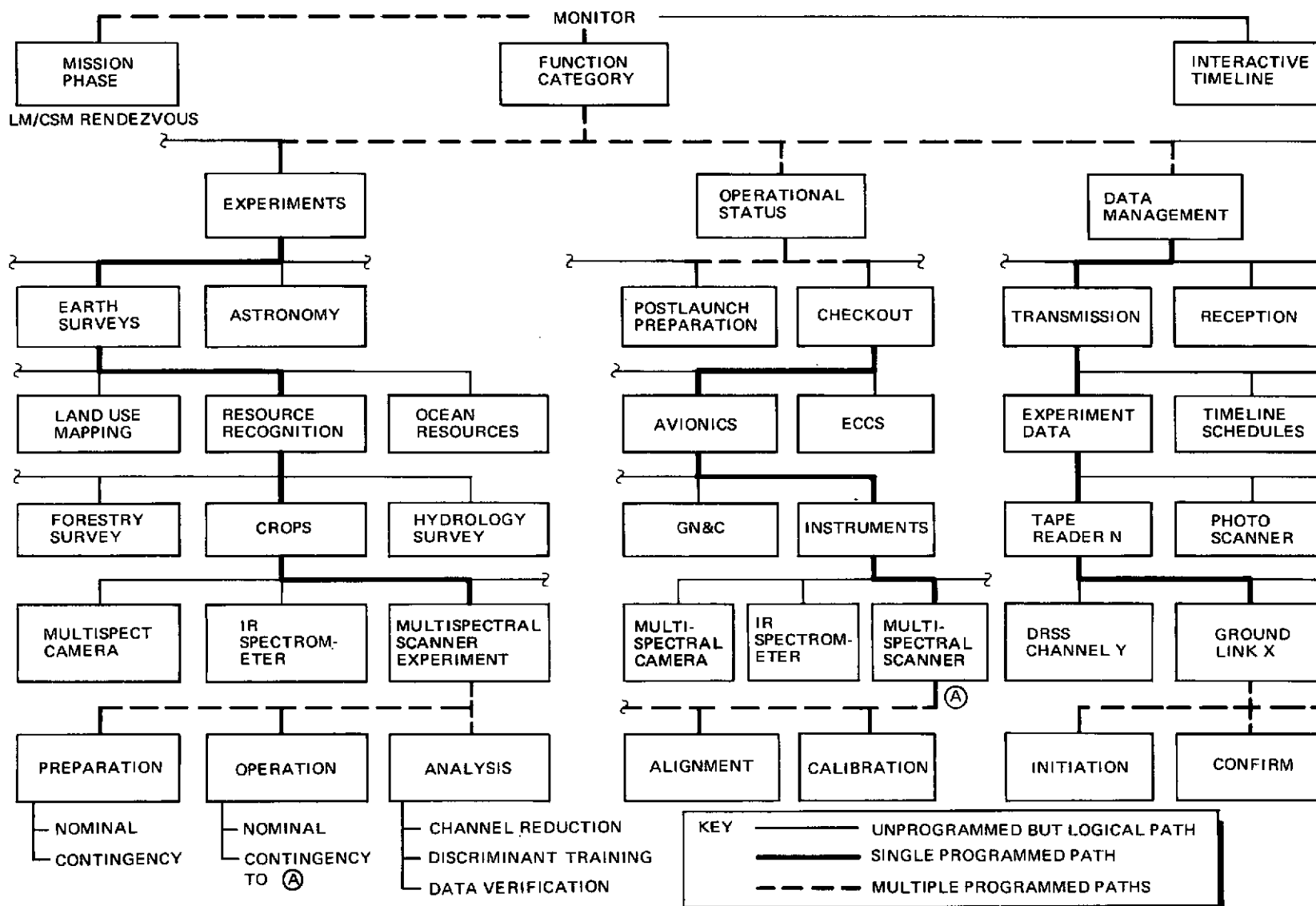
Figure 4-5 depicts the tree structure for the demonstration. The entire scenario is in Appendix D.

The primary task developed for the demonstration was the data analysis task. Execution of the task covers three time-sequential activities which occur after collection of multispectral scanner data during fly-over of a ground truth site.

In order to analyze crops in other areas, the onboard classifier program must be trained as to which multispectral scanner channels are most effective in recognizing crops. Additionally, the classifier can be "trained" to minimize the number of channels required for a specific crop. This training reduces the runtime of the program while maintaining a desirable level of accuracy.

The first activity in data analysis is to input to the classifier an accurate description of the ground truth site that contains crop boundaries and landmarks. Figure 4-6 is a facsimile of a ground truth site descriptive photo. The digitized description of this area is then input to the classifier program to establish a reference for comparing channel output.

During the second phase of data analysis, the classifying ability of each channel, by crop, is determined and the rankings presented on the PKD display. From the rankings and given the number of channels to use, the classifier discrimination algorithm can be defined. Figure 4-7 represents the CRT display of a two-channel classification of soil versus other crops. By defining a straight line which separates the crop's soil from the other data points and retaining the equation of that line, the classifier can then be used on an unknown site for detecting soil.



KEY

- UNPROGRAMMED BUT LOGICAL PATH
- - - SINGLE PROGRAMMED PATH
- · · MULTIPLE PROGRAMMED PATHS

Figure 4-4. Crew/Computer Communication Experiment Control Study

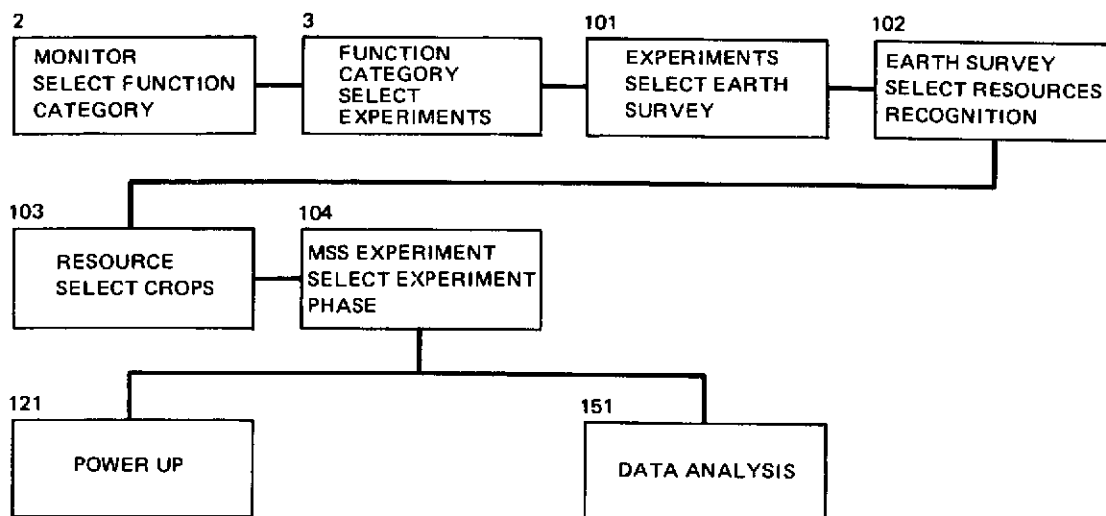


Figure 4-5. Experiment Control Scenario

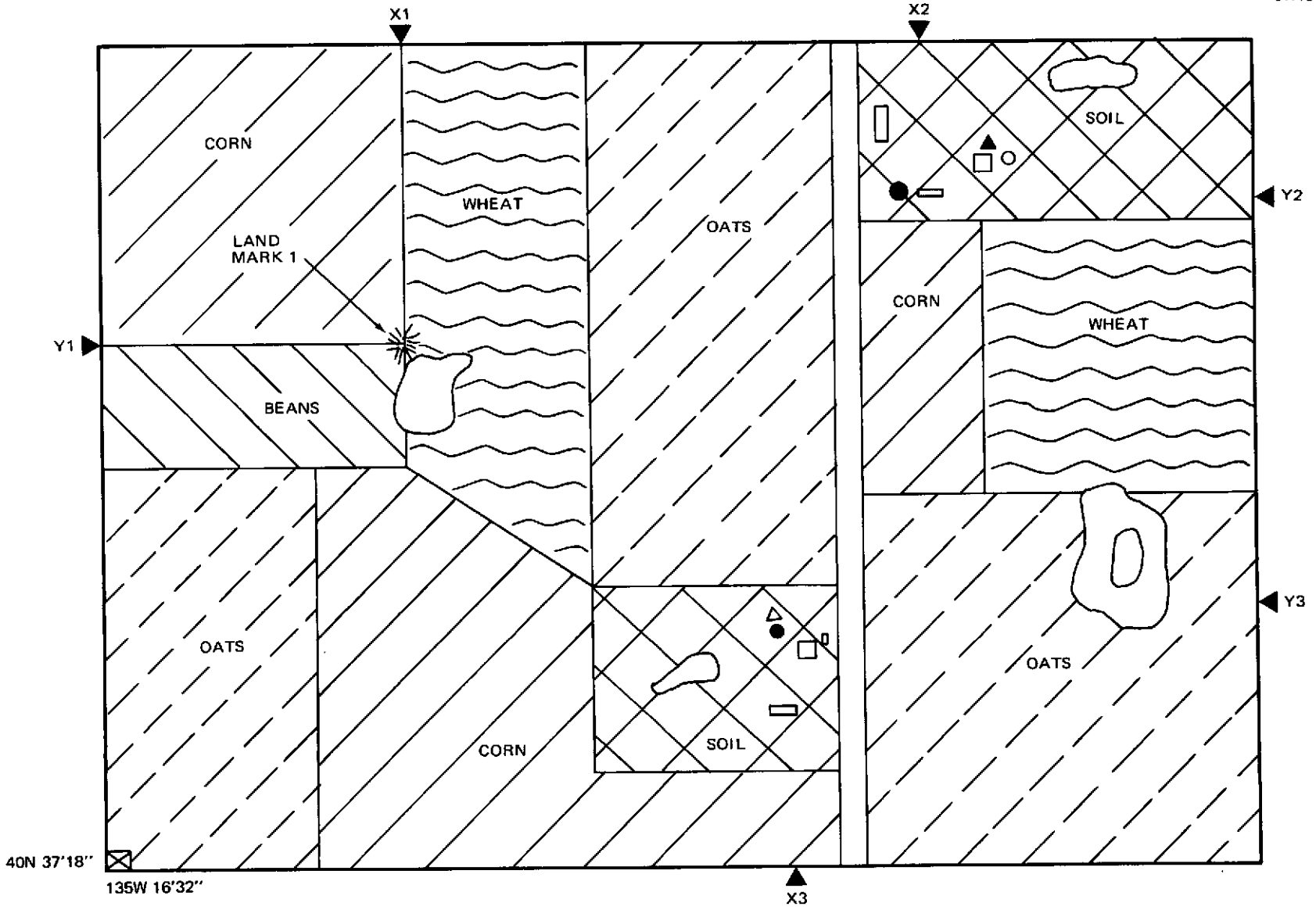
A previously trained classifier can be re-optimized during the third data analysis activity. Essentially, the classifier is re-optimized to use fewer channels for each crop to reduce computation time or more channels are used on certain crops to improve accuracy.

4.7 RECOMMENDATIONS AND CONCLUSIONS

Phase A, Part III of this study was a successful extension of the effort performed in Phase C, Part I. The results of this study continued to prove the efficiency and flexibility in crew/computer communications methods with a device such as the PKD.

Demonstrations of the experiment control application were considered successful and commensurate with the results of the rendezvous demonstrations. Implementation of the complex data analysis task certainly illustrated the capability of the hardware and support software.

123



40N 37'18"

135W 16'32"

Figure 4-6. Truth Site Photo F1234.5

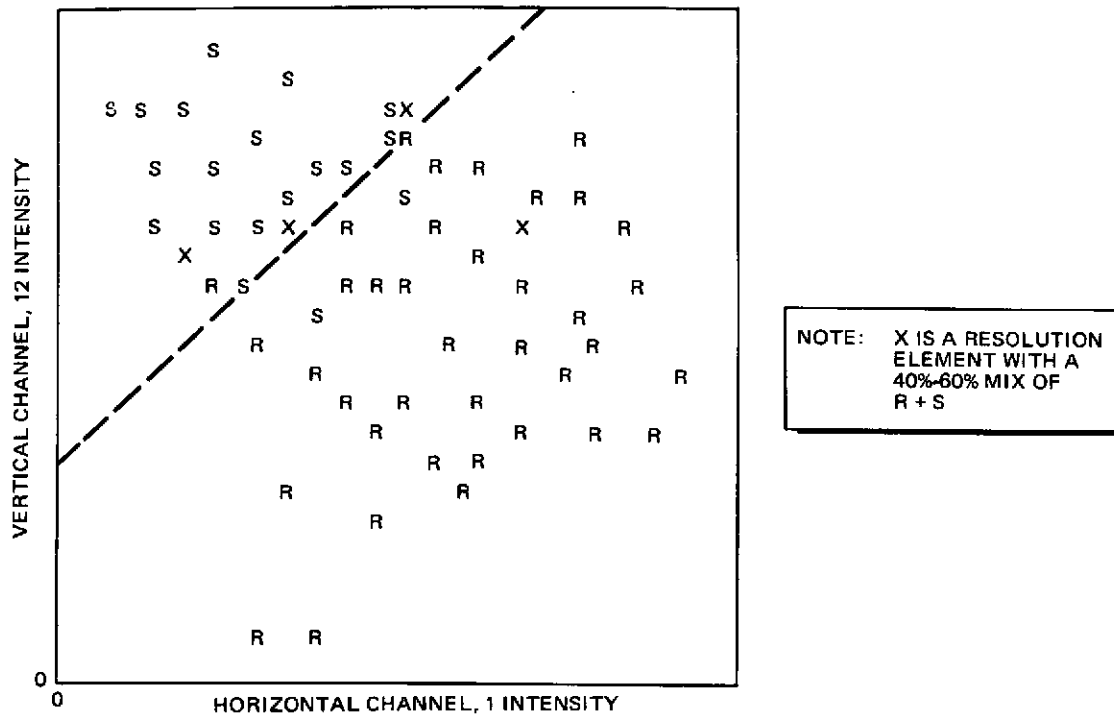


Figure 4.7. Initialization of Discriminant Trainer, Crop Soil (S) versus Remaining Crops (R)

The upgrading of the table-driven execution control program and the revised display file generation program paves the way for continuing the development of new applications. The use of an execution control program with the expanded number of built-in functions will reduce the need for special-purpose software in future activities. Additionally, since the routine was coded using structured program techniques, it can be easily rewritten for other computers and be easily understood by new users.

It is recommended that additional applications be developed for ground activities such as preflight checkout and logistics functions. It is also recommended that PKD-based communication be implemented in development facilities which will support future space efforts.

Section 5

RECOMMENDATIONS FOR FUTURE DEVELOPMENT

Further research on the speech recognition system is recommended in the following areas: (1) decreasing system response time, storage requirements, and cost; (2) adding a reject class; (3) adding higher-speed automatic level control; (4) improving extraction and identification of plosives; and (5) improving performance in a voice environment. The system should also be utilized with a large speaker population.

Research in this field should also be extended to cover recognition of words spoken over the telephone and in continuing conversation, while the vocabulary is gradually increased.

We further recommend that NASA channel the results of this study into other applications, including spacecraft control, medical diagnosis, and computerized commercial areas.

The spacecraft of the future will have increased automation. Man will perform fewer and fewer manual tasks, but he will continue to initiate execution of automatic procedures by the spoken word or by physical action. Improvements in man's ability to interact with the spacecraft computer which supports him will relieve him from tedious, repetitive operations and yet will increase the number of tasks he can perform. Devices such as the PKD can serve as a prompter and tutorial aid, thus eliminating the need for memorizing a large number of detailed procedures. All phases of experiment control can be enhanced by direct verbal selection of control routines, computational routines, and stored image data.

The application of structured vocabulary techniques and the word recognition system to the biomedical field would standardize and speed up physician-directed diagnosis. It would also aid in training the medical student by

presenting him with visual cues to aid him in selecting the logical paths that lead to a correct diagnosis. These technologies promise to increase the quality of patient diagnosis and reduce costs by minimizing physician time as well as by reducing the patient time and inconvenience.

The application of these technologies in the commercial area is almost unlimited. For example, the following areas would be ideal environments: (1) language translation, (2) store inventory, (3) purchase orders, (4) air traffic control, and (5) automatic checkout. These are only a few areas that can be enhanced by the effective application of structured vocabularies, the programmable control display, and the word recognition system.

REFERENCES

1. R. K. Potter, G. A. Kopp, and Harriett Green Kopp, Visible Speech. Dover Publications Inc., New York, N. Y. 1966.
2. G. Fairbanks, "Recent Experimental Investigations of Vocal Pitch in Speech," The Journal of the Acoustical Society of America, No. 11, pp 457-466, 1940.
3. James L. Flanagan, "Automatic Extraction of Formant Frequencies from Continuous Speech," Journal of Acoustical Society of America, Vol. 28, No. 2, pp 110-117, Jan 1956.
4. T. J. Edwards, A Speaker-Dependent Near Real-Time Discrete Word Recognition System, Master's Thesis at the University of Southern California, accepted October 1973.
5. R. L. Miller, "Nature of Vocal Cord Wave," Acoustical Society of America J., No. 31, pp 667-677, 1959.
6. N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," J. Acoustical Society of America, Vol. 19, pp 90-119, 1947.
7. H. F. Fletcher, Speech and Hearing in Communication, D. Van Nostrand Company (Princeton, N. J.), 1953.
8. Gordon E. Peterson and Harold L. Barney, "Control Methods Used in the Study of Vowels," Journal of Acoustical Society of America, Vol. 24, No. 2, pp 175-184, March 1952.
9. M. B. Herscher and T. B. Martin, RCA Engineer. RCA, January 1965.

BIBLIOGRAPHY

Crew/Computer Communications Study

1. Astronaut/Computer Communications Study – Final Report. McDonnell Douglas Astronautics Company, MDAC G0985, March 1971.
2. H. R. Luxenburg, and Rudolph L. Kuehn. Display Systems Engineering. McGraw-Hill Book Company (New York), 1968.
3. Apollo 15 – CSM 112 – Basic CSM G&C Checklist. NASA-MSC, DCW, March 22, 1971.
4. Skylab Flight Plan – Preliminary Reference Interim Revision. NASA-MSC, SL-2, SL-3, SL-4, March 1971.
5. Guidance System Operations Plan for Manned/CM Earth Orbital and Lunar Missions Using Program Colossus 2E. Charles Stark Draper Laboratory, Massachusetts Institute of Technology, R577, Cambridge, Massachusetts, July 1970.
6. Phase I – Crew Interface Definition Study – Pre-Simulation Report. McDonnell Douglas Astronautics Company, MDC E0484, October 1, 1971.
7. Space Shuttle Avionics Simulation Guidelines and Techniques Study. McDonnell Douglas Astronautics Company, MDC E0378, May 24, 1971.
8. Apollo Applications Baseline Reference Mission. NASA-MSC, MSC-KM-D-68-4A (AAP-1/AAP-2, AAP-3, and AAP-4), October 27, 1969.
9. Preliminary Program Plan for Concept Verification Test (CVT) Space Station Subsystem Breadboard. NASA-MSFC 50M75041, January 19, 1971.
10. Arthur D. Hall, A Methodology for Systems Engineering. D. Van Nostrand Company, Inc. (New Jersey), 1962.
11. Apollo 15, Mission J-1 – CSM Rendezvous Procedures (Final). McDonnell Douglas Astronautics Company – Houston Operations, MSC-04496, June 30, 1971.
12. Basic CSM G&C Checklist – Apollo 16 and 17, All Launch Dates. NASA-MSC, December 8, 1971.

PRECEDING PAGE BLANK NOT FILMED

13. CSM Procedures for Orbital Phase of Skylab 1/2 Mission. NASA/MSC, Basic Issue, MSC-05885, December 31, 1971.
14. Displays and Controls Functional Requirements Specification. McDonnell Douglas Astronautics Company, MDC E0490, October 1, 1971.
15. Francis F. Martin, Computer Modeling and Simulation. John Wiley & Sons, Inc. (New York), 1968.
16. LT-19 Multi-Station Teletype Control Instruction Manual. Digital Equipment Corporation, DEC-09-H2AA-D, Maynard, Massachusetts 1970.
17. User Handbook - PDP-9. Digital Equipment Corporation, F-95, Maynard, Massachusetts, 1968.
18. Advanced Monitor Software System for PDP-15/20/30/40. Digital Equipment Corporation, DEC-15-MR2B-D, Maynard, Massachusetts, 1970.
19. Programmer's Reference Manual - Macro-15 Assembler. Digital Equipment Corporation, DEC-15-AMZA-D, Maynard, Massachusetts, 1969.
20. SGEN Utility Program, Digital Equipment Corporation, DEC-15-YWZA-DN3, Maynard, Massachusetts, 1970.
21. PDP-15 Systems - Reference Manual. Digital Equipment Corporation DEC-15-BRZC-D, Maynard, Massachusetts, 1971.
22. PIP - Advanced Monitor Utility Program. Digital Equipment Corporation, DEC-15-YWZB-DN9, Maynard, Massachusetts, 1971.
23. FORTRAN IV Programming Manual - PDP-15 Systems. Digital Equipment Corporation, DEC-15-KFZA-D, Maynard, Massachusetts, 1969.
24. Programmer's Reference Manual - PDP-15 Utility Programs. Digital Equipment Corporation, DEC-15-YWZA-D, Maynard, Massachusetts 1969.

Experiment Control Study

1. Preliminary Edition of Reference Earth Orbital Research and Applications Investigation (Blue Book), NASA NHB 7150. 1, January 15, 1971.
2. Earth Orbital Experiment Program and Requirements Study. McDonnell Douglas Astronautics Company, December 1970 for NASA-Langley Research Center, under Contract NAS1-9464.
3. Shuttle Orbital Applications and Requirements Study. McDonnell Douglas Astronautics Company, December 1971 for NASA-MSFC under Contract NAS8-26790.
4. Skylab Experiment Operations Handbook, Vol. I, Experiment Descriptions, November 19, 1971; Vol. II, Experiment Operational Procedures, January 3, 1972, NASA-Manned Spacecraft Center, Houston, Texas.
5. The Application of Crew Computer Communications to Experiment Operations, McDonnell Douglas Astronautics Company, Technical Proposal 71W-684-2Q, September 1972.
6. Operating, Maintenance and Handling Procedures, Volume II, Flight Operating Procedures for S192 Multispectral Scanner. Honeywell Contract for NASA-MSFC, No. NAS9-11196, July 28, 1972.
7. Crew/Computer Communications Study (Annual Progress Report) McDonnell Douglas Astronautics Company, Report No. MDC G4263, October 1972.
8. A. D. Bond, and R. J. Atkinson, An Integrated Feature Selection and Supervised Learning Scheme for Fast Computer Classification of Multi-Spectral Data. Presented to University of Tennessee Space Institute Conference on Earth Resources Observation and Information System Analysis, March 13-14, 1972 (work done under Contract NAS8-21805).