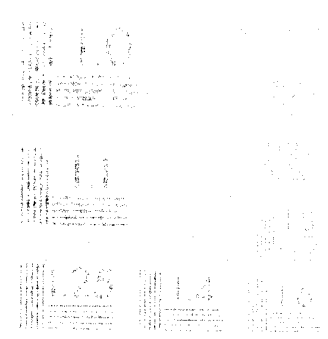


NCJRS

NCJRS is a non-profit organization that provides information and assistance to law enforcement agencies and the public. We are committed to providing high-quality, cost-effective services to our clients.

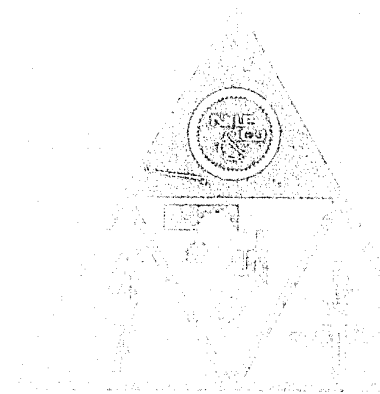


NCJRS is a non-profit organization that provides information and assistance to law enforcement agencies and the public. We are committed to providing high-quality, cost-effective services to our clients.

NCJRS is a non-profit organization that provides information and assistance to law enforcement agencies and the public. We are committed to providing high-quality, cost-effective services to our clients.

U.S. DEPARTMENT OF JUSTICE
LAW ENFORCEMENT ASSISTANCE ADMINISTRATION
NATIONAL CRIMINAL JUSTICE REFERENCE SERVICE
WASHINGTON, D.C. 20531

BP92H



U.S. DEPARTMENT OF JUSTICE
LAW ENFORCEMENT ASSISTANCE ADMINISTRATION
NATIONAL CRIMINAL JUSTICE REFERENCE SERVICE

The Reference Corporation

Aerospace Report No.
ATR-77(7617-07)-1

EQUIPMENT SYSTEMS IMPROVEMENT PROGRAM

SPEAKER IDENTIFICATION
PROGRAM 7907
FINAL REPORT

Law Enforcement Development Group
THE AEROSPACE CORPORATION
El Segundo, California

January 1977

Prepared for
National Institute of Law Enforcement
and Criminal Justice
LAW ENFORCEMENT ASSISTANCE ADMINISTRATION
U.S. DEPARTMENT OF JUSTICE

Contract No. J-LEAA-025-73

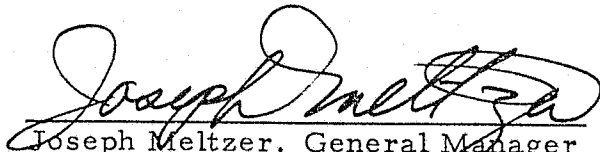
This project was supported by Contract Number J-LEAA-025-73 awarded by the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice, under the Omnibus Crime Control and Safe Streets Act of 1968, as amended. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

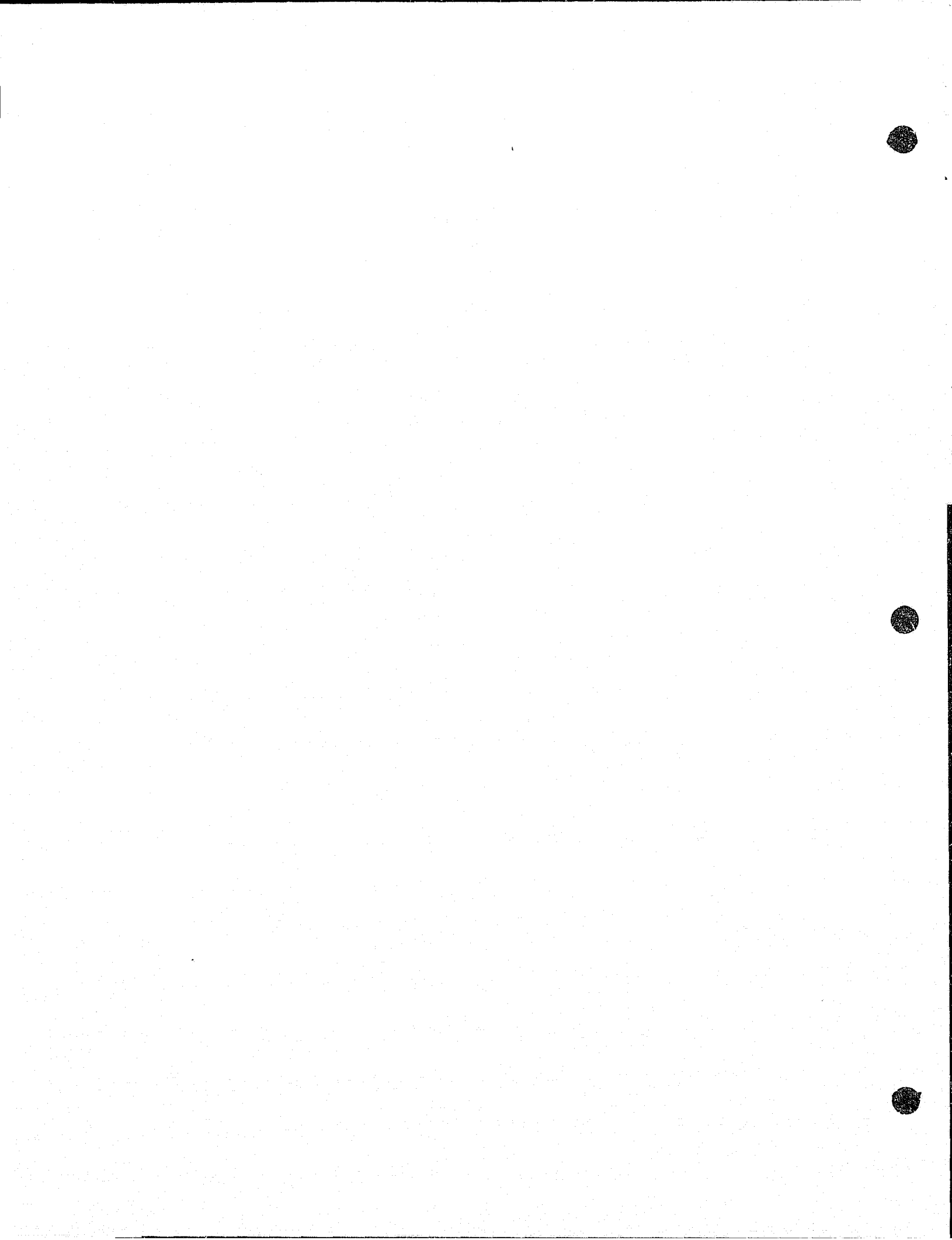
Report No.
ATR-77(7617-07)-1

EQUIPMENT SYSTEMS IMPROVEMENT PROGRAM

SPEAKER IDENTIFICATION, PROGRAM 7907,
FINAL REPORT

Approved


Joseph Meltzer, General Manager
Law Enforcement and
Telecommunications Division



ABSTRACT

This document is the final report of the efforts by The Aerospace Corporation to develop, for the Law Enforcement Assistance Administration, a computer-assisted speaker identification system for use in investigation, as well as in courtroom testimony, and to investigate other applications of speaker identification technology. Most of the effort was subcontracted to Rockwell International, which designed, fabricated, and tested the Semi-Automatic Speaker Identification System.

The report presents a general description of the design and operation of the Semi-Automatic Speaker Identification System, a history of the speaker identification program, the problems encountered during the testing of the system, and recommended design improvements, some of which were approved for implementation before the program was terminated by the customer.



CONTENTS

ABSTRACT	v
EXECUTIVE SUMMARY	xiii
1. INTRODUCTION	1
2. STATUS OF SEMI-AUTOMATIC SPEAKER IDENTIFICATION SYSTEM	5
3. GENERAL DESCRIPTION OF SEMI-AUTOMATIC SPEAKER IDENTIFICATION SYSTEM	7
3.1 Speech Process	7
3.2 Use of Voiceprints	10
3.3 Semi-Automatic Speaker Identification System	12
3.4 Operation of the Semi-Automatic Speaker Identification System	14
3.5 Performance of Semi-Automatic Speaker Identification System	23
4. HISTORY OF PROGRAM	29
4.1 Development of the Semi-Automatic Speaker Identification Program	30
4.2 Speaker Data Base and Comparison Algorithm	31
4.3 Laboratory Test	36
4.4 Pilot Test	37
4.5 Potential Solutions to System Problems	40
5. SYSTEM PROBLEMS AND RECOMMENDED IMPROVEMENTS	49
6. TASK OPTIONS	61
6.1 Channel Equalization	61
6.2 Enhancement of Semi-Automatic Speaker Identification System	62
6.3 Forensic Feasibility Study	63

6.4	Dynamic Feature Extraction (Digital)	66
6.5	Dynamic Feature Extraction (Analog)	67
6.6	Hybrid Speaker Identification System	68
6.7	Semi-Automatic Speaker Identification System Data Base	69
6.8	Enhancement of Recordings with Disturbances	70
6.9	Optimization of System Operation	71
6.10	Data Base Standard	71
7.	CONCLUSIONS	75
	NOTES	77
APPENDICES		
A.	PROGRAM DOCUMENTS AVAILABLE THROUGH THE NATIONAL CRIMINAL JUSTICE REFERENCE SERVICE	81
B.	SOFTWARE OVERVIEW SPECIFICATION FOR SEMI- AUTOMATIC SPEAKER IDENTIFICATION SYSTEM	85
C.	EQUIPMENT LISTING AND MANUALS FOR SEMI- AUTOMATIC SPEAKER IDENTIFICATION SYSTEM	169

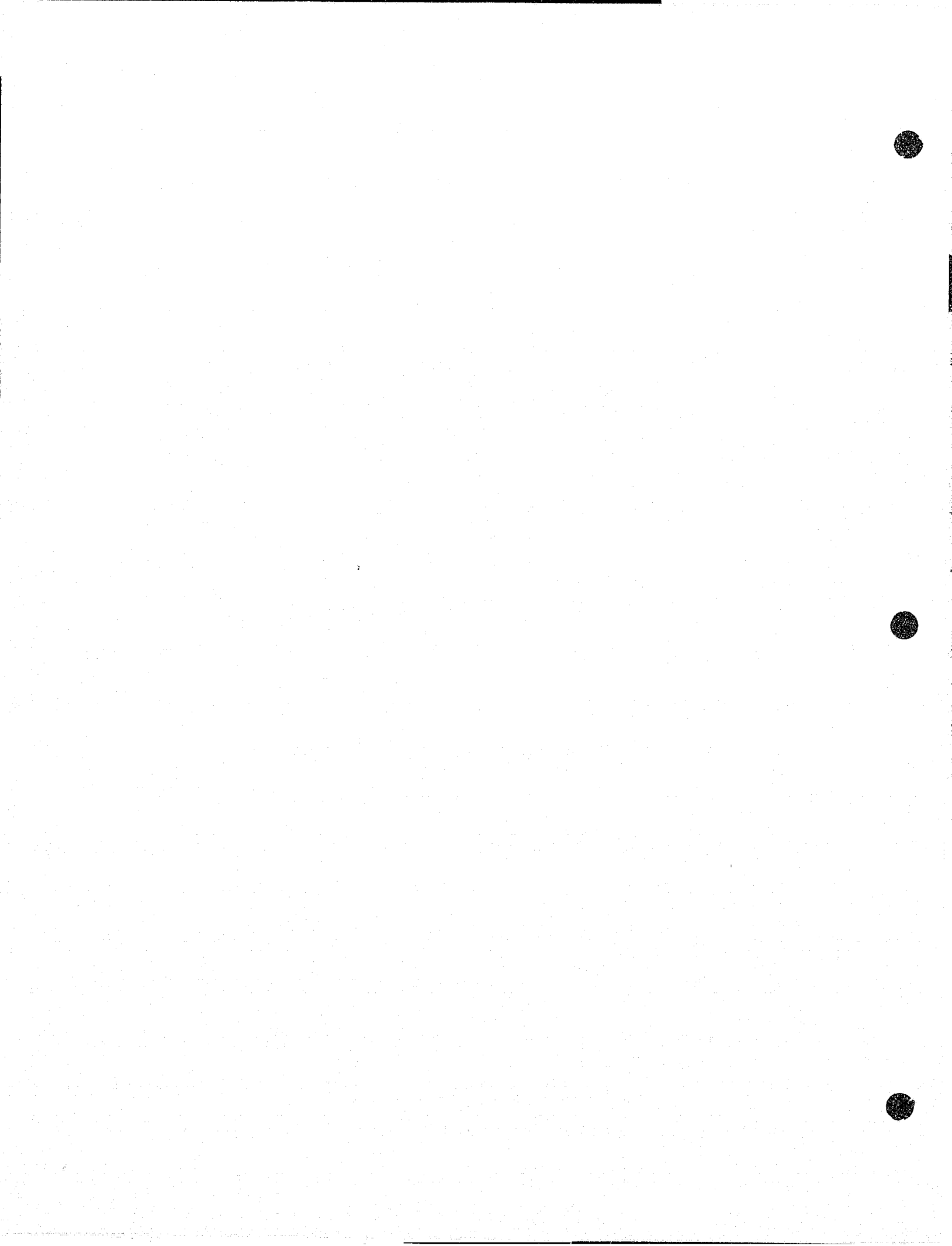
ILLUSTRATIONS

1.	Linear Model of the Speech Process	8
2.	Spectral Transformation	9
3.	Labeled Spectrogram of "This is a Voiceprint"	11
4.	Frequency Domain Features	13
5.	Speech Intensity Waveform	13
6.	Semi-Automatic Speaker Identification System	15
7.	System Operation	17
8.	Example of Alphabetic Transcription	18
9.	Computer-Aided Voice Comparison	19
10.	Semi-Automatic Speaker Identification System Identification Process	22
11.	Typical Evaluation Test Results	24
12.	Example of Confidence Estimates on Semi-Automatic Speaker Identification System Statistics	27
13.	Recording Data Base under Laboratory Conditions	32
14.	Semi-Automatic Speaker Identification System	33
15.	Examples of Comparison Features	33
16.	No-Decision Probability for Speaker Identification	35
17.	Voiceprint Examiners and Aerospace Program Manager at Los Angeles Police Department Pilot Test Site	38
18.	Frequency Response of Telephone Channels	39
19.	Formant Trajectories	41
20.	Formant Trajectories of Unknown Speaker	43

21.	Formant Trajectories of Two Exemplars	45
22.	New Design Improvement Algorithm Tested on Semi- Automatic Speaker Identification System at Aerospace . . .	47
23.	Examples of Measured Frequency Responses of Telephone Channels	51
24.	Speech Signal Passed Through Telephone Channel	51
25.	Long-Term Speech Spectra of Three Speakers Reading Different Texts	53
26.	Pictorial Explanation of Channel Equalization	53
27.	Noise Can Affect the Feature Values of the Semi-Automatic Speaker Identification System	54

TABLES

1.	Functional Comparison of Voiceprint Examination and Computer-Assisted Speaker Identification Techniques	14
2.	Selected Phoneme Types	16
3.	Sample Printout of Semi-Automatic Speaker Identification System Performance Statistics	25
4.	Speaker Dependent (Dynamic) Features of Phrase: "How Many Diesel Guided Missile Submarines?"	43
5.	Specifications for Data Base Standard.	73
6.	Summary of Tasks	74



EXECUTIVE SUMMARY

The Aerospace Corporation, under contract to the National Institute of Law Enforcement and Criminal Justice of the Law Enforcement Assistance Administration, technically monitored and provided system engineering capability for the speaker identification program from FY 73 to FY 76. This report presents the objectives, the accomplishments, and the problems encountered with the program.

The prime objective of the program was to develop a computer-assisted speaker identification system for use in investigation, as well as in courtroom testimony, and to investigate other applications of speaker identification technology. Subcontracts were awarded to Rockwell International for the development of the Semi-Automatic Speaker Identification System. Based on the recommendations from the Institute-funded research by Stanford Research Institute and Texas Instruments, Rockwell was directed to synthesize, under laboratory conditions, a speaker comparison algorithm that operates on the steady-state features of speech. The speaker identification system was then evaluated using laboratory-derived data and found to produce speaker identification accuracies in excess of 97 percent across a broad variety of texts and voice characteristics whenever ten phonetic events were present in the voice recordings. The computer-aided speaker identification system was designed to be operated by the voiceprint examiner or one who is similarly trained to locate sounds from voice spectrograms and to phonetically describe sounds accurately by listening.

After limited testing under extended laboratory conditions, the Semi-Automatic Speaker Identification System appeared invariant across the standard English, Black Urban, and Chicano female dialects, speech with simulated stress, and speech with nasal disguises. There was some inconsistency in the results when comparing speakers with the Black Urban male dialect, Chicano male dialect, or Black Urban male disguises. The most

significant result was that the telephone channel response was found to dominate speaker comparisons.

The Semi-Automatic Speaker Identification System then underwent a pilot test in a voiceprint laboratory where three voiceprint examiners were trained to operate the system. They processed criminal evidence and simulated recordings on the system. The pilot test demonstrated shortcomings in the operation of the speaker identification system, each accompanied by recommended design improvements. The major problem encountered was, again, that the telephone channel response dominates the speaker comparison. The other problems encountered were operator variability and inefficient man-machine interaction. Because of the presence of the overriding telephone channel problem and the limitation of available funds, the anticipated problem area of noise was not evaluated. Several improvements to the system were made, but others, including the system incorporation of a channel filter to suppress the channel effect, were beyond the scope of the subcontract.

Supported by successful laboratory demonstrations at Rockwell, other companies, and universities, Aerospace offered potential solutions to the channel problem. Aerospace defined a Design Optimization Requirements Definition with a recommended short-term study to be made on the Semi-Automatic Speaker Identification System to determine if there was a feasible solution to the problems of noise and channel effects and other undesirable parameters such as phase distortion and recorder clipping. The speaker identification program was terminated by the Law Enforcement Assistance Administration before the feasibility study was commenced.

CHAPTER 1. INTRODUCTION

In recent years, speaker identification through the use of voice spectrograms has become an important source of evidence in investigative activities and in criminal court proceedings. Speaker identification techniques currently in use are hampered by time-consuming, manual methods. The unscientific nature of existing methods has also led to controversy over their admissibility in courts of law.

In an effort to overcome these drawbacks, the National Institute of Law Enforcement and Criminal Justice has supported a number of efforts to evaluate the effectiveness of speaker identification techniques and promote the development of improved methods and equipment. This report describes the research, design, development, and testing of the Semi-Automatic Speaker Identification System and other related activities monitored or conducted by The Aerospace Corporation for the Institute.

The first effort sponsored by the National Institute of Law Enforcement and Criminal Justice was a study in 1968 conducted by Professor Oscar Tosi of the Michigan State University. After one month of training, 29 examiners were tested with spectrograms from 250 different speakers in a large variety of tests (34,992 trials). The results of the experiment showed significant examiner error rates.

In an effort to reduce the error incidence, further Institute support was provided to develop improved machine-assisted speaker identification systems. Michigan State Police, sponsored by the National Institute of Law Enforcement and Criminal Justice Grant NI-71-079G, subcontracted with Stanford Research Institute and Texas Instruments for machine-assisted speaker identification research. Both Stanford Research Institute and Texas Instruments, each using recorders and other general purpose

equipment as a base, developed systems for extracting features from speech and processed those features with digital computers. Their results were comparable and indicated that (a) the technique supplements information available from the voice spectrogram and warrants further effort, and (b) additional research on feature extraction and analysis is desirable. The research did not develop a set of optimum voice features for machine-assisted speaker identification, however, nor did it acquire sufficient evidence to demonstrate that the results of computer analysis can be used as courtroom evidence.

Based on the desire to improve the use and validity of voice spectrograms in law enforcement, the Institute awarded a contract to The Aerospace Corporation in FY 73 to initiate a comprehensive program on speaker identification. The objectives of the program were to:

- o Conduct investigations to improve voice spectrogram technology and properly validate areas for its applications.
- o Provide an interim semi-automated speaker identification system, along with statistical evidence of its capabilities.
- o Provide a voice classification system to allow the search of large files for identification purposes.
- o Provide new techniques for voice identification.

It was expected that the program would extend over several years and that most of the effort would be subcontracted. The design, fabrication, and testing of the Semi-Automatic Speaker Identification System were performed. However, the program was canceled by the Institute in FY 76 before the start of the design optimization phase of the program where problems in the system encountered during testing were to be minimized.

At the beginning of the speaker identification program, Rockwell International was awarded a subcontract to synthesize, under laboratory conditions, a speaker comparison algorithm that operates on the steady-state features of speech. This design was based on the recommendations from the Institute-funded research by Stanford Research Institute and Texas Instruments. The system performed well under laboratory conditions, but encountered a major problem with telephone recordings. It was subsequently concluded that the features (measurable quantities) of steady-state speech vary significantly with the frequency response of the telephone channel. The speaker identification program was then terminated. The cancellation followed an Institute approval to study the feasibility of reducing or modifying the system to accurately process criminal evidence, which is recorded under conditions less than ideal.

However, an algorithm to minimize this most prohibitive problem was tested on a limited number of recordings, and the results were very promising. Ironically, the feasibility study included the incorporation of this algorithm in the Semi-Automatic Speaker Identification System, which would have made the system operational in a forensic laboratory with criminal recordings.

The status of the Semi-Automatic Speaker Identification System program at the time the program was terminated is detailed in Chapter 2 of this report.

Chapter 3 provides a general description of the function, design, and operation of the Semi-Automatic Speaker Identification System. This description provides the reader with an understanding of the system capabilities and performance.

Chapter 4 gives the history of the Law Enforcement Assistance Administration's and Aerospace's role in the speaker identification program. Program accomplishments and achievements highlight this chapter.

The existing problems encountered in the operation of the Semi-Automatic Speaker Identification System are described and discussed in Chapter 5. A technical analysis and interpretation of the problems is included. Solutions to each problem and system improvements are recommended.

Chapter 6 addresses the task options for the most feasible and efficient plans on any possible future effort to continue the development of the Semi-Automatic Speaker Identification System toward the program's objectives.

A summary of the program and recommendations for additional development are presented in the conclusions in Chapter 7.

Appendix A lists references that can be found in the National Criminal Justice Reference Service.

Appendix B contains the Semi-Automatic Speaker Identification System software overview specifications, which were not previously documented.

The components of the Semi-Automatic Speaker Identification System and the equipment manuals are listed in Appendix C.

CHAPTER 2. STATUS OF SEMI-AUTOMATIC SPEAKER IDENTIFICATION SYSTEM

The Semi-Automatic Speaker Identification System is a computer graphics console on which an operator can process voice recordings to identify speakers. The essential function of the operator is to locate and identify the different phonemes of speech that meet the system's requirements. Since the sonogram (voiceprint) of each speech segment is displayed for phoneme selection, it is recommended that the operator be a voiceprint examiner or experienced in speech analysis.

The speaker identification system was designed and tested under laboratory conditions, using a large data base composed of sound-booth recordings of male General American English dialect speakers. When ten different types of phonemes are found in the pair of quality speech utterances to be compared, accurate identification is found 97 percent of the time.

The results of limited testing indicate that the system's speaker comparison algorithms are appropriate for comparing female General American English dialect speakers, female Black Urban dialect speakers, female Chicano dialect speakers, speech with simulated stress, and speech with nasal disguises. It appeared, however, that the comparison algorithm would have to be redesigned in order to compare male Black Urban dialect speakers, male Chicano dialect speakers, and male Black Urban dialect disguises.

A number of tests revealed a very prohibitive problem in the system with respect to processing forensic evidence. Telephone response characteristics dominate the speaker comparison results in the Semi-Automatic Speaker Identification System. This problem is inherent, but was unforeseen, because the speaker comparison algorithm is based on the steady-state segments of phonemes as opposed to the transient portions of speech. Analysis showed that the telephone response can make the same speaker look more different than different speakers. Since nearly all crimes of voice are

perpetrated over the telephone, the system in its present state will have little use in identifying criminals.

Another problem with the processing of forensic evidence can be the level of noise in the recording. A controlled experiment indicated that the speaker identification system would produce correct results when the signal-to-noise ratio of the recording is 10 dB or higher. Incorrect results were found when the signal-to-noise ratio was only 5 dB.

Altogether, the Semi-Automatic Speaker Identification System has not been adequately tested on recordings affected by the following parameters: various recording equipment (portable tape recorders, pocket recorders, inexpensive recorders, battery recorders); noise (street noise, restaurant noise, cross-talk, electrical disturbances); echos; nonlinearities (automatic gain control, recorder amplitude clipping, phase distortion); disguises; stress; and dialects.

The speaker identification system is without many of the software modifications that were recommended for facilitating the operation and efficiency of the system. The nominal processing time for comparing two 20-second recordings is 4 hours. A major inefficiency is that the accidental hitting of the RETURN key on the input terminal can void whatever was processed on that recording.

The Semi-Automatic Speaker Identification System is installed at The Aerospace Corporation. The system has been operational until recently when a failure occurred in the keyboard input operation. It is expected that this problem will be solved shortly by the manufacturer of the computer console. As an aid to learning the operation of the system, a training manual has been written.

CHAPTER 3. GENERAL DESCRIPTION OF SEMI-AUTOMATIC SPEAKER IDENTIFICATION SYSTEM

The Semi-Automatic Speaker Identification System analyzes speech samples to identify and extract speaker-dependent features and to subsequently perform a statistical comparison of the features from different samples. The purpose of the system is to enable law enforcement personnel to compare the recorded voice of a criminal (e.g., from a bomb threat recording) with recorded voice samples from suspects to identify the perpetrators of crimes. The system uses a minicomputer and associated peripherals to accept analog speech signals for processing and statistical comparison. In the criminal speech sample processing operation, specific phonetic events that have been found to have a high degree of discriminating power are identified and labeled by the operator, using an interactive graphic display terminal. When a suspect sample is obtained, the same phonetic events are selected for processing. In the comparison phase, each selected event from the criminal sample is compared with a like event from the suspect sample. The points of comparison are well defined and yield quantitative results on a repeatable basis.

3.1 Speech Process

In our communication oriented society, speech evidence is becoming more and more available to the crime investigator. Often the speech itself is the crime, as in a telephone bomb threat or extortion by wire.

Speech is usable for identification because it is a product of the speaker's individual anatomy and linguistic background. When air is expelled from the lungs, it passes through the glottis, which is the opening bounded on either side by the vocal folds. When the vocal folds are drawn together and air from the lungs is forced through them, they vibrate, making a buzzing sound. The waveform of the glottal source is a series of pulses¹ as shown in Figure 1. This sound is modified as it passes through the vocal tract, which is the tube formed principally by the pharyngeal cavity (throat) and the

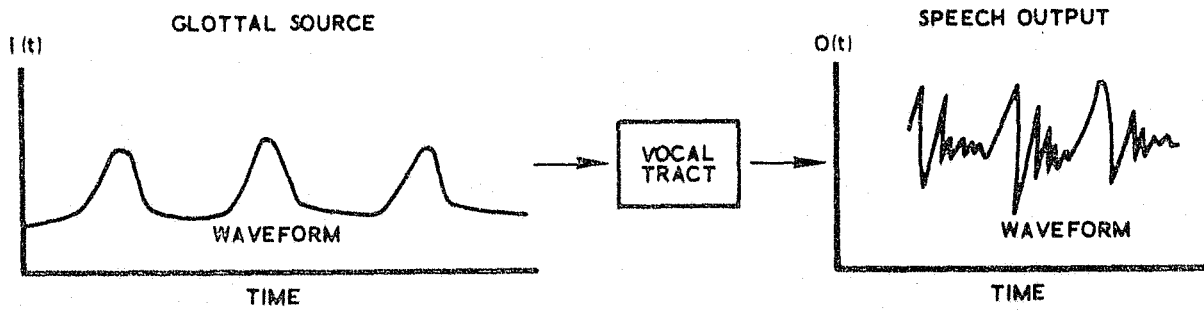
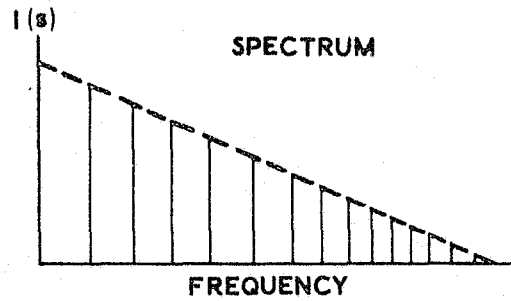


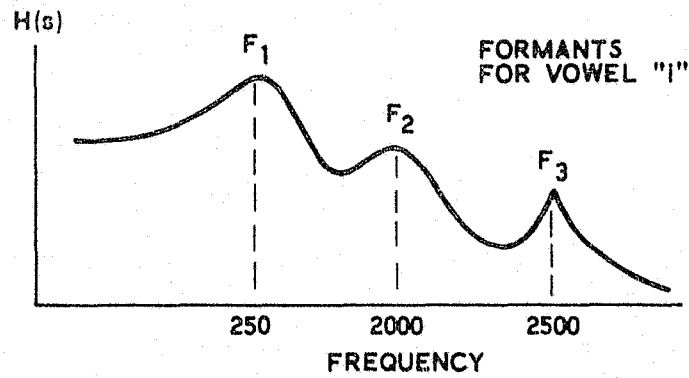
Figure 1. Linear Model of the Speech Process

oral cavity. The sound emanating from the vocal tract will be distinctly different from the initial buzz and will have a complex waveform as shown in Figure 1. The shape of the vocal tract serves to concentrate sound energy at certain frequencies and reduce it in others. Figure 2 illustrates how the spectrum of the glottal source is modified by the vocal tract. A transfer function can be used to describe the relationship between any input signal applied to the vocal tract and the resulting output signal. In transferring the acoustic energy from the glottis to the lips of the speaker, the vocal tract selectively emphasizes certain portions of the glottal spectrum in accordance with the particular transfer function it has at that point in time. During the speech, the shape of the vocal tract is continuously modified by movements of the tongue, lips, and other vocal organs. Thus, the quality of the speech sounds (phonemes) a speaker produces represents the size and shape of his vocal organs and the way he uses them in speaking. Speech characteristics vary from speaker to speaker. The effect is termed interspeaker (between speakers) variability. Speech analysis also reveals variability when the same speaker utters a given sound several times. This is called intraspeaker (within speaker) variability. Intraspeaker variability arises because the physiological activity necessary to make speech sounds need not be exactly controlled to effect adequate communication.

GLOTTAL SOURCE



VOCAL TRACT TRANSFER FUNCTION



SPEECH OUTPUT

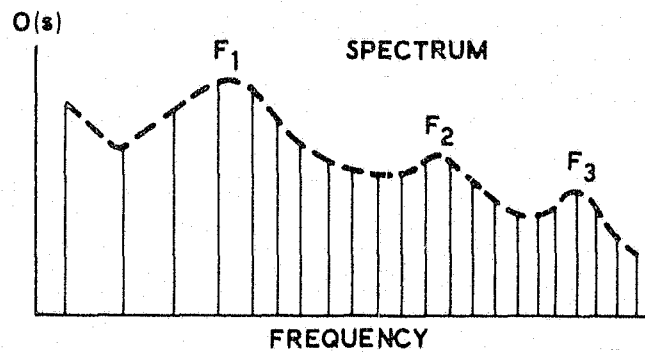


Figure 2. Spectral Transformation

The speech signal produced by a given individual is affected by both the organic characteristics of the speaker (in terms of vocal tract geometry) and learned differences due to ethnic or social factors.

From an individual's speech signal, a detailed spectral analysis can be performed to determine the shape of the vocal tract transfer function. Since the vocal tract transfer function exhibits several peaks that correspond to the natural frequencies of the vocal tract (formants), measurements of the properties of this function can provide indications of the unique manner in which a specific individual produces a given sound. Numerous such measurements, when properly combined, can provide information regarding the identity of the speaker.

The reliability of a speaker identification approach is related to the degree to which the interspeaker variability can be maximized relative to the intraspeaker variability.

3.2 Use of Voiceprints

A voiceprint (also called a sonogram) is a three-dimensional graph representing time, frequency, and intensity of speech sounds and is technically known as an acoustic spectrogram.² These characteristics, as illustrated in Figure 3, are depicted as follows:

- Time is represented from left to right along the dimension of the horizontal axis.
- Frequency is represented along the dimensions of the vertical axis, with height along the axis proportional to frequency.
- Intensity is represented along the dimension of a gray scale in which darkness is proportional to intensity. Thus, dark areas represent regions of relatively intense sound energy.

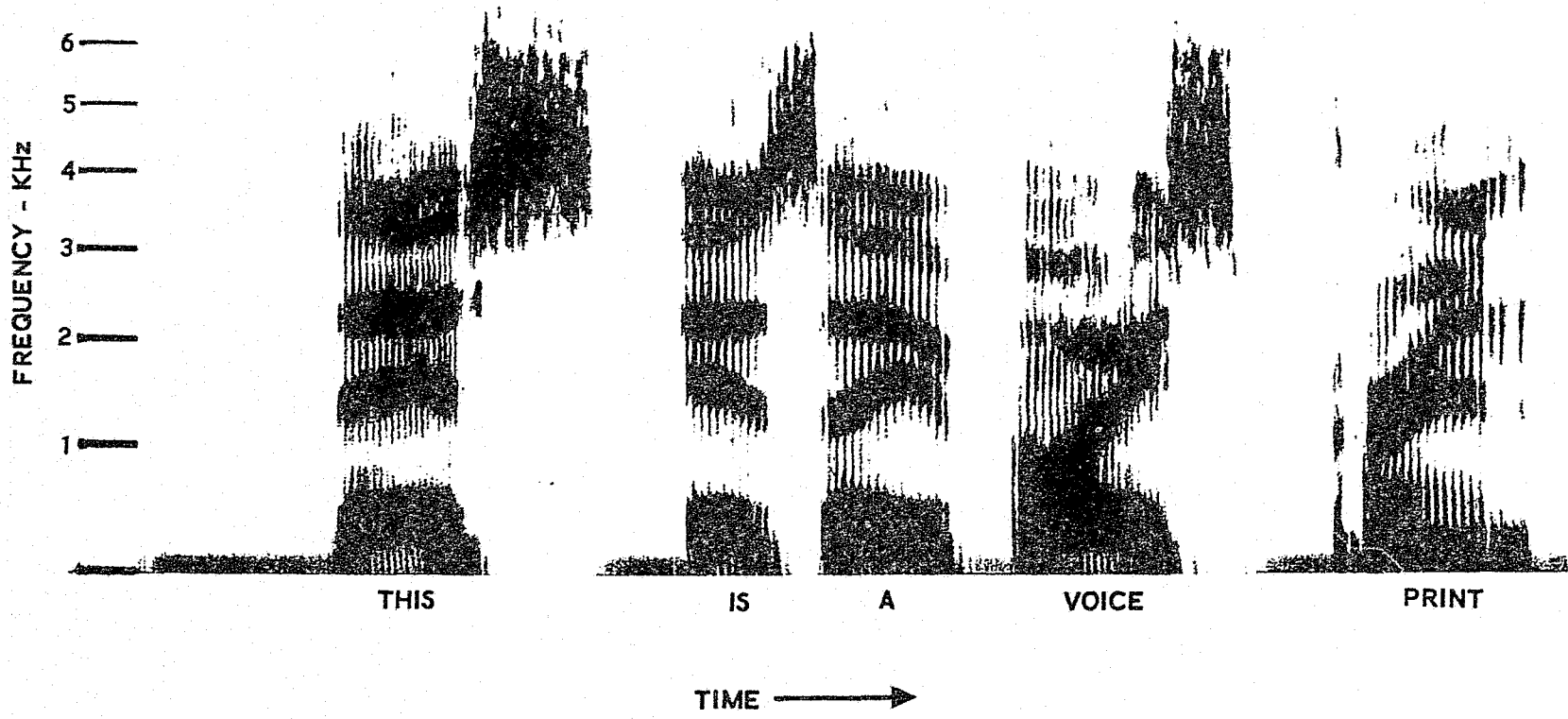


Figure 3. Labeled Spectrogram of "This is a Voiceprint"

In a way, the voiceprint is a representation of the vocal tract transfer function as it changes during the production of speech.

Voiceprints have been the subject of considerable controversy in the speech science community. Experiments aimed at establishing the accuracy of speaker identification using voiceprints have not resulted in widespread acceptance of the technique. The spectrographic identification of a voice by a trained observer appears to rely on a broad assessment of loosely defined points of similarity rather than a carefully specified set of objectively defined spectrographic attributes. This makes replication of experimental results by independent investigators highly unlikely and, therefore, prevents the acceptance of the technique as a recognized scientific procedure.³

Despite these shortcomings, the use of voiceprints for speaker identification is expected to increase over the next few years as police and criminalistics laboratory personnel are being trained in this technique. Recent judicial decisions regarding the admissibility of evidence and the manner in which suspects are interrogated have forced police agencies to rely more on new techniques and equipment for investigative purposes.

3.3 Semi-Automatic Speaker Identification System

The basic function of the Semi-Automatic Speaker Identification System is to extract speaker-dependent parameters or features from the speech signal and subject them to a statistical analysis. The basic techniques employed in the system were initially developed under a Law Enforcement Assistance Administration grant to the Stanford Research Institute and Texas Instruments Corporation.^{4,5} The major concept employed is that the steady-state regions of phonemes should be used for feature extraction since they are more speaker-dependent than the transient regions of speech.

The features extracted by the Semi-Automatic Speaker Identification System are directly related to either the power spectrum density function or the speech intensity time waveform of the steady-state segment

of selected phonemes. Formant frequencies, formant bandwidths, power spectrum density amplitudes, and power spectrum density slopes are features extracted from the power spectrum density curve, as shown in Figure 4.

The pitch period, the density of zero crossings, and linear prediction coefficients are features extracted from the speech-intensity time waveform, as shown in Figure 5. The linear prediction coefficient parameters are determined from a linear combination of discrete values of the amplitude of the waveform.

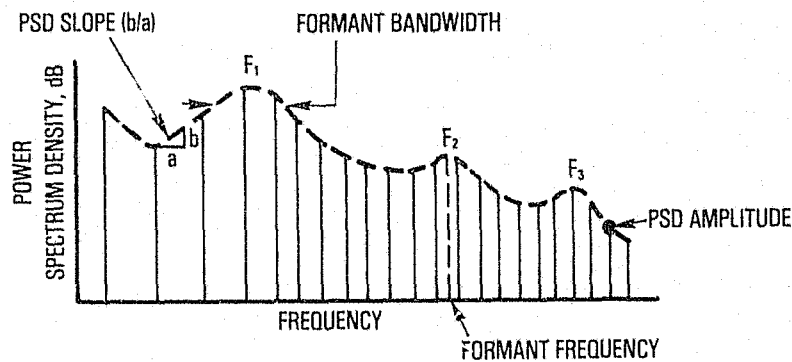


Figure 4. Frequency Domain Features

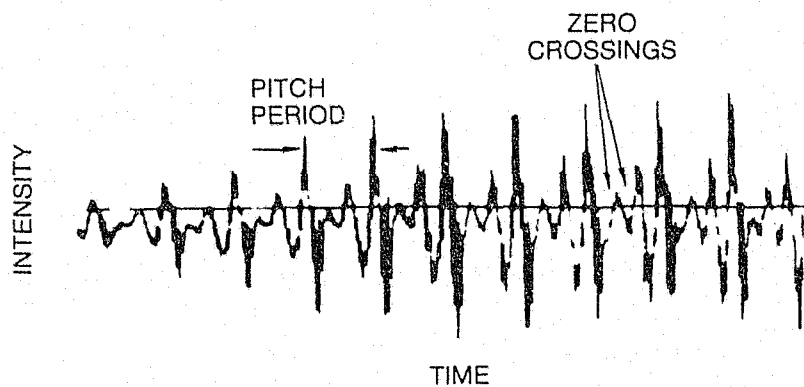


Figure 5. Speech Intensity Waveform

The computer system approach to speaker identification has two distinct advantages over the voiceprint examination technique:

- In programming the computer to make the identification, the expert is forced to clearly define what he considers to be significant similarities between the voices.
- Once the computer is programmed, the actual identification will be made on a consistent, objective basis.

Table 1 summarizes the basic functional differences between the two approaches.

Table 1. Functional Comparison of Voiceprint Examination and Computer-Assisted Speaker Identification Techniques

Voiceprint Examination	Computer-Assisted Technique
1. Decisions are subjective and dependent upon the individual examiner's expertness.	1. Decisions are objective — repeatable results may be obtained with different examiners.
2. Decisions are based upon loosely defined points of similarity.	2. Measures are made of differences between well-defined features of each voice sample.
3. Comparisons are qualitative in nature.	3. Quantitative measurements form the basis of comparisons.
4. Effects of distortion, disguise, etc., are difficult to assess.	4. Effects can be measured and the degree of confidence in the decision can be adjusted quantitatively.

3.4 Operation of Semi-Automatic Speaker Identification System

The system configuration is shown in Figure 6. It utilizes a general purpose computer coupled with high speed data processing and pattern recognition algorithms specially designed for the speaker identifica-

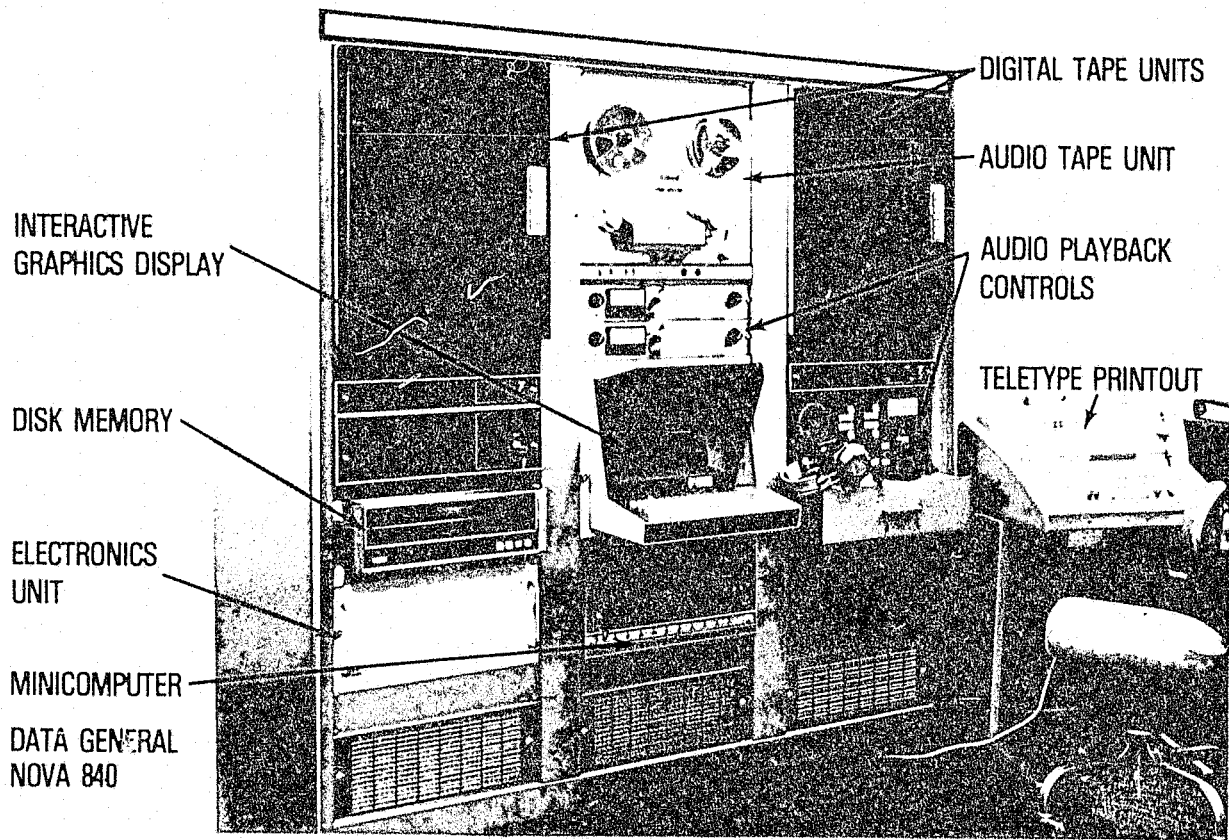


Figure 6. Semi-Automatic Speaker Identification System

tion task. By a combination of operator and computer functions, the parts of the speech sample that best contribute to speaker discrimination are selected and compared with other samples. The selected phoneme (speech sound) types that the system processes are listed in Table 2. On the basis of these comparisons, the computer can measure the degree of similarity between a criminal sample (e.g., from a bomb threat recording) and a sample from a suspect.

Table 2. Selected Phoneme Types

Alphaphonetic Symbol	Class	Example
MX	Nasal	<u>m</u> oon
NX	↓	<u>n</u> o
NG	↓	si <u>ng</u>
EE	Vowel	<u>e</u> ve
IX	↓	<u>i</u> t
EH	↓	<u>m</u> et
AH	↓	<u>a</u> sk
AA	↓	<u>f</u> ather
AW	↓	<u>a</u> ll
UX	↓	<u>u</u> t
UU	↓	<u>b</u> oot
UH	↓	<u>u</u> p
ER	↓	<u>b</u> ird

Figure 7 illustrates the overall operation of the system.⁶

Criminal speech samples from police station monitors, and covert recordings or authorized wire taps, are processed and stored on digital magnetic tape. In the processing operation, specific phonetic events that are known to have a high degree of discriminating power are identified and labeled. When a suspect sample is obtained, the same phonetic events are selected for processing. In the comparison phase, each selected event from the criminal sample is compared with a like event from a suspect sample. The points of comparison are well defined and yield quantitative results. The system is thus able to generate accurate and objective results on a repeatable basis.

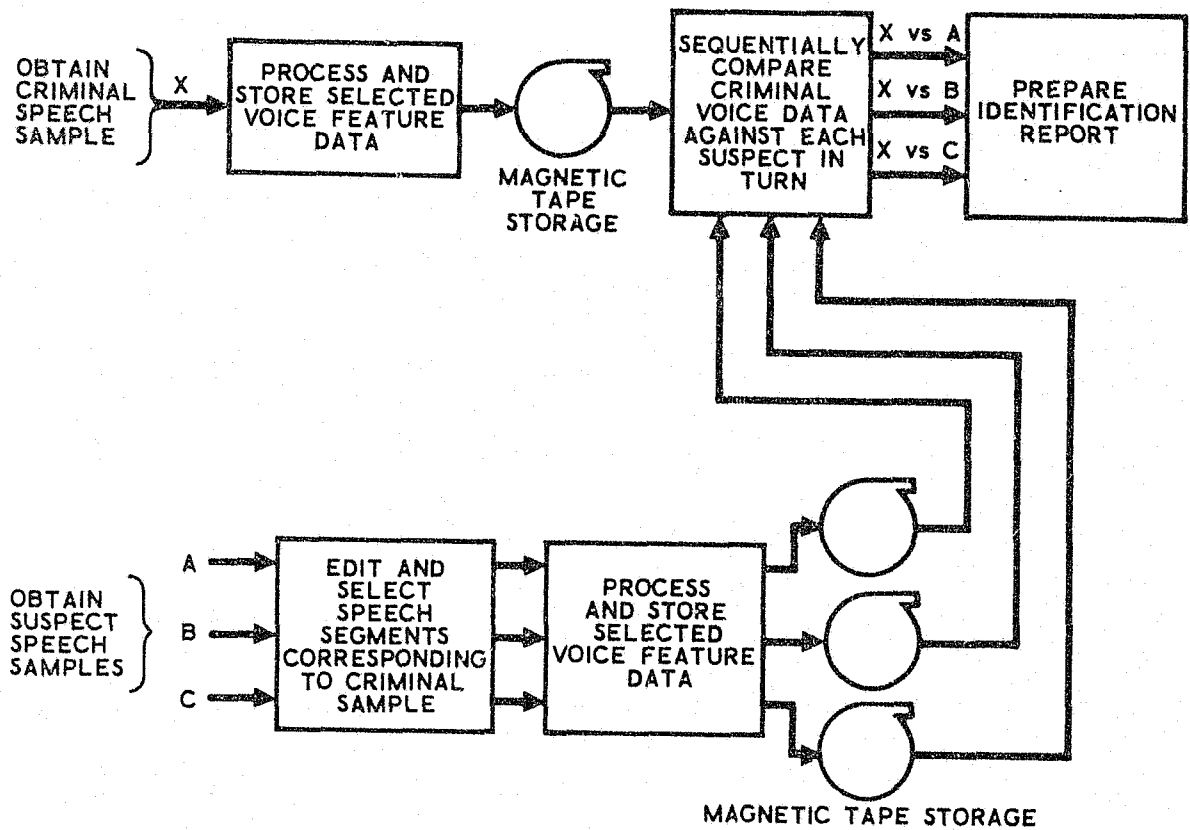


Figure 7. System Operation

Prior to using the equipment, the operator listens to and writes down the words spoken in the speech sample. He then prepares an alphabetic transcription, which separates the words into their phonetic parts. Figure 8 is an example of such a transcription. For example, the word "six" would be transcribed as /SX IX KX SX/, where each pair of letters in the transcription denotes a particular phonetic event. By examining the transcription, the operator identifies the phonetic events that are useful for comparison purposes. For the word "six," the "IX" phonetic event would likely be selected for comparison. In general, phonetic events representing vowel or nasal sounds have been found to be most useful for speaker identification.

ORIGINAL TEXT

HAVE THE MONEY READY BY SIX

ALPHAPHONETIC TRANSCRIPTION

| HX AH VX | DH UH | MX UH NX EE | RX EH DX IX | BX AA IX | SX IX KX SX |

Figure 8. Example of Alphaphonetic Transcription

After the operator identifies the phonetic events to be analyzed by the computer, he inputs the speech segment selected for processing into the system. The speech is sampled and digitized by the system, and the resulting digital data are stored in the system. The sampling rate in the prototype system is 6800 samples per second, which is great enough that no useable information in a telephone bandwidth sample is lost. The present prototype configuration allows input of up to 276 seconds of speech from each speaker.

When the operator is satisfied that the speech data have been correctly entered into the system, he directs the system to transform the data into a representation that shows the frequency distribution of the speech signal at each instant of time. This representation is called a sound spectrogram or sonogram. The operation of the system is illustrated in Figure 9. The operator can cause the sonogram to be displayed on the system graphics terminal. Each screen-full, or frame, of the display represents 1.1 seconds of speech data and, typically, will contain from one to four phonetic events useful in the comparison. The upper portion of Figure 9 shows a typical sonogram display.

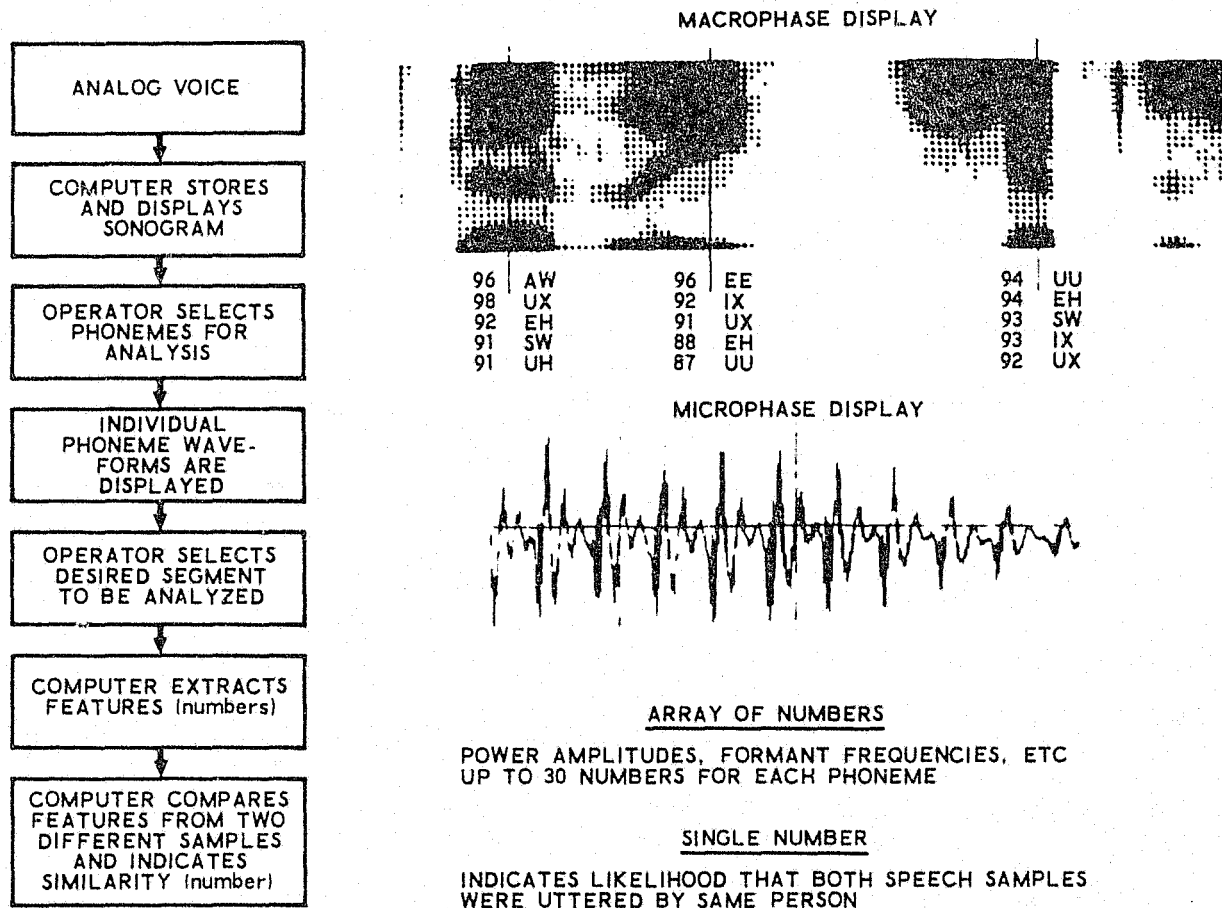


Figure 9. Computer-Aided Voice Comparison

The sonogram display is used in the portion of the interactive labeling procedure known as the macrophase. During this procedure, the operator identifies and labels, on the interactive terminal, the phonetic events that are to be used in the speech sampling comparison. Events are pointed out and identified by the use of a thumbwheel control that positions an interactive graphic cursor (electronic crosshair) superimposed over the sonogram. Through keyboard input, the operator controls audio playback, as well as up to 30 seconds of the entire speech segment that was input. A further aid in identifying and labeling phonetic events is the capability to compare selected segments of the sonogram against a reference inventory. When the operator commands a comparison by typing "C" on the keyboard, the computer displays the alphaphonetic names of the five phonetic types whose spectra correlate best with the spectrum of the speech segment pointed to by the cursor. The correlation values are also shown as numbers between 0 and 99.

Since individual phonetic event characteristics are dramatically affected by the phonemes adjacent to the target event, both the target event and the two adjacent events are labeled and subsequently used in the event comparison. The three events thus labeled are referred to as a phonetic triad. Tentative acceptance of an event is made by labeling and numbering the phonetic triad in which the desired event is centered. The operator will label all phonetic events of interest in a given sonogram frame. After he has finished the macrophase for a sonogram frame, he signals that fact to the computer through the keyboard. The computer then automatically proceeds to the microphase of labeling. In the microphase, a 100-millisecond segment of the speech waveform is displayed for each of the events labeled in the macrophase. The microphase serves two purposes. First, it allows the operator to reconsider his macrophase selections by viewing and listening to the speech waveform in selected short segments. Secondly, having confirmed his selection, the operator must use the graphic cursor to mark off

three consecutive pitch periods of the speech waveform for each selected event. This is required to ensure precision in the pitch-synchronous spectral analysis calculation that is subsequently performed.

After sequencing through the microphase display for each macrophase selection, the system returns to the macrophase, and the operator may label another sonogram frame. The alternation between the two phases of labeling continues until the operator has labeled every event of interest in the speech sample.

After labeling, the computer proceeds to compute the measurements or features on each labeled event that will be used for comparison. For each of the 13 event types that are allowed in the prototype system, there is a unique set of 30 features. Once the features have been calculated, the voice sample can be compared with any other voice sample that has been similarly processed.

The detailed comparison process is shown in Figure 10. Each speech sample goes through the same series of steps whereby the sample is digitized, sonograms and other displays are generated, and phonetic events are selected. For the speech sample of speaker A, the selected events could be designated 1A, 2A, 3A . . . Each event will produce a set of 30 features. Event 1A will thus produce features 1A1, 1A2, 1A3 . . . 1A30, as shown. Event 1B from speech sample B will likewise produce feature set 1B1, 1B2, 1B3 . . . 1B30. The two feature sets are combined in a manner that produces a distance measure set. The manner in which the numerical distance is derived is such that the widest separation between different speakers is achieved while maintaining the smallest distance between different utterances by the same speaker. A distance measure is obtained for each pair of selected phonetic event triads in the sample. Only like events are compared since the suspect exemplar is the same as the criminal utterance. Finally, the various distance measures are combined to arrive at an overall speaker

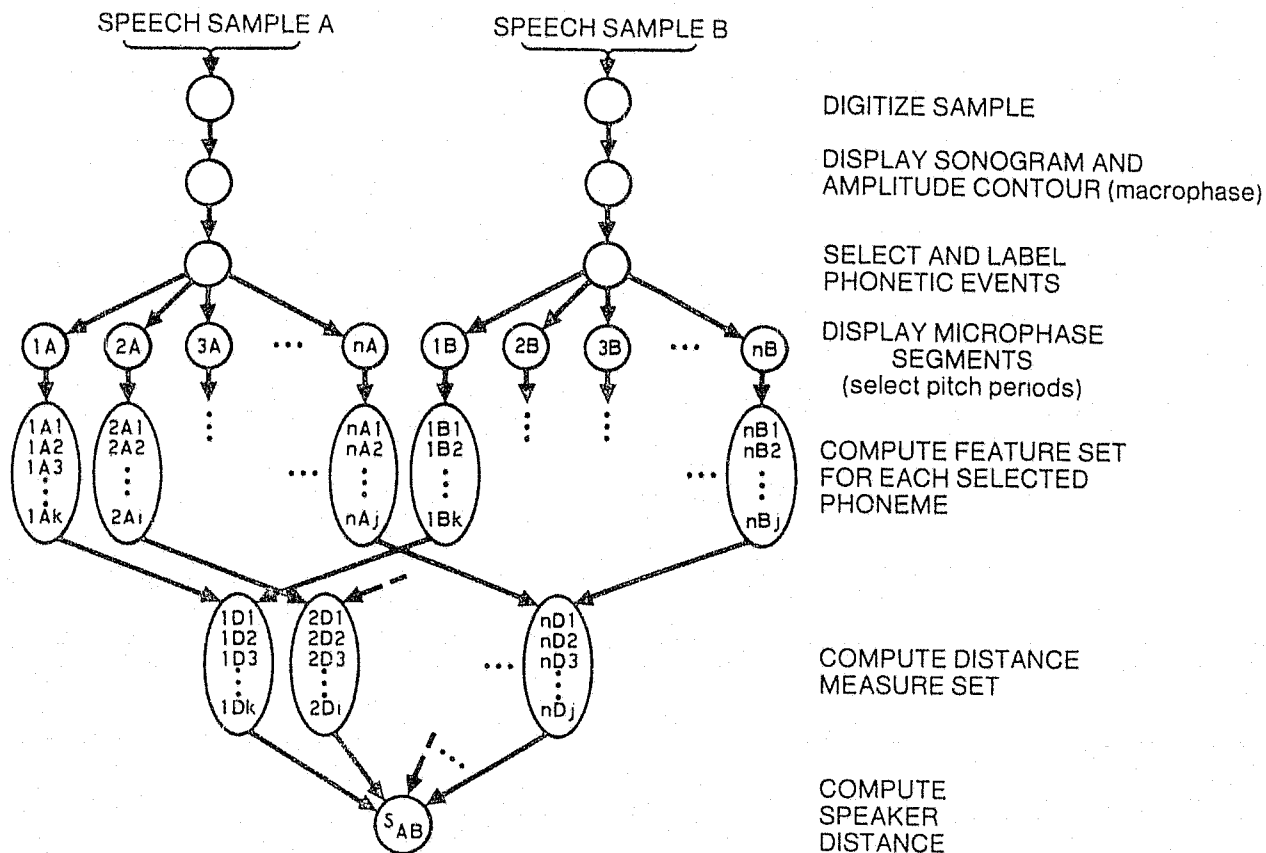


Figure 10. Semi-Automatic Speaker Identification System Identification Process

distance for the two samples. As before, the method of combination is selected to maximize the system's speaker discrimination capabilities.

To interpret the meaning of the distance measure, the operator is supplied with a set of statistical performance data which tabulates the speaker distance values obtained when comparisons were made of speech from a representative sample of speakers. The comparisons were made between utterances made by the same person on different occasions (to obtain intraspeaker distance measures) and between utterances made by different speakers (to obtain interspeaker distance measures). Since the numerical values of the similarity measures obtained from voice sample comparisons are strongly dependent upon the types of phonetic events in the samples, the comparisons were made independently for every possible combination of selected phonetic events that could occur in a speech sample. The operator, therefore, consults the specific table of performance statistics corresponding to the set of events that occur in the speech sample being compared.

3.5 Performance of Semi-Automatic Speaker Identification System

Figure 11 shows the results of the speech sample comparisons for two different sets of phonetic events. The curves on the left were obtained when utterances containing only the phonetic event type IX (as in "six") were compared. On the right side, curves are presented that were obtained when utterances containing eight phonetic event types were compared. The top curves show the frequency of occurrence of speaker distance values for intra- and interspeaker comparisons. As the curves indicate, intraspeaker comparisons tend to result in smaller values of speaker distance than do interspeaker comparisons. Also, comparisons made with more phonetic events tend to result in better speaker discrimination, as indicated by the smaller degree of overlap between the intraspeaker and interspeaker curves when more events are employed.

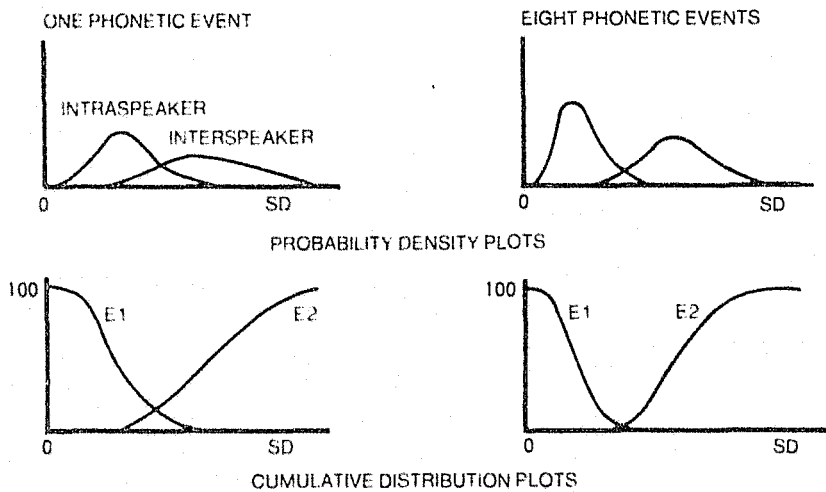


Figure 11. Typical Evaluation Test Results

The lower set of curves in Figure 11 represents the cumulative error distribution that would result if specific values of speaker distance were used in making decisions regarding whether two voice samples came from the same or different speakers. The curves are plots of the areas found under the "tail" of the probability density curves for specific values of speaker distance. If it were postulated that two speech samples can be considered to come from the same person if their comparison results in less than a given threshold value of speaker distance, then E2 would be the probability that a false identification will be made when the two samples come from different speakers. Similarly, E1 is the probability that an error will be made when the two samples come from the same person, if the speaker distance is greater than the given threshold value. The cumulative errors, as well as the values for the probability density, are listed in tables for each value of speaker distance, as shown in Table 3. Also tabulated is lambda, which is a measure of the likelihood that two speech samples from the same speaker will have the same speaker distance; or SD listed in the table. Lambda is computed from the formula $E1/E2$, but is only allowed to have values in the range between 0.01 and 100 since values outside of this range are difficult to interpret.

Table 3 relates to one specific set of phonetic event types. There is a unique table for every possible set of event types that can be encountered by the system. If an operator compares two speech samples containing the set of phonetic events listed at the bottom of the table (i. e., NG as in saying, EE as in feet, IX as in mix, UH as in but, ER as in hurt, EH as in met, AA as in father, and UX as in foot), he will enter this particular table with his computed value of the speaker distance between the two samples. For example, if the computed speaker distance (SD) was 13, the operator would find the closest values of SD listed in the table; in this case, 13.07. For this SD value, the table lists a value of 9 for E1; 1 for E2; and 8.18 for lambda. If the speech samples satisfy the same general conditions for which the performance statistics were calculated (e. g., General American English spoken by a male), the operator can make the following statements for the given sample:

- A speaker distance of this value or smaller will occur 91 percent of the time with speech samples from the same person.
- A speaker distance of this value or smaller will occur only 1 percent of the time with speech samples taken from different persons.
- It is about eight times more likely that the samples came from the same speaker than from different speakers.

The speaker data base used in the Semi-Automatic Speaker Identification System development is the largest of its type in existence. The fact that it is a sample of the total population, however, results in some statistical uncertainty in applying the results to the general population. Standard statistical techniques (such as those used by insurance companies and pollsters) were used to compute the confidence levels of the performance statistics, using worst-case assumptions. Figure 12 shows the upper and

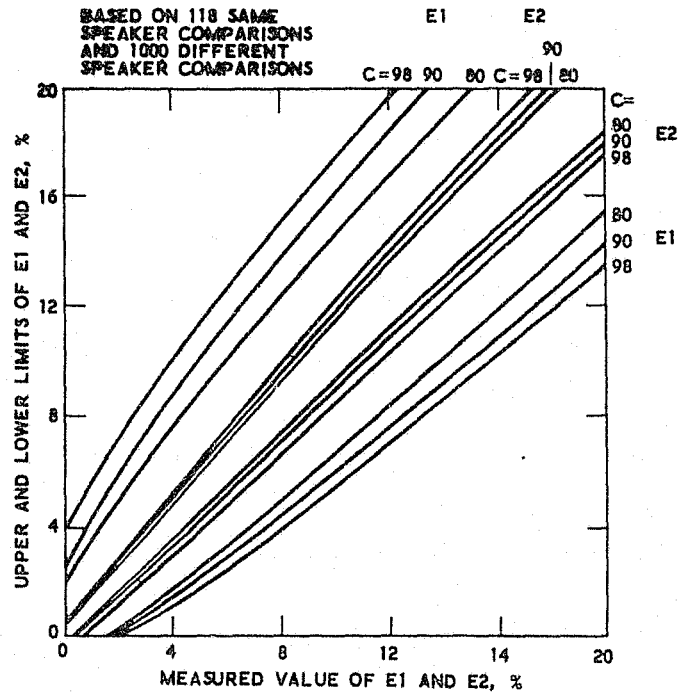


Figure 12. Example of Confidence Estimates on Semi-Automatic Speaker Identification System Statistics

lower limits to the E1 and E2 error probabilities at three different confidence levels. If the previous values of E1 and E2 are used, the figure shows that it can be stated with a 98 percent confidence that the true value for E1 is between 4.5 and 16 percent, and the true value for E2 is between 0 and 2 percent. These error limits indicate to the operator the statistical uncertainty associated with the measurements made by the system. These data allow the operator to assign a "weight" to the system results when he is using these results to arrive at a decision regarding the voice samples under consideration.

This system offers a greater degree of flexibility than any previous approach. It performs comparison on diverse, frequently occurring phonetic events and will analyze these phonetic events in any of a large number of combinations. The system was designed to use the best natural skills of the operator and the computer to arrive at a numerical, repeatable, and

objective conclusion. The goal of this approach was to overcome many of the objections that hamper the use of voice identification as courtroom evidence.

CHAPTER 4. HISTORY OF PROGRAM

Speaker identification through the use of voice spectrograms has become an important source of evidence in investigative activities and in criminal court proceedings. The current identification techniques, however, are hampered by time-consuming, manual methods as stated in Chapter 1.

In an effort to promote the development of improved methods and equipment, the National Institute of Law Enforcement and Criminal Justice has supported a number of efforts to evaluate the effectiveness of speaker identification techniques. The results of studies, conducted by Stanford Research Institute and Texas Instruments, for machine-assisted speaker identification research were comparable and indicated that the method looked promising. It was recommended that a comparison algorithm be designed using steady-state features of speech.

Based on the desire to improve the use and validity of voice spectrograms in law enforcement, the Institute awarded a contract to The Aerospace Corporation in FY 73 to initiate a comprehensive program on speaker identification. The objectives of the program were to:

- Conduct investigations to improve voice spectrogram technology and properly validate areas for its applications.
- Provide an interim semi-automated speaker identification system, along with statistical evidence of its capabilities.
- Provide a voice classification system to allow the search of large files for identification purposes.
- Provide new techniques for voice identification.

It was expected that the program would extend over several years and that most of the effort would be subcontracted.

4.1 Development of the Semi-Automatic Speaker Identification Program

Near the end of FY 73, Aerospace solicited proposals for the research and development of a Semi-Automatic Speaker Identification System. Rockwell International was selected as the subcontractor out of nine competitors. Aerospace also solicited sources to conduct a voiceprint evaluation test program and recommended that a subcontract be awarded during the following year to conduct the test design, test operations, and data analysis of the voiceprint technique.

In addition to the subcontracting support, selected in-house studies were also undertaken by Aerospace. The theory and status of voiceprint technology was reviewed, and a "Voiceprint Applications Manual" was prepared and published.² The purpose of this manual is to upgrade voiceprint practice by giving potential users an understanding of the principles of voiceprint analysis and knowledge of correct practices in collecting and submitting voice samples for evaluation.

A second in-house effort undertaken at the direction of the Institute was a system study analyzing the recording of illegal telephone calls. The results of this study were also published.⁷ It was determined that compact, portable recording equipment costing about \$700 per unit could be installed on the premises of a person receiving illegal telephone calls and could provide recordings acceptable as court evidence. However, a system analysis investigating the factors involved in recording at a customer's telephone or at the local telephone exchange showed that central recording can be done more cheaply. Aerospace also recommended that no further Institute-sponsored effort be expended at that time on telephone recording projects associated with speaker identification.

The focus of the FY 74 activity was the development of the Semi-Automatic Speaker Identification System.^{6, 8, 9, 10} The Rockwell contract called for definition within 60 days of certain analytical tasks that

were to be added to the basic contract. The intent of this added analytic effort was to provide for a more effective and useful hardware development. A subcontract change was negotiated to include these additional analytic tasks. Involved was the collection and utilization of a larger data base, including Black Urban and Chicano dialects, and telephone channel conditions, from diverse caller locations. The recording of the data base is shown in Figure 13.

4.2 Speaker Data Base and Comparison Algorithm

By June 1974, the hardware shown in Figure 14 had been assembled and tested. Subsequent contractor activity was devoted to the development of complementary operational software, including generation of a speaker data base containing the variations among different speakers. The data base also included temporal variation for the same speaker, dialect differences, channel variations, and the effects of co-articulation on individual speech variation. These data were generated to conduct laboratory testing and provide the base for statistically establishing the accuracy of individual voice identification tests.

The speaker comparison algorithms were developed using the best 30 of 165 identified features in each of thirteen vowel and nasal steady-state sounds. Figure 15 gives an example of four comparison features extracted from the power spectrum density of a phonetic event. These four examples are the slope, bandwidth, second formant,^{*} and amplitude. Other feature examples would be pitch period, zero crossings, and linear prediction coefficients from the speech intensity waveform (intensity versus time).

* A formant is a natural frequency of the sound produced by the vocal organs. It varies with individuals for a given phonetic event and with events for a given individual.

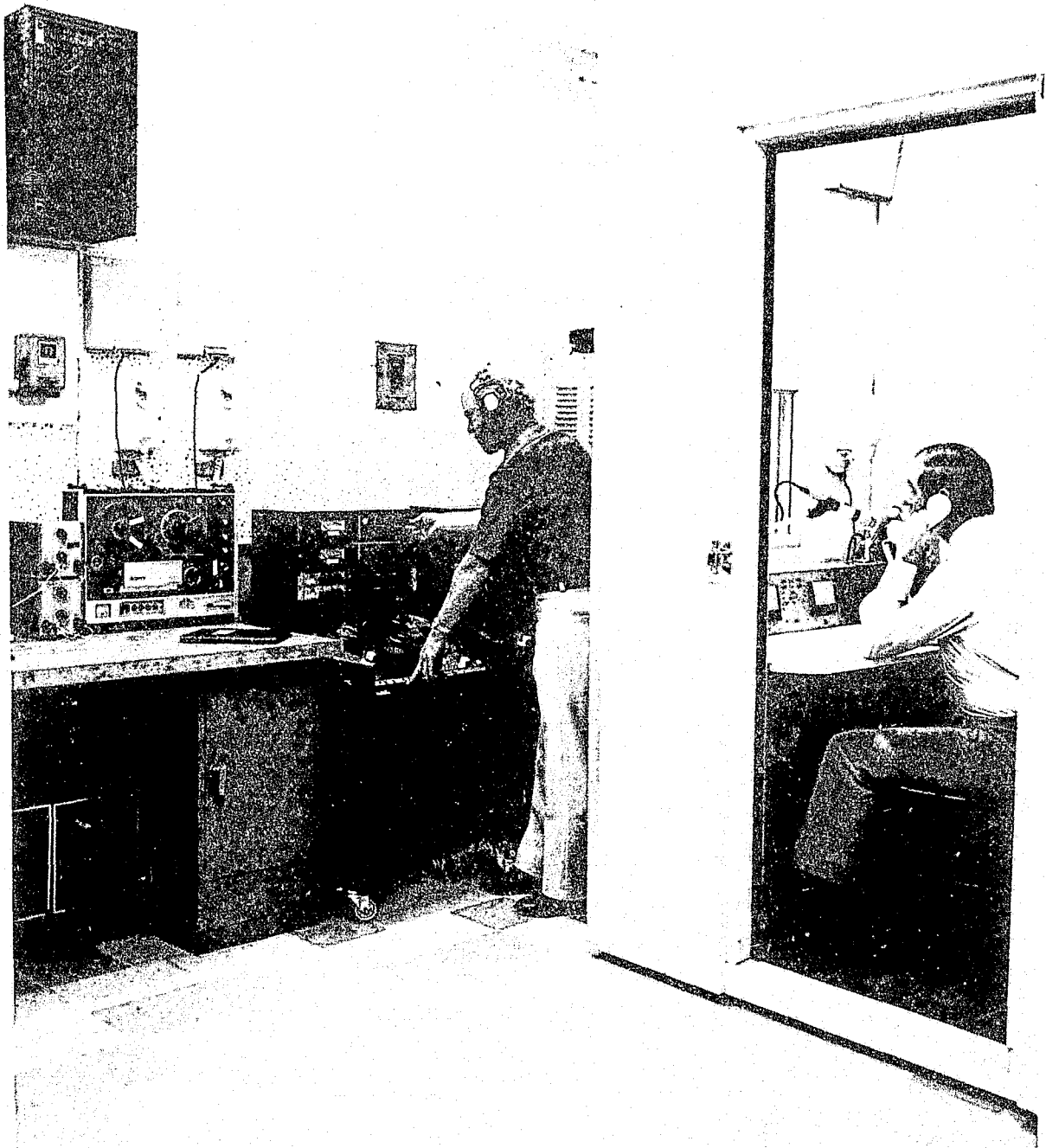


Figure 13. Recording Data Base under Laboratory Conditions

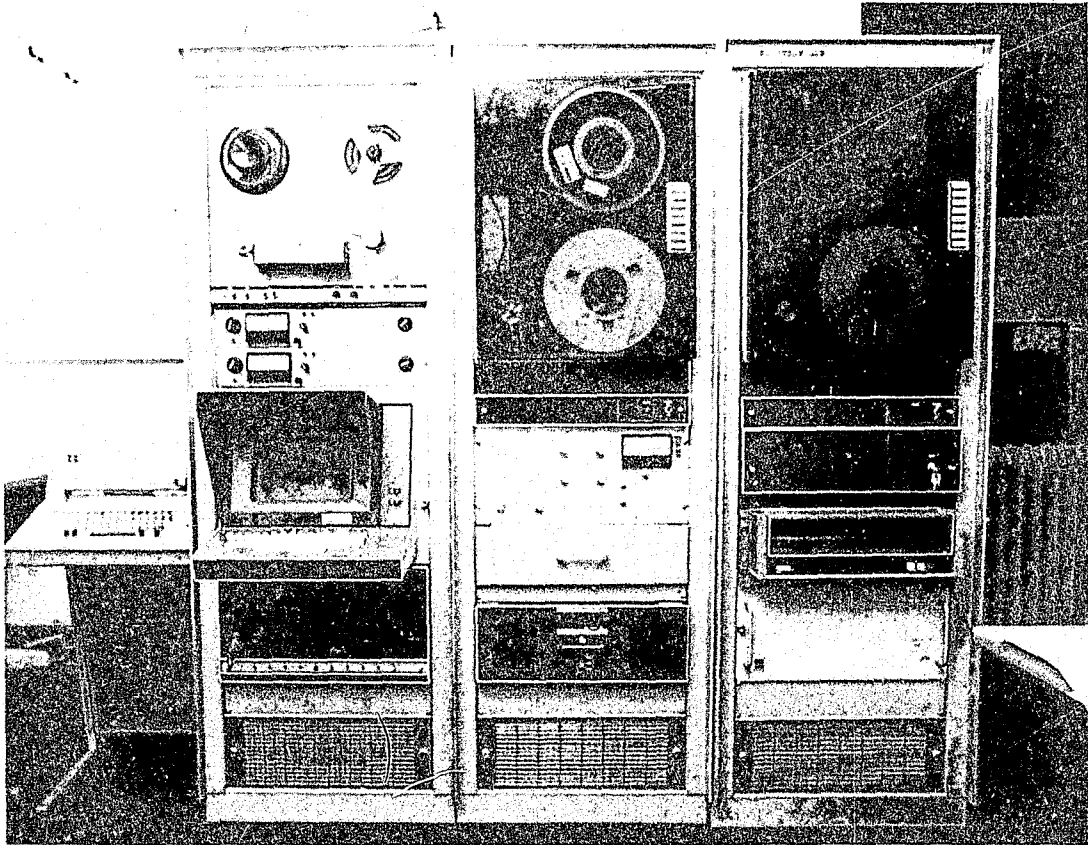


Figure 14. Semi-Automatic Speaker Identification System

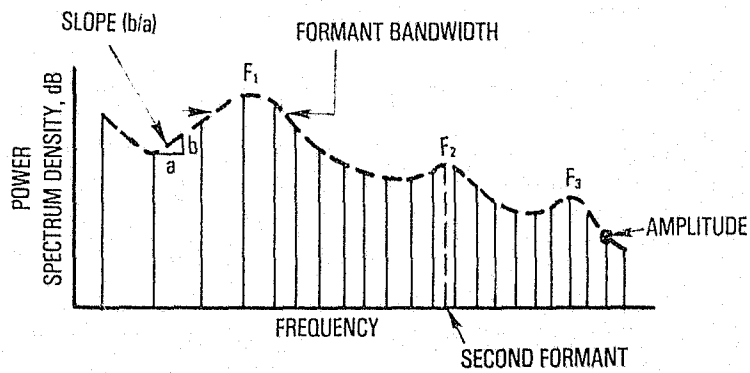


Figure 15. Examples of Comparison Features

Supporting FY 74 Aerospace effort included an independent assessment of the feasibility and concepts upon which the semi-automatic system development is based.^{11, 12} From the data base assembled by Rockwell, Aerospace analyzed the algorithm used to select and process specific phonetic events. A high-speed computer coupled with a stepwise discriminant analysis program was used to select the best subset of features for each such event. It was concluded that, by utilizing the best features, a decision based on a single phoneme can achieve an accuracy of 60 percent in selecting one out of 25 speakers. If six phonemes are used in making the decision, an accuracy very close to 100 percent can be achieved. Noise, distortion, and other influences occurring in the field are expected to influence this accuracy. Nevertheless, the discriminating power of the features chosen was confirmed, and the overall design of the system was validated.

In addition to support for the Semi-Automatic Speaker Identification System, an Aerospace study was also made of means for extending the courtroom use of voiceprints as evidence. Methods for obtaining test data and the statistical basis for a test program to ensure test result acceptability were considered. A concept development plan was subsequently published, and Aerospace continued its effort to subcontract the Voiceprint Validation Test.

In FY 75, the portion of the data base previously reserved was used to obtain system performance data. Using the high-speed data processing facility in their speech recognition laboratory, Rockwell exercised the feature extraction and comparison algorithms with a total of approximately 12 million separate speech sample comparisons. The data obtained verified the effectiveness of the system software and provided a statistical reference whereby the measured quantity of difference between speech samples may be compared to equivalent measures obtained from the sample population of speakers. In this manner, the likelihood that two

speech segments were uttered by the same or different people may be calculated. Figure 16 shows plots of performance data for several sets of possible phonetic event categories and the decision limitations on the system by the availability of comparable sounds (events) between two speech samples. The performance data are expressed in terms of the speaker distance (calculated by the system) which is a composite measure of the degree of similarity between two speech samples. The larger the speaker distance, the more dissimilar the two samples are.

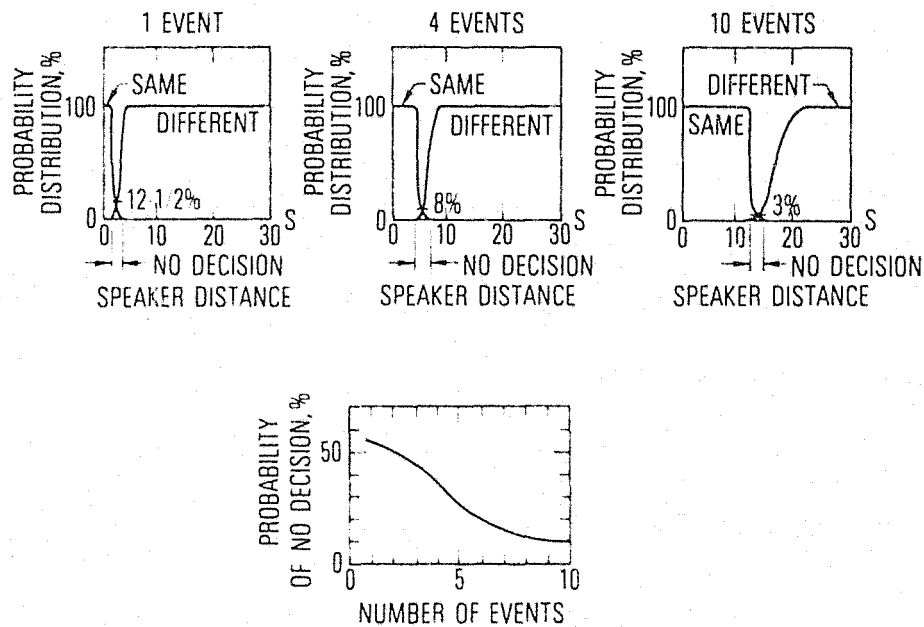


Figure 16. No-Decision Probability for Speaker Identification

Since a total of 1023 sets of the ten possible event categories can occur in a given speech sample, the performance data are described by a set of 1023 such plots. As Figure 16 shows, the system calculates significantly smaller speaker distances for repeated utterances by the same speaker than for utterances by different speakers. The figure also shows that the separation of the values of speaker distances for same or different

speakers increases as the number of events compared increases. The degree of overlap is such that, if decisions regarding a match or an elimination were made on the basis of at least a 99 percent probability of the choice being correct, the percentage of cases in which no decision could be made (the uncertainty region) is as shown in the lower curve of Figure 16.

4.3 Laboratory Test

Concurrent with the completion of the development effort on the brassboard Semi-Automatic Speaker Identification System, a follow-on subcontract was awarded to Rockwell International for the conduct of laboratory and pilot field test activities. The purpose of the laboratory test effort was to provide information regarding the performance of the system when processing speech samples different from those of the data base used in the development phase.¹³ These samples included female voices of various dialects, speech containing effects of intentional disguises and emotional stress, and speech recorded under simulated real-world conditions.

The results of the laboratory test were statistically insignificant because of the limited number of samples processed. However, the Semi-Automatic Speaker Identification System appeared invariant across female standard speech, Black Urban and Chicano female dialects, nasal disguises, and speech with simulated stress. The system appeared variant with Black Urban and Chicano male dialects and with Black Urban disguises. The most significant result was that the telephone channel response dominated the speaker comparisons. This shortcoming meant that a pair of recordings by a given speaker, one made directly from a tape recorder and another made over the telephone, look more unlike to the speaker identification system than would two different speakers.

An actual case sample from the police department of a major city was processed on the Semi-Automatic Speaker Identification System at the request of the Institute. The noise background and unlike texts violated current system constraints but pointed to improvements required to the system.

Test planning was completed for the pilot field test of the brassboard system with a law enforcement group having voiceprint capability and interest. The pilot test plan was prepared by Aerospace and approved by the Institute.

Other Aerospace activities included conduct of a study to assess the potential applications of the computer-aided speaker identification system.

A final report was prepared which covered present and future uses of voice identification in the law enforcement and criminal justice community, and problems with current methods.¹⁴ Estimates were made regarding how the use of voice identification can be expected to change as a consequence of supplementing the manual examination of spectrograms with the computer-aided system. Technical requirements were defined for applications such as computer access security, area access security, identity verification for check and credit card usage, and remote identity verification by police in the field. The principal technical factors addressed were verification accuracy of the equipment when used for automatic identification and the requirements for generating and maintaining a data base of individual speech characteristics for use in identification.

4.4 Pilot Test

In FY 76, the Semi-Automatic Speaker Identification System was pilot tested by three voiceprint examiners in the Voice Identification Laboratory of the Los Angeles Police Department, as shown in Figure 17.¹⁵

The pilot test began with a one-week training session for the participants on the operation of the system.¹⁶ The three voiceprint examiners then individually processed speaker identification cases on the system on a part-time basis over the following five months.



Figure 17. Voiceprint Examiners and Aerospace Program Manager at Los Angeles Police Department Pilot Test Site

In all, 22 cases were processed, including operator variability test cases, cases from actual forensic evidence, and simulated cases generated by the Association of Official Analytical Chemists for evaluating and subsequently endorsing voiceprint examiners. Of the 40 pairs of speaker comparisons processed, 13 comparisons yielded incorrect results on the system. For the forensic cases, the results of comparison by the voiceprint method were considered the correct results.

The major problem encountered during the pilot test was the same as that encountered during the laboratory test phase: the telephone channel response dominates the speaker comparison. Figure 18 shows how the telephone response can differ from that which was used in the system development.

During development, a common telephone simulator was used for all laboratory recordings. Certain features used in developing the speaker comparison algorithms are dependent upon amplitude, as was seen in Figure 15. The channel effects can thus have significant distortion effect, as can be seen at the higher frequencies in Figure 18.

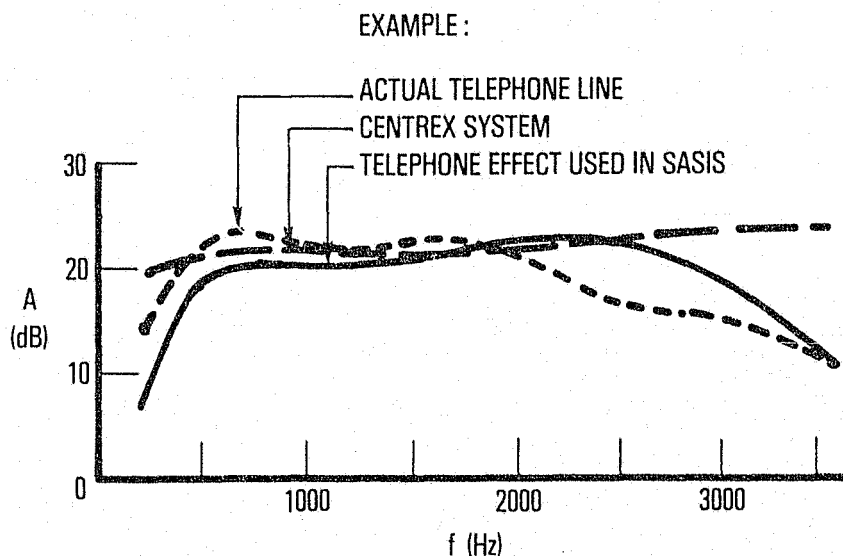


Figure 18. Frequency Response of Telephone Channels

Approximately half of the incorrect results were due to the channel effect and half were due to operator errors made in operating the computer-aided system. The operator errors can be avoided in the future, based on software modifications made after the pilot test to simplify operator input and recall. Thus, the pilot test achieved its primary objective of evaluating new operator interactions with the system and determining operational limitations of the system. The anticipated problem of noise effects was not evaluated during the pilot test since the voiceprint examiner casework contained recordings with signal-to-noise ratios of 10 dB or greater. Preliminary evaluation of the noise problem on some San Diego case tapes early in the year indicated that a signal-to-noise ratio of at least 4 to 6 dB was needed.

4.5 Potential Solutions to System Problems

As discussed above, the pilot test demonstrated other shortcomings in the operation of the Semi-Automatic Speaker Identification System, each accompanied by recommended design improvements. Some recommendations, including improvements of the man/machine interaction, were made under the remaining funds on the Rockwell subcontract; but others, including the system incorporation of a channel filter to suppress the channel effect, were beyond the scope of the subcontract. Because of the presence of the overriding telephone channel problem and the limitation of available funds, the anticipated problem area of noise was again not evaluated. It was believed, however, that preprocessing concepts using noise filters could alleviate the problem for correlated noise, echos, and frequency varying hums when it was formally addressed.

Aerospace conducted an independent investigation of solutions to the telephone channel problem. A number of companies and universities having potential solutions were contacted, and two of these potential solutions looked promising. The first was a channel deconvolution algorithm, which is a computer algorithm that filters the channel response from

digitized speech. Since this algorithm is a preprocessor, the present speaker identification system design would be unaltered.

The second potential solution was to redesign the decision algorithm to include a dynamic feature of speech, in addition to the steady state features. The dynamic feature considered was the center frequency trajectory of the first, second, and third formants, as shown in Figure 19. It was believed that the time waveforms of the formant frequencies were relatively invariant with telephone channels, and it was known that they were speaker-dependent.

Simulated cases of telephone recordings that were generated by Aerospace were processed by the University of Utah, Westinghouse Electric, and the System Development Corporation on their speech systems.

The University of Utah, the original developer of the channel filter, calls its algorithm "Blind Deconvolution." The recordings were filtered by Blind Deconvolution and sent to Rockwell for speaker comparison on the Semi-Automatic Speaker Identification System. The results from

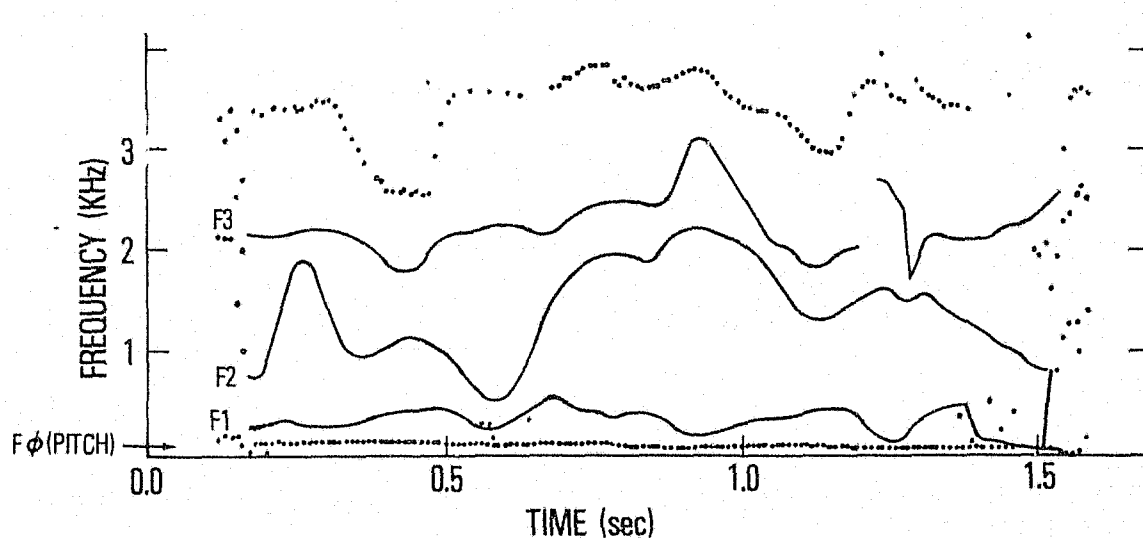


Figure 19. Formant Trajectories

this limited experiment were that deconvoluted recordings of duration of ten seconds or more yielded correct results, whereas the unprocessed telephone recordings yielded incorrect results. Deconvoluted recordings of five seconds of speech did not show improved results, however.

Westinghouse Electric has developed an automatic speaker verification system that has an analog circuit which tracks the center formant frequencies in real time, and a simple comparison algorithm. They processed ten closed cases, each composed of an unknown telephone recording and three known sound-booth recordings. These cases were simulated to be compatible with the system's current constraint that speech input be composed of only three isolated words. Of the 30 speaker comparisons made, the Westinghouse system made 24 correct, 4 marginal, and 2 incorrect determinations.

The System Development Corporation, under a Department of Defense contract, has developed a computer program that enables a digital system like the Semi-Automatic Speaker Identification System to track formant frequencies of speech waveforms as was shown in Figure 19. The time-varying waveforms of formant frequencies of telephone recordings were shown to be identical to that of direct recordings for the same utterances, demonstrating the system's invariance with the response of the telephone. Only a speaker comparison algorithm is needed to extend this system into a speaker identification system.

Professor Peter Ladefoged of the Phonetics Laboratory of the University of California at Los Angeles provided an example of such an algorithm. An utterance of the phrase, "How many diesel-guided missile submarines?" was recorded by an unknown speaker. Professor Ladefoged identified ten segments of the formant trajectory of the utterance, as shown in Figure 20, that have speaker-dependent features. Table 4 describes these dynamic features.

Table 4. Speaker Dependent (Dynamic) Features of Phrase:
 "How Many Diesel Guided Missile Submarines? "

PROPERTY OBSERVED IN "UNKNOWN" FORMANT TRACK	SCORE FOR "KNOWN" FORMANT TRACK	
	BILL	ROLLIN
1. SLOPE OF F2 ASPIRATION IN HH OF "HOW" IS NEGATIVE	10	0
2. DURATION OF AW IN "HOW" IS MUCH LONGER THAN DURATION OF M IN "MANY"	9	2
3. F1 OF AW IN "HOW" HAS AN AREA WITH CONSTANT NEGATIVE SLOPE	8	1
4. F1 SHOWS A SHARP TRANSITION FROM AW IN "HOW" TO M IN "MANY"	9	3
5. F2 IS NOT DISCONTINUOUS IN THE N IN "MANY"	10	0
6. F2 SLOPE CHANGES ABRUPTLY FROM IY IN "DIESEL" TO Z IN "DIESEL"	9	6
7. Z IN "DIESEL" IS DEVOICED	10	10
8. F2 OF L IN "DIESEL" IS ABOVE 1000 Hz AT ITS END	10	5
9. F2 TRANSITION OUT OF G IN "GUIDED" IS STRICTLY NEGATIVE IN SLOPE	10	4
10. F2 SHOWS CLEAR MINIMUM BETWEEN SEGMENTS OF OPPOSITE SLOPE IN AY IN "GUIDED"	10	5
TOTAL POSSIBLE SCORE = 100	95	36

-43-

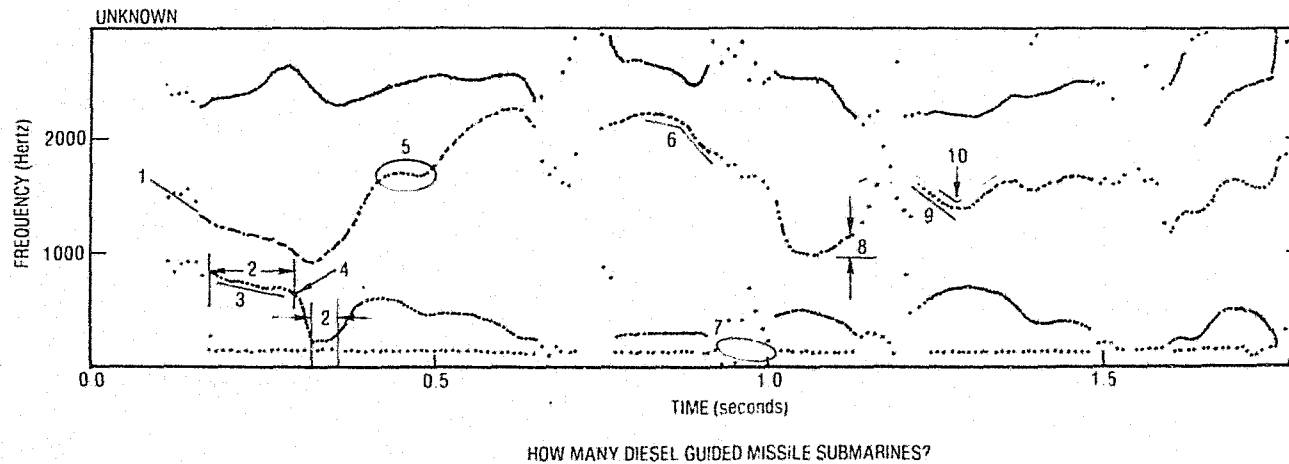


Figure 20. Formant Trajectories of Unknown Speaker

Recordings of the utterance were made of two known speakers, Bill and Rollin. Figure 21 shows the time-varying waveforms of the formant frequencies of the given utterances, along with the location of the dynamic features. The features of the known speakers were compared with those of the unknown speaker, and a score of 0 to a maximum of 10 was given for each feature and speaker. The System Development Corporation arrived at the actual numbers in Table 4, using their judgment of the visual comparison.

With ten features, the maximum score attainable is 100. When the features of Bill were compared with those of the unknown, a score of 95 resulted. When the features of Rollin were compared with the unknown, a score of 36 resulted. The unknown speaker was actually Bill, as the algorithm indicated. Two other exemplars were compared and yielded scores of 55 and 61.

This example illustrates that the concept of a speaker identification algorithm using dynamic features can be extended to apply to general speech and can be developed as a computer-aided method of speaker identification.

Supported by these investigations, Aerospace defined a Design Optimization Requirements Definition and recommended that a short-term feasibility study be made on the Semi-Automatic Speaker Identification System. The study, which was designed on successful laboratory demonstrations, would determine if there was a feasible solution to the problems of noise and channel effects and other undesirable parameters, such as phase distortion, and recorder clipping. With the Institute's approval for this study, procurement packages were approved and sent to Rockwell, Westinghouse Electric, and the System Development Corporation to perform various tasks.

Under subcontract to Rockwell, the University of Utah was to develop a channel deconvolution algorithm for recordings of duration of

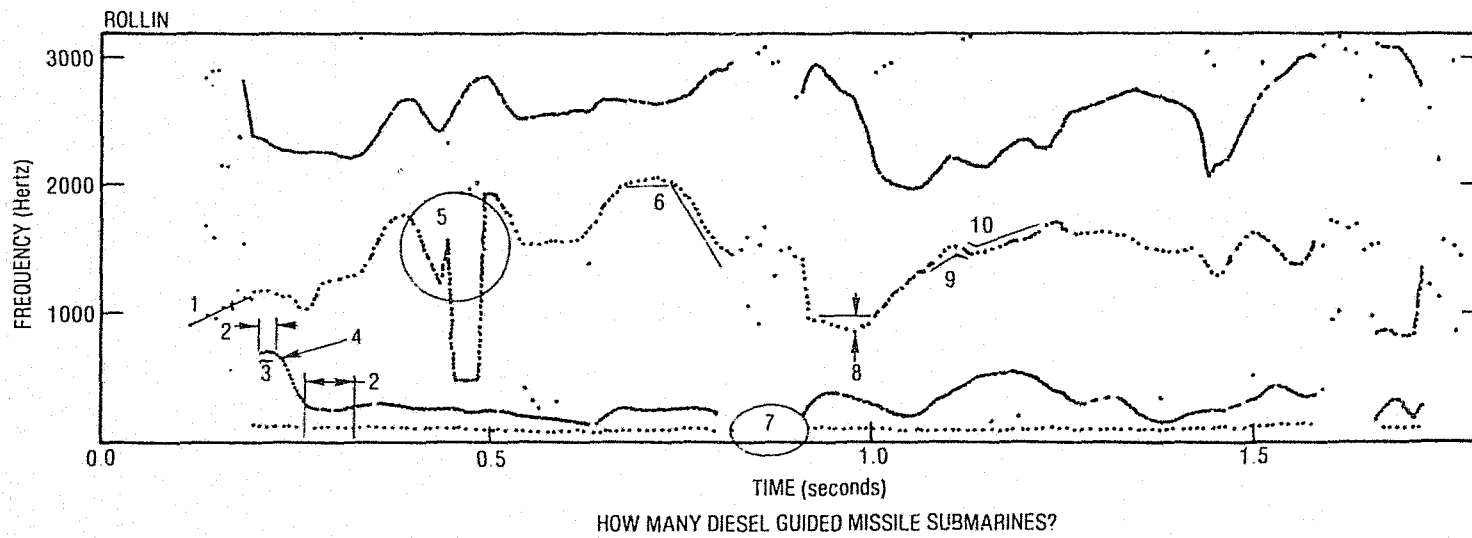
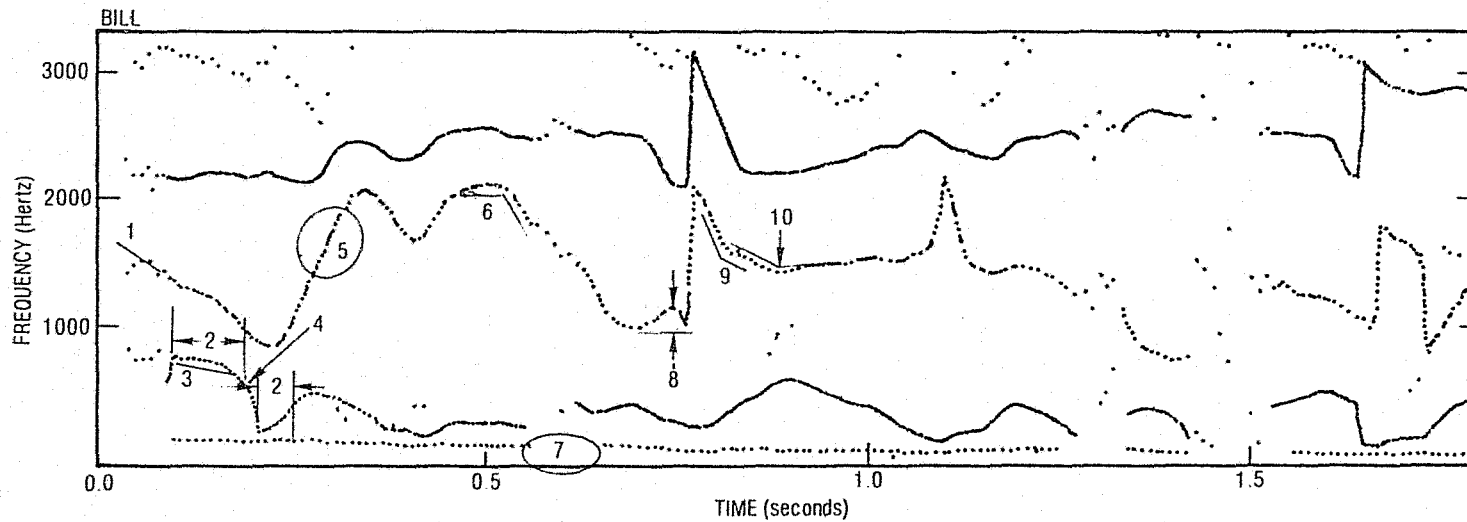


Figure 21. Formant Trajectories of Two Exemplars

ten seconds or less. Rockwell was to test this algorithm along with the previously developed Blind Deconvolution algorithm on simulated telephone recordings. In addition, Rockwell was to test its innovative speech enhancement algorithms, which were presented at the 1976 Carnahan Conference on Crime Countermeasures, on simulated noise recordings and also generate a forensic data base which could be used as a standard in testing speaker identification systems and methods. The algorithms of the noise and channel filters and the forensic data base were to be delivered to Aerospace.

Westinghouse Electric was to test its speaker verification system on the forensic data base and other simulated recordings to evaluate the performance in speaker identification using dynamic features.

The System Development Corporation was to extend their formant tracking system to include a speaker comparison algorithm and then to test the system on the forensic data base and other simulated recordings. These algorithms were to be delivered to Aerospace.

At the conclusion of the Rockwell subcontract and as part of this feasibility study, the Semi-Automatic Speaker Identification System was moved from the Rockwell facilities to the Aerospace facilities. Figure 22 shows the system installed at Aerospace. As part of this system handover, Aerospace initiated formal documentation of the system's software and initiated software modifications to facilitate the operation and troubleshooting of the system. Preparation was made also for the tasks of incorporating the channel and noise filter algorithms and the formant tracking algorithm into the speaker identification system.

The speaker identification program was canceled by the Law Enforcement Assistance Administration in August 1976, before subcontracts had been let for the system short-term feasibility study. Aerospace was tasked to close out the program and prepare a final report, including completion of the formal software documentation.

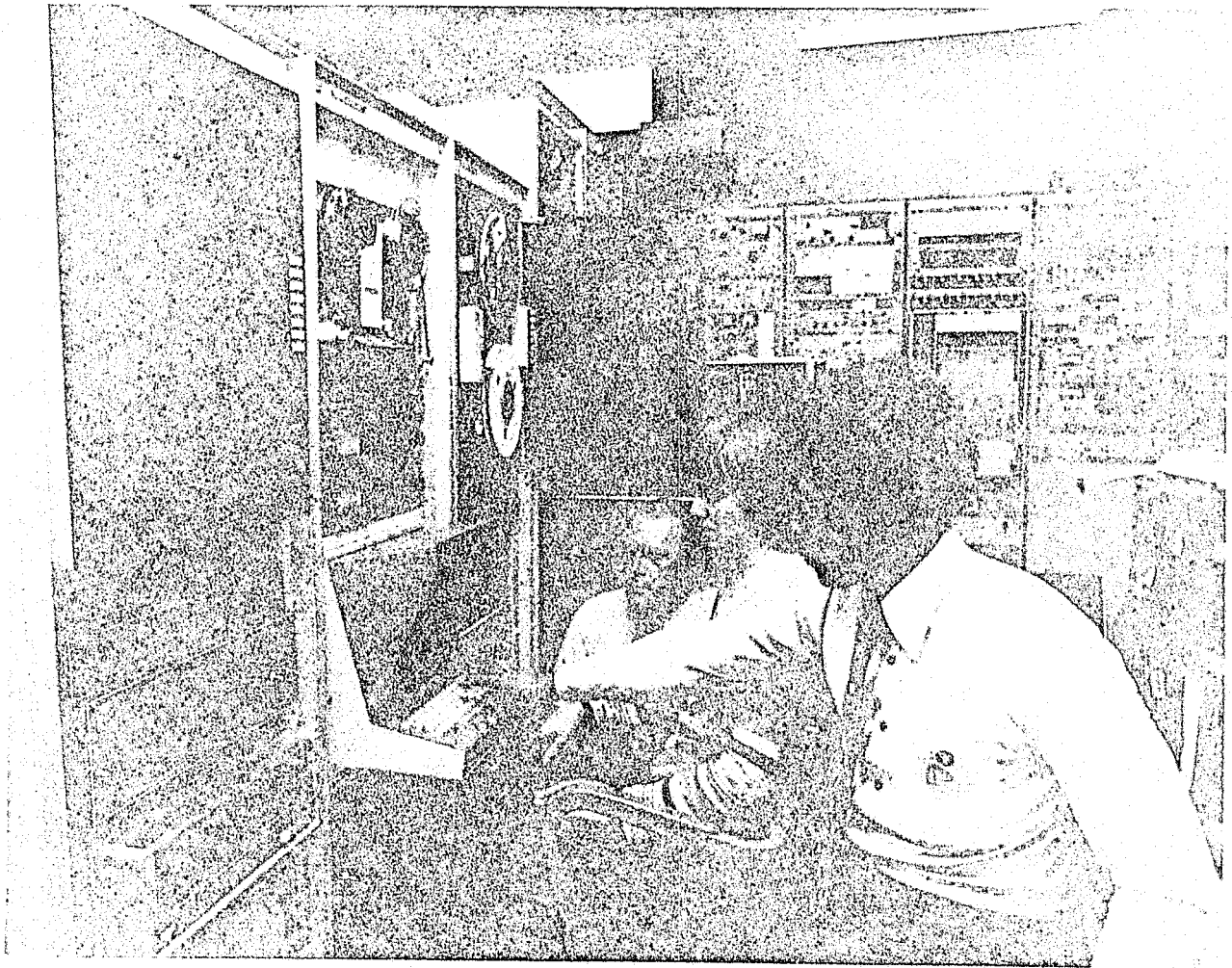


Figure 22. New Design Improvement Algorithm Tested on
Semi-Automatic Speaker Identification System
at Aerospace



CHAPTER 5. SYSTEM PROBLEMS AND RECOMMENDED IMPROVEMENTS

A system for performing statistically valid speaker comparisons based on laboratory-type speech exists as a result of the speaker identification program. The laboratory test and pilot test have revealed problems that the Semi-Automatic Speaker Identification System will encounter when processing recordings of voices different from the male General American English dialect and conditions less desirable than the quality recording studio. The following conclusions were drawn from these tests:

The system appears invariant across female General American English dialect speakers, female Black Urban dialect speakers, female Chicano dialect speakers, speech with simulated stress, and speech with nasal disguises. The system shows variance with male Black Urban dialect speakers, male Chicano dialect speakers, and with male Black Urban disguises.

The system demonstrated that it could be successfully operated by trained forensic technicians in a police laboratory environment. A set of operator interface software changes were recommended during the pilot field test and were incorporated into the system software. The major difficulty encountered by the Semi-Automatic Speaker Identification System in processing real-world data was separating the spectral variations associated with different channels and different speakers.

A second problem encountered in the pilot field test was that of operator reluctance to consistently follow the operating procedure with the care and diligence required. In several cases, the operator performance was very good; however, knowledge of the channel equalization problem and the absence of Rockwell supervision during some cases contributed significantly to this problem.

The problems encountered in the system fall into essentially two categories: software problems and algorithm design problems. A software

problem is one that impedes efficient or satisfactory operation of the system by the investigator. An algorithm design problem is one that affects the accuracy of the system's results. The algorithm design problems and recommended solutions or improvements will be discussed first.

The effects manifested on the acoustic recording other than the speaker's own voice are referred to here corporately as channel and noise effects. These effects may result from a variety of sources in addition to the communications channel; however, passive analysis does not permit readily identifying and separating each source. The type of effects witnessed on real-world data include spectral modification, stationary additive noise, reflections (echos), and nonlinear distortions. The first three effects appear addressable using conventional techniques. However, the nonlinear distortion, which fortunately represents only a small portion of the problem, cannot be readily addressed in a passive environment.

A major factor detrimentally affecting the Semi-Automatic Speaker Identification System operation with real-world data is the effect of the channel on the speech signal being recorded. The telephone channel usually does not have a flat spectral transfer function, as shown in Figure 23, but often displays a highly irregular transfer function (2 to 4 dB standard deviation according to Bell System Technical Reference 41005). Speech signals passing through different telephone channels are, therefore, modified by the channels, causing exactly the same acoustic sounds to display wide spectral variations, as shown in Figure 24. Since the system bases most of its feature analysis on spectral measurements and computes speaker distance based on the difference in the spectrum measurements, differences in channel transfer functions can manifest themselves in the same manner as differences in speakers and produce large values of speaker distances.

A procedure for implementing an interactive channel equalizer on the Semi-Automatic Speaker Identification System, using existing hardware with supplemental channel analysis and equalizing software, is discussed in

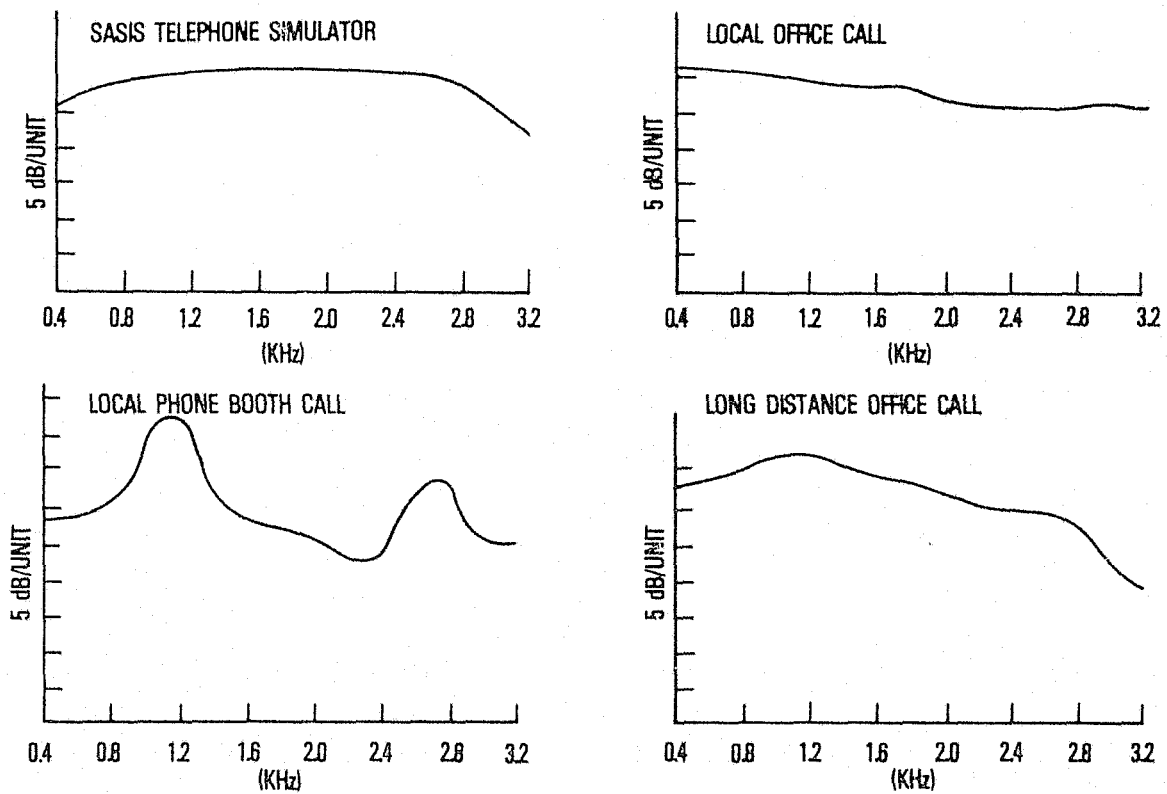
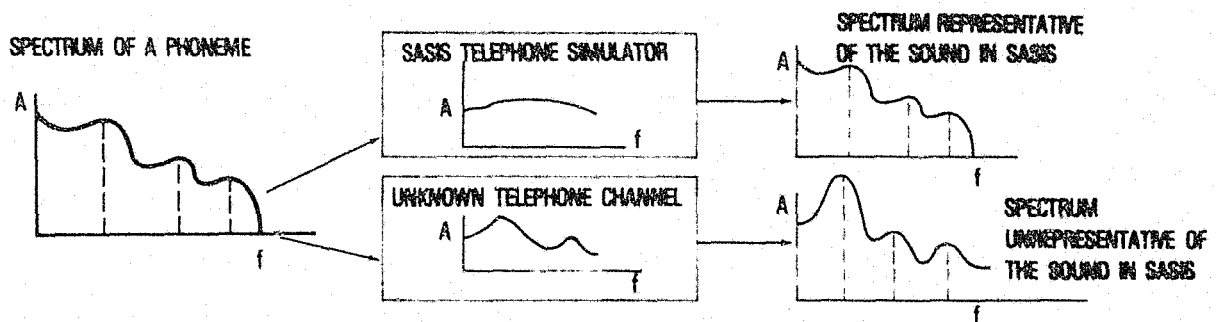


Figure 23. Examples of Measured Frequency Responses of Telephone Channels



● TELEPHONE CHANNEL CAN ALTER IDENTIFICATION FEATURES OF SPEECH

EXAMPLES:

- FORMANT FREQUENCIES
- PSD AMPLITUDES
- PSD SLOPES

Figure 24. Speech Signal Passed Through Telephone Channel

the Laboratory Test Report.¹³ This procedure is based on a channel model involving stationary spectral modification, additive noise, and echoes.

Briefly, the method of channel equalization is explained as follows:

- Channel equalization is an algorithm that removes the effects of the response of an unknown channel from a recording. It works on the assumption that the channel spectrum is constant over the period of speech and that a person's long-term speech spectrum is constant for random text speech. Figure 25 shows the slight variation in the measured long-term spectra of different speakers.
- First, the voice recording is segmented into small intervals (0.5 second) of speech, as indicated in Figure 26. Then the power spectrum density of each segment is measured and averaged. The unknown channel spectrum is then found by taking the ratio of the average just obtained with the known long-term speech spectrum. The original recording is channel equalized by filtering it with the inverse of the channel spectrum function.

The effects of various types and levels of noise on the speaker distance measure is poorly understood. The experiments reported in the Rockwell Final Report indicate a wide range of effects, depending upon the noise properties. The speaker distance measure, when noise is present on the acoustic signal, cannot currently be interpreted in terms of the analytical studies⁸ (Paul, et al., 1974) performance statistics.

Needed is an evaluation of the effects of additive shaped Gaussian noise, harmonic-rich 60-Hz noise, street noise, and conventional "restaurant effect" noise, all at several levels. Figure 27 shows how noise with a relatively constant power spectrum density can affect the features of the Semi-Automatic Speaker Identification System. Such experiments would render a

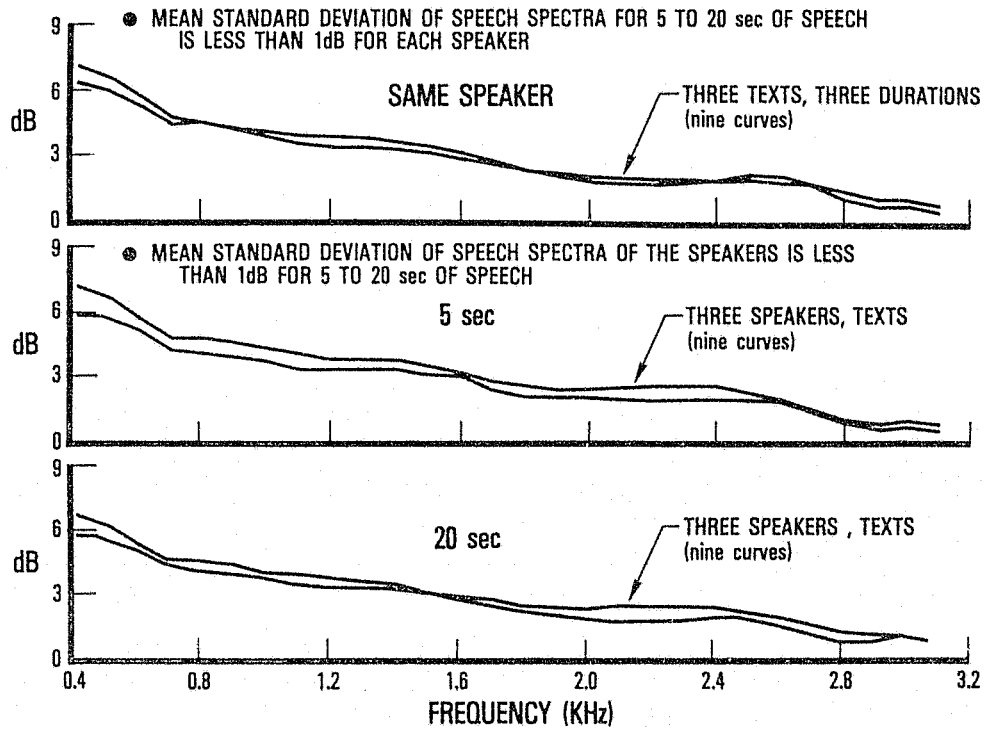


Figure 25. Long-Term Speech Spectra of Three Speakers Reading Different Texts

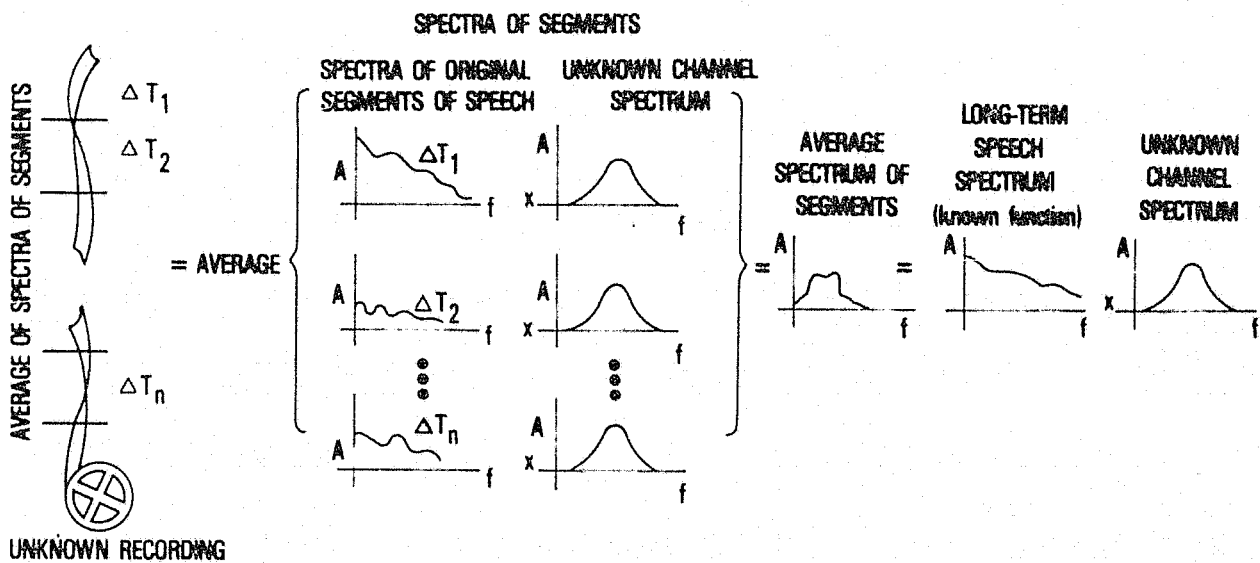


Figure 26. Pictorial Explanation of Channel Equalization

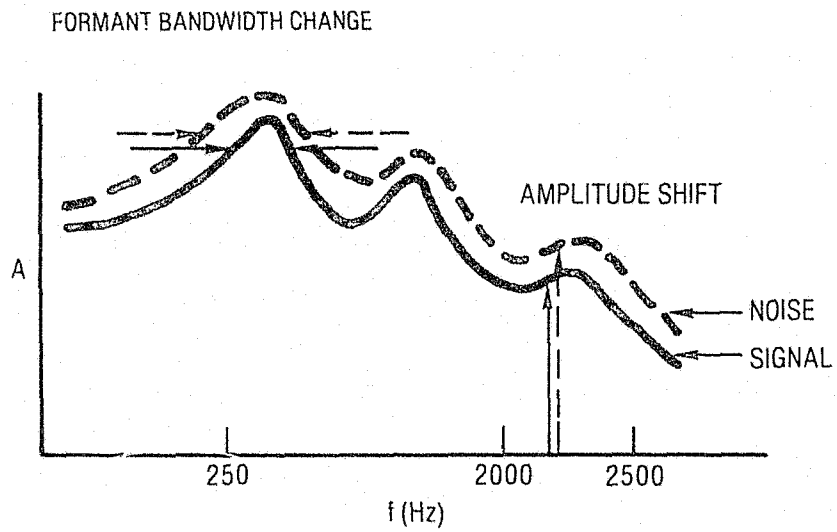


Figure 27. Noise Can Affect the Feature Values of the Semi-Automatic Speaker Identification System

measure of system sensitivity to different types of noise and would lend insight into corrective procedures.

Another recommended improvement to the Semi-Automatic Speaker Identification System is modification of the decision basis. The current procedure for performing speaker comparison requires that an overall speaker distance measure be computed from all the labeled events. A set of subjective criteria exists for selecting events, but is probably inadequate for the real-world environment. An alternative procedure for making an identification using the Semi-Automatic Speaker Identification System appears plausible. This procedure would require that a speaker distance measure be computed individually for each labeled event, and, if N good matches occur, then a positive identification would be rendered. This procedure is similar to the procedure followed by many voiceprint examiners in which a match of ten words is considered a sufficient criterion for identification. This interpretation procedure would make use of the objective qualities of the Semi-Automatic Speaker Identification System and would permit rejecting dissimilar properties that occur due to disguises, reading effects, colds, etc.

An improvement to desensitize the linear predictor coefficients in the feature set of the speaker identification algorithm would be to have the level of the speech input automatically controlled.

Dynamic speech or suprasegmented features are an obvious consideration for improving the Semi-Automatic Speaker Identification System.

The acoustic speech signal contains additional information about speaker identity other than that extracted by the steady-state-sensitive features of the Semi-Automatic Speaker Identification System. The addition of suprasegmental features would reduce the system's dependence on absolute spectral parameters. Spectral parameters are far more sensitive to the spectral modifications that are produced by telephone channels than the

suprasegmental features. Voiceprint analysis relies heavily on these suprasegmental features. The additional features could include such measurements as formant trajectories, vowel-consonant transitional information, and perceptual correlates.

Another improvement is the normalization of the operational amplitude of the system. In the current analysis of the Semi-Automatic Speaker Identification System, audio data are amplitude normalized at the sentence level in the time domain, using a combination peak and average magnitude normalization procedure. This procedure works well with laboratory data by closely matching the amplitude levels of two sentences and preserving the intra-sentence stress levels.

In an operational environment such as the field test, audio data are not structured on a uniform sentence basis, and stress levels are not always identical with the same speaker. This is due to different acoustic environments, emotional stress, etc. Moreover, random segments may be edited from either the basic or query utterance due to noise or other undesirable extraneous audio. It, therefore, seems plausible to employ an alternative amplitude normalization procedure on the isolated phonetic event.

The procedure proposed is to take a time-domain in a three-pitch-period segment and to normalize the energy in this signal to a reference value. Two unknowns immediately ensue. First, the effect on feature values is unknown, since the normalization process neutralizes amplitude stress information. Second, system performance statistics are probably modified, since the input data have been modified.

This improvement is to introduce event-level amplitude normalization and to determine the effect on both feature selection and overall system performance statistics. Given that the system is available with both normalization procedures, a comparison can be effectively derived on a limited data base of operational data.

An area that has not been investigated is the nasal and glide environments of speech for labeling on the Semi-Automatic Speaker Identification System. The current analysis constraints require that selected phonetic events for analysis must be free of noise, from the same texts, and not in a nasal or glide environment. According to a study, conducted by Mr. George Papcun of the Phonetics Laboratory of the University of California at Los Angeles, on vowel environments in operational telephone data, over 40 percent of candidate classification events possess a nasal or glide in their immediate context. Since telephone threats, bomb calls, etc., are usually short in duration and are often accompanied with extraneous noises, an initial elimination of acceptable quality data by 40 percent seems undesirable.

It is felt that the effect of these environments on long vowels, where target positions have sufficient time to be achieved, is minimal. In fact, several of these environments exist in Data Base 4 (recordings used to synthesize the identification algorithms). The effects, however, are not known, and it is desirable to perform experimental analysis to quantify these effects.

A procedure for determining the effects of these nonstandard environments on the Semi-Automatic Speaker Identification System operation is to select data recorded from 25 or more speakers from either Data Base 1 or 2, or a combination of the two. These data inherently have a large sample of glide/nasal environments. After quantifying these environments, selected events from these already digitized data can be labeled to form a new data base. It is expected that the data base size could be as small as 1000 events, if well planned. The system statistics could then be recomputed based on nonstandard environments and compared with the standard environment statistics. Should statistics differ significantly, one of these approaches could be selected:

- Remove from consideration nonstandard environments.
- Develop a second set of statistics to be used with these environments.

- Develop a normalization procedure at either the feature level or the distance measure level to normalize the effect of these environments to that of standard phonetic environments.

The evaluation of the effect of operator variations could reveal additional areas of improvement. To better interpret the results and be aware of the speaker distance variation attributable to the operator, it is appropriate to measure the effects of operator variations. As reported in the Rockwell Final Report,¹⁵ the variations due to inter- and intra-operator labeling are minimal. Because the number of triads in the experiment is insufficient to generalize, additional investigation is needed.

Also, very little is known of the effect on the Semi-Automatic Speaker Identification System operation of varying the position of the cursor in isolating the phonetic event. In a sustained vowel, greater than ten pitch periods are available from which the operator selects three subjectively, and differences can occur in segmentation between two operators. It is suspected that these differences will not affect performance significantly.

To measure the effects of cursor variation, it is recommended that an experiment be conducted with a set of phonetic event tokens that vary in duration from four to ten-plus pitch periods. Each event will be labeled and isolated more than once, each time selecting a different, but not necessarily disjoint, set of pitch periods. An intra-speaker versus inter-speaker comparison can thus be made measuring the degree of variation directly attributable to modification in cursor position (event boundaries).

The major function of the operator is to locate and identify for the computer the sounds that are to be analyzed. This operation is called, of course, labeling. A set of enhancements to the interactive graphics labeling procedure is recommended in order to improve overall operator performance. A means of automatically scaling the intensity symbols on the screen

based on displayed signal level would be useful. Other more technical recommendations are listed in the Rockwell Final Report.¹⁵

The other category of improvements is software improvement. These recommendations are rather complex, but are explained in the Rockwell Final Report. However, some of the recommendations improve the speed of the system computation and remove pitfalls, but are not failproof, while others make the operation of the system easier and clearer for the examiner.



CHAPTER 6. TASK OPTIONS

The Semi-Automatic Speaker Identification System, in its present stage of development, cannot be used in criminal investigation except in a limited number of cases recorded under conditions that resemble laboratory conditions. At the time the speaker identification program was discontinued, a system design optimization task had been scheduled that potentially would greatly broaden the category of cases that could be processed on the speaker identification system. This chapter describes the various task options that can be implemented if the speaker identification program is revived in order to progress, in various degrees, toward the objective of the program (Table 6, page 74). These tasks are discussed below and include their costs in 1976 dollars.

6.1 Channel Equalization

This task option involves the incorporation of a channel equalization algorithm into the Semi-Automatic Speaker Identification System. A limited evaluation of the algorithm that suppresses the effects of any channel response on a tape recording was performed with the cooperation of the University of Utah. The algorithm, called "Blind Deconvolution,"¹⁷ was developed by Professor Tom Stockham, and it appeared successful in removing errors due to telephone channel effects when deconvoluted tapes of duration of ten seconds or more were subsequently processed on the Semi-Automatic Speaker Identification System. However, further development is necessary in order to produce an algorithm that is equally successful on shorter segments.

An equivalent algorithm has been developed by Rockwell International and by the Bell Telephone Laboratories. Rockwell's algorithm is contained in their Interactive Digital Filtering System package that was described at the 1976 Carnahan Conference on Crime Countermeasures.¹⁸ A description of the Bell Laboratories algorithm has also been published.¹⁹

The channel equalization algorithm is a preprocessor that operates on the voice recordings. It can be programmed directly in the system's memory without affecting the existing algorithms. However, the speaker distance statistics would have to be recomputed to account for the channel equalization. With a channel equalization algorithm, most of the criminal cases, previously excluded because they involved telephone recordings, can now be processed on the system.

The estimated budget for this task is as follows:

•	Development of channel equalization algorithm	\$30,000
-	Short segment (less than 10 seconds) deconvolution algorithm	
-	Documentation of the Blind Deconvolution and short segment algorithms	
•	Integration of algorithms into Semi-Automatic Speaker Identification System and system checkout	30,000
•	System test	<u>20,000</u>
		<u>\$80,000</u>

6.2 Enhancement of Semi-Automatic Speaker Identification System

This task option enhances the Semi-Automatic Speaker Identification System with preprocessing algorithms that can suppress the two major problems in processing criminal recordings of channel and noise effects. The channel equalization algorithm and the noise suppression algorithm can be incorporated into the Semi-Automatic Speaker Identification System with its present memory capabilities. The speaker comparison statistics would have to be rederived and the system checked-out and tested.

The estimated budget for the task is as follows:

•	Development of channel equalization algorithm	\$30,000
-	Short segment (less than 10 seconds) deconvolution algorithm	
-	Documentation of the Blind Deconvolution and short segment algorithms	
•	Optimization of noise suppression algorithm	50,000
•	Integration of algorithms into Semi-Automatic Speaker Identification System	40,000
-	System checkout	
-	System documentation	
•	System test	<u>40,000</u>
		<u>\$160,000</u>

6.3 Forensic Feasibility Study

The channel equalization task option does not address the question of whether the sole use of dynamic features of speech would make a more accurate speaker identification system or, at least, augment the capabilities of the present system. The forensic feasibility study option provides for conducting several short studies on the feasibility of developing computer-aided speaker identification systems that operate on dynamic or steady-state features of speech. Each system would be tested on a data base of recordings with text and various undesirable parameters typically found in criminal recordings. The results of these tests would help to support the decision to forego any further development of the system or to develop one or a combination of the systems.

In the feasibility study option, the Semi-Automatic Speaker Identification System would be tested with telephone recordings preprocessed by the channel equalization algorithms and with noisy recordings preprocessed by a noise filter package. For example, a number of noise cancellation and channel equalization procedures have been developed by Rockwell International in its Interactive Digital Filtering System¹⁸ under internal funding, with

applications directed toward restoration of degraded voice recordings. These procedures have been highly successful in improving the intelligence of voice recordings exposed to severe degradation factors. These available speech filtering algorithms, though not optimized for the Semi-Automatic Speaker Identification System, may provide the basis for substantial reduction of the degradations in system performance attributable to noise and channel effects.

A forensic data base would be generated to be used in the final test for feasibility of computer-aided speaker identification systems. The data base would consist of cases of recordings made under controlled conditions with various texts and parameters typical of criminal evidence. The parameters would be combinations of the following: telephone calls (local, toll, telephone booth, long distance); noise (background disturbances of talking, music, activity, or street noise; cross-talk; static; or echos); dialects (Black Urban, Chicano, female); stress; and nonlinearities (automatic gain control, phase distortion, amplitude clipping). This data base could be used as a standard test base for systems and techniques in voice identification.

The forensic data base would be used to test systems that extracts dynamic features of speech for speaker comparison. The Westinghouse Corporation has developed, with its internal funds, a speaker identification system that incorporates several novel design concepts. The most important concept is a circuit that uses a combination of frequency locked loops and phase locked loops for the real-time tracking of the first and second formants of voice phonemes. Because formant frequencies are known to be relatively insensitive to telephone channels, Westinghouse has processed a series of simulated cases that contained telephone recordings. In this short experiment, the Westinghouse system achieved favorable results.

The System Development Corporation has developed a digital formant tracker that is programmed in its interactive computer system. The development of this algorithm was part of a 5-year program which began in 1971 and was supported by the Advanced Research Projects Agency. The overall intent of this research effort was to develop technologies for improved man-machine interaction and for new data management capabilities. Because it is well known that formant trajectories are speaker-dependent, as well as channel insensitive, formant trajectories are considered a source for speaker identification features that are invariant to the spectral effects of the telephone channel.

Other sources of formant trackers are the Speech Communication Research Laboratory, the University of California at Los Angeles, the Bell Telephone Laboratories, the Stanford Research Institute, Lincoln Laboratories, and Bolt, Beranek and Newman, Inc.

The estimated budget for the task is as follows:

•	Generate forensic data base	\$40,000
•	Evaluate Semi-Automatic Speaker Identification System	140,000
-	Test channel equalization	
-	Test noise filter package	
-	Test Semi-Automatic Speaker Identification System with forensic data base	
•	Evaluate dynamic system	90,000
-	Adapt dynamic system for speaker identification	
-	Test system on forensic data base	
		<u>\$270,000</u>

6.4 Dynamic Feature Extraction (Digital)

This task option is concerned with redesigning the algorithms of the Semi-Automatic Speaker Identification System to extract dynamic features of speech for speaker comparisons. A digital algorithm would be defined which tracks the time-waveforms of the formant frequency. Digital formant trackers have been developed by a number of companies and universities, as indicated in Section 6.3. From these trajectories, a set of features that prove to be speaker dependent will be selected from a large set of candidate features, including some of those looked for by voiceprint examiners. Because the Semi-Automatic Speaker Identification System data base was structured to emphasize the steady-state portions of speech, it contained few diphthongs or phoneme transitional regions that predominate speech. Consequently, a new data base must be generated. The Semi-Automatic Speaker Identification System hardware is digital and, therefore, needs no modification.

The estimated budget for the task is as follows:

• Data base	\$60,000
• Speaker identification algorithm	100,000
- Optimization of a formant tracking algorithm	
- Development of a speaker comparison algorithm	
• Implementation of speaker identification algorithm in Semi-Automatic Speaker Identification System	80,000
- System checkout	
- System documentation	
• System test	40,000
• System modifications	40,000
	<u>\$320,000</u>

6.5 Dynamic Feature Extraction (Analog)

This task option is also concerned with redesigning the algorithms of the Semi-Automatic Speaker Identification System to extract dynamic features of speech for speaker comparison. The Westinghouse Corporation has developed an electronic circuit that automatically produces the time-waveform of the first and second formant frequencies from speech input. Additional effort would be needed to develop the capability to track the third formant frequency. From these trajectories a set of features that prove to be speaker dependent would be selected from a large set of candidate features, including some of those looked for by voiceprint examiners and those used in Westinghouse's speaker verification system. A new data base would be generated for this digital system also. Because the Semi-Automatic Speaker Identification System hardware is digital, the system would be modified to incorporate the analog formant tracker.

The estimated budget for the task is as follows:

• Data base	\$40,000
• Speaker identification algorithm	180,000
- Development of tracker of third formant frequency	
- Development of speaker comparison algorithm	
• Incorporation of speaker identification algorithm into Semi-Automatic Speaker Identification System	80,000
- System checkout	
- System documentation	
• System test	40,000
• System modifications	<u>40,000</u>
	<u>\$380,000</u>

6.6 Hybrid Speaker Identification System

In this task option, the Semi-Automatic Speaker Identification System would be modified to extract features of the dynamic segments of speech as well as of the steady-state portions of speech for speaker comparison. The digital formant tracker would be adopted over the analog tracker to minimize cost. The speaker identification system would be enhanced with the channel equalization and noise suppression algorithms as described in Section 6.2 and would be modified to track formant frequencies as described in Section 6.4. A data base with adequate phoneme transitional regions for dynamic feature extraction would be generated. A hybrid speaker identification algorithm would be developed that would, in an optimal sense, combine the steady-state and dynamic features of speech to produce a speaker distance and the associated conditional probabilities that the pair of speakers compared are the same or different. The system would be checked out and tested on an independent data base, and improvements would be made to the system software where warranted.

The estimated budget for this task is as follows:

• Channel equalization algorithm development	\$30,000
- Short segment (less than 10 seconds) deconvolution algorithm	
- Documentation of the Blind Deconvolution and short segment algorithms	
• Noise suppression algorithm optimization	50,000
• Data base	60,000
• Dynamic speech features	80,000
- Optimization of a formant tracking algorithm	
- Selection of speaker dependent features	
• Hybrid speaker identification algorithm development	80,000

• Integration of algorithms into Semi-Automatic Speaker Identification System	\$120,000
- System checkout	
- System documentation	
• System test	60,000
• Software modifications	<u>60,000</u>
	<u>\$540,000</u>

6.7 Semi-Automatic Speaker Identification System Data Base

This task is concerned with the re-examination of the data base that was generated for the Semi-Automatic Speaker Identification System. The data base used in the Semi-Automatic Speaker Identification System was a large collection of utterances of sentences that contained a high density of stressed vowels and nasals (the steady-state segments of the sounds used in the Semi-Automatic Speaker Identification System comparison algorithms).

Several aspects of the data base have not been investigated. One aspect is concerned with the variance of the system's results when comparing utterances of the same speaker recorded 2 weeks apart versus 10 minutes or 6 months apart. A second aspect is concerned with using the average of two or more utterances repeated in succession for the exemplar recording instead of only a single utterance. Another aspect is concerned with the fact that the utterances comprising the data base were read aloud instead of spoken spontaneously.

In addition, very limited data were recorded of female dialects, Black Urban and Chicano male dialects, disguises, and talkers with stress. As reported earlier,¹³ the results were attained with little confidence. Therefore, a large data base would be generated to provide sufficient statistics to interpret the speaker comparisons of a dialect different from male General American English.

The estimated budget for the task is as follows:

• Variance in Semi-Automatic Speaker Identification System results	\$60,000
- With delay in repeating utterances	
- With averaging repeated utterances	
- With spoken and read utterances	
• Expansion of data base	180,000
- Females	
- Black Urban and Chicano dialects	
- Disguises	
- Stress	
	<hr/>
	<u>\$240,000</u>

6.8 Enhancement of Recordings with Disturbances

This task is concerned with suppression of the undesirable effects that are sometimes present on voice recordings used as criminal evidence. In addition to minimizing the channel and noise effects as discussed in Section 6.2, nonlinear disturbances such as automatic gain control, amplitude clipping, and phase distortion would be addressed also. An algorithm would be developed, tested, and optimized to enhance recordings with these disturbances prior to processing the recordings on the Semi-Automatic Speaker Identification System. Since the enhancement algorithm would be a preprocessor, it can be used with a speaker identification system based on steady-state sounds, dynamic sounds, or the combination of the two.

The estimated budget for the task is as follows:

• Development of channel equalization algorithm	\$30,000
- Short segment (less than 10 seconds) deconvolution algorithm	
- Documentation of the Blind Deconvolution and short segment algorithms	

• Optimization of noise suppression algorithm	\$50,000
• Development of techniques to suppress nonlinear disturbances	160,000
- Automatic gain control	
- Amplitude clipping	
- Phase distortion	
• Integration of algorithms into Semi-Automatic Speaker Identification System	60,000
- System checkout	
- System documentation	
• System test	<u>60,000</u>
	<u>\$360,000</u>

6.9 Optimization of System Operation

This task is concerned with the improvement of the efficiency of the man-machine interface of the Semi-Automatic Speaker Identification System. There are several modifications that can be made in the system software and the system hardware to facilitate the manual operation of the speaker identification system and to increase the accuracy of the operators. These recommendations were listed in the Rockwell Final Report.¹⁵

The estimated budget for the task is as follows:

• Hardware design changes in Semi-Automatic Speaker Identification System	\$40,000
• Software design changes	<u>160,000</u>
	<u>\$200,000</u>

6.10 Data Base Standard

This task addresses the need for a set of recordings made under forensic conditions of simulated criminal cases to be used as a standard for testing computer-aided speaker identification systems. The data

base would consist of ten simulated cases of various texts. Each case would comprise a criminal utterance and four sets of exemplar repeated utterances. Also, each exemplar speaker would speak his or her utterance three consecutive times. The recordings within each case would be of a common text, with a different text for each case.

The texts would be typical of actual forensic evidence with durations of approximately 5, 10, or 20 seconds, including the parameters indicated in Table 5.

The estimated budget for the task is \$28,000.

Table 5. Specifications for Data Base Standard *

Case	Duration (sec)	Parameters for Unknown Recording	Parameters for Exemplar Recordings
1	10	Long distance telephone Static noise Echo	Clean
2	20	Chicano dialect speaker Local telephone. Recorder with amplitude clipping	Chicano dialect speakers
3	10	Clean	Clean
4	10	Local home telephone Stress	AGC recording
5	5	Local home telephone (through two exchanges) Background noise	Background noise
6	10	Local home telephone Stress	Clean
7	5	Local home telephone (through two exchanges) Background noise	Clean
8	20	Female General American English speaker Toll telephone call	Female GAE speakers
9	10	Black Urban dialect speaker Open phone booth call Street noise	Black Urban dialect speaker
10	20	Disguises male General American English speaker Open telephone in restaurant Background noise AGC recording	Clean

* All speakers are male standard, except as noted.

Table 6. Summary of Task Options

Task Options	Suppress Channel Effect	Suppress Noise Effect	Other System Improvement	Estimated Cost
Channel equalization	X			\$80,000
Semi-Automatic Speaker Identification System enhancement	X	X		\$160,000
Forensic feasibility study	X	X		\$270,000
Dynamic feature extraction (digital)	X			\$320,000
Dynamic feature extraction (analog)	X			\$380,000
Hybrid speaker identification system	X	X		\$540,000
Semi-Automatic Speaker Identification System data base			X	\$240,000
Enhancement of recordings with disturbances	X	X	X	\$360,000
Optimization of system operation			X	\$200,000
Data base standard				\$28,000

CHAPTER 7. CONCLUSIONS

The Aerospace Corporation was under contract to the National Institute of Law Enforcement and Criminal Justice of the Law Enforcement Assistance Administration to technically monitor and provide system engineering capability for the speaker identification program from FY 73 to FY 76. The objectives and accomplishments of the speaker identification program were reported in this document, with emphasis on the development of the Semi-Automatic Speaker Identification System.

The objective of the program was to develop a computer-assisted speaker identification system for use in investigation as well as in courtroom testimony, and to investigate other applications of speaker identification technology. The Semi-Automatic Speaker Identification System was designed and optimized using laboratory data and tested both on simulated criminal cases recorded under ideal conditions and on actual criminal evidence. The results were very good for the simulated cases but poor for the actual cases. The laboratory designed system encountered a prohibitive problem when telephone recordings were processed. The speaker identification program was terminated by the Law Enforcement Assistance Administration before the system could be modified.

This final report on speaker identification has detailed the history of the program from its inception to termination and has presented a general description of the design and operation of the Semi-Automatic Speaker Identification System. The system problems that were encountered during testing were explained and supported by experimental analysis. Investigations and recommendations of potential solutions to these problems are the basis for the set of task options described in Chapter 6.

As mentioned above and emphasized throughout this report, the prohibitive problem of the channel effect prevents the Semi-Automatic Speaker Identification System from being operational in a forensic

laboratory on criminal evidence. The problem is that the response of the telephone channel dominates speaker comparisons. The features of the speaker identification system are very channel dependent, as was demonstrated during the system tests, which is an unfortunate phenomenon uncovered about steady-state features of speech that is insignificant with dynamic features of speech. With the correction of this problem, the Semi-Automatic Speaker Identification System will be at least operational in a forensic laboratory on criminal evidence. Thus, the incorporation of a channel equalization algorithm into the system is recommended as a necessary task to be performed if the program is continued.

Since the channel equalization algorithm filters the channel response because it is time invariant, the algorithm also filters noise that is stationary and correlated. Thus, with the inclusion of a few additional speech enhancement algorithms, the problems of both noise and channel effects can be suppressed.

Consequently, it is recommended that Task Option 6.2 (Enhancement of the Semi-Automatic Speaker Identification System) stated in Chapter 6 be the next task in any renewed funding of the speaker identification program. The estimated cost of optimizing, incorporating, checking, testing, and documenting these enhancement algorithms is \$160,000.

NOTES

1. M.H.L. Hecker, "Speaker Recognition - An Interpretive Survey of the Literature," American Speech and Hearing Association Monograph No. 16, January 1971.
2. "Voiceprint Applications Manual," Report No. TOR-0073(3654-06)-1, The Aerospace Corporation, El Segundo, California (July 1973).
3. "Voiceprint Identification," Georgetown Law Journal, Vol. 61, Issue 3, February 1973.
4. R.W. Becker, et al., "A Semi-automatic Speaker Recognition System," Final Report, Project 1363, Stanford Research Institute, Menlo Park, California (August 1972).
5. G.D. Hair and T.W. Rekieta, "Speaker Identification Research," Final Report, Texas Instruments, Inc., Dallas, Texas (August 1972).
6. "Semi-Automatic Speaker Identification System (SASIS) Final Report," Report No. C74-1185/501, Rockwell International, Anaheim, California (December 1974).
7. "Voiceprint Validation Test," The Aerospace Corporation, El Segundo, California (21 September 1973).
8. "Semi-Automatic Speaker Identification System (SASIS) Analytical Studies Final Report," Report No. C74-1184/501, Rockwell International, Anaheim, California (December 1974).

9. P.K. Broderick, et al., "Semi-automatic speaker identification system," Proceedings, Carnahan Conference on Crime Countermeasures, University of Kentucky, Lexington, 7-9 May 1975.
10. J.E. Paul, Jr., et al., "Development of analytical methods for a semi-automatic speaker identification system," Proceedings, Carnahan Conference on Crime Countermeasures, University of Kentucky, Lexington, 7-9 May 1975.
11. "Preliminary Applications Survey for Semi-Automatic Speaker Identification System (SASIS)," Rockwell International, Anaheim, California (5 April 1974).
12. "Preliminary Investigation of Applications of the Computer-Aided Speaker Identification System," Report No. ATR-74(7905)-1, The Aerospace Corporation, El Segundo, California (June 1974).
13. "Semi-Automatic Speaker Identification System (SASIS) Laboratory Test Report," Report No. C75-701/501, Rockwell International, Anaheim, California (August 1975).
14. "Applications of Semi-Automatic Speaker Identification Techniques," Report No. ATR-75(7907)-1, The Aerospace Corporation, El Segundo, California (March 1975).
15. "Semi-Automatic Speaker Identification System (SASIS) Final Report," Report No. C76-96-501, Rockwell International, Anaheim, California (2 February 1976).
16. "Semi-Automatic Speaker Identification System (SASIS) Training Manual," Report No. C75-623/501, Rockwell International, Anaheim, California (July 1975).

17. T.G. Stockham, Jr., et al., "Blind deconvolution through digital signal processing," Proceedings of the Institute for Electrical and Electronics Engineering, Vol. 63, No. 4, pp. 678-692.

18. J.E. Paul and V.A. Vitols, "Restoration of degraded audio recordings," Proceedings, Carnahan Conference on Crime Countermeasures, University of Kentucky, Lexington, 5-7 May 1976.

19. F. Itakura, "Minimum prediction residual principle applied to speech recognition," Institute of Electrical and Electronics Engineering, Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-23, No. 1, February 1975, pp. 67-72.



APPENDIX A. PROGRAM DOCUMENTS AVAILABLE THROUGH THE
NATIONAL CRIMINAL JUSTICE REFERENCE SERVICE

The following publications, which are also listed in the Notes section, were prepared as part of, or as a consequence of, the Speaker Identification program. These articles and reports can be found in the National Criminal Justice Reference Service in Washington, D.C., 20531.

1. "Voiceprint Applications Manual," The Aerospace Corporation, Report No. TOR-0073(3654-06)-1, prepared for the Law Enforcement Assistance Administration, July 1973.

The manual provides the reader who has only limited experience with voiceprints an understanding of the principles of voiceprint analysis and knowledge of correct practices in collecting and submitting voice samples for evaluation.

2. "Voiceprint Validation Test," The Aerospace Corporation, 21 September 1973.

This report describes a test designed to replicate the forensic application of voiceprint identification as much as possible so that methods of identification, test variables, and recording conditions are similar to those encountered by the forensic examiner.

3. "Preliminary Applications Survey for Semi-Automatic Speaker Identification System (SASIS)," Rockwell International, prepared for The Aerospace Corporation, 5 April 1974.

This survey contains results of the study to obtain detailed information regarding the potential application of a computer-aided speaker identification system by local law enforcement agencies.



CONTINUED

1 OF 2

4. "Preliminary Investigation of Applications of the Computer-Aided Speaker Identification System," The Aerospace Corporation, Report No. ATR-74(7905)-1, prepared for the Law Enforcement Assistance Administration, June 1974.

The report documents an investigation of the present and future uses of voice identification in the law enforcement and criminal justice community and projects the nature and scope of the potential applications of a computer-aided speaker identification system.

5. "Semi-Automatic Speaker Identification System (SASIS) Analytical Studies Final Report," Rockwell International Report No. C74-1184/501, prepared for The Aerospace Corporation, December 1974.

The report describes the technical studies that were carried out to develop the mathematical and experimental techniques used in the Semi-Automatic Speaker Identification System where segments of speech are extracted from two speech utterances and are computer-analyzed to yield a statistical measure indicating whether or not the utterances were said by the same or different speakers.

6. "Semi-Automatic Speaker Identification System (SASIS) Final Report," Rockwell International Report No. C74-1185/501, prepared for The Aerospace Corporation, December 1974.

The report describes the Semi-Automatic Speaker Identification System and the objective, background, and scope of its development. The operator qualifications and the forensic application of the Semi-Automatic Speaker Identification System are explained, detailed descriptions of the system design and functional operations are given, and the relationship between this computer-aided approach and the voiceprint method is discussed.

7. "Applications of Semi-Automatic Speaker Identification Techniques," The Aerospace Corporation, Report No. ATR-75(7907)-1, prepared for the Law Enforcement Assistance Administration, March 1975.

The paper is a tutorial description of the Semi-Automatic Speaker Identification System, its purpose, design, and operation.

8. Broderick, P.K., et al., "Semi-Automatic Speaker Identification System," Carnahan Conference on Crime Countermeasures, May 1975.

The paper presents a summary of the analytical studies leading to the development of the Semi-Automatic Speaker Identification System.

9. Paul, Jr., J.E., et al., "Development of Analytical Methods for a Semi-Automatic Speaker Identification System," Carnahan Conference on Crime Countermeasures, May 1975.

The report documents a study of the potential uses of techniques for speaker identification through computer analysis of voice samples. Related efforts in automatic personal identification, using fingerprints, voice samples, etc., are surveyed and the fundamental techniques being used are discussed.

10. "Semi-Automatic Speaker Identification System (SASIS) Training Manual," Rockwell International Report No. C75-623/501, prepared for The Aerospace Corporation, July 1975.

The manual is a comprehensive document which covers the basic aspects of hardware and software operations, offers appended material detailing these operations, and is written to serve both tutorial and operational reference requirements.

11. "Semi-Automatic Speaker Identification System (SASIS) Laboratory Test Report," Rockwell International Report No. C75-701/501, prepared for The Aerospace Corporation, August 1975.

The report describes the laboratory test which showed that the speaker identification algorithm designed on male General American English dialect appeared consistent with certain other types of speech samples (i. e., General American, Black Urban, and Chicano female dialects, simulated stress, and nasal disguises).

12. "Semi-Automatic Speaker Identification System (SASIS) Final Report," Rockwell International Report No. C76-96-501, prepared for The Aerospace Corporation, 2 February 1976.

The report describes the pilot test, details the results and explains the problem areas. The major problem encountered during the pilot test was the same encountered during the laboratory test phases: the telephone channel response dominates the speaker comparison. However, half of the incorrect results were due to the channel effect and half were due to avoidable errors made in operating the computer-aided system. Except for the prohibitive problem of the telephone channel effect, the results of the pilot test were favorable.

APPENDIX B

SEMI-AUTOMATIC SPEAKER IDENTIFICATION SYSTEM (SASIS)
SOFTWARE OVERVIEW SPECIFICATION



TABLE OF CONTENTS

1.0	Introduction	89
2.0	System Architecture	90
2.1	The SASIS System Tree	90
2.2	DISPATCH	92
2.3	Major Module Trees	99
2.4	Module Communications	123
2.5	A Computational View of SASIS Data Flow	124
3.0	Data Management	127
3.1	Data Management Software	128
3.2	Data Structures	131
4.0	System Creation and Backup Procedures	140
4.1	Creating Execute Modules	140
4.2	Creating the Operational System	142
4.3	Experimental Variations	143
4.4	System Backup Procedures	144
5.0	Support Software	148
5.1	Utility Programs	148
5.2	Diagnostic and Test Programs	153
5.3	Technical Management Programs	156
6.0	A Software Development Philosophy	160
6.1	Making Software Changes	160
6.2	Disk Backup	161
6.3	Development and System Log	161
7.0	List of Unfinished Tasks	162

8.0	Recommendations	163
8.1	Additional Hardware	163
8.2	Additional Software	163
8.3	Additional Documentation	163
9.0	Related Documents	164

1.0 INTRODUCTION

The Semi-Automatic Speaker Identification System is a research and development tool designed to support the investigation of speech comparison and identification techniques. Potentially, it is also an investigative tool for use in law enforcement and criminal justice. The achievement of that potential depends upon the complete, unambiguous audit trail output record which would be acceptable in legal proceedings. Given a valid set of speech comparison techniques, such an audit trail facilitates replication of results as compared to subjective voice-print methods. Other system features such as duplicative magnetic tape data records add to the reliability of the system and minimize the probability of discrediting evidence in court over minor technicalities such as a parity error. In summary, the Semi-Automatic Speaker Identification System is an applied research tool with requisite features for use in real world applications.

This document may be used to provide an overview of the structure of the Semi-Automatic Speaker Identification System software and, in particular, those areas where potential enhancements are most likely. In addition, procedures are defined for software system creation, modification, and management of software changes. Support software including utilities, diagnostics, and technical management programs are documented from the user's viewpoint. Emphasis is placed on why they were written and how they can best be used. The first section describes the overall structure of the Semi-Automatic Speaker Identification System.

2.0 SYSTEM ARCHITECTURE

In this section, the major Semi-Automatic Speaker Identification System modules are displayed in tree structures. The SASIS Training Manual¹ may be used as an ancillary reference to this section. The relationship between the software structure and application are broached in the document. No attempt is made to include a complete set of flowcharts. However, in order to facilitate understanding the Semi-Automatic Speaker Identification System, a number of flowcharts have been included where the need is anticipated.

2.1 The SASIS System Tree

SASIS execution is initiated from the DISPATCH program from which the operator calls various major modules to input, label, and perform calculations on the comparison of speech.* The DISPATCH program and the major modules are shown in tree form below:

* In Data General nomenclature, a source program (e.g., DISPATCH) has no extension or an SR extension. The relocatable binary output by compiler and/or assembler has an .RB extension (e.g., DISPATCH.RB). When linked together with other .RB's, the executable form of a program has an .SV extension (e.g., DISPATCH.SV). The operator can call an .SV file from the CLI and .SV files call other .SV files on a roll-in-roll-out basis. .SV can call directoried overlay programs identified by the .OL extension. These "roll" into a specified area of memory and do not require a complete core "swap."

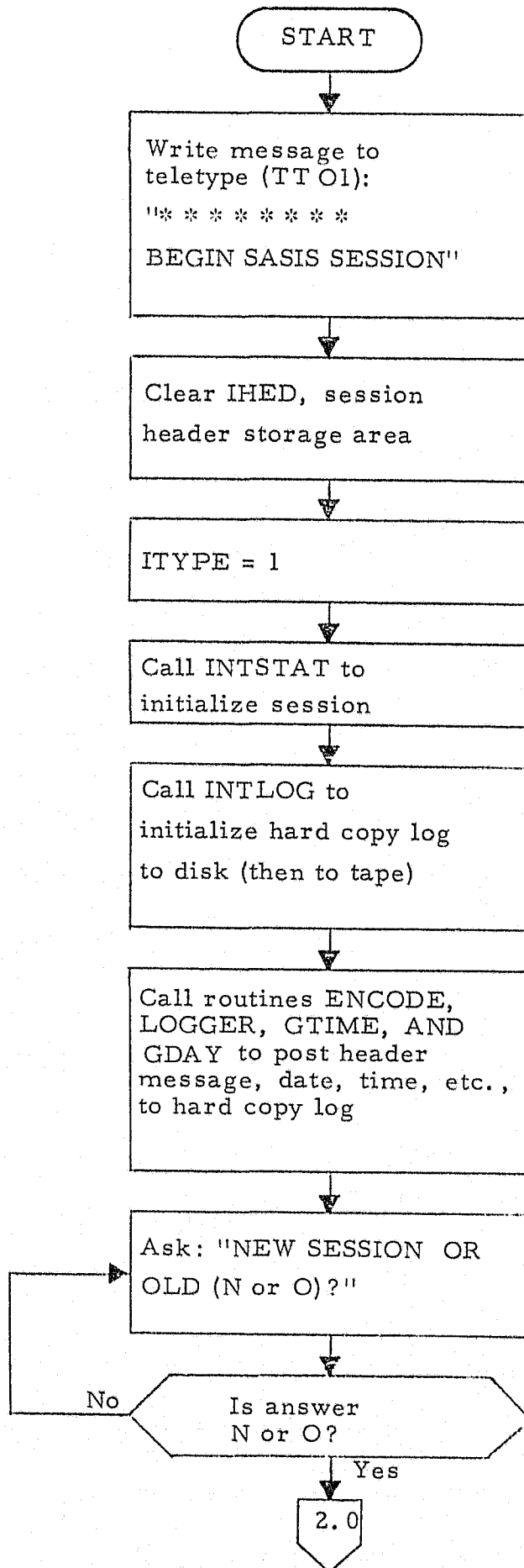
DISPATCH.SV

TERMINATE.SV
SPINP.SV
WSDF.SV
LABEL 11.SV
LABEL 12.SV
WRFILE.SV
XFEAT.SV
COMPAR.SV

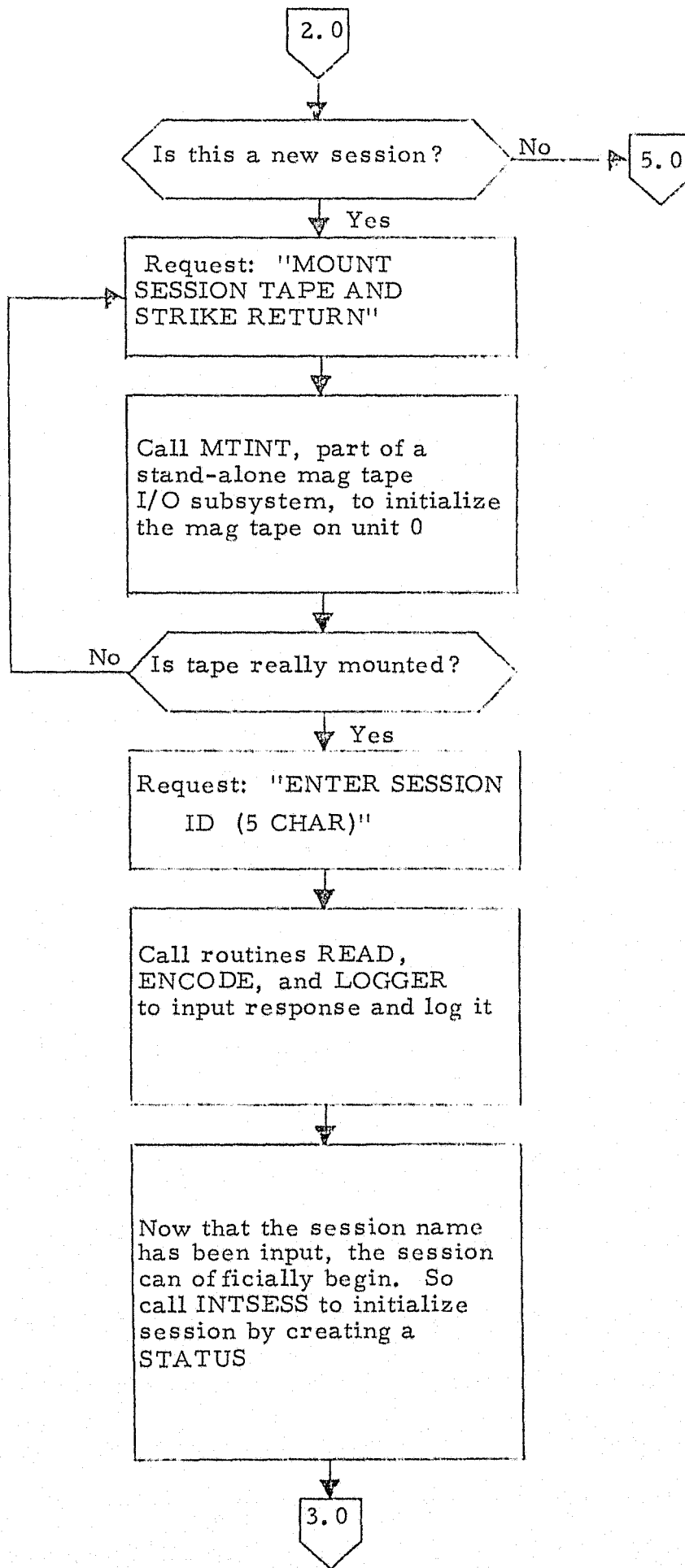
Module functions may be summarized as follows:

<u>Name</u>	<u>Function</u>
DISPATCH	System executive. Operator initiates calls to other modules from DISPATCH.
TERMINATE	System bookkeeping at the end of a session.
SPINP	Speech input (not under RDOS I/O control).
WSDF	Writes speech input from disk to tape.
LABEL 11, LABEL 12	Facilitates segmentation of speech into contextual speech events.
WRFILE	Writes session files except speech to tape and performs bookkeeping functions.
XFEAT	Extracts features from an audio source labeled by LABELXX.
COMPAR	Compares speech samples from labeled samples from two audio sources.

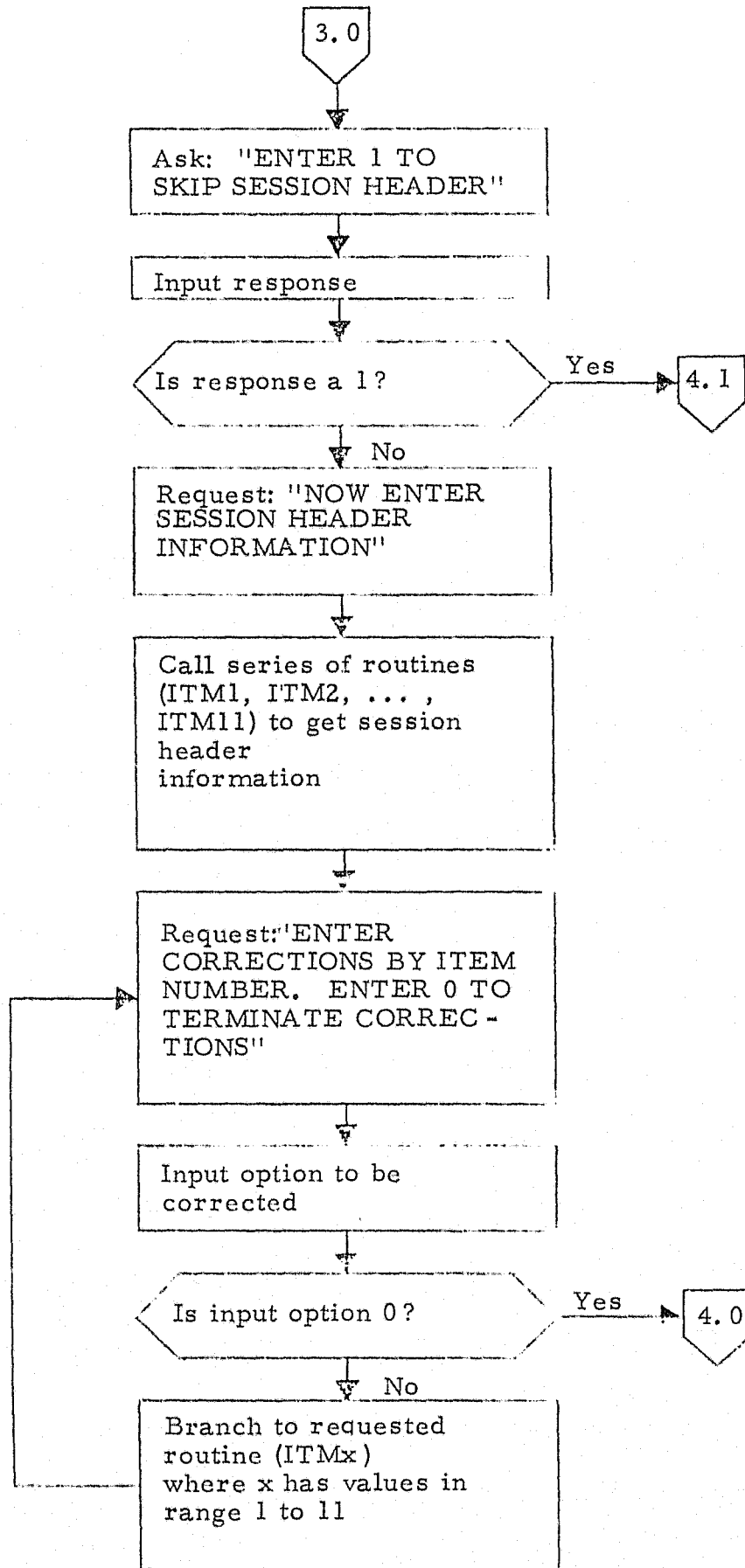
DISPATCH



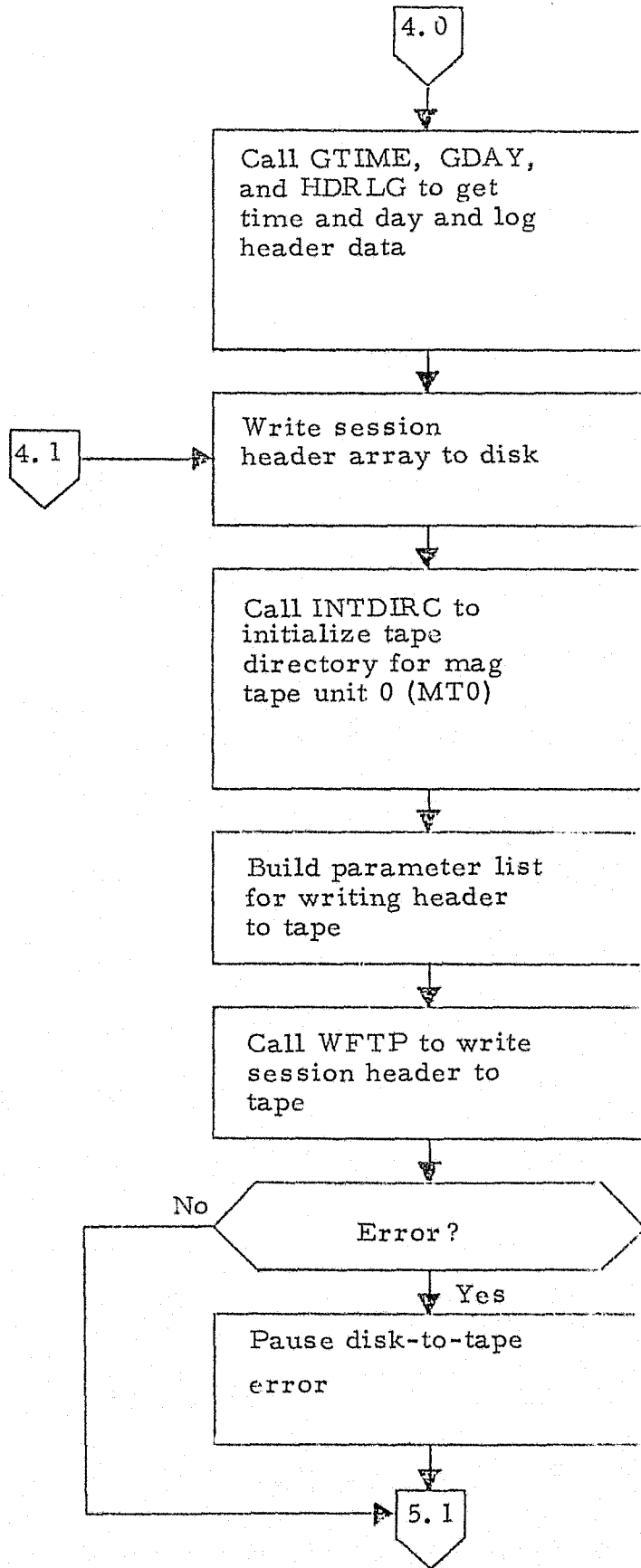
DISPATCH (Contd)



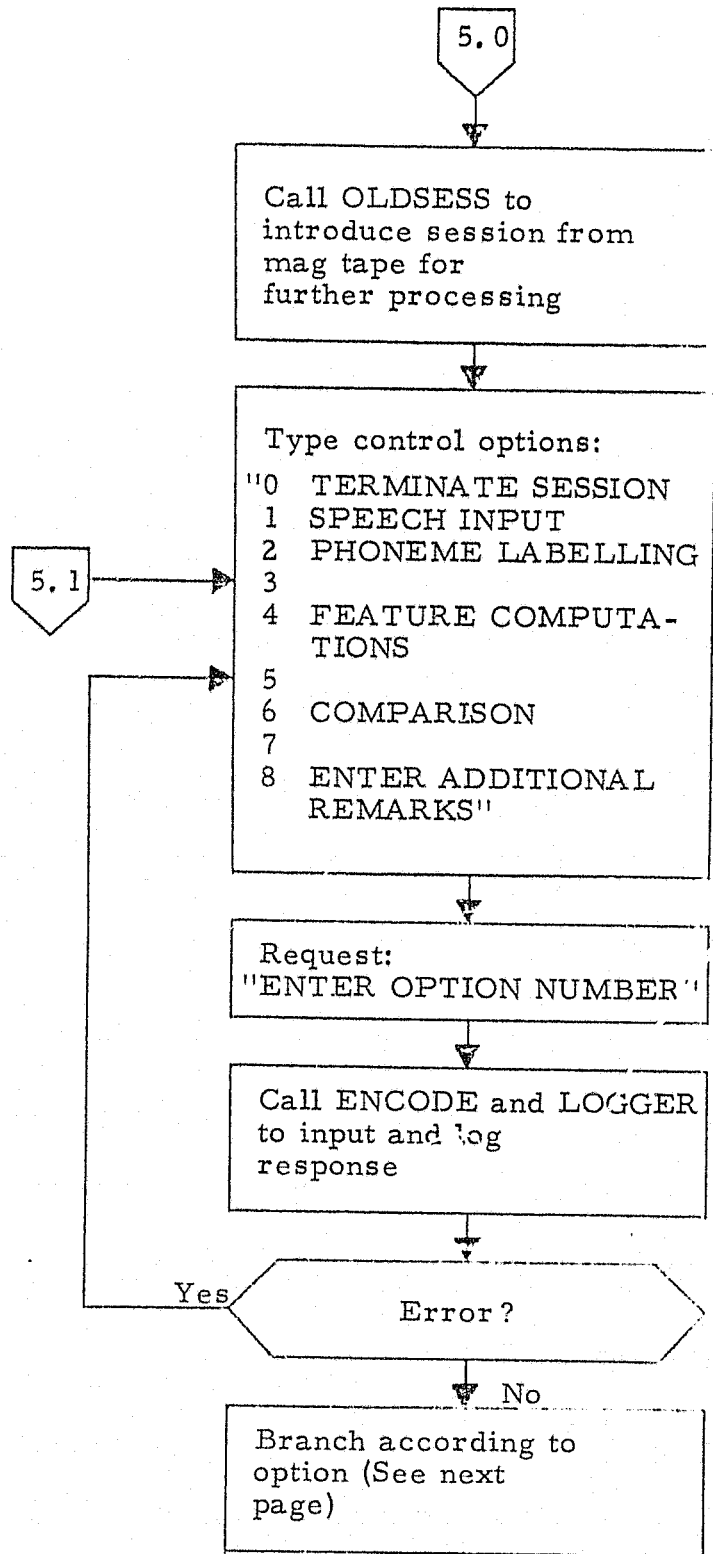
DISPATCH (Contd)



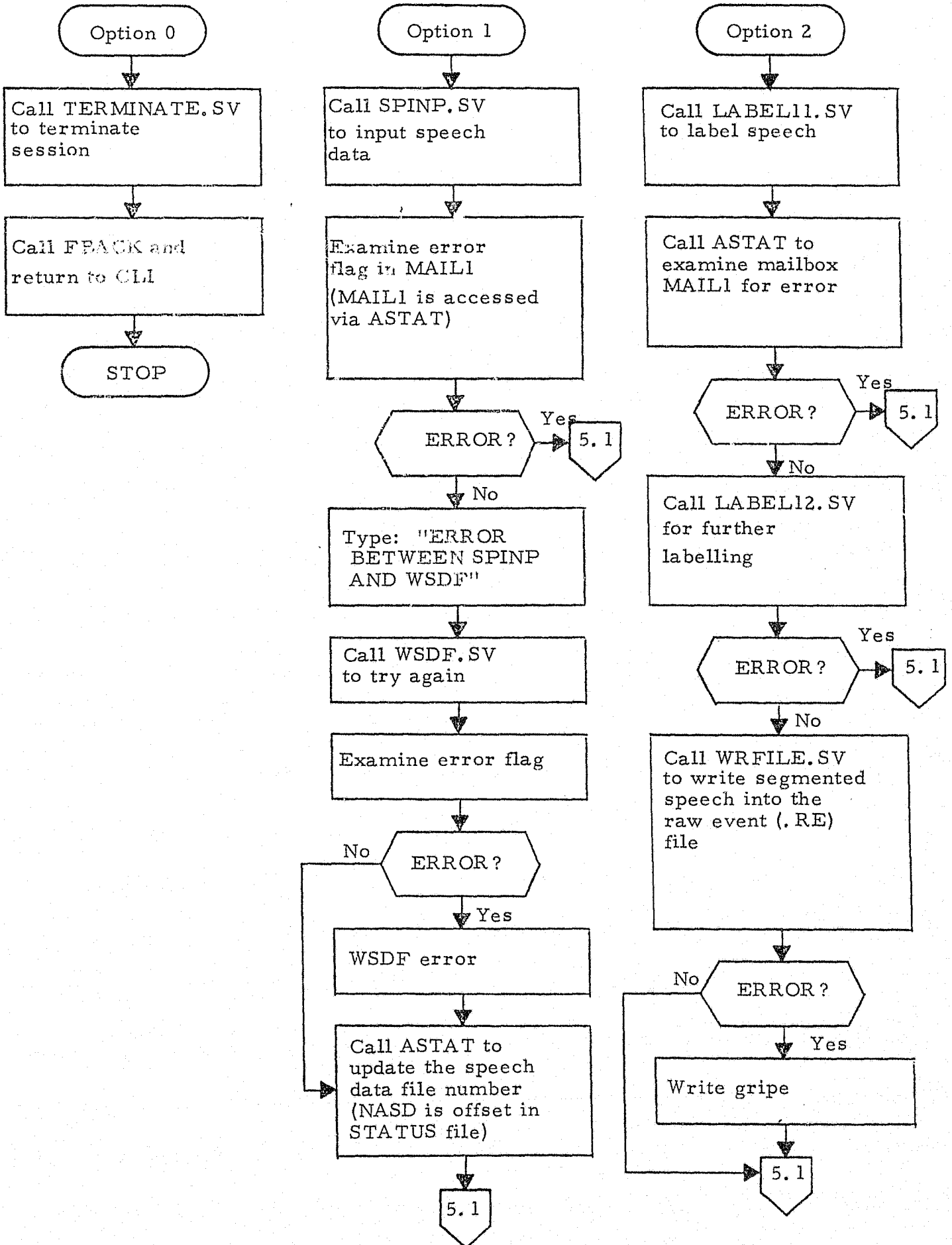
DISPATCH (Contd)



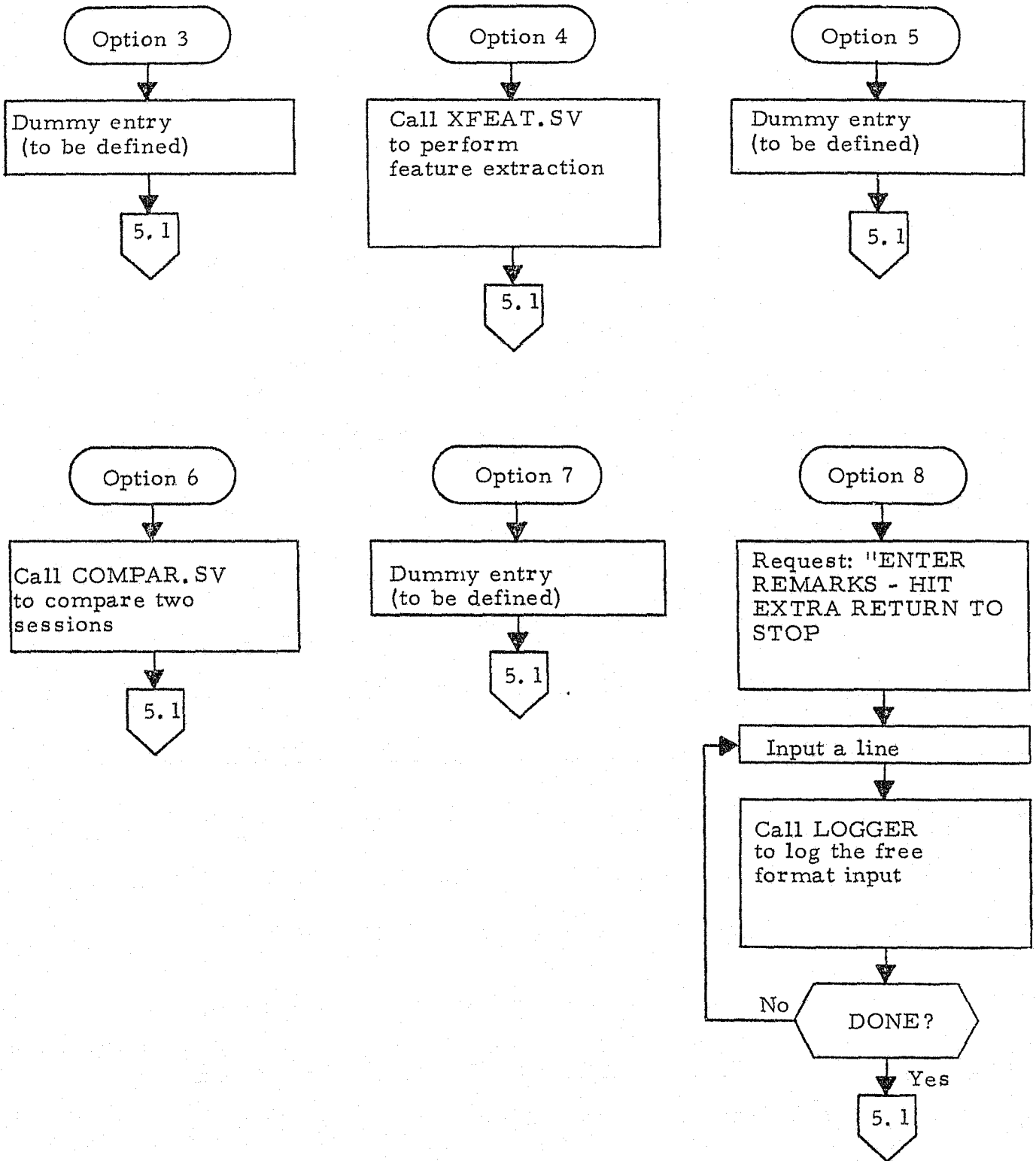
DISPATCH (Contd)



DISPATCH (Contd)



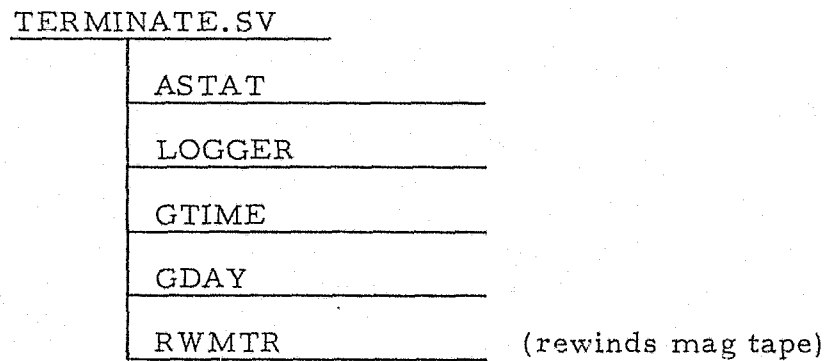
DISPATCH (Contd)



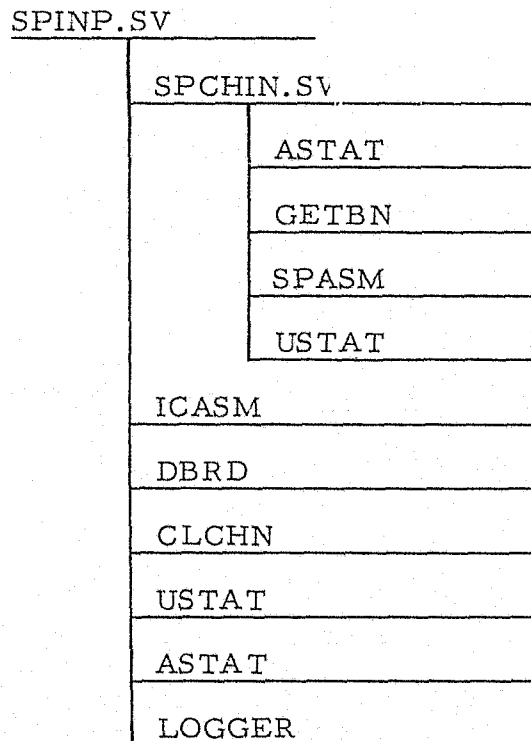
2.3 Major Module Trees

The significant routines and/or swap modules called by each of the major modules called by DISPATCH are defined in trees in this paragraph. Not all subroutines called are shown. For a comprehensive list see the program listings or TREE, produced by TREEGEN and described in paragraph 5.3 below. When a swap is called its routines are shown.

2.3.1 TERMINATE Module



2.3.2 Speech Input Module (SPINP)



WSDF

ICASM
ASTAT
GDAY
GTIME
OFWT
WARC
DBRD
WFMRK
CFWT
CLCHN
USTAT
SDCLR

LABEL11.SV

INITT

VALSS

LOGGER

USTAT

ASTAT

RETRFILE.SV

ASTAT

RFRTN

USTAT

RWMTR

ISTAT

AMPSON.SV

ASTAT

ICASM

DBRD

CLCHN

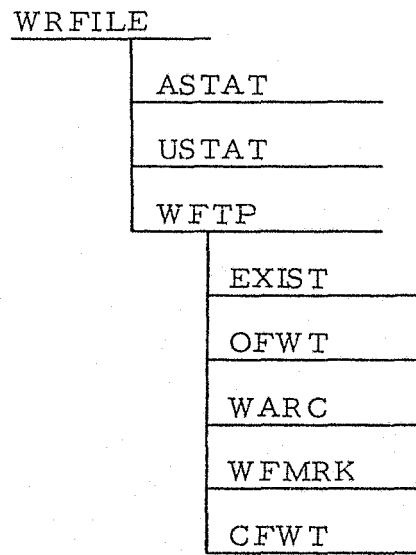
SONO

LABEL12.SV

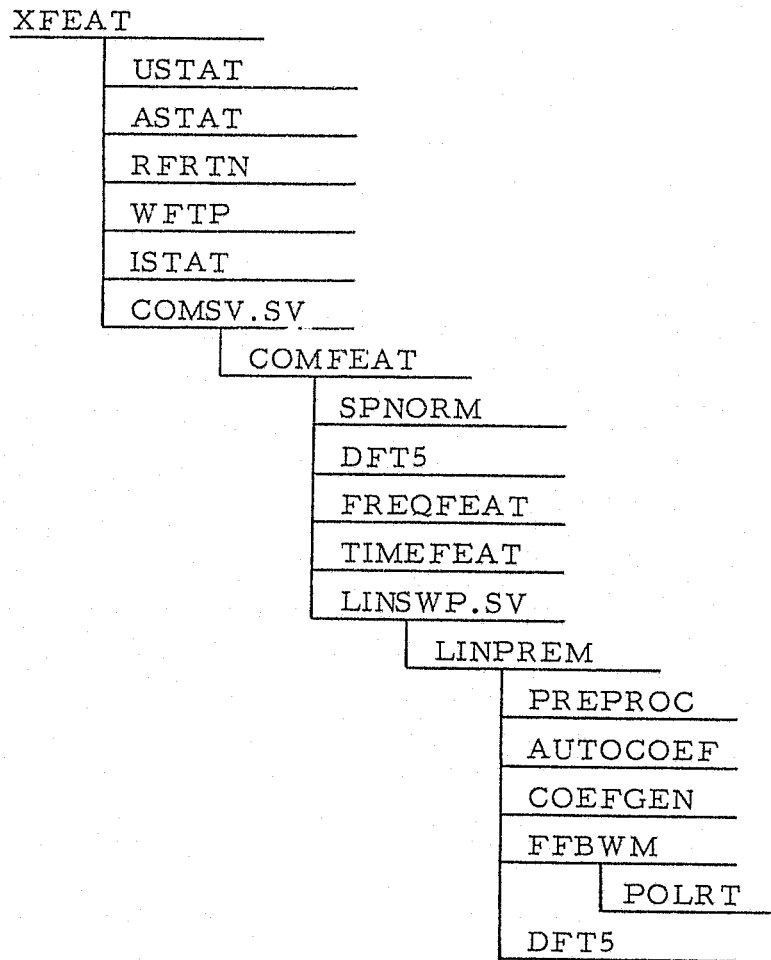
ASTAT
USTAT
DNXFM
LBMAC
MOVABS
ANMODE
VECMOD
DCURSR
DCORR
DRWT
PBACK
GETBN
PLASM
DRWABS
TINPUT
AWT
LOGGER
LBMIC
ICASM
INITT
DBRD
DRHMS
MOVABS
ANMODE
ZCROS
DRWABS
PBACK

2.3.5

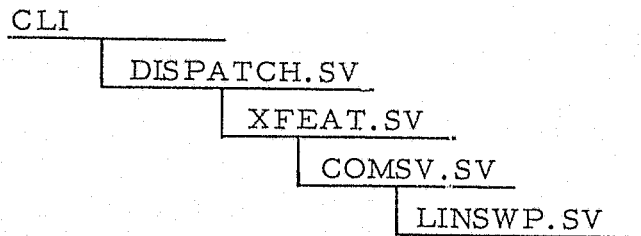
Write File from Disc to Tape (WRFILE.SV)



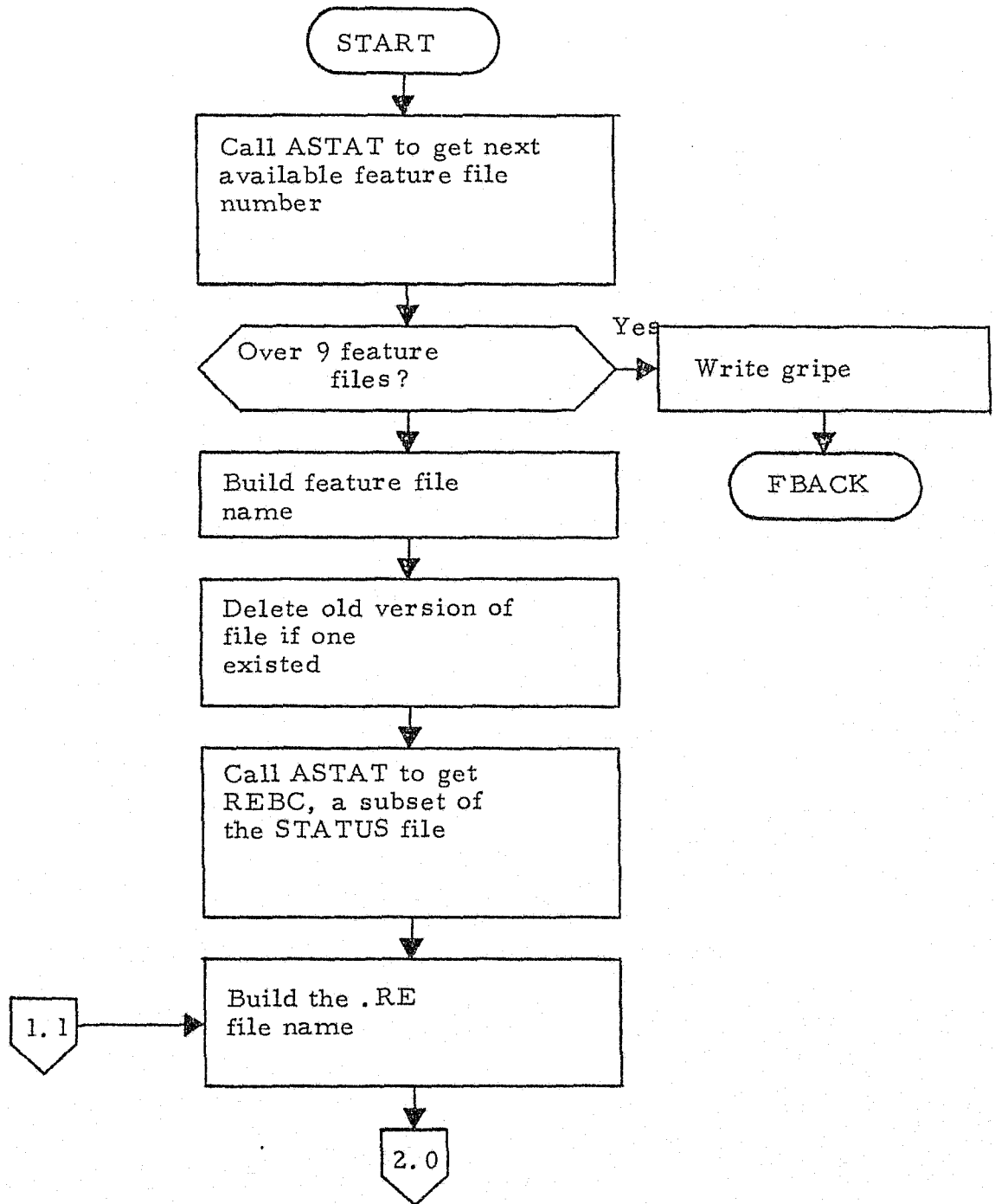
Feature Extraction Module (XFEAT.SV)



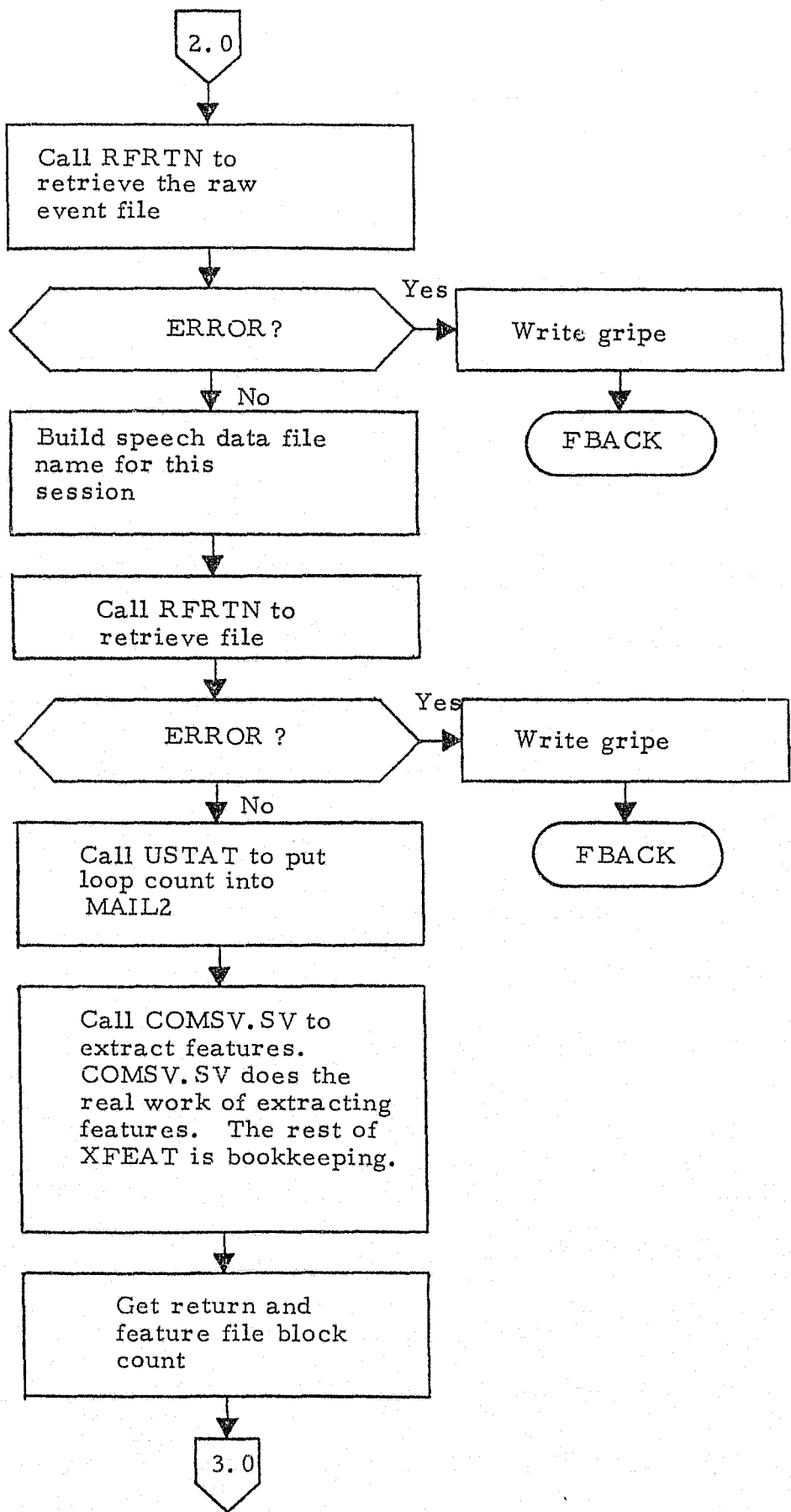
Note: LINSWP.SV is the deepest that SASIS swaps. Pictorially:



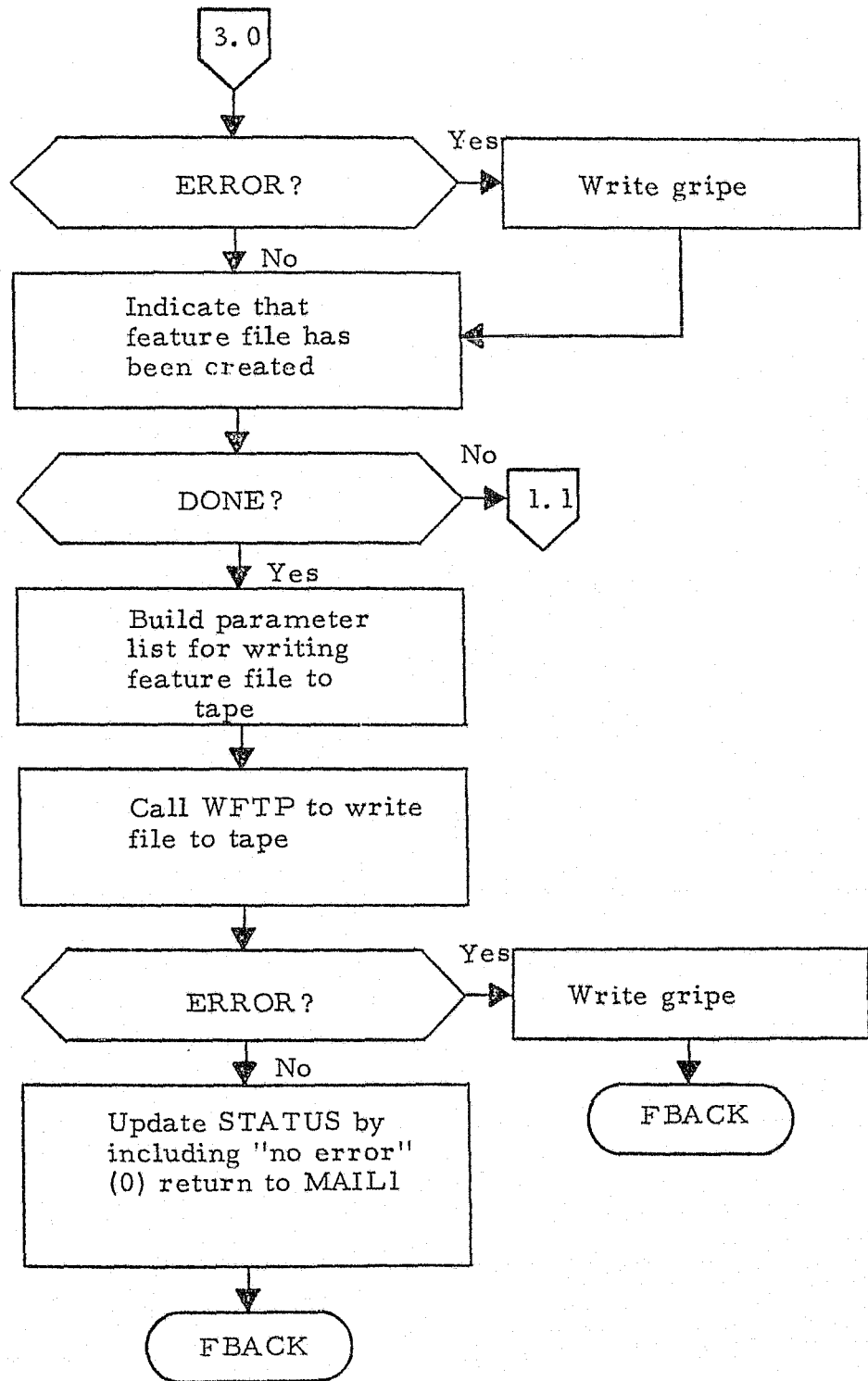
XFEAT



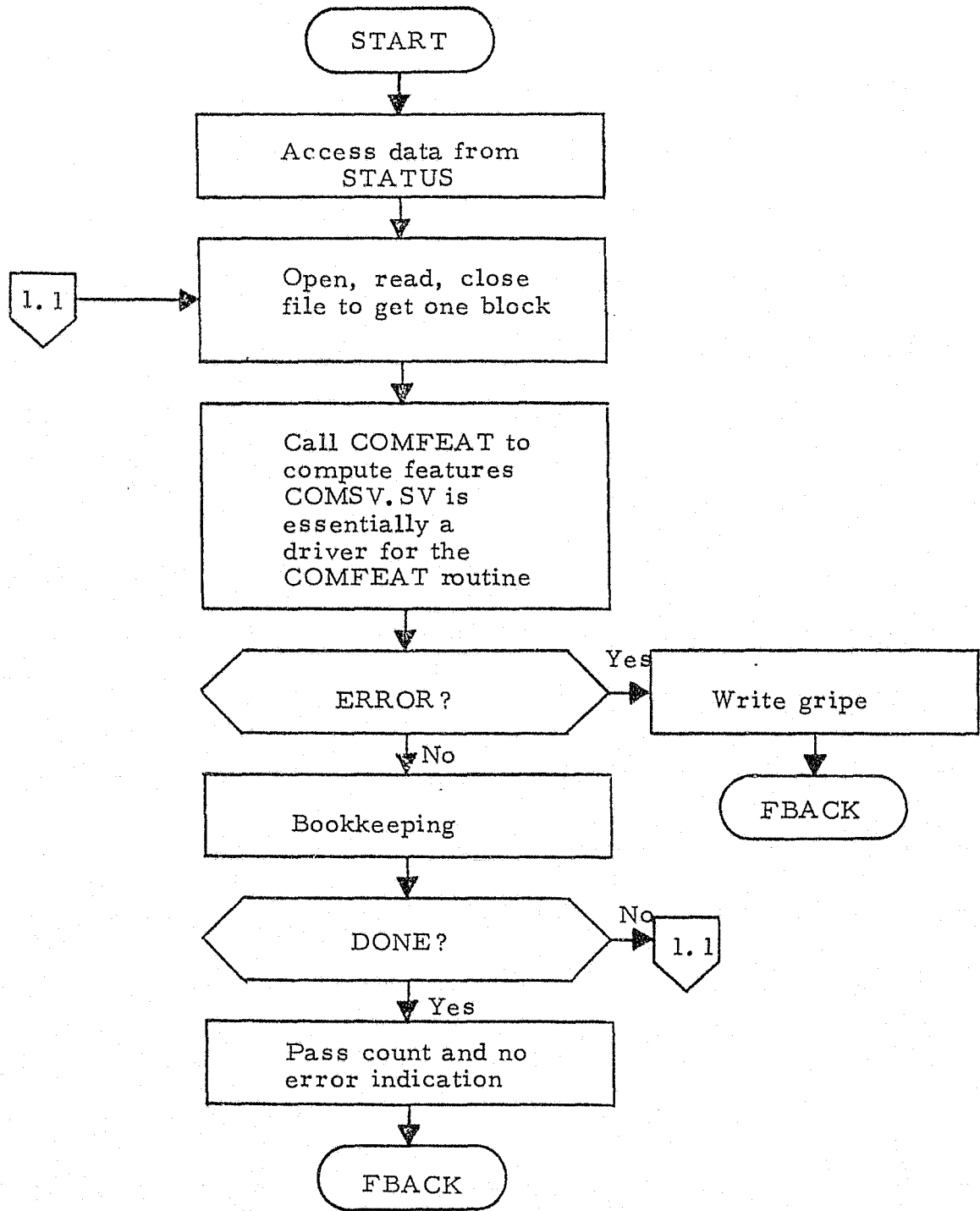
XFEAT (Contd)



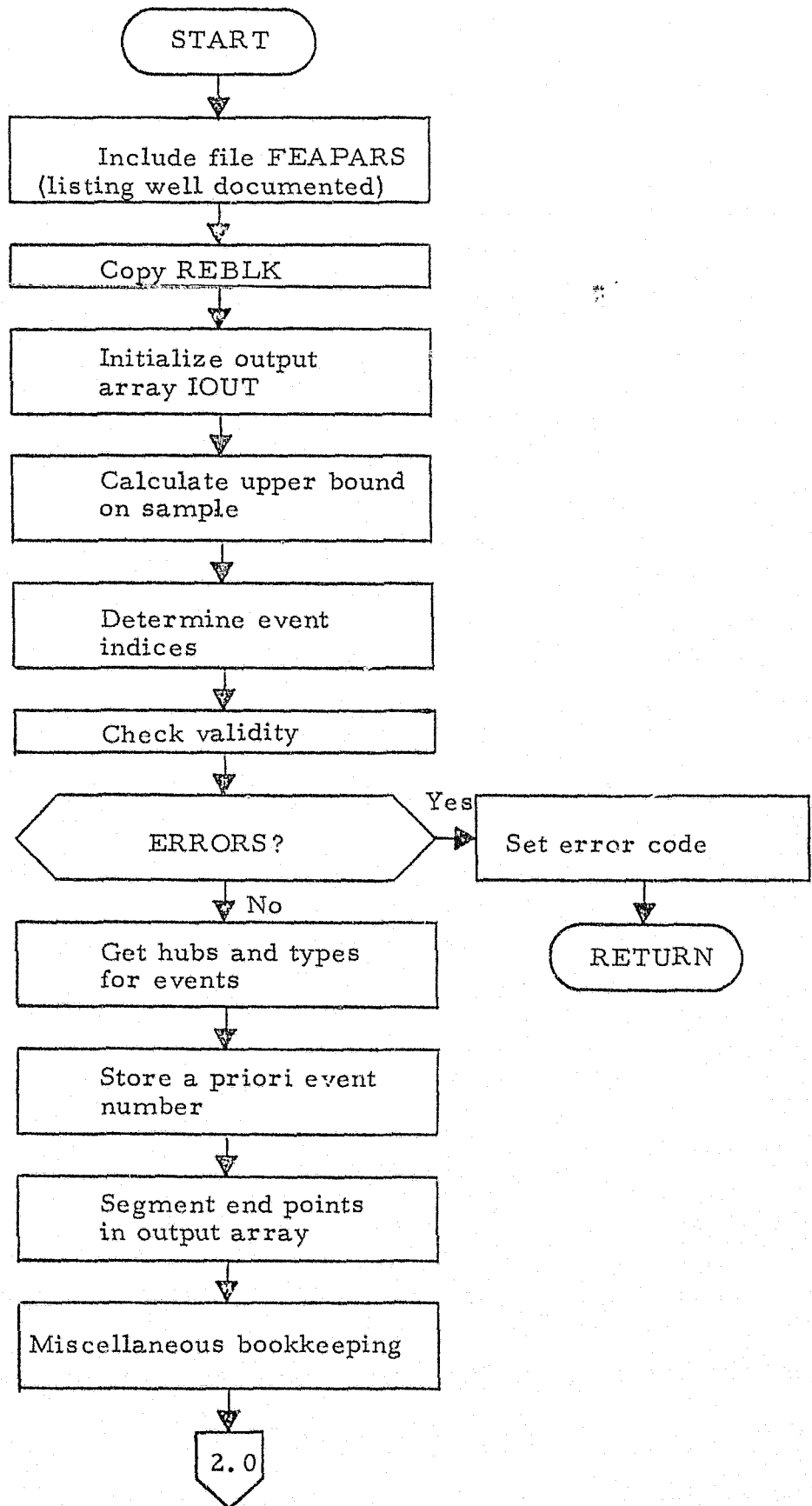
XFEAT (Contd)



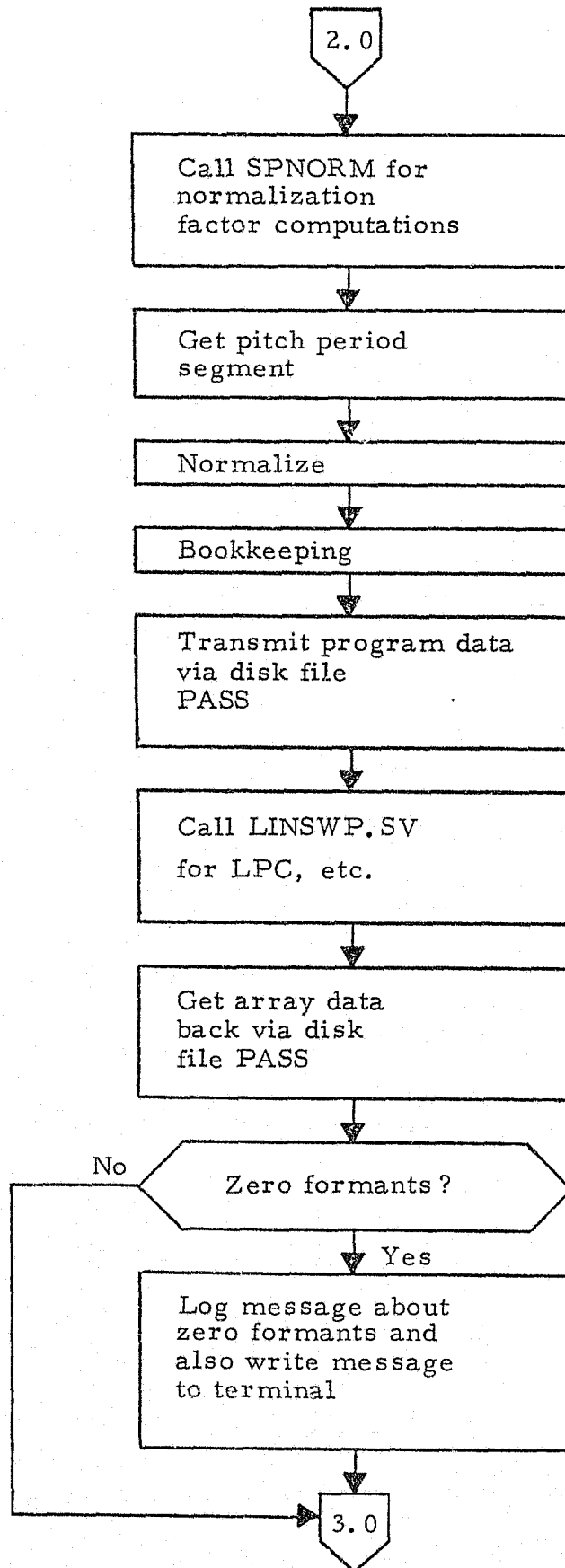
COMSV



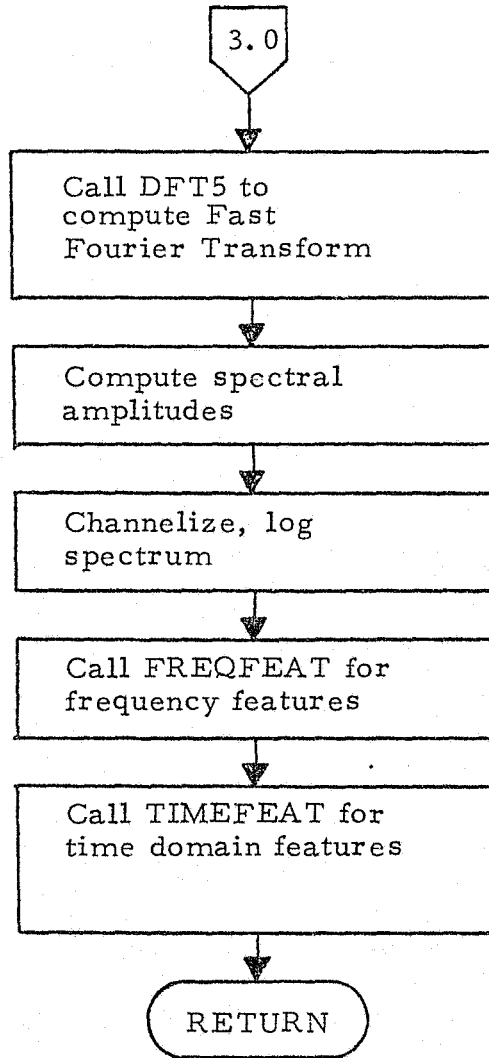
COMFEAT



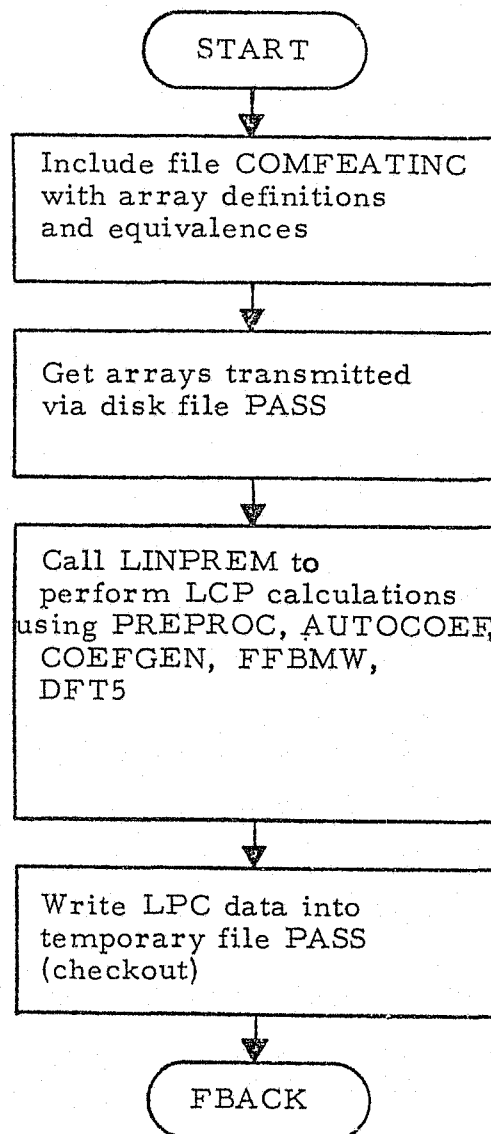
COMFEAT (Contd)

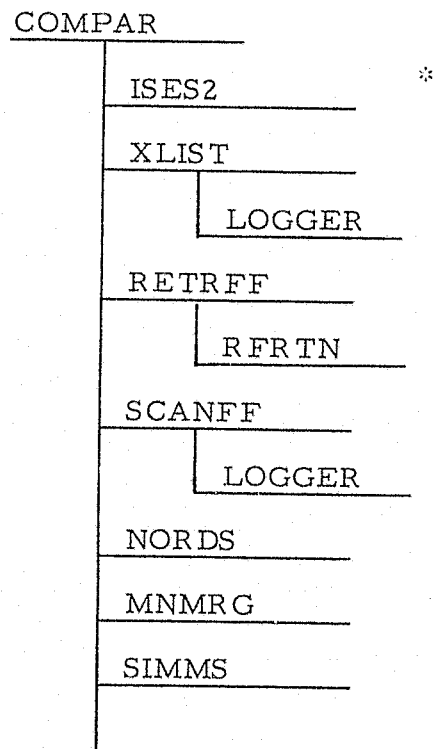


COMFEAT (Contd)



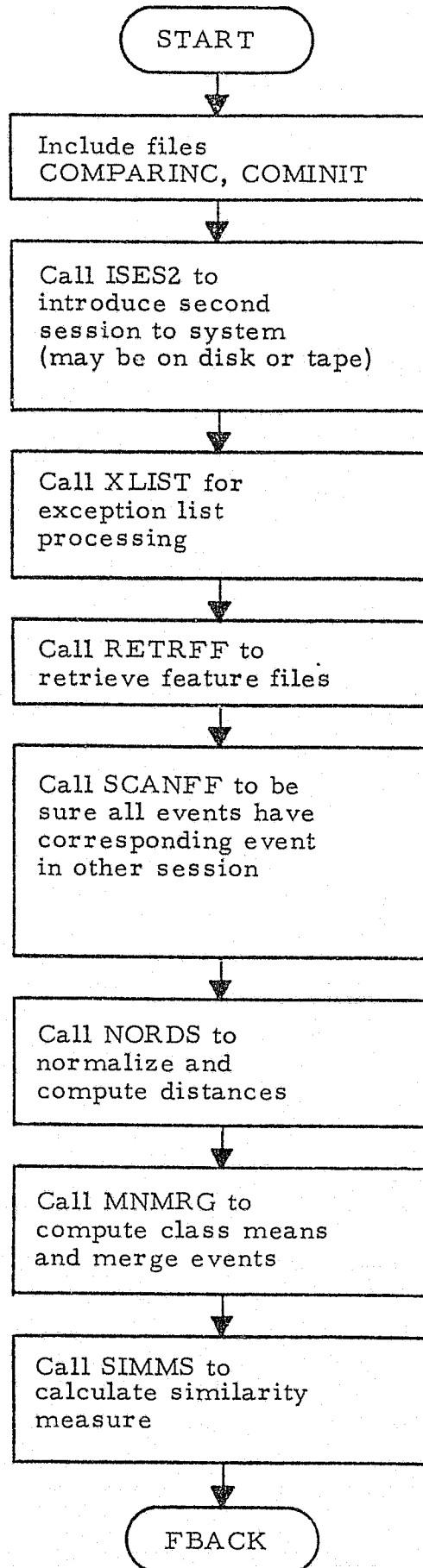
LINSWP



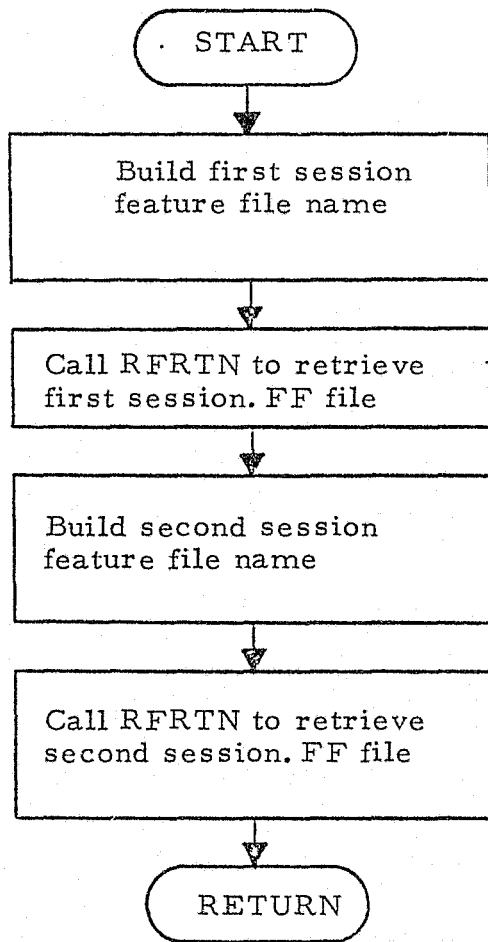


* ISES2 introduces a second session to SASIS. ISES2DEMO substitutes for ISES2. It assumes files are already on disk.

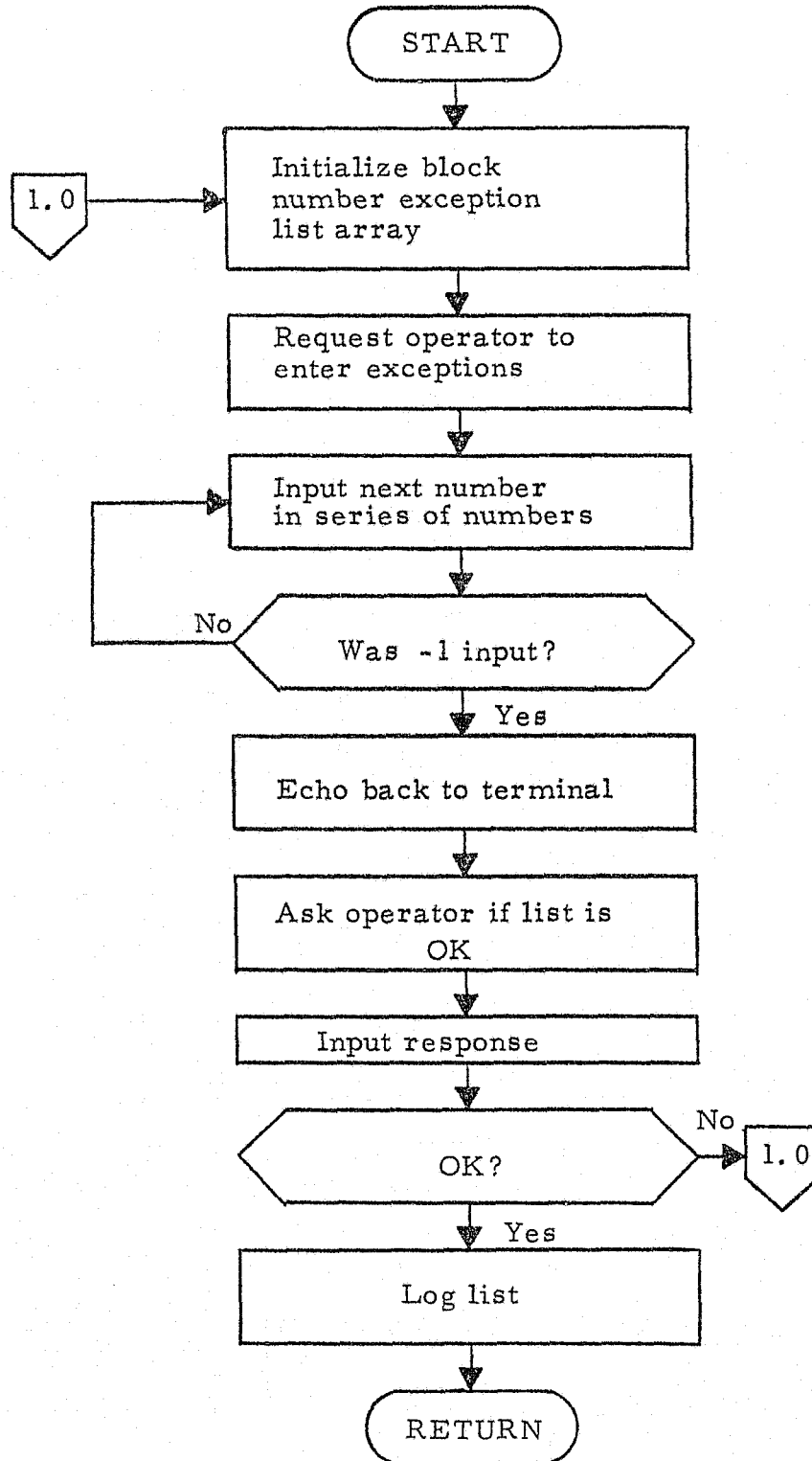
COMPAR



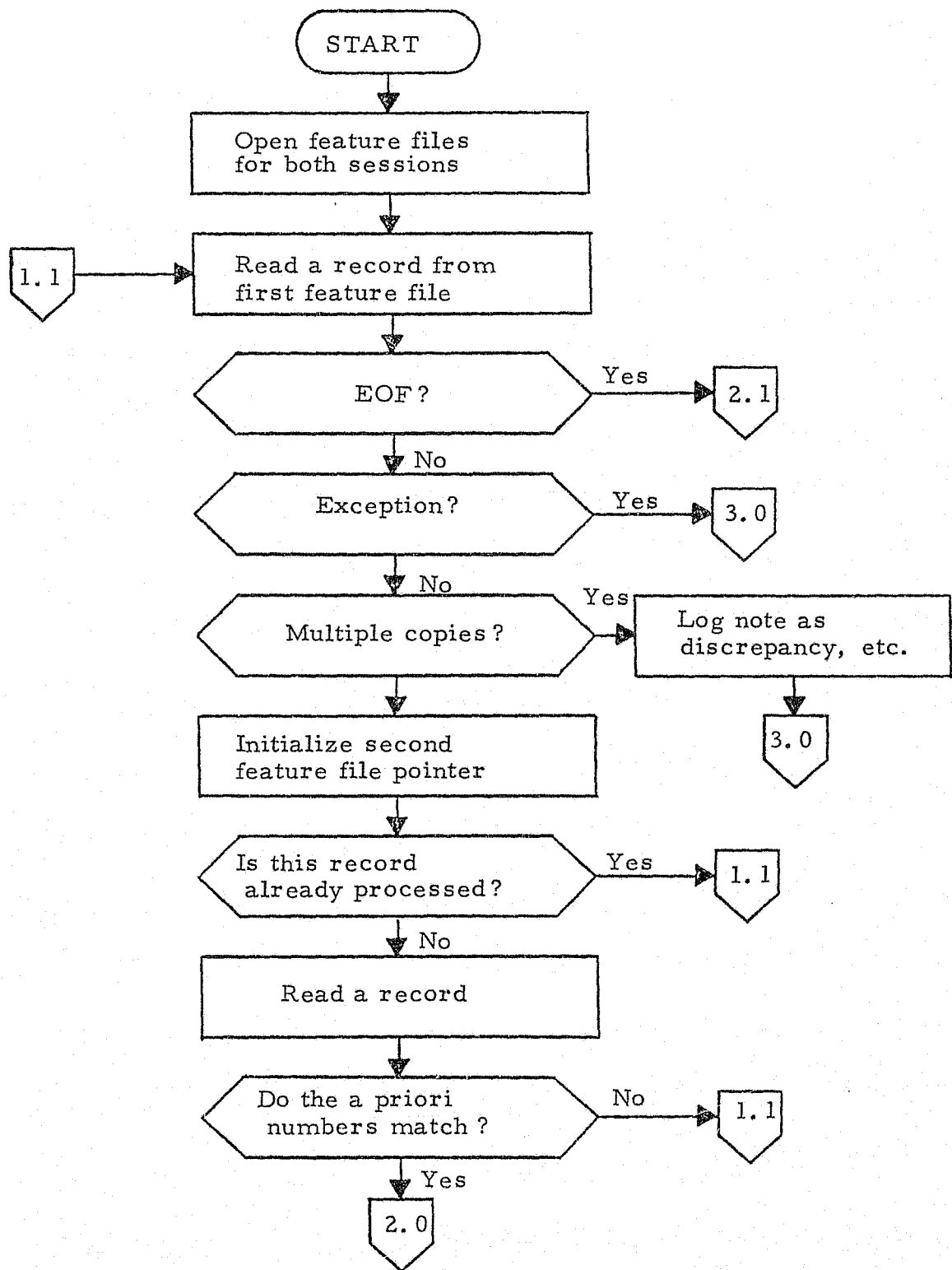
RETRIEVE FEATURE FILE ROUTINE (RETRFF)



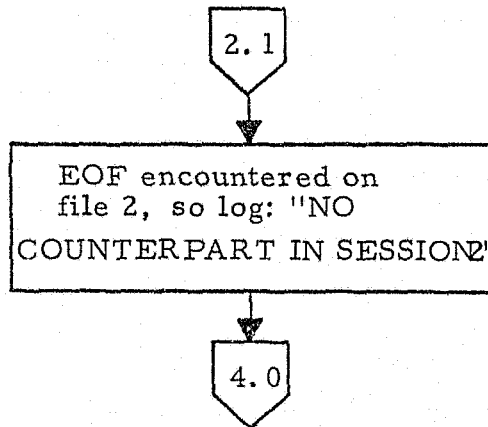
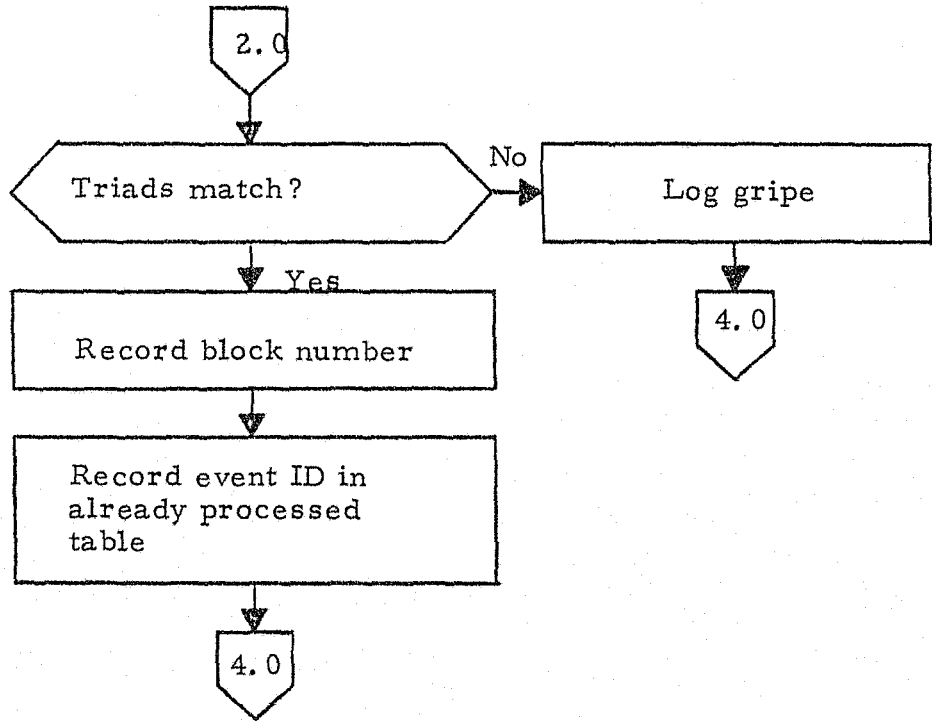
XLIST



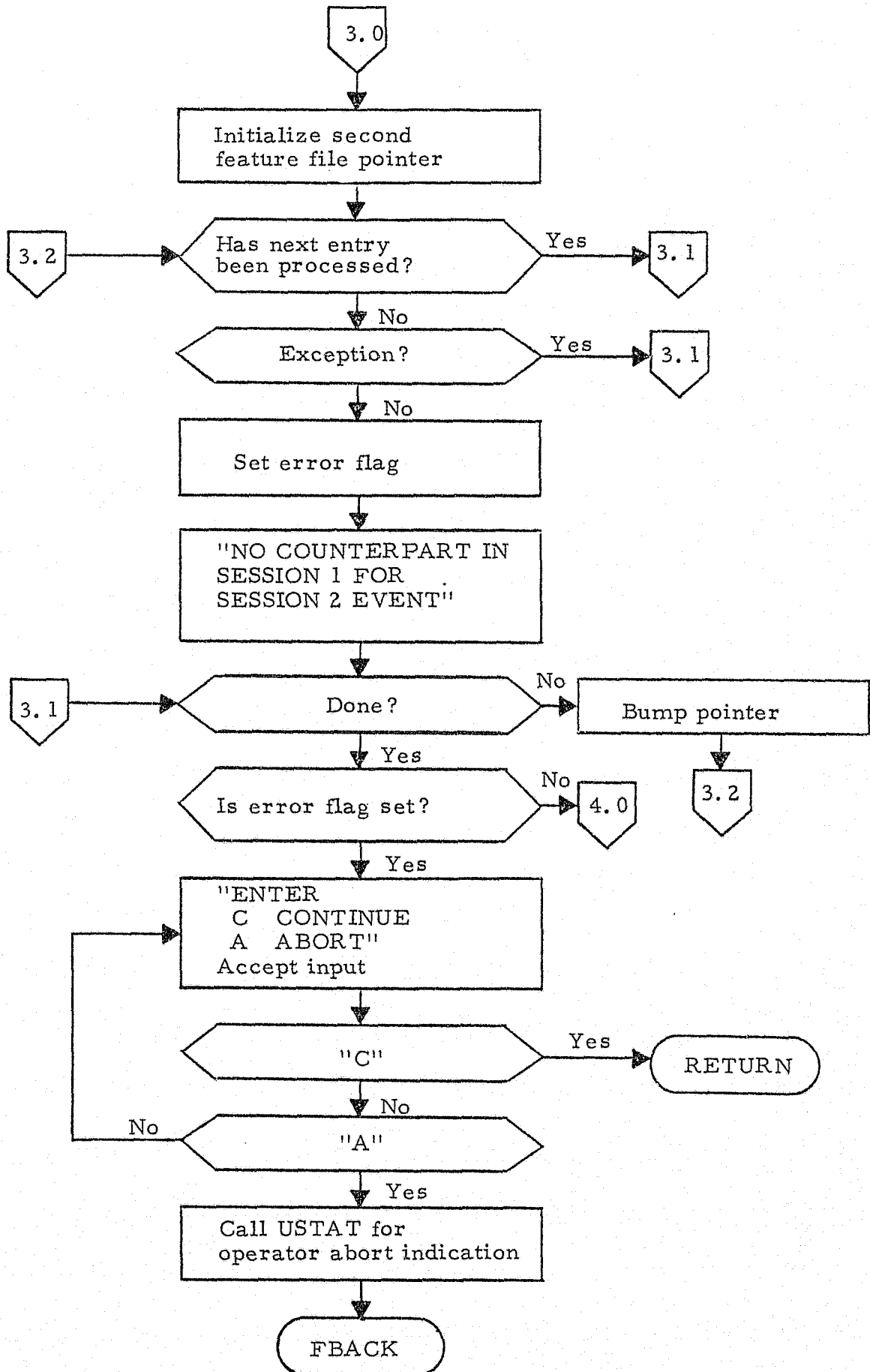
SCAN FEATURE FILE (SCANFF)



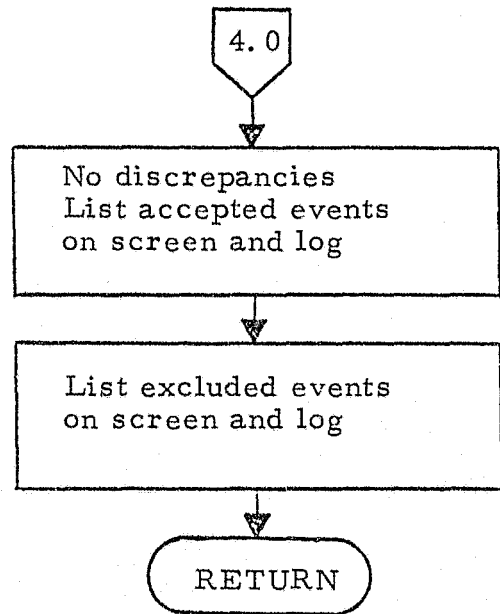
SCANFF (Contd)



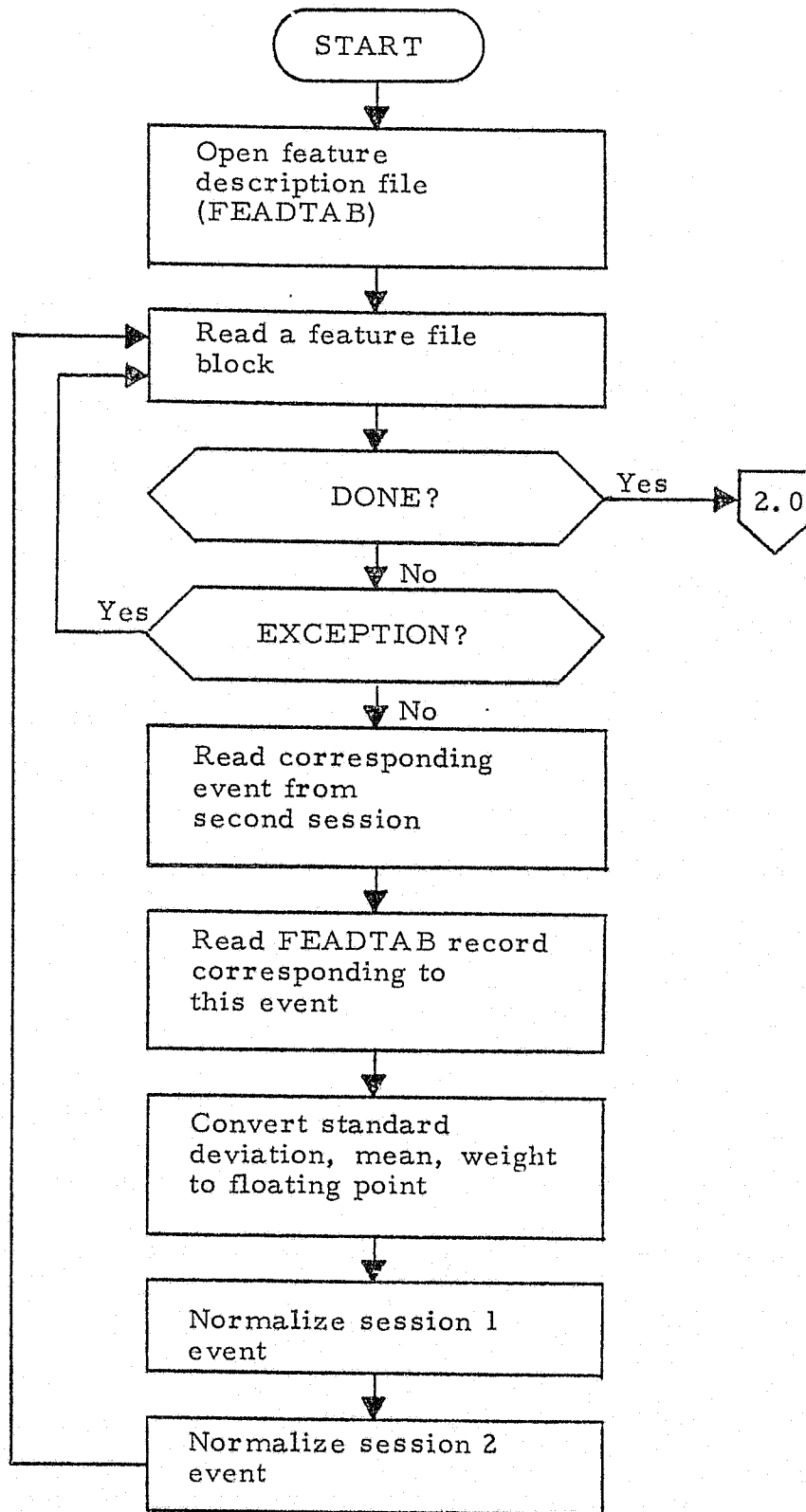
SCANFF (Contd)



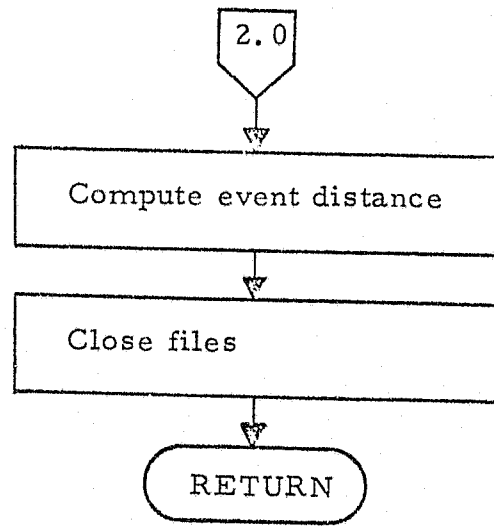
SCANFF (Contd)



NORMALIZE AND COMPUTE DISTANCE (NORDS)



NORDS (Contd)



2.4 Module Communications

The storage capacity required for operating SASIS was accommodated by segmentation into modules (swaps, i. e., .SV files). Data communication between swaps via memory is not possible as it is between subroutines within a swap because the swapping process is a complete "roll-in-roll-out" of memory. SASIS uses a disk to circumvent the data communication problem and enhance the memory. Various disk files were created to support SASIS. Since STATUS already existed, setting aside a few "mailbox" entries *e. g., MAIL1) in STATUS for passing parameters, error codes, etc., provided a straightforward data communication method.

One modules became so large that a module (LINSWP.SV) had to be extracted from it. Since the routines called in LINSWP (see paragraph 2.3.6) process large arrays which the calling module must access, a temporary disk file called "PASS" is used to transfer the arrays from module to module.

The only recommendation on the further use of the STATUS mailboxes or of files, such as and including PASS, is to avoid interfering with current usages. Only after a careful study of the software listings can one be assured of this.

2.5 A Computational View of SASIS Data Flow

This section focuses on the computations performed on speech data by SASIS. The tree diagram is given in paragraph 2.0, and the figure below. The operator inputs speech by initiating a call to SPINP from DISPATCH. Up to 68 tracks (3072 words each) of data (roughly 30 seconds) of contiguous speech can be input. SPINP scans the data to protect against saturating the screen. The number of samples over 0.85 of the maximum value is computed for this purpose to determine the likelihood of saturation. Nothing is done to the data itself.

The next processing occurs in the LABEL module (actually two .SV files) where the speech events are identified. An amplitude contour and spectrogram of the speech data are computed and presented to the operator to support the labeling process. The output of the labeling sequence is a file of speech event definitions which are listed by the sample numbers that delimit speech events. Scaling is performed prior to display.

When all events have been identified in one to nine audio samples (i.e., maximum total speech would be 9x30 or 270 seconds), the operator computes a set of feature values for each event by calling the feature extraction module XFEAT. XFEAT produces a file (-.FF) containing the values of the computed features. The format of that file is shown in paragraph 3.2. When execution XFEAT is completed, all computations describing a session are complete. The procedure is then repeated for another speech sample which is to be compared with the first. The comparison of the two samples is then initiated.

COMPAR performs the comparison between sessions. Any discrepancies between event lists (e.g., an event in one session may have no counterpart in the other) are processed by the operator. The features of each event are then normalized and distances computed by NORDS as follows.

1. A working array WORK1 of normalized feature values is computed for each event:

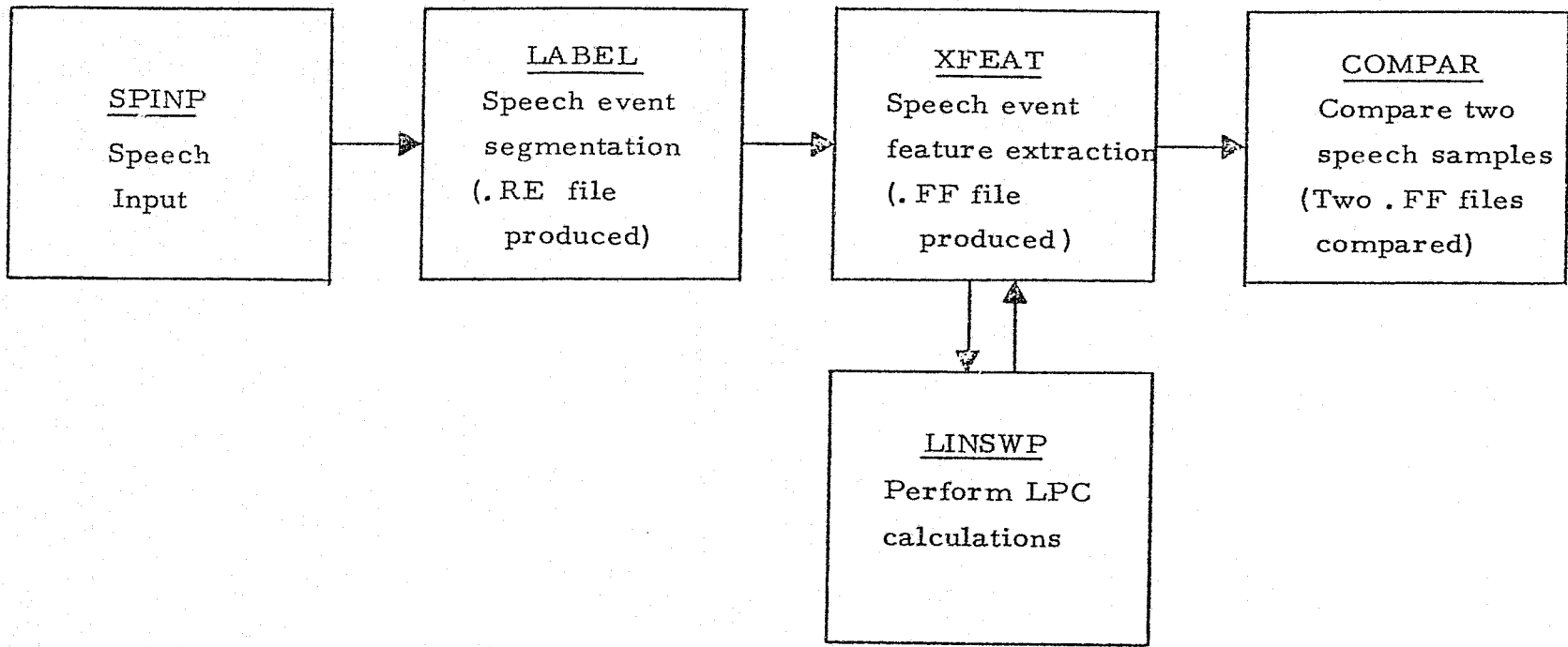
$$\text{WORK1}(K) = \frac{64.0}{\text{SDEV}} \times (X - \bar{X}) + 127.0$$

where: SDEV is the standard deviation of the feature and \bar{X} is the mean value. WORK1 values less than 0 are set to 0; values greater than 255 are set equal to 255.

2. The distance measures are computed from the normalized session 1 and 2 values X_1 and X_2 .

$$\text{DIST} = \sum \text{WT}^2 \times (X_1 - X_2)^2$$

where: WT is the weight associated with the feature. NORDS retrieves the means, standard deviations and weights from FEADTAB. MNMRG takes the distances for the 13 possible event classes and merges them into 10. SIMMS computes a similarity measure from the merged events using SIMTAB and then computes and outputs a page number for table look-up by the operator.



HEURISTIC FLOW BETWEEN MAJOR MODULES

3.0 DATA MANAGEMENT

The design goal of the SASIS management subsystem (DMS) was the same as any data management facility, i. e., to support storing and retrieving data in an efficient and orderly manner. The particular directions SASIS DMS took were dictated by the application and by characteristics and certain limitations of RDOS. Three of the latter should be explained at the outset.

First, the speech sampling rates required by SASIS are too great for RDOS. * There is simply too much overhead in RDOS to handle the A-D. SASIS, therefore, takes control of I/O processing during speech input, inputting from the A-D and writing directly to a directoried disk file. Since the speech data are directoried under RDOS after I/O control is returned to RDOS, the data can be accessed via RDOS I/O facilities.

Secondly, it would have been possible to use RDOS magnetic tape I/O facilities except for the fact that all files including tape units are closed during program swaps. The overhead resulting from frequent tape rewinding and repositioning would have been prohibitive. Therefore, stand-alone mag tape software was created to circumvent this operating system feature.

Third, when required, system facilities existed via assembly language calls but not via FORTRAN calls, driver routines written in assembly language and called from FORTRAN were implemented. A number of these, as well as other design "diddes" were found necessary. As revisions of RDOS are released, undoubtedly some of these problems will be resolved leaving the present version redundant and somewhat awkward.

* There is a potentiometer on the A-D board which can be adjusted with a small screwdriver. The potentiometer controls the clock rate. The range of the clock rate is higher than the 6800 samples per second rate required by SASIS. Therefore, the clock rate is set to $(6800 \times 2) = 13,600$ samples per second and every other sample is used giving an effective rate of 6800 samples/second.

For continuity, the various routines are described below. They are grouped by function.

3.1 Data Management Software

This software description is divided into four functional groups: executive, disk-tape transfer, logging, and speech input.

3.1.1 Executive Level DMS

The INTSTAT, ISTAT, USTAT, and ASTAT routines create and maintain the STATUS file which contains the next available file numbers for various file types and the status of active mag tapes such as their position. STATUS is also used as a mailbox to transmit information (parameters, error flags, etc.) between swaps. Each of these routines references STDESC as an INCLUDE file for symbolic offsets into STATUS. The structure of STATUS and these offsets are shown in the DM specification.

3.1.1.1 Initialize STATUS File (INTSTAT)

INTSTAT deletes STATUS if it existed from a previous session. It creates and initializes STATUS including the values of the next available file numbers (e.g., SPEECHDATA) before returning to the calling routine.

3.1.1.2 Access STATUS File (ASTAT)

ASTAT accesses the contents of STATUS. Its calling sequence is CALL ASTAT(IX,ARRAY,ICNT,NUNIT), where IX is the symbolic entry number, ARRAY is the return array, ICNT is the number of words to return in ARRAY, and NUNIT refers to the mag tape unit. When referring to entries not tape unit specific, use NUNIT=0.

3.1.1.3 Update STATUS File (USTAT)

USTAT is the complement of ASTAT. Instead of a block being read from STATUS to an array, an array is written into STATUS. The calling sequence (except for the subroutine name) is the same.

3.1.1.4 Increment STATUS (ISTAT)

ISTAT is equivalent to an ASTAT and USTAT call for updating single parameters. The calling sequence for ISTAT is CALL ISTAT(IX,INCRE,NUNIT) where entry IX is to be incremented by INCRE. NUNIT has the same meaning and use as in ASTAT and USTAT.

3.1.2 Disk-Tape, Tape-Disk Transfers

The DMS is designed for maximum ease of file retrieval. To the applications program, the location of a file is irrelevant. If on tape it is transferred to disk. If on disk already the file is simply opened. Since tape and disk file formats are almost identical, transfer between the media is straightforward. Only the SPEECHDATA file is handled as a special case because there are no headers on the disk version. Those routines called by application level software are explained.

3.1.2.1 Write File To Tape (WRFILE)

WRFILE is called as a swap. It is in essence a driver for WFTP (Write File to Tape). Parameters shown below are passed to WRFILE via the STATUS file mailbox MAIL1 using USTAT.

CALL USTAT(WORD, A, COUNT)

Where WORD is the first word to be updated in the data block within the disk block, A is the array containing the data block, and COUNT is the block word count.

3.1.2.2 Write Speechdata to Tape (WSDF)

WSDF was written to supplement WRFILE because of the special requirements of the SPEECHDATA file. The standard header present on the disk version of .RE and .FF files does not reside on SPEECHDATA. Therefore the speechdata must be reformatted before it is written to tape.

3.1.2.3 Retrieve File (RFRTN)

RFRTN is called to retrieve a file in the event it isn't disk resident. It scans TDIRCZ or TDIRC1 as appropriate to determine the mag tape file number. The tape is repositioned if necessary and the file read to disk.

SASIS accesses and creates and maintains various data structures during a session. A complete list of these and a description of their use and/or a cross reference to such a description is included.

<u>Name</u>	<u>Description</u>
STATUS	A working file created at session initialization maintained and accessed by most SASIS modules during the session. Used primarily for tape management file name construction and parameter passing. See paragraph 3.1.1.
TDIRCZ/TDIRC1	Simple directories of (named) files on SASIS session tapes. The order of the entries is the order of the files on tape. TDIRCZ is a directory to the mag tape on unit 0 (the session being processed) and TDIRC1 to the tape on unit 1 (the session being compared). See paragraph 3.1.2 ff.
SHEDR.	The session header contains descriptions and parameters relevant to the entire session. See below for format.
SPEECHDATA	A 68 track contiguous file created and given a permanent attribute at disk pack initialization. (See CSDFL program description under paragraph 5.1.) SPINP calls SPASM to input data from the A-D to SPEECHDATA. PBACK calls PLASM to play it back through the D-A. SDCLR fills SPEECHDATA with zeros. Other routines can access SPEECHDATA contents directly.

<u>Name</u>	<u>Description</u>
-.RE	The .RE extension defines a raw event file consisting of 42 word records each of which define a labelled speech event. Among other information (see below for format) the raw event record contains the beginning and ending sample numbers of the pitch period and the sample number of the middle of the 100 ms. segment containing the event. LABEL creates this file.
-.FF	The .FF extension defines the feature file. Each 768 word record corresponds to a .RE record. XFEAT creates this file from the .RE file. It computes features for the event and stores them in the FF block.
FEADTAB	The Feature Description Table consists of thirteen 121 word records. Each record consists of the event class number and thirty 4 word feature descriptor blocks. These descriptor blocks contain the feature number, standard deviation, mean, and weight for that feature. These values are stored as 16 bit integers and are scaled at execution time by NORDS.
SIMTAB"N"	A set of Similarity Tables are also on disk. N corresponds to the number of event classes available for comparison between the two sessions. The file name is computed by SIMMS before the file is opened and accessed by SIMMS to determine the similarity measure.
PASS	See paragraph 2.4.

The format of the FF record is shown below. To maintain consistency between the analytic studies software and the SASIS operational system, the number of the feature is its entry number in the FF record. The reason the record is so long is that 768 words is an even 3 sectors on a 12 sector/track disk. There is surplus capacity for enhancement.

RAW EVENT FILE RECORD FORMAT

0	FILE TYPE (=5)
1	
2	
3	FILE
4	NAME
5	
6	
7	
8	MONTH
9	DAY
10	YEAR
11	HOUR
12	MINUTE
13	SAMPLE NO OF MIDDLE
14	OF 100 MS SEGMENT (HMS)
15	(IN DOUBLE PRECISION
16	FORMAT)
17	SAMPLE NO OF BEGINNING
18	OF PITCH PERIOD SEGMENT
19	
20	
21	SAMPLE NO OF END OF
22	PITCH PERIOD SEGMENT
23	
24	
25	LEFT EVENT OF TRIAD
26	ASCII LABEL
27	MIDDLE EVENT LABEL
28	
29	RIGHT EVENT LABEL
30	
31	EVENT NO W. I. SUBSESSION
32	A PRIORI EVENT NO.
33-	EVENT
	ORTHOGRAPHY
41	

SESSN-M-KKKK.RE

FEATURE FILE RECORD FORMAT

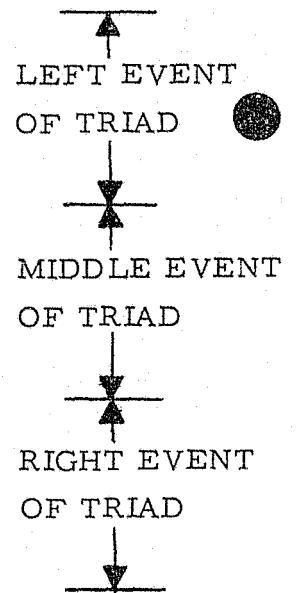
1	FILE TYPE (=6)
2	FILENAME
3	
4	
5	
6	
7	
8	
9	
10	EVENT ID

FEATURE FILE RECORD FORMAT (Contd)

Word

81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104

BEG SAMPLE NUMBER (FROM REBLK (18:21))
END SAMPLE NUMBER (FROM REBLK (22:25))
SESSION ID
EVENT ID



Note: According to FEADTAB 105-210 not used.

FEATURE FILE RECORD FORMAT (Contd)

Word

248

249

250

251

252

253

↓
635

CROSSINGS PER PERIOD
NUMBER OF ZERO CROSSINGS

↑
TIME FEATURES

↓
LPC

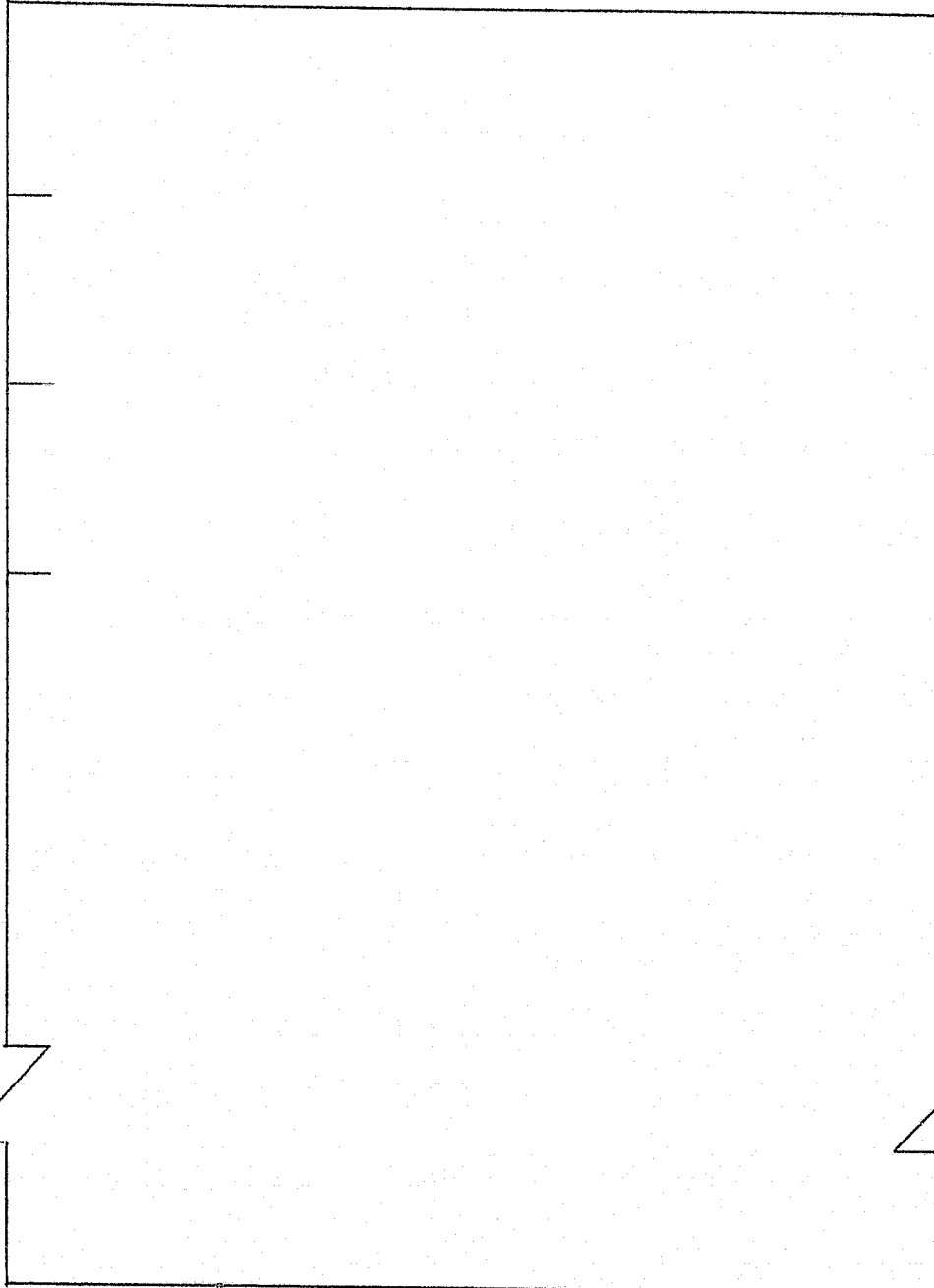
Note: 315-509 not used, 567-768 not used.

WORD

1
2
3
4
5

EVENT NUMBER
FEATURE NUMBER
STANDARD DEV
MEAN
WEIGHT

} 30 of these blocks



121

FEATURE DESCRIPTION TABLE FORMAT

The FEADTAB disk file consists of 13 such tables, one for each event class.

4.0 SYSTEM CREATION AND BACKUP PROCEDURES

This section explains how to create a SASIS operational system starting with source programs only. Although some of this material is covered by RDOS manuals, the redundancy seems justified by the goal to bring programmers new to SASIS and RDOS "on board" quickly. Concurrent reading of the RDOS manual section on the RLDR command will be helpful.

4.1 Creating Executive Modules

The output of a successful compilation by the FORTRAN compiler is a relocatable binary file. Relocatable binary files are linked by the relocating loader into executable modules. These are executed from the RDOS CLI (the Command Line Interpreter) is in control when the system comes up. Control is returned to the CLI when any program is completed, or as a swap from another executable program depending on the context in which the module is used. For example,

1. Compile source program BILL:

```
FORTRAN BILL (CR) *
```

2. Link load BILL with FORTRAN 5 run time routines:

```
RLDR BILL FORT5.LB
```

3. Execute BILL by simply entering its name from CLI:

```
BILL(CR)
```

4. Or suppose BILL was also called from JOE as a swap. Here is JOE:

```
      .  
      .  
      .  
CALL FSWAP (BILL.SV)  
      .  
      .  
      .
```

* (CR) indicates CARRIAGE RETURN key is pressed.

There are a number of .SV modules in SASIS and the number of subroutines each calls is nontrivial. Therefore, "canned" procedures have been defined to facilitate creation or recreation of the .SV modules when only the source files are available. For example, there may be need to recreate the system because of a failure (e. g., disk crash) or enhancement or other changes are made to the source. The text of these canned procedures constitutes Appendix A. The steps in using them follow.

Creating SASIS Executable Modules

1. Compile and/or assemble those source programs for which no .RB files exist. If .RB's for the entire system must be generated, using the SETUPTRANS utility is recommended. See paragraph 5.1 for a description of its use.
2. Recreate the SASIS system libraries using the CREATELIBS procedure (enter: @CREATELIBS@(CR) to the CLI) or if only a few .RB's are new, use the applicable (nested) procedure called by CREATELIBS. If none of the routines are included in the new .RB's, go to step 3.
3. To create the executable modules from the new library:
 - a. If all modules are to be created enter: @ALLSAS@.
 - b. If only a few are to be created, use the specific procedures called by ALLSAS.
 - c. To create the support modules (utilities, etc.), enter @ALLOTHERS@.

Since there is insufficient space on a disk cartridge for source .RB and .SV files plus the data generated by one or more SASIS sessions, the executable modules must be transferred to another disk. The procedure for this is defined in the next section.

4.2

Creating the Operational System

This procedure assumes two disk cartridges - a "Development Pack" (DP) which contains source, .RB, and .SV files and an "Operational Pack" (OP) on which SASIS is to reside and be used. If the system only has the operating system, follow these steps.

1. Transfer the necessary files from the DP to the OP by mounting a scratch mag tape on mag tape unit zero and entering:
@TRANSFER@(CR).
2. Delete all user programs from the OP using CHATR to "depermanentize" if necessary and the DELETE/A/V --(CR) command.
3. Load the SPEECHDATA creation software from file 0 of the transfer tape:

```
INIT MT0(CR)
```

```
LOAD/A/V MT0:0(CR)
```

Create SPEECHDATA using the procedures outlined in paragraph 5.1.8ff for the CSDFL program.

4. Then load the remaining files by entering

```
LOAD/A/V MT0:1(CR)
```

If the operational pack has SPEECHDATA, etc., already resident, delete the old versions of the .SV files to be transferred and then load them from the transfer tape.

4.3 Experimental Variations

There are requirements which necessitate substitution of an operational system module with an experimental, demonstration or other version. Such temporary substitutions are best accomplished by procedures separate from the normal operational procedures. The module used to verify the accuracy of SASIS with the analytic studies software illustrates this principle.

The assembly language-based speech input routine (SPASM) samples data from the A-D converter. A version of SPASM called SPASMSUB was created to input data from a mag tape created on the analytic studies hardware instead of from the A-D. A canned procedure called SPCHSUBCR creates the version of SPINP containing SPASMSUB instead of SPASM. The SPINP.SV file was transferred to the OP via tape per the procedure described above by first deleting the regular SPINP.SV and loading the experimental version:

```
DELETE/A/V SPINP.SV  
INIT MT0  
LOAD/A/V MT0:0SPINP.SV
```

4.4 System Backup Procedures

Two facilities for backing up files from disk to tape are available. The first uses the RDOS DUMP and LOAD commands which transfer named (i. e., directoried) files between disk and tape. The second is a stand-alone (i. e., non-RDOS) program that copies complete disk cartridges physically. Named files, stored randomly under RDOS, are transferred not as logical entities but as they reside randomly on 406 disk tracks, each consisting of 3072 16-bit words. The advantage of the latter method is, of course, the relative speed with which track images can be transferred. The disadvantage lies in lack of safety and compatibility. If a tape generated by the stand-alone facility develops a parity error and that error is in part of the disk's directory, no recovery is possible (partly because the RDOS directory is singly linked). Transferring from one disk to another is usually straightforward with the stand-alone facility. If a disk develops a bad sector, RDOS will avoid it. However, the stand-alone facility cannot avoid a bad sector. Therefore, transferring from a disk with a good sector "X" to one with a bad sector "X" is not possible. With those precautions and limitations in mind, let us proceed to the use of the backup facilities.

4.4.1 RDOS - Supported Backup

Refer to the RDOS manual for details on DUMP and its counterpart, LOAD. One conservative approach in backing up files is to make two copies of all files on each of two mag tapes or at least to alternate between two backups so that if one goes bad, at least there is another. Saving marked-up listings between backups allows you to re-edit source programs in a reasonably short time. Frequency of backup depends partly on the stability of the system. If any development is underway, backup once a day is minimal.

4.4.2 DGC West Tape Disk Utility

This stand-alone program is quite simple. No source was available since the program is not supported by DG as a product. Two procedures are included here.

4.4.2.1 Procedure for Writing Tape to Disk

1. Mount mag tape on unit 0.
2. Put 100022 in switches.
3. Press STOP, RESET, PROGRAM LOAD.
4. The following will appear on the terminal:
FULL (0) OR PARTIAL (1)
5. Enter: 1(CR) and the utility will be loaded and the following messages will appear. Enter the underlined data, where N is the number of the file containing the disk image to be loaded:

RTOS REV 3.00

DGC/WEST TAPE DISK UTILITY

VERSION 1 7/1/74

ENTER DISK TYPE (0=DIABLO, 1=2314)? 0(CR)

ENTER INPUT FILE (DP0, DP1, OR MTX:XX)

MT0:N(CR)

ENTER OUTPUT FILE (DP0, DP1 OR MTX:XX)

DP0(CR)

DPCOPY COMPLETED

ENTER DISK TYPE (0=DIABLO, 1=2314)?

4.4.2.2

Procedure for Backing up Disk to Tape

For each copy of the disk, perform steps 1 - 5.

1. Mount tape on unit 0.
2. Put 100022 in switches
3. Press STOP, RESET, PROGRAM LOAD.
4. Operator responses are underlined:

FULL (0) OR PARTIAL (1) 1(CR)

FROM MT0: 1(CR)

RTOS REV 3.00

DGC/WEST TAPE DISK UTILITY

VERSION 1 7/1/74

ENTER DISK TYPE (0=DIABLO, 1=2314) 0(CR)

5. The following messages will appear. Respond as underlined, where N is the number of the tape file to contain a copy of the disk in track image format.

ENTER INPUT FILE (DP0, DP1, OR MTX:XX)

DP0(CR)

ENTER OUTPUT FILE (DP0, DP1 OR MTX:XX)

MT0:N

4.4.2.3 Procedure for Creating the Basic Track Image Tape

1. Mount master tape (i. e., with bootstrap load and utility) on unit 0.
2. Mount scratch tape with write ring on unit 1.
3. Enter on keyboard:

INIT MT0

INIT MT1

XFER MT0:0 MT1:0(CR)

XFER MT0:1 MT1:1(CR)

RELEASE MT0

RELEASE MT1

At this point the mag tape loader and utility have been transferred to files 0 and 1 of the new tape.

Alternatively the DGC/WEST TAPE DISK UTILITY can be used.

1. Perform steps 1 and 2 above.
2. Bootstrap the tape. Specify MT0:0 as the input file and MT1:0 as the output file.
3. Rewind both tapes. Bootstrap again. Specify MT0:1 as the input file and MT1:1 as the output file.

5.0 SUPPORT SOFTWARE

A number of programs were written to support the SASIS Operational System software and hardware. For convenience, these are divided into three categories: utilities; diagnostic and test; and technical management routines. Each program writeup includes a description and usage instructions. The reader should refer to the program listings for detail. See Appendix B for the location of the listings.

5.1 Utility Programs

5.1.1 Time the A-D Clock (TADCLOCK)

5.1.1.1 Description

The operator uses TADCLOCK to measure the A-D sample rate. At present, the SASIS software samples at $(6800 \times 2) = 13,600$ samples per second, but ignores every other sample. Thus, the effective sample rate is 6800/second. TADCLOCK asks the operator to hit a key to begin and a key to end the timing interval. That interval can be measured with a wristwatch. Note that the error associated with using a wristwatch decreases with a larger interval. For example, support an operator can be accurate to .5 second. If he uses a 1 minute time interval, the error is .8 percent. With a five minutes interval (i. e., $300 \pm .5$ sec), the error is reduced to .2 percent.

5.1.1.2

Sample Usage

Operator inputs are underlined in the following example:

TADCLOCK(CR)

THIS IS TADCLOCK

STRIKE ANY KEY TO START COUNTING CONVERSIONS

(SP)*

COUNTING STARTED. TO STOP COUNTING, STRIKE ANY KEY.

(SP) ENTER ELAPSED TIME BETWEEN STRIKING KEYS

IN SECONDS.

10.

ENTER DESIRED** SAMPLE RATE

IN SAMPLES/SECOND

13600.(CR)

SAMPLING RATE= 13418.0

PERCENTAGE OF ERROR= 1

STOP

R

* (SP) indicates space bar

** Or expected

5.1.2 Karman's Special Utility (KARSPEC)

5.1.2.1 Description

This utility uses MASTERLIST (or one of its variations such as MASTERALPH) to produce a canned procedure file applicable to all referenced source files. Both a preamble string and post-amble string are defined by the operator through the keyboard. The reader should refer to Appendix B for a listing of MASTERLIST.

Originally this utility was used to print all FORTRAN and assembly language routines by entering PRINT as the preamble and the null string as a post-amble.

5.1.3 SEL86 to Data General Source Transfer Program (SELDGSRC)

5.1.3.1 Description

SELDGSRC reads a file of source records in ASCII-coded "40A2" format. This was written to transfer source programs to the Nova because a card reader was not configured on the system.

5.1.4 Writeout Source Files (WRITOUT)

5.1.4.1 Description

WRITOUT performs the inverse function of SELDGSRC. It writes those source files specified by the operator to tape. The record format is EBCDIC-coded 40A2 (two bytes per 16 bit word). WRITOUT terminates each file with an EOF mark.

5.1.5 Set Up Translation (SETUPTRANS)

5.1.5.1 Description

SETUPTRANS creates from MASTERLIST a canned procedure file (TRANSLATE) of compilation and assembly commands which when executed generate a complete set of .RB files. To execute the procedure enter:
@TRANSLATE@.

5.1.6 Tape Summary Program (SUMMAR)

5.1.6.1 Description

SUMMAR reads a SASIS session tape mounted on unit 0. Simply mount the tape and enter: SUMMAR(CR). The file names on the tape will be output to the terminal.

5.1.7 Logical Block to Physical Address Conversion Program (LOGBLOCK)

5.1.7.1 Description

LOGBLOCK accepts from the keyboard in octal format a logical disk block number. It converts it to a physical disk address and outputs it to the terminal. The motivation for this program is straightforward. The CLI LIST command optionally outputs a file's starting logical block number.* Since we needed to know the corresponding physical disk address and it was somewhat tedious to calculate manually, the LOGBLOCK program was developed.

LOGBLOCK(CR)

ENTER BLOCK NUMBER IN OCTAL 7312(CR)

BLOCK 3786** IS CYL 157 TRK 1 SEC 6

STOP

R

* LIST, E

** Decimal

5.1.8 Create Speech Data File (CSDFL)

5.1.8.1 Description

CSDFL evolved from the need to create the 68 track contiguous file for inputting speech (SPEECHDATA) beginning on a track boundary. Since there wasn't (and apparently still isn't) a way of doing this with a CLI command, CSDFL was written. CSDFL fills up available disk storage with dummy files one sector in length. When one of these files is created at the address desired for SPEECHDATA, it is deleted and SPEECHDATA is created with a length of 68 tracks. The @ESCDFL@ procedure which executes CSDFL includes a command to assign SPEECHDATA the permanent attribute.

5.1.9 Mag Tape File Name Lister (LISTNAME)

5.1.9.1 Description

LISTNAME writes to the terminal the names of the files as it reads through a SASIS session tape.

5.2 Diagnostic and Test Programs

5.2.1 Analog Linearity Test Program (ANSUM)

This program tests the linearity of the A-D using the D-A. Since the offset between the two devices is typically not zero, it can be determined by using the WRAP program. ANSUM requests that value from the operator, accepts it, then outputs each possible value on the D-A, inputs it on the A-D and tallies the result. A table recording the results is printed at the end of the test. (N.B. The A-D and D-A must be connected during this test. If they aren't, the program may blow up.)

5.2.2 Analog Through Test (ATHRU)

ATHRU inputs a value via the terminal and outputs it through the D-A which is connected to the A-D. The A-D inputs the value and compares it to the value output.

5.2.3 Output Switch Value Through D-A (DACSW)

5.2.3.1 Description

DACSW continuously monitors the sense switches and outputs the value of the switches to the D-A. DACSW is terminated by a console interrupt ("CNTRL-A").

5.2.4 Output Tone Through D-A (DACTONE)

5.2.4.1 Description

DACTONE uses the 256 word file TONE to output an audio sine wave through the D-A. TONE was not on any of the disk packs transferred from Rockwell.

5.2.5 Analog Wrap Around Test (WRAP)

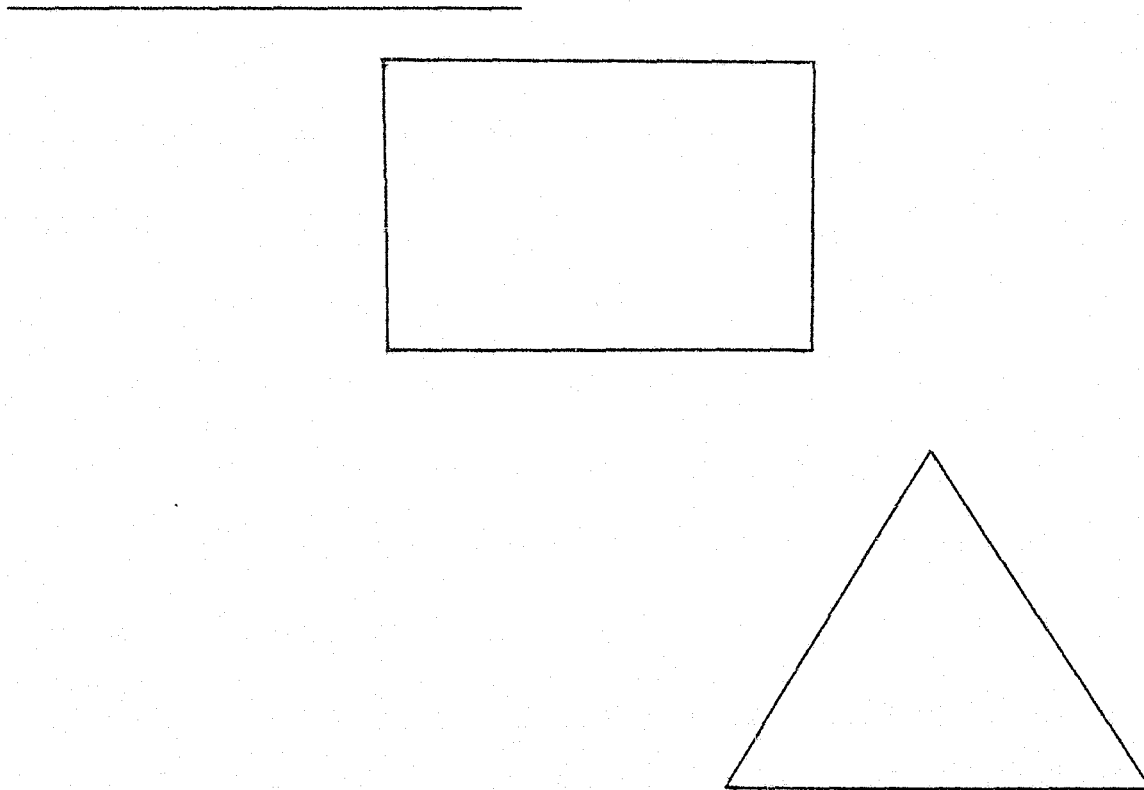
5.2.5.1 Description

WRAP exercises the analog subsystem with the A-D and D-A connected with a cable.

5.2.6 Graphics Pattern Test (GRAFIC)

5.2.6.1 Description

GRAFIC outputs simple geometric patterns to the terminal. The program exits via console interrupt (CNTRL-A). The pattern is shown in the figure below.



To exit enter CNTRL-A.

5.2.7 Alpha Write Through Test (WRAFIC)

5.2.7.1 Description

The WRAFIC test writes a line to the graphics terminal in write-thru mode. a set of characters is presented:

```
!"#$%&'()*+,-./0123456789:;<=>?@ABCDE
```

```
FGHIJKLMNPOQRSTUVWXYZ[\]^_`abcdefghi
```

To terminate the program press the space bar. The screen will be "cleaned" five times and these messages will appear:

```
STOP KEYBOARD
```

```
R
```

5.3 Technical Management Programs

5.3.1 Verify the Existence of Programs in MASTERLIST (EXISTENCE)

5.3.1.1 Description

EXISTENCE attempts to find each file referenced in MASTERLIST.

If a file cannot be found, an exception message identifying the file by name is output to the terminal. The operator simply enters: EXISTENCE(CR) to the CLI.

5.3.2 Librarian (LIBR)

5.3.2.1 Description

LIBR is called via the MARION procedure (enter: @MARION@ (CR)).

It edits one or more source files with a comment line that indicates when the program was compiled and the initials of the responsible person. In the following example operator responses are underlined:

@MARION@ (CR)

INPUT YOUR INITIALS AND PROGRAM NAME

RNK ENTREGEN*

CREATING TODAY'S DEV LOG. IS YESTERDAY'S ON HARD COPY? **

ENTER NO OF HARD COPIES DESIRED***

0. (CR)

ENTER COMPILATION OPTION

N NO COMPILATION

C TO COMPILE

B COMPILATION WITH /B LISTING TO \$LPT

X TO COMPILE WITH /X OPTION

N (CR)

MORE? (Y OR N)

Y (CR)

INPUT YOUR INITIALS AND PROGRAM NAME

RNK NCOMP (CR)

* There is exactly one space expected between the mandatory three initials and filename.

** A reminder question. No operator reply is expected.

*** Output to line printer. Current configuration does not support printer.

ENTER NO OF HARD COPIES DESIRED.

0.(CR)

ENTER COMPILATION OPTION

- N NO COMPILATION
- C TO COMPILE
- B COMPILATION WITH /B
- X TO COMPILE WITH /X OPTION

N(CR)

MORE? (Y OR N)

Y(CR)

INPUT YOUR INITIALS . . .

RNK SCAN(CR)

.
. .
.

MORE? (Y OR N)

N(CR)

R

At this point the today's log was typed out. Today's date was 8/26/75.

TYPE LOG0826(CR)

ENTREGEN	COMPILED	8/26/76 AT 10:12 BY RNK
NCOMP	COMPILED	8/26/76 AT 10:17 BY RNK
SCAN	COMPILED	8/26/76 AT 10:17 BY RNK

5.3.3 Sort MASTERLIST into Alphabetical Order (MASTERSORT)

5.3.3.1 Description

MASTERSORT reads the entire MASTERLIST disk file into core, sorts the filenames into alphabetical order and outputs the resulting array to a source file named MASTERALPH. That file can be output to a hard copy device or to tape via WRITOUT for listing on another system. Some of what MASTERSORT was designed to do is performed by RDOS Rev 3 in its LIST/S option and LOG/ENDLOG facilities.

5.3.4 Generate Software Tree (TREEGEN)

5.3.4.1 Description

TREEGEN reads the FORTRAN source programs listed in MASTERLIST. It scans for the string "CALL." It finds the program name following the CALL and outputs it to a source file named TREE. TREEGEN is not sophisticated; it does not eliminate redundancies. It does not alphabetize the names of called routines. TREEGEN outputs the source line when an FSWAP or INCLUDE string is found in the line.

6.0

SOFTWARE DEVELOPMENT PHILOSOPHY

Performing and managing software development requires well defined policies and procedures. This section describes the facilities provided by Data General and by the SASIS development group. Whatever variations are exercised on the recommendations set down here, the goal should be a clear audit trail so that a technically competent contributor generally familiar with the purpose of SASIS could come in "cold" and be able to proceed backwards from where the project is to where it came from.

6.1

Making Software Changes

The SASIS source programs were created and can be modified using the EDIT utility. A manual describing EDIT is available from Data General. After editing is complete, the source program can be compiled from the CLI using the FORTRAN FILENAME(CR) command or by using a special utility created for SASIS named MARION. See paragraph 5.3.2 for a description of MARION. Successive EDIT-compilations for source programs should be filed by program so that a history is maintained. There are numerous benefits to an historical record. However, for brevity only two will be discussed. If a software change causes a system problem unforeseen at the time the change was planned and implemented, having the original (i. e., working version) on a listing will help the programmer analyze the source of the induced problem. Second, if one programmer makes a change without creating a public record and the new software of another programmer is affected, much time can be wasted before either the second programmer discovers the problem on his own or the first programmer offers the required insight. A final comment is that managing a large system is extremely difficult, if not impossible, without adequate controls if for no other reason than staff turnovers constitute substantial management risk.

6.2 Disk Backup

This subject was addressed earlier. Backup of disk should be performed often during development activity. An up-to-date label should be attached to each tape identifying: (1) the disk pack; (2) content summary by tape file number; and (3) the date and time.

It is prudent to be conservative with respect to backup procedures. During heavy development activity, backing up a developing disk pack-to-tape three times a day would be normal practice. The backup tape might have both source files and all files, that is:

```
INIT MT0
DUMP/A/V MT0:0 —. —(CR)
DUMP/A/V MT0:1 —.(CR)
```

6.3 Development and System Log

It is extremely helpful to a development team for each member to keep a log of his activities. Busy people, by necessity, forget many details after a few days or weeks. A development log compensates for this.

System failures, quirks, subtle features, modifications, etc., that are more general than the specific development activities of a team member are recorded in a system log. Such a log is invaluable when hardware failures are intermittent and/or symptoms are subtle and elusive. It also helps to record minor troubles so that a list can be quickly compiled when it is necessary to perform trouble-shooting on a major problem.

7.0 LIST OF UNFINISHED TASKS

The tasks are not necessarily listed in ordered priority:

1. It must be determined which copy of certain source files is the latest. Two copies of DISPATCH exist on mag tape file 2 disk backup. Another earlier (I hope) version resides on file 4.
2. WRITOUT should output a header on every page or on at least every file defining program name, date, and any information the operator wishes to output.
3. MASTERLIST should be expanded to include all source files and include fields to segment the list into categories. MASTERALPH, EXISTENCE, etc., must be changed.
4. The ACCEPT statement problem must be solved.
5. The documentation on Textronix supplied FORTRAN software may be available from Textronix if not shipped with system.
6. The ACCEPT statements, even when working, are unforgiving of operator input errors. They should be replaced with regular I/O statements and thorough editing logic.
7. Not covered in this spec are descriptions of how the software represents speech to the operator. Methods of smoothing the amplitude contour, etc., are not detailed.
8. Of particular interest to Aerospace are the formants. How they are computed, why they are normalized by division by 4, 10, and 16 is not clear. FFBWM calculates the formants. They are stored in words 214, 215, and 216 of the FF file.

8.0 RECOMMENDATIONS

8.1 Additional Hardware

It is clear that a number of magnetic tapes will be required to backup disks and output SASIS sessions. Additional disk packs would be desirable, would save time, and be useful for disk diagnostic purposes.

8.2 Additional Software

It would be desirable to have these items:

1. Complete set of paper tape diagnostics for the Nova system. Be sure that the system exercisor is included.
2. Having the Nova SYSGEN software would facilitate the creation of a disk pack from scratch. About 1/2 hour or less is required.

8.3 Additional Documentation

The need for additional software documentation will be a function of: (1) the completeness of this document used in conjunction with other documents as listed in Section 9.0 below; and (2) the efforts of the programming staff to interact with RDOS facilities and SASIS materials, listings, and documentation.

Adding new computation algorithms will be straightforward. The area of concern is the labeling process. LABEL11 and LABEL12 are non-trivial modules. If larger segments of speech are to be compared and thus identified as "macro events," LABEL11 and LABEL12 must be reworked.

A-D	Analog to digital
Assembler	A program which translates input symbolic codes into machine instructions suitable for input to a linker or loader
Block	A group of consecutive locations in disk storage which are generally 255 or 256 words in length on the NOVA 840
CLI	The Command Line Interpreter (CLI) is a system program that accepts command lines from the console and translates the input as commands to the operating system. The CLI acts as an interface between the user at the console and the RDOS system.
D-A	Digital to analog
DELETE	A CLI command which deletes a file or a series of files
DP0	Reserved file name for the disk unit
DUMP	A CLI command which dumps files from disk onto a file or device, such as a magnetic tape
EXTENSION	A file name extension is a name that can be appended to a file name. It is a string of alphabetic characters, but only the first two are considered significant.
FF	Abbreviation for feature file and also the name extension (.FF) of a feature file
File	Any collection of information or any device receiving or providing information

File name A file name consists of any number of alphabetic characters, numerals, or the \$ character, but the system considers only the first ten significant. All devices and disk files have file names.

FORTTRAN A CLI command which performs a FORTRAN 5 compilation

I/O Synonymous with input-output which is a general term for the equipment used to communicate with a computer and the data involved in the communication

INIT A CLI command used to initialize a magnetic tape or device

Input The information or data transferred from an external storage medium (for example, a magnetic tape or disk) into the internal storage of a computer. In the SASIS system, the speech data is an example of input.

LIST A CLI command which lists file directory information

LOAD A CLI command which reloads files that have been dumped with the DUMP command

Module A portion or segment of a larger program

MTn Reserved file name for magnetic tape transport n, where n is 0 or 1

Operating System A set of software furnished by the computer manufacturer to control management aspects of the computer operation. The operating system is differentiated from user applications software, which is applications task oriented.

Output The information transferred from the internal storage of a computer to an external storage medium or any device outside of the computer. In the SASIS system, the teletype printout is an example of output.

Overlay A disk file called by a root program which remains core resident. Overlays permit the overwriting of a portion of core with disk file images, which is helpful when large programs are to be executed within limited core. Overlays have name extensions of .OL.

R A symbol which appears on the terminal screen indicating that CLI is ready to accept commands

RDOS The Real Time Disk Operating System (RDOS) is the operating system for the NOVA 840 computer

RE Abbreviation for raw event and also the name extension (.RE) of a raw event file

RELEASE A CLI command which releases a device from the system

Relocatable Binary File A computer program after it has been assembled. Relocatable binary files are input to the relocatable loader. They have file name extensions of .RB.

Relocatable Loader The relocatable loader loads and relocates a program at absolute locations producing a core image file (also called a save file)

RLDR A CLI command which performs a relocatable load

Save file When a source program has been assembled and loaded, it is a save file (also called a core image file) ready for execution. Save files have name extensions of .SV.

Spooling (Simultaneous peripheral operation on-line) Spooling is a method of handling low-speed I/O devices commonly implemented in operating systems to increase throughput. Spooling increases throughput because the central processing unit

spends less time waiting for input data to be delivered to, or output data to be taken from, its buffers.

Swap

Swaps are save files containing core images of total user address space. They permit the overwriting of resident core images with disk file images. Swapping occurs when a program executing under the operating system suspends its own execution and invokes another program or another segment of itself that exists as a save file on disk. The calling program is referred to as executing at a higher level in the system than the called program, or program swap, which is referred to as running at a lower level. The called program may, in turn, invoke another program swap. The operating system provides for up to five levels of program swaps.

XFER

A CLI command used to copy the contents of a file to another file

@

Paired @ signs around a file name are understood to represent the contents of the file rather than the file name itself

\$TTI

Reserved file name for the terminal keyboard

\$TTO

Reserved file name for the terminal screen



APPENDIX C. EQUIPMENT LISTING FOR
SEMI-AUTOMATIC SPEAKER IDENTIFICATION SYSTEM



Table C-1. Semi-Automatic Speaker Identification System Equipment List

Item	Quantity	Part No.	Description	Supplier
1	1	8299	NOVA 840 minicomputer with 32K core	Data General
2	1	8206	Power monitor and auto restart	" "
3	1	8207	Hardware multiply/divide	" "
4	1	8208	Automatic program load	" "
5	1	8020	Floating point processor	" "
6	2	4007	I/O interface subassembly	" "
7	2	4010	Teletype I/O interface	" "
8	1	4010C	Teletype Model 35 KSR	" "
9	1	4008	Real time clock	" "
10	1	4011	Paper tape reader control	" "
11	1	6013	High-speed paper tape reader	" "
12	1	4030	Magnetic tape control	" "
13	2	4030J	Magnetic tape transport (Wang 45 LPS)	" "
14	1	4046	Disc control	" "
15	1	4047	Disc adapter and power supply	" "
16	1	4047A	Disc drive (Diablo Model 31)	" "
17	1	4140	A/D subsystem	" "
18	1	4180	D/A subsystem	" "
19	1	—	Three-bay rack cabinet	Electro-Rack
20	1	R4012	Computer graphics display terminal (rack mounting version)	Tektronix
21	1	Option 4	TTY Port/NOVA	"
22	1	CM018-0074-00	Variable speed vector generator	"
23	1	018-0069-00	Accessory motherboard	"
24	1	062-1427-01	Plot-10 minicomputer NOVA software	"
25	1	016-0304-00	Viewing hood	"

Semi-Automatic Speaker Identification System Equipment List (cont.)

Item	Quantity	Part No.	Description	Supplier
26	1	016-0291-00	Copyholder	Tektronix
27	1	—	Extender cable and connector	"
28	2	—	Low-pass filter	T. T. Electronics
29	1	AG440B	Audio tape recorder, 2 channel, 1/2 and 1/4 track, with servo capstan and scrape flutter idler	Ampex
30	1	AV-NS-7	Headphones	Superex
31	—	—	Various components for audio control panel, including amplifiers, meters, loudspeaker	Various
32	—	—	Magnetic tapes and magnetic discs for data recording and storage	Various

Table C-2. Manuals for Semi-Automatic Speaker
Identification System Components

1. Ampex Recorder AG-440B and AG-445B, Recorder and Reproducer Operator's and Maintenance Manual
2. DSI Series 30 Disk Drives Maintenance Manual
3. Tektronix 4012 Computer Display Terminal
4. Tektronix 4010 Teletype Port Interface
5. Teletype Model 35 Technical Manual, Volumes I and II
6. Teletype Model 35 Parts Manual
7. Teletype Model 35 Motor Manual
8. Data General Corporation Technical Manuals 800, 840
9. Data General Corporation Integrated Circuit User's Guide
10. Data General Corporation Engineering Specification
11. Data General Corporation Instructions and Reference Card
12. Data General Corporation: How to Use the Nova Computers
13. Data General Corporation Paper Tape Reader
14. Data General Corporation 4046 Moving Head Disk Controller, Volumes 1 and 2
15. Wangco Inc., Model 10 NRZI Magnetic Tape Transport Operator's and Maintenance Manual
16. Schematic Diagrams: Floating Point Unit, CPU, A/D-D/A Converter, Paper Tape Reader, Teletype Machine, Magnetic Tape, Console, Audio Control Unit, Disk

END